11-11-2016

Measure and Integration

Jose L. Menaldi
Wayne State University, menaldi@wayne.edu

Recommended Citation
Menaldi, Jose L., "Measure and Integration" (2016). Mathematics Faculty Research Publications. 22.
https://digitalcommons.wayne.edu/mathfrp/22

This Book is brought to you for free and open access by the Mathematics at DigitalCommons@WayneState. It has been accepted for inclusion in Mathematics Faculty Research Publications by an authorized administrator of DigitalCommons@WayneState.
Measure and Integration\textsuperscript{1}

\textsc{Jose-Luis Menaldi}\textsuperscript{2}

Current Version: 11 November 2016\textsuperscript{3}
First Version: 2014 \textsuperscript{4}

\textsuperscript{1}©Copyright 2014. No part of this book may be reproduced by any process without prior written permission from the author.

\textsuperscript{2}Wayne State University, Department of Mathematics, Detroit, MI 48202, USA (e-mail: menaldi@wayne.edu).

\textsuperscript{3}Long Title. \textit{Measure and Integration: Theory and Exercises}

\textsuperscript{4}This book is being progressively updated and expanded. If you discover any errors or you have any improvements to suggest, please e-mail the author.
# Contents

## Preface

## Introduction

### 0 Background

0.1 Cardinality .................................................. 1  
0.2 Frequent Axioms .............................................. 3  
0.3 Metrizable Spaces ............................................ 5  
0.4 Some Basic Lemmas .......................................... 8

### 1 Measurable Spaces

1.1 Classes of Sets ............................................... 13  
1.2 Borel Sets and Topology ..................................... 20  
1.3 Measurable Functions ........................................ 23  
1.4 Some Examples ................................................ 25  
1.5 Various Tools ................................................ 27

### 2 Measure Theory

2.1 Abstract Measures ............................................. 33  
2.2 Caratheodory’s Arguments ................................... 37  
2.3 Inner Approach ............................................... 48  
2.4 Geometric Construction ...................................... 53  
2.5 Lebesgue Measures .......................................... 56

### 3 Measures and Topology

3.1 Borel Measures ............................................... 71  
3.2 On Metric Spaces .............................................. 75  
3.3 On Locally Compact Spaces .................................. 80  
3.4 Product Measures ............................................ 86

### 4 Integration Theory

4.1 Definition and Properties ................................... 93  
4.2 Cartesian Products ............................................ 103  
4.3 Convergence in Measure ..................................... 108  
4.4 Almost Measurable Functions ............................... 114
4.5 Typical Function Spaces ........................................ 121

5 Integrals on Euclidean Spaces .................................... 127
  5.1 Multidimensional Riemann Integrals .............................. 127
  5.2 Riemann-Stieltjes Integrals ..................................... 135
  5.3 Diadic Riemann Integrals ....................................... 145
  5.4 Lebesgue Measure on Manifolds .................................. 148
  5.5 Hausdorff Measure ............................................... 155
  5.6 Area and Co-area Formulae ...................................... 161

6 Measures and Integrals ............................................. 171
  6.1 Signed Measures .................................................. 171
  6.2 Essential Supremum .............................................. 177
  6.3 Orthogonal Projection ........................................... 183
  6.4 Uniform Integrability ............................................ 186
  6.5 Representation Theorems ....................................... 204

7 Elements of Real Analysis ......................................... 209
  7.1 Differentiation and Approximation .............................. 210
  7.2 Partition of the Unity ........................................... 217
  7.3 Lebesgue Points .................................................. 219
  7.4 Functions of one variable ....................................... 225
  7.5 Lebesgue Spaces .................................................. 233
  7.6 Trigonometric Series ............................................ 239
  7.7 Some Complements ............................................... 244

Appendix - Solutions to Exercises ................................. 253
  1 Measurable Spaces ................................................. 253
  2 Measure Theory .................................................... 267
  3 Measures and Topology ............................................ 303
  4 Integration Theory ................................................. 313
  5 Integrals on Euclidean Spaces .................................... 341
  6 Measures and Integrals ............................................ 351
  7 Elements of Real Analysis ....................................... 359

Notation ................................................................. 391

Bibliography ............................................................ 395

Index ...................................................................... 403
Preface

This project has several parts, of which this book is the first one. The second part deals with basic function spaces, particularly the theory of distributions, while part three is dedicated to elementary probability (after measure theory). In part four, stochastic integrals are studied in some details, and in part five, stochastic ordinary differential equations are discussed, with a clear emphasis on estimates. Each part was designed independent (as possible) of the others, but it makes a lot of sense to consider all five parts as a sequence.

The last two parts are derived from a previous course to supplement stochastic optimal control theory, while the first three parts of these lectures begun when preparing and teaching a short course in Elementary Probability with a first introduction to measure theory at the University of Parma (Italy) during the Winter Semester of 2006. Later preparing to teach our regular Real Analysis series at Wayne State University during the academic year 2007, and after reviewing many books with suitable material, I decided to enlarge and to complete my notes instead of adopting one of books commonly used here and there. Clearly, there are many excellent books from which a (two semesters) Real Analysis course can be taught, but perhaps each instructor has a unique opinion and a particular selection of topics. Nevertheless, most of instructors will agree (in principle) with a selection of sections included in this book.

As mentioned earlier, this course grew out of an interest in Probability, but without rushing throughout the measure and integration (theory), what in most cases is the difference between students in analysis with a pure interest versus a more applied orientation. Thus, the reader will note a subtle insistence in the extension of measures form a semi-ring, in some general properties of measures on topological spaces (a chapter that can be skipped during the first reading) and in various spaces of measures and measurable functions. In a way, the approach is to cover first the essential and to deal later with the complements. Most of the style is formal (propositions, theorems, remarks), but there are instances where a more narrative presentation is used, the purpose being to force the student to pause and fill-in the details.

Practically, there are no specific section of exercises, giving to the instructor the freedom of choosing problems from various sources (and according to a particular interest of subjects) and reinforcing the desired orientation. There is no intention to diminish the difficulty of the material to put students at ease, on the contrary, all points are presented as blunt as possible, even sometimes
shorten some proofs, but with appropriate references. For instance, we assume that most of Chapter 0 (the background) is somehow known, even if not all will be actually needed, but it is preferable to warn earlier the reader of the deep analysis ahead. The first three chapters (1, 2 and 4) give the basic stuff about abstract measure and integration, while Chapter 5 and 6 complement the material in two opposite directions. Chapter 3 is a little more demanding. Finally, in Chapter 7, we are ready to see the results of the theory.

This book is written for the instructor rather than for the student in a sense that the instructor (familiar with the material) has to fill-in some (small) details and selects exercises to give a personal direction to the course. It should be taken more as Lecture Notes, addressed indirectly (via an instructor) to the student. In a way, the student seeing this material for the first time may be overwhelmed, but with time and dedication the reader can check most of the points indicated in the references to complete some hard details. Perhaps the expression of a guided tour could be used here.

In the appendix, all exercises are re-listed by section, but now, most of them have a (possible) solution. **Certainly, this appendix is not for the first reading,** i.e., this part is meant to be read after having struggled (a little) with the exercises. Sometimes, there are many ways of solving problems, and depending of what was developed “in the theory”, solving the exercises could have alternative ways. The instructor will find that some exercises are trivial while other are not simple. It is clear that what we may call “Exercises” in one textbook could be called “Propositions” in others. This part one has a large number of exercises, but as the material get more complicated (i.e., in several chapters in parts two and three), a few or not at all exercises are given.

The combination of parts I, II, and III is neither a comprehensive course in measure and integration (but a certain number of generalizations suitable for probability are included), nor a basic course in probability (but most of language used in probability is discussed), nor a functional analysis course (but function spaces and the three essential principles are addressed), nor a course in theory of distribution (but most of the key component are there). One of the objectives of these first three books is to show the reader a large open door to probability (and partial differential equations), without committing oneself to probability (or partial differential equations) and without ignoring hard parts in measure and integration theory.

Michigan (USA),

Jose-Luis Menaldi, June 2015
Introduction

As indicated by the title, these lecture notes concern measure and integration theory. The objective is the $d$-dimensional Lebesgue integral, but in going there, some general properties valid for measures in metric spaces are developed. Instead of taking a direct way to reach our goal (after a background chapter), we prefer a more systematic approach in which family of sets and measurable functions are presented first, without a direct motivation. Chapter 2 (abstract measures) begins with the classic Caratheodory construction (with the typical example of the Lebesgue measure), followed by the inner measure approach and a little of the geometric construction. Chapter 3 (can be skipped in the first reading) continues with Borel measures (which can also be used to establish more specific properties on the Lebesgue measure). Basic on integrals is developed in Chapter 4, where the convergence theorem are obtained, and in Chapters 5 and 6, we give some complement on measure and integration theory (including Riemann-Stieltjes integrals and signed measures). Finally, in Chapter 7, we consider most of the useful results applicable to Euclidean $d$-dimensional spaces. As much as possible, each section is kept independent of other sections in the same chapter.

For instance, to have a quick historical evolution of ideas within the integrals (from Cauchy to Lebesgue), the interested reader may take a look at the book Burk [23, Chapter I] or Chae [25, Chapter I and II]. Perhaps checking most of Gordon [54], the reader may find a detailed discussion on various type of integrals. Now, before going into more details and as a preview of what is to come, a quick discussion on discrete measures (including the typical convergence theorems) is in order.

Essentially based on property of the sup, recall that for any series of nonnegative real numbers $\sum_{i=1}^{\infty} a_i$, with $a_i \geq 0$, the sum $\sum_{i=1}^{\infty} a_i$ is the $\sup\{\sum_{i=1}^{n} a_i : n \geq 1\}$ and therefore: (1) if $\iota$ is a bijective function from the positive integers into themselves then $\sum_{i=1}^{\infty} a_i = \sum_{i=1}^{\infty} a_{\iota(i)}$; (2) if $I_1, I_2, \ldots$ is a partition (finite or non) of the positive integers then $\sum_{i=1}^{\infty} a_i = \sum_{n} \sum_{i \in I_n} a_i$. It is convenient to use a discrete example as a motivation for our discussion.

Let $\Omega$ be a non empty set and denote by $2^\Omega$ the family of all the subsets of $\Omega$, and then, choose a finite or at most countable subset $I$ of $\Omega$ and a sequence of strictly positive real numbers $\{a_i : i \in I\}$. Consider $m : 2^\Omega \to [0, \infty]$ defined by $m(A) = \sum_{i \in I} a_i \mathbb{1}_A(i)$, where $\mathbb{1}_A(i) = \mathbb{1}_{\{i \in A\}}$ is equal to 1 only if $i \in A$ and zero otherwise. Note that initially, all $a_i$ are nonnegative real numbers, but we
can include the symbol $+\infty$ preserving the linear order and the meaning of the series, i.e., all series have nonnegative terms, a converging series has a finite value and a non convergence (divergent) series has the symbol $+\infty$ as its value.

Therefore we have by definition (1) $m(\emptyset) = 0$ and the following property (so-called $\sigma$-additivity), (2) if $A = \bigcup_{i=1}^{\infty} A_i$ with $A_i \cap A_j = \emptyset$ for any $i \neq j$, then $m(A) = \sum_{i=1}^{\infty} m(A_i)$.

This function (defined on sets) $m$ is called a discrete measure, the set $I$ is the set of atoms and $a_i$ is the measure (or weight) of the atom $i$. Clearly, to define $m$ we need only to know the values $m(\{i\})$ for any $i$ in the finite or countable set $I$. An element $N$ of $2^\Omega$ is called negligible with respect to $m$ if $m(N) = 0$. In the case of discrete measures, any subset of $N$ of $\Omega \setminus I$ is negligible. If $m(\Omega) = 1$ then we say that $m$ is a discrete probability measure.

A function $f : \Omega \to \mathbb{R}$ is called integrable with respect to $m$ if the series $\sum_{i \in I} |f(i)| m(\{i\}) = \sum_{i \in I} |f(i)| a_i$ converges, and in this case the integral with respect to $m$ is defined as the following real number:

$$\int f \, dm = \int f \, dm = \sum_{i \in I} f(i) m(\{i\}) = \sum_{i \in I} f(i) a_i.$$ 

Even if the series diverges, if $f$ is nonnegative then we can define the integral as above (a nonnegative number when the series converges or the symbol $+\infty$ otherwise). Next, a function $f$ is called quasi-integrable with respect to $m$ if either the positive part $f^+$ or the negative part $f^-$ is integrable. The class of integrable functions is denoted by $\mathcal{L}$ or $\mathcal{L}(\Omega, 2^\Omega, m)$ if necessary.

A couple of properties are immediately proved for the integral:
(1) if $c \in \mathbb{R}$ and $f, g \in \mathcal{L}$ then $cf + g \in \mathcal{L}$ and $\int (cf + g) \, dm = c \int f \, dm + \int g \, dm$;
(2) if $f \leq g$, quasi-integrable then $\int f \, dm \leq \int g \, dm$;
(3) $f \in \mathcal{L}$ if and only if $|f| \in \mathcal{L}$ and in this case $|\int f \, dm| \leq \int |f| \, dm$;
(4) if $f \geq 0$ and $\int f \, dm = 0$ then $f = 0$ except in a negligible set.

There are three main ways of taking limit inside the integral for a sequence $f_n$ of functions:
(a) Beppo Levi’s monotone convergence: if $0 \leq f_n \leq f_{n+1}$ and $f(x) = \lim_n f_n(x)$ for any $x \in \Omega$ then $\int f \, dm = \lim_n \int f_n \, dm$;
(b) Fatou’s lemma: suppose that $f_n \geq 0$ and $f(x) = \lim inf_n f_n(x)$ then $\int f \, dm \leq \lim inf_n \int f_n \, dm$;
(c) Lebesgue’s dominated convergence: if $|f_n| \leq g$ with $g \in \mathcal{L}$ and $f(x) = \lim_n f_n(x)$ exists for any $x \in \Omega$ then $\int f \, dm = \lim_n \int f_n \, dm$.

Essentially, any one of the three theorems can be deduced from any of the others. For instance, use (a) on $g_n(x) = \inf_{k \geq n} f_k(x)$ to get (b) and use (b) on $g \pm f_n$ to deduce (c). To prove (a), for any number $C < \int f \, dm$ there exists a finite set $J \subset I$ of atoms such that $\sum_{i \in J} f(i) m(\{i\}) > C$. Since $J$ is finite, for every $\varepsilon > 0$ there exists $N = N(\varepsilon, C, J)$ such that $\sum_{i \in J} f_n(i) m(\{i\}) > C - \varepsilon$ for every $n \geq N$. Because $f_n \geq 0$ and $C, \varepsilon$ are arbitrary we deduce $\int f \, dm \leq \lim_n \int f_n \, dm$, and the equality follows.

For instance, if $\Omega = \{1, 2, \ldots\}$ is the set of strictly positive integers then $m(\{i\}) = 2^{-i}$ defines a probability measure. Consider the sequence of functions
\{f_n\} given by \(f_n(x) = 2^n \mathbb{1}_{\{n=x\}}\). We can verify that each \(f_n\) is integrable, and that the sequence \(\{f_n\}\) converges to the function identically zero but \(\int f_n \, dm = 1\) for every \(n\), i.e., we cannot use (a) or (c) and the inequality in (b) is strict.

Negligible sets (or sets of measure zero) do not play an essential role for discrete measures since there is a largest set of measure zero, namely \(\Omega \setminus I\), i.e., for a given discrete measure \(m\) with atoms \(I\) we may ignore the complement of \(I\). However, in general, negligible sets are a fundamental part, i.e., almost everything happens except a set of measure zero, referred to as \textit{almost everywhere} (a.e.) or \textit{almost surely} (a.s.) when a probability is used.

The generalization of these arguments is the basis of measure and integration theory. However, before being able to reach the classic example of the Lebesgue-Borel measure and the extension of the Riemann-Stieltjes integral, several points should be discussed.

Certainly, there are many (text)books on measure theory and integration at various levels of difficulties (too many to make a non-exhaustive reasonable list), but let me mention that the classic textbooks Halmos [57] and Natanson [86] (among others) were available (besides lecture notes) for me (when studying this subject for the first time). Now, for the (student) reader that just want to take a (serious) panoramic view, (among other) the textbooks Bass [8] and Pollard [90] are an excellent option; and in Richardson [93, Chapter 1, pp. 1–9], a quick history on the subject is given. Perhaps the reader may be interested in a concrete view (and less traditional) of the material, e.g. Swartz [113, Chapters 1–4, pp. 1–164]. In several parts of this ‘lecture notes’, there are precise references to (text)books (with pages) that the reader may check to enlarge details of proofs or viewpoints (and discussions) on the various aspects of measure (and integration) theory covered in this manuscript.
Chapter 0

Background

Before going into some details and without really discussion set theory (e.g. see Halmos [58]) we have to consider a couple of issues. On the other hand, we need to recall some basic topology, e.g. most of what is presented in the preliminaries section of Yosida [122, Chapter 0, pp. 1–22]. Alternatively, the reader may check part of the material in Royden [98, Chapters 1,2,7,8 and 9] or Dshalalow [36, Part I, pp. 1–200]. Certainly, being familiar with most of the material developed in basic mathematical analysis books (e.g., Apostol [4] or Hoffman [63] or Rudin [99]) yields a comfortable background, while being familiar with the material covered in elementary mathematical analysis books (e.g., Kirkwood [70] or Lewin [76] or Trench [116]) provides an almost sufficient background. It not the intention to write a comprehensive treatment of Measure Theory (as e.g., the five volumes of Fremlin [46]), but the reader may find a little more than expected. Reading what follows for the first time could be very dense, so that the reader should have some acquaintance with most of the concepts discussed in this preliminary chapter. Clearly, not every aspect (of this preliminary chapter) is needed later, but it is preferred to face these possible difficulties now and not later, when the actual focus of interest is revealed.

0.1 Cardinality

We want to count the number of element of any set. If a set is finite, then the number of elements (also called the cardinal of the set) is a natural or nonnegative integer number (zero if the set is empty). However, if the set is infinite then some consideration should be made. To define the cardinal of a set in general, we say that two sets have the same cardinal or are equipotent if there exits a bijection between them. Since equipotent is a reflexive, symmetric and transitive relation, we have equivalence classes of sets with the same cardinal. Also we say that the cardinal of a set $A$ is not greater than the cardinal of another set $B$ if there is a injective function from $A$ into $B$, usually denoted by $\text{card}(A) \leq \text{card}(B)$, and we can show that $\text{card}(A) \leq \text{card}(B)$ and $\text{card}(B) \leq$
card(A) imply that A and B have the same cardinal. Similarly, if A and B have the same cardinal then we write card(A) = card(B), and also card(A) < card(B) with obvious meaning.

We use the symbol \( \aleph_0 \) (aleph-nought) for the cardinal of \( \mathbb{N} \) the set of natural (or nonnegative integer) numbers, and we show that \( \aleph_0 \) is first nonfinite cardinal. Any set in the class \( \aleph_0 \) is called countable infinite or denumerable, while countable sets, may be finite or infinite. With time, the two names countable and denumerable are used indistinctly and if necessary, we have to specify finite or infinite for countable sets. It can be shown that the integer numbers \( \mathbb{Z} \) and the rational numbers \( \mathbb{Q} \) are both countable sets, moreover, the union or the finite (Cartesian) product does not change the cardinality of infinite sets, e.g., if \( \{A_i : i \geq 1\} \) is a sequence of countable sets \( A_i \) then \( \bigcup_{i=1}^{\infty} A_i \) and \( \prod_{i=1}^{n} A_i \) are also countable, for any positive integer \( n \). However, we can also show that the cardinal of \( 2^\mathbb{A} \), the set of the parts of a nonempty set \( A \) (i.e., the set of all subsets of \( A \), which can be identified with the product \( \{0,1\}^A \) has cardinal strictly greater than card(\( A \)). Nevertheless, if \( A \) is an infinite set then the set composed by all subsets of \( A \) having a finite number of element (called the finite-parts of \( A \)) have the same cardinal as \( A \). Indeed, if \( A = \{1,2,\ldots\} \) then the set \( 2^A \) of the finite-parts of \( A \) can be represented (omitting the empty set) as finite sequences \( a = \{a_1,\ldots,a_n\} \) of elements \( a_i \) in \( A \). Thus, if \( \{2,3,5,7,\ldots,p_i,\ldots\} \) is the sequence of all prime numbers then for each \( a \) there is a unique positive integer \( m = 2^{a_1}3^{a_2}5^{a_3}7^{a_4}\cdots p_n^{a_n} \), and because the factorization in term of the prime numbers is unique, the mapping \( a \mapsto m \) is one-to-one, i.e., \( 2^A \) is countable.

Representing real numbers in binary form, we observe that \( 2\{0,1,\ldots\} \) has the same cardinal as the real numbers \( \mathbb{R} \), which is strictly greater than \( \aleph_0 \). The cardinality of \( \mathbb{R} \) is called cardinality of continuum and denoted by \( 2^{\aleph_0} \). However, we do not know whether or not there exists a set \( A \) with cardinal \( \aleph \) such that \( \aleph_0 < \aleph \) and \( \aleph < 2^{\aleph_0} \). Anyway, it is customary to use the notation \( \aleph_1 = 2^{\aleph_0} \), \( \aleph_2 = 2^{\aleph_1} \), and so on.

The (generalized) continuum hypothesis states that for any infinite set \( A \) there is no set with cardinal strictly between the cardinal of \( A \) and the cardinal of \( 2^A \). In particular for \( \aleph_0 \) and \( 2^{\aleph_0} \), this assumption (so-called continuum hypothesis) has an equivalent formulation as follows: the set \( \mathbb{R} \) can be well-ordered in such a way that each element of \( \mathbb{R} \) is preceded by only countably many elements, i.e., there is a relation \( \leq \) satisfies (a) for any two real numbers \( x \) and \( y \) we have \( x \leq y \) or \( y \leq x \) or \( x = y \) (linear order), (b) for every real number \( x \) we have \( x \leq x \) (reflexive), and (c) every nonempty subset of real numbers \( A \) has a first number, i.e., there exist \( a_0 \) in \( A \) such that \( a_0 \leq a \) for any \( a \) in \( A \) (well-ordered), and the extra condition (d) for every real number \( x \) the set of real numbers \( y \leq x \) is a countable set. It can also be proved that the cardinal number of \( \mathbb{R} \) (which is called the continuum cardinal) is indeed the cardinal of \( 2^\mathbb{N} \) (i.e., the part of \( \mathbb{N} \)). Furthermore, the continuum hypothesis (i.e., there is not cardinal number between \( \aleph_0 \) and the continuum) is independent of the axioms of set theory (i.e., it cannot be proved or disproved using those axioms).

For instance, the reader may check the books Ciesielski [26], Cohen [27], Gol-
drei [53], Moschovakis [84] or Smullyan and Fitting [107], for a comprehensive treatment in cardinality and set theory axioms. Also Pugh [91, Chapter 1, pp. 1–50] or Strichartz [111, Chapter 1, pp. 1–24] is a suitable initial reading.

0.2 Frequent Axioms

Sometimes, we need to differentiate sets with the same cardinal based on other characteristics, e.g., a natural order of numbers or the natural inclusion for collection or family of sets.

A partially ordered set \((X, \preceq)\) is a set \(X\) (family of sets) and a relation \(\preceq\), which is transitive (\(a \preceq b\) and \(b \preceq c\) imply \(a \preceq c\)) and antisymmetric (\(a \preceq b\) and \(b \preceq a\) imply \(a = b\)). An order \(\preceq\) (on a set \(X\)) is called (1) linear (or total, and the set \(X\) is linearly ordered) if for every \(a\) and \(b\) in \(X\) we have either \(a \preceq b\) or \(b \preceq a\), and (2) well-order (and the set \(X\) is well-ordered) if (1) holds and any nonempty subset \(A\) of \(X\) has a minimum element, i.e., there is \(a\) in \(A\) such that \(a \preceq a\), for every \(a\) in \(A\).

Certainly (several) well-order \(\preceq\) can be given to a finite set, and any subset of integer numbers with a finite infimum inherited a well-order from the integer numbers. A typical situation is to partially order a collection of sets with the inclusion. Note that the natural order of the real numbers \(\mathbb{R}\) is a linear order, but not a well-order. Also, the set \(\mathbb{R}^I\) of all real-valued functions on a set \(I\) (of more than one element), with the natural partial order \((x_i) \preceq (y_i)\) if \(x_i \leq y_i\) for every \(i\) in \(I\), is not a linear order.

In a partially ordered set not all elements are comparable, i.e., we may have two elements \(a\) and \(b\) such that neither \(a \preceq b\) nor \(b \preceq a\). Thus, given a partially ordered set \((X, \preceq)\) and a subset \(A \subset X\), we say that \(x\) in \(X\) is an upper bound of \(A\) if \(a \preceq x\) for every \(a\) in \(A\), and if \(x\) belongs to \(A\) then \(x\) is the maximum element of \(A\). Maximum has little use for partially ordered set, instead, we say that \(m\) in \(A\) is a maximal element, chain of \(A\) if for any \(a\) in \(A\) such that \(m \preceq a\) we have \(a = m\) (i.e., \(m\) is larger or equal to any other element \(a\) in \(A\) that can be compared with \(m\)). A chain in \(X\) is a subset \(C \subset X\) such that \(\preceq\) becomes a linear order on \(C\).

There are several equivalent ways of expressing the well-ordering principle (or axioms), e.g.,

**Hausdorff Maximal Principle:** Every partially ordered set \((X, \preceq)\) has a maximal chain, i.e., a subset \(C\) of \(X\) such that \((C, \preceq)\) is a linearly ordered set and \((C', \preceq)\) is not a linearly ordered set, for any subset \(C'\) strictly containing \(C\).

**Zorn’s Lemma:** Every nonempty partially ordered set has a maximal element if any chain has an upper bound.

**Zermelo’s Axiom:** Every set can be well-ordered, i.e., if \(X\) is a set, then there is some well-order \(\preceq\) on \(X\), i.e., \(\preceq\) is a linear order (all elements in \(X\) are comparable) and every nonempty subset of \(X\) has a first element.

A typical use is when a construction of sets satisfying some properties is
partially ordered (e.g., by the inclusion), and we deduce the existence of a 
maximal set satisfying those properties.

Related to the above assumptions, but independent from other axioms of the 
set theory, is the so-called Axiom of Choice (AoC), which can also be expressed 
in various equivalent ways, e.g.,

AoC Form (a): The Cartesian product of any nonempty family of nonempty 
sets must be a nonempty set, i.e., if \( \{A_i : i \in I\} \) is a family of sets such that 
\( I \neq \emptyset \) and \( A_i \neq \emptyset \), for any \( i \in I \) then there exits at least one choice \( a_i \in A_i \), for 
any \( i \in I \).

AoC Form (b): If \( \{A_i : i \in I\} \) is a family of arbitrary nonempty disjoint sets 
indexed by a set \( I \), then there exists a set consisting of exactly one element form 
each \( A_i \), with \( i \in I \).

All these axioms come into play when dealing with uncountable sets.

Beside using cardinality to classify sets (mainly sets involving numbers), 
we may push further and classify well-ordered sets. Thus, similarly to the 
cardinality, we say that two well-ordered sets \( (X, \preceq) \) and \( (Y, \preceq) \) have the same 
ordinal if there is a bijection between them preserving the order. Thus, to each 
well-ordered set we associate an ordinal (an equivalence class). It clear that 
finite ordinals are the sets of natural numbers \( \{1, 2, \ldots, n\} \), \( n = 1, 2, \ldots \), with 
the natural order, and the first infinite ordinal is the set of natural numbers 
\( \mathbb{N} = \{1, 2, \ldots\} \) (or equivalently any infinite subset of integer number with a 
finite infimum). Each ordinal has a next ordinal, i.e., given an ordinal \( (X, \preceq) \) we define \( (X + 1, \preceq) \) to be \( X + 1 = X \cup \{\infty\} \), with \( \infty \not\in X \) and \( x \preceq \infty \), for every 
\( x \) in \( X \). However, each ordinal not necessarily has a previous (or precedent) 
ordinal, e.g., there is not an ordinal \( X \) such that \( X + 1 = \mathbb{N} \).

Similarly to cardinals, we say that the ordinal of a well-ordered set \( (X, \preceq) \) is 
not greater than the ordinal of another well-ordered set \( (Y, \preceq) \) if there is a 
injective function from \( X \) into \( Y \) preserving the linear order, usually denoted by 
\( \text{ord}(X) \leq \text{ord}(Y) \). Thus, we can show that if \( \text{ord}(X) \leq \text{ord}(Y) \) and \( \text{ord}(Y) \leq \text{ord}(X) \) then \( (X, \preceq) \) and \( (Y, \preceq) \) have the same ordinal. Moreover, there are many 
properties satisfied by the ordinal (e.g., see Brown and Pearcy [21, Chapter 5, 
pp. 80–95], Kelley [67, Appendix, 240–281]), we state for future reference

**Ordinal Order:** The set (actually class or family of sets) of all ordinals is well-
ordered (the above order denoted by \( \leq \) and the strict order by \( < \), i.e., the \( \leq \) 
and the \( \neq \)), namely, every nonempty subset (subfamily) of ordinals has a first 
ordinal. Moreover, given an ordinal \( x \), the set \( \{y < x\} \) of all ordinal strictly 
precedent to \( x \) has the same ordinal as \( x \).

For instance, we may call \( \omega_0 \) the first uncountable ordinal, i.e., any ordinal 
\( \omega < \omega_0 \) is countable (finite or infinite). It is clear that there are plainly of 
ordinals between \( \mathbb{N} \) and \( \omega_0 \). Thus, transfinite induction and recursion can be 
used with ordinals, i.e., first for any element \( a \) of a well-ordered set \( (X, \preceq) \) define 
the initial segment of \( X \) determined by \( a \), i.e., \( I(a) = \{x \in X : x \preceq a, x \neq a\} \),
to have the
0.3. Metrizable Spaces

Recall that a topology $\tau$ on a nonempty set $X$ is a collection (or class or family) of “open” subsets of $X$, such that (1) $X$ and $\emptyset$ are open sets, (2) any union of open sets is an open set, (3) any finite intersection of open sets is an open set. It is not necessary to describe the whole family of open sets $\tau$, usually it suffices to give a base for the topology $\tau$, i.e., a subfamily $\beta \subset \tau$ of open sets such that for any open set $O \in \tau$ and any point $x$ in $O$ there is a member $B \in \beta$ in the base satisfying $x \in B \subset O$. Also, if $\tau_1$ and $\tau_2$ and two topology on $X$ then $\tau_1$ is stronger (or finer) than $\tau_2$, or equivalently, $\tau_2$ is weaker (coarser) than $\tau_1$ if $\tau_1 \subset \tau_2$, i.e., every open set in $(X, \tau_1)$ is an open set in $(X, \tau_2)$. Thus, closed sets are complements of open sets, and the (sequential) convergence (also cluster, interior, boundary, compact, connect, dense, etc) is then defined. Clearly, an abstract space with a topology is called a topological space. Actually, remark that we use only with topological spaces where points are separated closed sets, i.e., Hausdorff spaces, so that these properties are implicitly assumed everywhere (even if it is seldom restated) in the text, unless explicitly mentioned otherwise.

A metric on a set $X$ is a function $d : X \times X \to \mathbb{R}$ satisfying for every $x$ and $y$ in $X$ the following conditions: (a) $d(x, y) \geq 0$ and $d(x, y) = 0$ if and only if $x = y$, (b) $d(x, y) = d(y, x)$ and (c) $d(x, y) \leq d(x, z) + d(z, y)$ for every $z$ in $X$. The couple $(X, d)$ is called a metric space, which becomes a topological space with the open sets defined by means of the (base of) open balls $B(x, r) = \{y \in X : d(x, y) < r\}$, for any $x$ in $X$ and any $r > 0$. On the other hand, a metrizable space is a topological space $X$ in which a metric can be defined (but not really used) so that $(X, d)$ has a topology equivalent to the initial one given on $X$ (where the topology has a simple characterization).

In a metric space, for every $x$ in $X$ the countable family of balls $B(x, 1/n)$, $n = 1, 2, \ldots$, forms a neighborhood-base at $x$ and so, the $(X, d)$ topology is first-countable or it satisfies the first axiom of countability. In a first-countable topology (in particular in a metric space), we can use only convergent sequences to define its topology, i.e., a subset $A$ of $X$ is closed for the topology induced by the metric $d$ if and only if for every sequence $\{a_n : n \geq 1\}$ of points in $A$ such

**Transfinite Induction Principle:** If a subset $A$ of a well-ordered set $X$ satisfies (for every $a$ in $X$) the condition “$I(a) \subset A \Rightarrow a \in A$” then $A$ is indeed the whole set, i.e., $A = X$. 

It is clear that if $a$ is the minimum element in $X \setminus A$ then by definition $I(a) \subset A$ and therefore $a$ belongs to $A$. A neat case is when the well-ordered set $X$ is actually the natural number, i.e., the ordinary mathematical induction. Similarly to the ordinary recursion argument, where a function $f$ on the natural numbers can be defined by specifying $f(0)$ and then defining $f(n)$ in terms of $f(0), \ldots, f(n - 1)$, we have the recursion principle for well-ordered set, e.g., see Dudley [37, Section I.3, pp. 12–15]. Alternatively, the reader may check Folland [45, Chapter 0, pp. 1–17], for a short and clean discussion on the above points. Also see Berberian [12, Chapter 1, pp. 1–85].
that \( d(a_n, a) \to 0 \) as \( n \to \infty \), for some \( a \) in \( X \), we have that the limit point \( a \) belongs to \( A \).

On the other hand, a topological space \( X \) is called \textit{separable} if there exists a countable dense subset \( Q \subset X \), i.e., \( Q \) is countable and its closure \( \overline{Q} \) is the whole space \( X \). Hence, a separable metric space \((X, d)\) is \textit{second-countable} or it satisfies the second axiom of countability, i.e., its contains a countable base, namely, the family of balls \( B(q, 1/n) \) with \( q \) in a countable dense set \( Q \) and \( n \geq 1 \). Actually the converse is also true, i.e., a metric space \((X, d)\) is second-countable if and only if it is separable, and similarly, a topological space with a countable base is separable and metrizable. Moreover any locally compact (or vector topological) space with a countable base is metrizable, but certainly, the converse is false. For instance, the reader may want to take a quick look at the book Kelley [67, Chapter 4, 112–134].

Recall that a topological space is called \textit{sequentially compact} if every sequence admits a convergent subsequence. Any sequentially compact metric space is separable, and for a sequential space (i.e., where convergent sequences are used to define its topology), compactness and sequentially compactness are equivalent.

Given a family \( \{(X_i, d_i) : i \in I\} \) of metric spaces the product space \( X = \prod_{i \in I} X_i \) is a topological space with the product topology which may not be metrizable. For a countable family \( I = \{1, 2, \ldots\} \), we may define the metric
\[
d(x, y) = \sum_{i=1}^{\infty} 2^{-i} \frac{d_i(x_i, y_i)}{1 + d_i(x_i, y_i)}, \quad \forall x = (x_i), y = (y_i),
\]
which induces an equivalent topology on \( X \), i.e., a countable product space \( X = \prod_{i=1}^{\infty} X_i \) is indeed metrizable with the above metric \( d \).

We may consider uniform continuity and Cauchy sequences in a metric space \((X, d)\). Thus, \((X, d)\) is \textit{complete} if any Cauchy sequence has a limit, i.e., if \( d(x_n, x_m) \to 0 \) as \( n, m \to \infty \) then there exists \( x \) such that \( d(x_n, x) \to 0 \) as \( n \to \infty \). If a space \((X, d)\) is not complete then we can completed it in the same way as we pass from the rational number to the real numbers. However, the concept of \textit{completeness} is not a topological property, i.e., on a given space we may have two metrics yielding equivalent topologies but only one of them is complete. Anyway, every compact metric space is complete. A complete separable metrizable (or separable completely metrizable) space is called a \textit{Polish space}, i.e., a separable topological space \( X \) with a metric yielding a complete metric space \((X, d)\). For instance, the space \( C(\mathbb{R}^d) \) of all real-valued continuous functions is a Polish space with the metric
\[
d(f, g) = \sum_{n=1}^{\infty} 2^{-n} \frac{\|f - g\|_n}{1 + \|f - g\|_n}, \quad \|f\|_n = \sup_{|x| \leq n} |f(x)|.
\]
In probability, the sample spaces are Polish spaces, most of the times, we use the space of continuous functions from \( \mathbb{R} \) into \( \mathbb{R}^d \) or the space of all cad-lag functions from \( \mathbb{R} \) into \( \mathbb{R}^d \), i.e., functions continuous from the right and having limits from the left.
A subset $K$ of a metric space $(X,d)$ is called **totally bounded** if for every $\epsilon > 0$ there exists a finite number of points $x_1, \ldots, x_n$ in $K$ such that $K \subseteq \bigcup_{i=1}^{n} B(x_i, \epsilon)$, i.e., any $x$ in $K$ is within a distance $\epsilon$ from the set $\{x_1, \ldots, x_n\}$. It is very instructive (but no simple) to show that a subset $K$ (of a complete metric space) is totally bounded if and only if the closure of $K$ is compact, e.g., Yosida [122, Section 0.2, pp. 13–15].

A **vector topological space** has a topology compatible with the vector structure, i.e., such that the addition and the scalar multiplication (of vectors) are continuous operations (e.g., check the books Kothe [73] or Schaefer [103]). An example is the so-called locally convex spaces, and better, a **normed space** $X$, which is vector space with a **norm**, i.e., a nonnegative function $\| \cdot \|$ defined on $X$ such that (a) $\|\lambda x\| = |\lambda| \|x\|$, for every $x$ in $X$ and $\lambda$ in $\mathbb{R}$, (b) $\|x+y\| \leq \|x\|+\|y\|$, for every $x$ and $y$ in $X$, and (c) $\|x\| \geq 0$ for every $x$ in $X$ and $\|x\| = 0$ only if $x = 0$. Given a norm, we define a metric $d(x, y) = \|x - y\|$ (but not any metric comes from a norm), which yields the topology.

In a normed space, any set that can be covered by a ball is called a **bounded set**. But, only on finite dimensional normed spaces, any closed and bounded set is necessarily compacts. A complete normed space is called a **Banach space**. The space $C_b(X)$ of all real-valued (or complex-valued) bounded continuous functions on a Hausdorff topological space $X$, with the sup-norm

$$\|f\| = \sup \{|f(x)| : x \in X\},$$

is a typical example of an infinite dimensional Banach space.

A topological space $X$ is said to be **locally compact** if every point has a compact neighborhood, i.e., an open set with compact closure. This implies that for every point $x$ in $X$ and any open set $U$ containing $x$ there is another open set $V$ containing $x$ such that $V \subset U$ and the closure $\overline{V}$ is compact. A locally compact Banach space is necessarily a space of finite dimension (i.e., homeomorphic to some $\mathbb{R}^d$, $d \geq 1$).

Again, better than a norm is an **inner or scalar product**, i.e., a bilinear maps $(\cdot, \cdot)$ from $X \times X$ into $\mathbb{R}$ satisfying (a) $(\lambda x+y, z) = \lambda(x, z)+(y, z)$, for every $x, y$ in $X$ and $\lambda$ in $\mathbb{R}$, (b) $(x, y) = (y, x)$, for every $x, y$ in $X$, and (c) $(x, x) \geq 0$ for every $x$ in $X$ and $(x, x) = 0$ only if $x = 0$. From an inner product we can define a norm $\|x\| = \sqrt{(x,x)}$, indeed, by considering the discriminant of the positive quadratic form $\lambda \mapsto (x+\lambda y, x+\lambda)$ we obtain the Cauchy inequality $|(x, y)| \leq \|x\| \|y\|$, for every $x$ and $y$, which yields the triangular inequality (b) for the norm. Certainly, not any norm comes from an inner product, indeed, a norm is derived form an inner product if and only if the parallelogram law is satisfied, i.e.,

$$\|x+y\|^2+\|x-y\|^2 = 2\|x\|^2+2\|y\|^2,$$

for every $x$ and $y$ in $X$, in which case the inner product is defined by the polarization identity $\|x+y\|^2 = \|x\|^2+\|y\|^2+2\Re(x, y)$. A complete normed space where the norm comes from an inner product is called a **Hilbert space**.

A typical infinite dimensional Hilbert space is $\ell^2$, the space of real-valued sequences $a = \{a_n\}$ satisfying $\sum_n a_n^2 < \infty$, with the inner product $(a, b) = \sum_n a_n b_n$. A more elaborate example is space $L^2(K)$, with $K$ a compact subset.
of $\mathbb{R}^d$, which is the completion of $C_b(K) = C(K)$, space of continuous functions, with the inner product

$$(f,g) = \int_K f(x)g(x)\,dx.$$ 

By means of the theory of the integral we are able to study in great detail spaces similar to this one.

Moreover, we may have complex Banach and Hilbert spaces, i.e., the vector space is on the complex field $\mathbb{C}$, and $|\lambda|$ denotes the modulus (instead of the absolute value) when $\lambda$ belongs to $\mathbb{C}$ for the condition (a) of norm. In the case of a inner product, the condition (b) becomes $(x,y) = \overline{(y,x)}$, where the over-line means complex conjugate, i.e., the application $(\cdot,\cdot)$ is sesquilinear (instead or bilinear) with complex values.

As seen later, general discussions of good measures are focus on a separable complete metrizable space (i.e., a Polish space), while general discussions of (nonlinear) functions are considered on locally compact spaces with a countable bases. As expected, the typical oversimplified example is $\mathbb{R}^n$.

For instance, the reader may take a look at DiBenedetto[32, Chapters 1 and 2, pp. 1–64] or Hewitt and Stromberg [62, Chapter 2, pp. 53–103] or Royden [98, Chapters 1 and 2, pp. 1–53] or Rudin [101, Chapter 1, pp. 3–40], for a quick review on topology and continuous functions. Also most of Brown and Pearcy [21], and Pugh [91, Chapter 2, pp. 51–115] are a suitable initial reading.

### 0.4 Some Basic Lemmas

There are a couple of very useful results, but for our treatment this may be considered “on the side”, such as

**Lemma 0.1** (Urysohn). *Let $A$ and $B$ be two nonempty, disjoint and closed sets in a metric space $(X,d)$. If $a$ and $b$ are two distinct real numbers then the function

$$x \mapsto g(x) = \frac{a + (b - a)d(x,A)}{d(x,A) + d(x,B)}$$

is continuous and $g(x) = a$ for any $x$ in $A$ and $g(x) = b$ for any $x$ in $B$.***

**Proof.** The distance from a point to a set is defined as $d(x,A) = \inf\{d(x,y) : y \in A\}$ and the triangular inequality yields that $|d(x,A) - d(y,A)| \leq d(x,y)$, which proves that the above function $g$ is continuous. \hfill $\Box$

**Proposition 0.2** (Tietze’s Extension). *Let $f$ be a bounded real-valued function defined on a closed subset $C$ of a metric space $(X,d)$. Then there exists a continuous extension $g$ of $f$ to the entire space $X$.***
Proof. A proof of Tietze’s Extension is based on the following construction, which shows that an extension \( g \) is uniform limit in \( X \) of the series \( \sum_k g_k \) of continuous functions.

First, with \( a = \sup_{C} |f(x)| \) define \( A = \{ x \in C : f(x) \leq \frac{-a}{3} \} \) and \( B = \{ x \in C : f(x) \geq \frac{a}{3} \} \) to find a continuous function \( g_1 : X \to [-\frac{a}{3}, a/3] \) satisfying \( |f(x) - g_1(x)| \leq 2a/3 \), for every \( x \) in \( C \).

Next with \( f_1 = f - g_1 \), \( a_1 = 2a/3 \) define \( A_1 = \{ x \in C : f_1(x) \leq \frac{-a_1}{3} \} \) and \( B_1 = \{ x \in C : f_1(x) \geq \frac{a_1}{3} \} \) to find a continuous function \( g_2 : X \to [-\frac{a_1}{3}, \frac{a_1}{3}] \) satisfying \( |f_1(x) - g_2(x)| \leq 2a_1/3 \), for every \( x \) in \( C \).

Finally, by induction it follows that \( |f(x) - \sum_{k=1}^{n} g_k(x)| \leq 2^n a 3^{-n} \), for every \( x \) in \( C \) and that \( |g_n(x)| \leq 2^{n-1} a 3^{-n} \), for any \( x \).

\[ \square \]

Statements in Lemma 0.1 and Proposition 0.2 are also valid for more general topological spaces, e.g., Hausdorff locally compact spaces and normal spaces.

Another key result frequently used, is Weierstrauss approximation theorem,

**Theorem 0.3 (Weierstrauss).** If \( f \) is a real-valued continuous function on the compact interval \([a, b]\) then there exits a sequence of polynomials \( \{p_n\} \) such that \( p_n(x) \to f(x) \) uniformly on \([a, b]\).

**Proof.** For a given real-valued continuous function \( f \) on \([a, b]\) the change of variable \( y = (x-a)/(b-a) \) reduces the discussion to the case where the compact interval \([a, b]\) is \([0, 1]\). Moreover, for \( f \) on \([0, 1]\) the change of function \( g(x) = [f(x) - f(0)] - x[f(1) - f(0)] \) yields \( g(0) = g(1) = 1 \). Briefly, it should be proven that for any real-valued continuous function on the real line \( \mathbb{R} \) which vanishes outside the interval \([0, 1]\) there exists a sequence of polynomials \( \{p_n\} \) such that \( p_n(x) \to f(x) \) uniformly on \([0, 1]\). Note that \( f \) is uniformly continuous on \( \mathbb{R} \).

Consider the polynomials \( q_n(x) = c_n(1 - x^2)^n, \ n = 1, 2, \ldots \), where the coefficients \( c_n \) are chosen so that

\[ c_n \int_{-1}^{1} (1 - x^2)^n \, dx = 1, \quad \forall n \geq 1. \]

Remark that the function \((1 - x^2)^n - 1 + nx^2\) vanishes at \( x = 0 \) and has derivative positive on the open interval \((0, 1)\) to deduce that

\[ (1 - x^2)^n \geq 1 - nx^2, \quad \forall x \in [0, 1]. \]

Hence, the calculation

\[ \int_{-1}^{1} (1 - x^2)^n \, dx = 2 \int_{0}^{1} (1 - x^2)^n \, dx \geq 2 \int_{0}^{1/\sqrt{n}} (1 - x^2)^n \, dx \geq 2 \int_{0}^{1/\sqrt{n}} (1 - nx^2) \, dx > \frac{1}{\sqrt{n}} \]

implies that \( 0 \leq c_n \leq 1/\sqrt{n} \), for every \( n \geq 1 \). Therefore, for any \( \delta > 0 \), the following estimate

\[ 0 \leq q_n(x) \leq \sqrt{n}(1 - \delta^2)^n, \quad \forall x \in [-1, 1], \text{ with } |x| \geq \delta, \]
holds true.

Next, define the polynomial

$$p_n(x) = \int_{-1}^{1} f(x+y)q_n(y)dy = \int_{0}^{1} f(z)q_n(z-x)dy, \quad \forall x \in [-1, 1].$$

By continuity, for every $\varepsilon > 0$ there exists $\delta > 0$ such that $|x - y| < \delta$ implies $|f(x) - f(y)| < \varepsilon$ and also, $|f(x)| \leq C < \infty$ for every $x$ in $[0, 1]$. Thus

$$|p_n(x) - f(x)| \leq \int_{-1}^{1} |f(x+y) - f(x)|q_n(y)dy \leq$$

$$\leq 2C \int_{-1}^{-\delta} q_n(y)dy + \varepsilon \int_{-\delta}^{\delta} q_n(y)dy + 2C \int_{\delta}^{1} q_n(y)dy \leq$$

$$\leq 4C \sqrt{n}(1 - \delta^2)^n + \varepsilon,$$

which shows that $p_n(x) \to f(x)$ uniformly on $[0, 1]$. \hfill \square

Revising the above proof, it should be clear that the same argument is valid when the function $f$ takes complex values and then the coefficients of the polynomials $p_n$ are complex.

• Remark 0.4. A vector space $F$ of real-valued functions defined a nonempty set $X$, is called an algebra if for any $f$ and $g$ in $F$ the product function $fg$ belongs also to $F$. A family $F$ of real-valued functions is said to separate points in $X$ if for every $x \neq y$ in $X$ there exists a function $f$ in $F$ such that $f(x) \neq f(y)$. If $X$ is a Hausdorff topological space then $C(X)$ the space of all continuous real-valued functions defined on $X$ is an algebra that separate points. The so-called Stone-Weierstrass Theorem states that if $K$ is a compact Hausdorff space and $F$ is an algebra of functions in $C(K)$ which separate points and contains constants functions, then for every $\varepsilon > 0$ and any $g$ in $C(K)$ there exists $f$ in $F$ such that

$$\sup \{|f(x) - g(x)| : x \in X\} = \|f - g\| < \varepsilon,$$

i.e., $F$ is dense in $C(K)$ for the sup-norm $\|\cdot\|$. Moreover, $K$ is stable under the complex-conjugate operation then the same result is valid for complex-valued functions, e.g., check the books DiBenedetto [32, Sections IV.16–18, pp. 199–203], Rudin [99, Chapter 7, 159–165] or Yosida [122, Section 0.2, pp. 8–11]. \hfill \square

A typical application of Zorn’s Lemma is the following:

**Lemma 0.5.** Any linear vector space $X$ contains a subset $\{x_i : i \in I\}$, so-called Hamel basis for $X$, of linear independent elements such that the linear subspace spanned by $\{x_i : i \in I\}$ coincides with $X$.

**Proof.** Consider the partially ordered set $S$ whose elements are all the subsets of linearly independent elements in $X$, with the partial order given by the inclusion. If $\{A_\alpha\}$ is a chain or totally ordered subset of $S$ then $A = \bigcup_\alpha A_\alpha$ is an upper bound, since any finite number of element in $A$ are linearly independent. Hence,
Zorn’s Lemma implies the existence of a maximal element, denoted by \( \{ x_i : i \in I \} \). Now, for any \( x \) in \( X \) the set \( \{ x_i : i \in I \} \cup \{ x \} \) has to be linearly dependent and so, \( x \) is linear combination of some finite number of elements in \( \{ x_i : i \in I \} \), as desired.

For instance, the interested reader may check the books by Berberian [12], DiBenedetto [32], Dshalalow [36], Dudley [37], Hewitt and Stromberg [62], Royden [98], among many others. On the other hand, the reader may take a look (again) at the books by Apostol [4], Dieudonné [33] or Rudin [99]. Essentially, as a background for what follows, the reader should be familiar with most of the basic material covered in Taylor [114, Chapters 1-3, pp. 1–176].
Chapter 1

Measurable Spaces

Discrete measures were briefly presented in the introduction, where we were able
to measure any subset of the given (possible uncountable) space $\Omega$, essentially,
by counting its elements. However, in general (with non-discrete measures),
due to the requirement of $\sigma$-additivity and some axioms of the set theory, we
cannot use (or measure) every element in $2^{\Omega}$. Therefore, a first point to consider
is classes (or family or collection or systems) of subsets (of a given space $\Omega$)
suitable for our analysis. We begin with a general discussion on infinite sets
and some basic facts about functions.

1.1 Classes of Sets

Let $\Omega$ be a nonempty set and $2^{\Omega}$ be the parts of $\Omega$, i.e., set of all subsets of $\Omega$.
Clearly, if $\Omega$ has $n$ elements then $2^{\Omega}$ has $2^n$ elements, but our interest is when $\Omega$
has an infinite number of elements, for instance if $\Omega$ is countable infinite (i.e., it
is in a one-to-one relation with the positive integers) then $2^{\Omega}$ has the cardinality
of the continuum. A class (collection or family or system) of sets is a subset of
$2^{\Omega}$, that by convenience, we assume it contains the empty set $\emptyset$. Note that $\emptyset \subset \Omega$
and $\emptyset, \Omega \in 2^{\Omega}$. Typical operations between two elements $A$ and $B$ in $2^{\Omega}$ are the
intersection $A \cap B$, the union $A \cup B$, the difference $A \setminus B$ and the complement
$A^c = \Omega \setminus A$. The union and the intersection can be extended to any number
of sets, e.g., if $A_i \in 2^{\Omega}$ for $i$ in some sets of indexes $I$ then we have $\bigcup_{i \in I} A_i$
and $\bigcap_{i \in I} A_i$. Sometimes, to simplify notation we write $A + B$ or $\sum_{i \in I} A_i$ (for
disjoint unions) to express the fact that $A + B = A \cup B$ with $A \cap B = \emptyset$ or
$\sum_{i \in I} A_i = \bigcup_{i \in I} A_i$ with $A_i \cap A_j = \emptyset$ if $i \neq j$.

Below we introduce a number of elementary (or intermediary) classes, which
are used later in the following Chapters, when the extension questions are ad-
dressed.

Definition 1.1. Given classes $\mathcal{P}$, $\mathcal{L}$, $\mathcal{R}$ and $A$ of subsets of $\Omega$, each containing
$\emptyset$, we say that

$\bullet$ $\mathcal{P}$ is a $\pi$-class if $A, B \in \mathcal{P}$ implies $A \cap B \in \mathcal{P}$,
• \( \mathcal{L} \) is a \( \ell \)-class (or additive class) if (a) \( A, B \in \mathcal{L} \) with \( A \cap B = \emptyset \) implies \( A \cup B \in \mathcal{L} \) and (b) \( A, B \in \mathcal{L} \) with \( A \subset B \) implies \( B \setminus A \in \mathcal{L} \).

• \( \mathcal{R} \) is a ring if \( A, B \in \mathcal{R} \) implies (a) \( A \setminus B \in \mathcal{R} \) and (b) \( A \cup B \in \mathcal{R} \).

• \( \mathcal{A} \) is algebra if (a) \( A \in \mathcal{A} \) implies \( A^c \in \mathcal{A} \) and (b) \( A, B \in \mathcal{A} \) implies \( A \cup B \in \mathcal{A} \).

Finally, a \( \pi \)-class \( \mathcal{S} \) is called (1) a semi-ring if \( A, B \in \mathcal{S} \) with \( A \subset B \) implies \( B \setminus A = \sum_{i=1}^{n} C_i \) with \( C_i \in \mathcal{S} \), (2) a semi-algebra if \( A \in \mathcal{S} \) implies \( A^c = \sum_{i=1}^{n} C_i \) with \( C_i \in \mathcal{S} \), and (3) a lattice if \( A, B \in \mathcal{S} \) implies \( A \cup B \in \mathcal{S} \).

Usually, a field is defined as an algebra, but the name \( \ell \)-class or additive class is not completely standard in the literature. By induction, a \( \pi \)-class is stable under the formation of finite intersections, and a lattice is also stable under the formation of finite unions. Similarly, \( \ell \)-classes are stable under the formation of monotone differences, rings are stable under the formation of all differences, and algebras are stable under the formation of complements. Since \( A \cap B = [A \cup B] \setminus [(A \setminus B) \cup (B \setminus A)] \), \( A \cap B = (A^c \cup B^c)^c \) and \( A \setminus B = (A^c \cup B)^c \), rings and algebras are also stable under finite intersections, and stable under the formation of complements implies stable under differences. Thus, any algebra is a ring, and every ring is simultaneously a \( \pi \)-class, an \( \ell \)-class, and a lattice. The equality \( A \cup B = ((A \setminus (A \cap B)) \cup B) \) implies that any \( \ell \)-class which is also a \( \pi \)-class is necessarily a ring. Also, a ring containing the whole space is indeed an algebra, and a lattice is not necessarily an \( \ell \)-class, e.g., the class \( \{\emptyset, F, \emptyset\} \) (with \( F \) not empty and different from \( \Omega \)) is a lattice but not an \( \ell \)-class.

From the definitions, it is clear that any interception of \( \pi \)-classes, \( \ell \)-classes, lattices, rings or algebras is again a \( \pi \)-class, an \( \ell \)-class, a lattice, a ring or an algebra. Therefore, given any subset \( \mathcal{G} \) of \( 2^\Omega \) we may define the \( \pi \)-class, \( \ell \)-class, lattice, ring or algebra generated by \( \mathcal{G} \), e.g., the algebra \( \mathcal{A}(\mathcal{G}) \) generated by \( \mathcal{G} \) is indeed the intersection of all algebras containing \( \mathcal{G} \).

For instance, if \( A \) and \( B \) are subsets of \( \Omega \) then the minimal classes \( \mathcal{P}, \mathcal{L}, \mathcal{R} \) and \( \mathcal{A} \) containing \( A \) and \( B \) (as the name suggests) are \( \mathcal{P} = \{\emptyset, A, B, A \cap B\} \), \( \mathcal{L} = \{\emptyset, A, B, A \setminus B \text{ if } B \subset A, B \setminus A \text{ if } A \subset B, A \cup B \text{ if } A \cap B = \emptyset\} \), \( \mathcal{R} = \{\emptyset, A, B, A \cap B, A \setminus B, B \setminus A, A \cup B, (A \setminus B) \cup (B \setminus A)\} \), and \( \mathcal{A} = \{\emptyset, A, B, A \cap B, A^c, B^c, A^c \cap B^c, A^c \cup B^c, A^c \cap B, A^c \cup B, A^c \cup B^c, A^c \cap B^c, A^c \cup B^c, \Omega\} \). Certainly, interesting cases are when an infinite number of subsets of \( \Omega \) are involved.

**Exercise 1.1.** Prove that the algebra \( \mathcal{A} \) (ring) generated by a \( \mathcal{S} \) semi-algebra (semi-ring) is the class of finite disjoint unions of sets in \( \mathcal{S} \), i.e., \( A \in \mathcal{A} \) if and only if \( A = \sum_{i=1}^{n} A_i \) with \( A_i \in \mathcal{S} \). Hint: prove first that the class of finite disjoint unions of sets in \( \mathcal{S} \) is stable under the formation of finite intersections.

Similarly, remark that lattice \( \mathcal{L} \) generated by a \( \pi \)-class \( \mathcal{P} \) is the class of finite unions of sets in \( \mathcal{P} \), i.e., \( A \in \mathcal{L} \) if and only if \( A = \bigcup_{i=1}^{n} A_i \) with \( A_i \in \mathcal{L} \). As seen later, a lattice of interest is the class \( \mathcal{L} \subset 2^\mathbb{R} \) of all finite unions of closed intervals, while a semi-ring of interest for us is the class \( \mathcal{S} \) of intervals of the form \((a, b] \), with \( a, b \) real numbers, where the previous Exercise 1.1 can be applied.
For instance, an carefully discussion on semi-rings can be found in Dudley [37, Section 3.2, pp. 94–101].

A convenient way of dealing with class $K \subset 2^\Omega$ is to identify a subset $A$ with the function $1_A : \Omega \rightarrow \{0, 1\}$ defined by $1_A(x) = 1$ if $x \in A$ and $1_A(x) = 0$ if $x \in \Omega \setminus A$. Thus, intersection corresponds to min (or multiplication), i.e., $1_A \cap 1_B = \min\{1_A, 1_B\} = 1_A 1_B$, while union corresponds to plus or rather max, i.e., $1_A \cup 1_B = \max\{1_A, 1_B\}$ and $1_{A \cup (B \setminus A)} = 1_A + 1_B \setminus 1_A$. In this context, the class $2^\Omega$ is identify to the space $X$ of function $1_A$ as above, which is an algebra with the addition defined as max and the multiplication defined as min, i.e., $(X, \max, \min)$ or $(\Omega, \cup, \cap)$ is so-called a Boole algebra. There are other alternatives, i.e., the symmetric difference could be used instead of max as the addition, e.g., see Berberian [11, Chapter 1], among others. For instance, an algebra class is a Boole sub-algebra, a $\pi$-class is a Boole multiplicative semi-group, and a lattice is also a Boole additive semi-group.

We begin with the following

**Proposition 1.2.** Let $K = \ell(\mathcal{P})$ be the smallest $\ell$-class containing a given $\pi$-class $\mathcal{P}$. Then $K$ is also the ring generated by $\mathcal{P}$. Moreover, if $\Omega \in K$ then $K$ is the smallest algebra containing $\mathcal{P}$.

**Proof.** For every $K \in K$ define the class of sets $\Phi_K = \{ A \in K : A \cap K \in K \}$. Clearly, (a) $A \in \Phi_K$ if and only if $K \in \Phi_A$, and (b) the relations $(B \setminus A) \cap K = (B \cap K) \setminus (A \cap K)$ and $(A \cup B) \cap K = (A \cap K) \cup (B \cap K)$ imply that $\Phi_K$ is an $\ell$-class for any fixed $K$.

In particular, if $K = P \in \mathcal{P}$ then $A \in \mathcal{P}$ implies $A \in \Phi_P$. Thus $\mathcal{P} \subseteq \Phi_P$ and because $K$ is the smallest $\ell$-class containing $\mathcal{P}$ we have $K \subseteq \Phi_P$. This is $K \in K$ implies $K \in \Phi_P$, or equivalently $P \in \Phi_K$, for every $P \in \mathcal{P}$. Hence $\mathcal{P} \subseteq \Phi_K$ and again, because $K$ is the smallest $\ell$-class containing $\mathcal{P}$ we have $K \subseteq \Phi_K$, but this time for every $K \in K$. This proves that for any $A, K \in K$ we have $A \cap K \in K$, i.e., $K$ is stable under finite intersections.

Finally, we conclude by making use of the relations $A \setminus B = A \setminus (A \cap B)$,

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

and the fact that a ring is an algebra if and only if it contains $\Omega$. $\square$

Usually, adding the prefix ‘$\sigma$-’ to a class of sets, we require ‘countable’ instead of ‘finite’ in number of operations allowed, e.g.,

**Definition 1.3.** A $\sigma$-algebra (or $\sigma$-field) $\mathcal{A}$ is a class containing $\emptyset$ which is stable under the (formation of) complements and countable unions, i.e., (a) if $A \in \mathcal{A}$ then $A^c \in \mathcal{A}$ and (b) if $A_i \in \mathcal{A}$, $i = 1, 2, \ldots$ then $\bigcup_{i=1}^\infty A_i \in \mathcal{A}$. Similarly, a $\sigma$-ring $\mathcal{A}$ is a non-empty class stable under differences and countable unions, i.e., (c) if $A, B \in \mathcal{R}$ then $A \setminus B \in \mathcal{R}$ and (b) as above. $\square$

A $\pi$-class stable under countable unions could be called as a $\sigma$-lattice, i.e., it could also be called a ‘countable topology’ (if it contains the whole space) since this means a class containing the empty set, stable under finite intersections and under countable unions. Similarly, sometimes a $\pi$-class stable under countable unions and countable intersections could be necessary, e.g., the minimal class of...
sets containing all open sets (or closed sets, of a given topology) and being stable under countable intersections. However, the interesting concepts of monotone class, and $\lambda$-class (or $\sigma$-additive class) are defined and discussed below.

The classes mostly used are the $\sigma$-algebras. Certainly, any ring (or algebra) with a finite number of elements is a $\sigma$-ring (or $\sigma$-algebra). Remark that if $K$ is a given arbitrary class, since the class $R$ (or $R_\sigma$) of all sets that are contained in a finite (or countable) union of sets in $K$ is a ring (or $\sigma$-ring), then $R$ (or $R_\sigma$) contains the ring (or $\sigma$-ring) generated by $K$.

**Exercise 1.2.** First show that any $\sigma$-algebra is a $\sigma$-ring and that a $\sigma$-ring is stable under the formation of countable intersections. Next, prove that an algebra $A$ (a ring $R$) is a $\sigma$-algebra (a $\sigma$-ring) if and only if $A$ ($R$) is stable under the formation countable increasing unions.

It is relatively simple to generate a $\pi$-class, an $\ell$-class, a ring or an algebra. For a given class $K$ we define $P_0 = L_0 = R_0 = K \cup \{\emptyset\}$, $A_0 = K \cup \{\emptyset, \Omega\}$, and by induction $P_i = \{A \cap B : A, B \in P_{i-1}\}$, $L_i = \{A \cup B \mid A \cap B = \emptyset$, and $A \setminus B$ if $B \subseteq A : A, B \in L_{i-1}\}$, $R_i = \{A \cup B, A \setminus B : A, B \in R_{i-1}\}$, and $A = \{A \cup B, A^c : A, B \in A_{i-1}\}$, for $i \geq 1$. Thus, the classes $P = \bigcup_{i=0}^\infty P_i$, $L = \bigcup_{i=0}^\infty L_i$, $R = \bigcup_{i=0}^\infty R_i$ and $A = \bigcup_{i=0}^\infty A_i$ are the smallest $\pi$-class, $\ell$-class, ring or algebra containing the class $K$. Essentially, the same arguments used with ordinal and transfinite induction yield the corresponding $\sigma$-classes, but we prefer to avoid these type of procedures.

It is clear that for a class with a finite number of elements (and only in this case), there is not difference between being either a $\sigma$-algebra (or $\sigma$-ring) and an algebra (or ring). The concept of monotone classes helps to clarify the distinction between algebras (or rings) and $\sigma$-algebras (or $\sigma$-rings). A monotone class (of subset of $\Omega$) is a subset $M$ of $2^\Omega$ stable under countable monotone unions and intersections, i.e., (a) $A_i \in M$, $A_i \subset A_{i+1}$, $i = 1, 2, \ldots$ then $\bigcup_{i=1}^\infty A_i \in M$ and (b) $A_i \in M$, $A_i \supset A_{i+1}$, $i = 1, 2, \ldots$ then $\bigcap_{i=1}^\infty A_i \in M$.

**Proposition 1.4** (monotone class). Let $K = M(R)$ be the smallest monotone class containing a given ring $R$. Then $K$ is also the $\sigma$-ring generated by $R$. Moreover, if $\Omega \in K$ then $K$ is the smallest $\sigma$-algebra containing $R$.

**Proof.** Most of the arguments are similar to those of Proposition 1.2.

For every $K \in K$ define the class of sets $\Phi_K = \{A \in K : A \setminus K, K \setminus A, A \cup K \in K\}$. Clearly, (a) $A \in \Phi_K$ if and only if $K \in \Phi_A$, and (b) the relations $(\bigcup_i A_i) \setminus K = \bigcup_i (A_i \setminus K)$, $(\bigcap_i A_i) \setminus K = \bigcap_i (A_i \setminus K)$, $K \setminus (\bigcup_i A_i) = \bigcup_i (K \setminus A_i)$, $K \setminus (\bigcap_i A_i) = \bigcap_i (K \setminus A_i)$, $(\bigcup_i A_i) \cup K = \bigcup_i (A_i \cup K)$ and $(\bigcap_i A_i) \cup K = \bigcap_i (A_i \cup K)$ imply that $\Phi_K$ is a monotone class for any fixed $K$.

In particular, if $K = R \in R$ then $A \in R$ implies $A \in \Phi_R$. Thus $R \subset \Phi_R$ and because $K$ is the smallest monotone class containing $R$ we have $K \subset \Phi_R$. This is $K \in K$ implies $K \in \Phi_R$, or equivalently $R \in \Phi_K$, for every $R \in R$. Hence $R \subset \Phi_K$ and again, because $K$ is the smallest monotone class containing $R$ we have $K \subset \Phi_R$, but this time for every $K \in K$. This proves that for any $A, K \in K$ we have $A \setminus K, K \setminus A, A \cup K \in K$, i.e., $K$ is a ring.
Finally, we conclude by noting that a \( \sigma \)-ring is a \( \sigma \)-algebra if and only if it contains \( \Omega \).

A common used of above Proposition 1.4 is in the form of a so-called **monotone argument** as follows: if a class of set \( \mathcal{R} \) is a ring and each element enjoys a certain property, and this property stable by countable monotone unions and countable monotone intersections, then this property holds true for the \( \sigma \)-ring generated by \( \mathcal{R} \).

- **Remark 1.5.** Combining Propositions 1.2 and 1.4, it is clear that if \( P \) is a \( \pi \)-class then \( M(P) \) is the smallest \( \sigma \)-algebra containing \( P \).

The notation \( \sigma(K) \) means the smallest \( \sigma \)-algebra containing a given class \( K \), or the \( \sigma \)-algebra generated by \( K \). It is clear that if \( K \) is finite then \( \sigma(K) \) is also finite.

**Exercise 1.3.** Since only countable operations are involved, we can convince oneself that if \( K \) has the cardinality of the continuum (or greater) then \( \sigma(K) \) preserves its cardinality. Make a formal argument to show the validity of the previous statement, e.g., see Rama [92, Section 4.5, pp. 110-112] where transfinite induction is used.

Let \( \mathcal{R} \) be the union of all \( \sigma \)-rings \( \mathcal{R}(\mathcal{E}_c) \) generated by a countable subclass \( \mathcal{E}_c \) of a given a class \( \mathcal{E} \) in \( 2^\Omega \) containing the empty set. Since \( \mathcal{R} \) is indeed a \( \sigma \)-ring, we have \( \mathcal{R} = \mathcal{R}(\mathcal{E}) \), the \( \sigma \)-ring generated by the whole class \( \mathcal{E} \). Thus, for a given \( A \) in \( \mathcal{R}(\mathcal{E}) \) there exists a countable subclass \( \mathcal{E}_c \) (depending on \( A \)) such that \( A \) belongs to \( \mathcal{R}(\mathcal{E}_c) \). If we can keep the same countable subclass for every set \( A \) then the \( \sigma \)-ring \( \mathcal{R}(\mathcal{E}) \) is called separable. Moreover, we say that a \( \sigma \)-algebra \( \mathcal{F} \) is **countable generated** or **separable** if there exists a countable class \( \mathcal{K} \) such that \( \mathcal{F} = \sigma(\mathcal{K}) \).

Frequently, the previous Propositions are combined in the so-called **argument of monotone class** as follows. A **\( \lambda \)-class** (or \( \sigma \)-additive class) is a subset \( \mathcal{D} \) of \( 2^\Omega \) stable under the formation of countable monotone unions, monotone differences and it contains \( \Omega \), i.e., (a) \( A_i \in \mathcal{D}, A_i \subset A_{i+1}, i = 1,2,\ldots \) then \( \bigcup_{i=1}^{\infty} A_i \in \mathcal{D} \), (b) if \( A, B \in \mathcal{D} \) with \( A \subset B \) then \( B \setminus A \in \mathcal{D} \) and (c) \( \Omega \in \mathcal{D} \). From the equality \( A + B = (A^c \setminus B)^c \) we deduce that a \( \lambda \)-class is stable under the formation of countable disjoint unions.

**Proposition 1.6** (**monotone argument**). Let \( \mathcal{D} \) be a \( \lambda \)-class and \( \mathcal{P} \) be a \( \pi \)-class. Then \( \mathcal{D} \) is a \( \sigma \)-algebra if and only if \( \mathcal{D} \) is also stable under finite intersections. Moreover, if \( \mathcal{P} \subset \mathcal{D} \) then \( \sigma(\mathcal{P}) \subset \mathcal{D} \).

**Proof.** To verify the first part, because \( \Omega \in \mathcal{D} \) we remark that \( \mathcal{D} \) is stable under complement. Next, we note that any countable union \( A = \cup_i A_i \) can be expressed as \( A = \cup_i B_i \), with \( B_n = \bigcup_{i=1}^{n} A_i = (\bigcap_{i=1}^{n} A_i^c)^c \) which satisfy \( B_i \subset B_{i+1} \), for every \( i \). So, if \( \mathcal{D} \) is also a \( \pi \)-class then \( B_n \) belongs to \( \mathcal{D} \), and \( \mathcal{D} \) is stable under countable union, i.e., a \( \mathcal{D} \) is indeed a \( \sigma \)-algebra.

For the second part, if \( \lambda(\mathcal{P}) \) denotes the smallest \( \lambda \)-class containing \( \mathcal{P} \) then, for every \( E \in \lambda(\mathcal{P}) \), define the class of sets \( \Phi_E = \{ A \in \lambda(\mathcal{P}) : A \cap E \in \lambda(\mathcal{P}) \} \).
An argument similar to those of Propositions 1.2 and 1.4 proves that $\Phi_E = \lambda(\mathcal{P})$ is a $\sigma$-algebra, i.e., $\lambda(\mathcal{P}) = \sigma(\mathcal{P})$. Therefore $\sigma(\mathcal{P}) \subset \mathcal{D}$. \hfill \qed

In several circumstances, we have a family $\mathcal{K}$ of subsets of $\Omega$ which contains a certain $\pi$-class $\mathcal{P}$ and has a prescribed property. If this prescribed property is true for the whole space $\Omega$ and is stable under monotone differences and countable monotone unions then every set in the $\sigma$-algebra generated by the $\pi$-class $\mathcal{P}$ has this property, i.e., $\sigma(\mathcal{P}) \subset \mathcal{K}$.

Exercise 1.4. A subset $\mathcal{D}$ of $2^{\Omega}$ containing the empty set is called a Dynkin class if $\mathcal{D}$ is stable under the formation of complements and countable disjoint unions. Prove that a Dynkin class $\mathcal{D}$ is also a $\lambda$-class. How about the converse? (e.g., see Bauer [9, Section I.2]). \hfill \qed

Exercise 1.5. Given a family $\{F_{i,j} : i \in I_j, j \in J\}$ of subsets of $\Omega$, verify the distributive formula

$$
\bigcup_{j \in J} \bigcap_{i \in I_j} F_{i,j} = \bigcap_{k \in \mathcal{K}} \bigcup_{j \in J} F_{i,j}^k \quad \text{and} \quad \bigcap_{j \in J} \bigcup_{i \in I_j} F_{i,j} = \bigcup_{k \in \mathcal{K}} \bigcap_{j \in J} F_{i,j}^k,
$$

where $K = \prod_{j \in J} I_j$, i.e., $\{i_j : j \in J\}$, and $F_{i,j}^k = F_{i,j}$. It is clear that if $J$ is finite and each $I_j$ is countable then $K$ is a countable set, however, if for instance, $I_j = \{0, 1\}$ for every $j$ in an infinite set of indexes $J$ then $K = \{0, 1\}^J$ is not a countable set of indexes. \hfill \qed

\begin{itemize}
  \item Remark 1.7. Recalling that $\bigcup$ denotes disjoint union of sets, for a given semi-ring (or ring or algebra) $\mathcal{E} \subset 2^X$, we consider the class $\mathcal{F} = \{ \sum_{i=1}^{\infty} E_k : E_k \in \mathcal{E} \}$ of subsets in $2^X$. First, if $A = \bigcup_{i=1}^{\infty} A_i$ with $A_i \subset A_{i+1}$ and $A_i$ in $\mathcal{E}$ then $A = \bigcup_{i=1}^{\infty} B_i$, with $B_1 = A_1$, $B_2 = A_2 \setminus B_1$, … , $B_n = A_n \setminus B_{n-1}$, and because $\mathcal{E}$ is a semi-ring, each $B_i$ is a finite disjoint union of elements of $\mathcal{E}$, i.e., $A$ is a countable disjoint union of sets in $\mathcal{E}$, which proves that $\mathcal{F} = \{ \bigcup_{k=1}^{\infty} E_k : E_k \in \mathcal{E} \}$. Now, if $F_j = \bigcup_i E_{i,j}$ then $F = \bigcup_j F_j = \bigcup_{i,j} E_{i,j}$ is a countable union of sets in $\mathcal{E}$ and therefore, $F$ belongs to $\mathcal{F}$, i.e., $\mathcal{F}$ is stable under countable unions. However, the distributive formula of Exercise 1.5 can only be used to show that $\mathcal{F}$ is stable under finite intersections, since $\bigcap_{j=1}^{\infty} \sum_{i=1}^{\infty} E_{i,j} = \sum_{k \in \mathcal{K}} \bigcap_{i,j} E_k$, where $K = I^J$, $E_k = E_{i,j}$, but $K$ is not a countable set of indexes. Thus, if $A = \bigcup_{i=1}^{\infty} A_i$ and $B = \bigcup_{j=1}^{\infty} B_j$ with $A_i$ and $B_j$ in $\mathcal{E}$ $A \setminus B = \bigcup_{i=1}^{\infty} \bigcap_{j=1}^{\infty} (A_i \setminus B_j)$, where each difference $A_i \setminus B_j$ is a finite disjoint union of elements in $\mathcal{E}$, but $\bigcap_{j=1}^{\infty} (A_i \setminus B_j)$ is not necessarily in $\mathcal{F}$, i.e., $\mathcal{F}$ may not be stable neither under countable intersection nor under differences, check the arguments in Exercise 1.1. Therefore $\mathcal{F}$, which is stable under countable unions and finite intersections, may be strictly smaller than the $\sigma$-ring generated by $\mathcal{E}$, see Exercise 1.7 below. \hfill \qed
\end{itemize}

Exercise 1.6. If $\mathcal{E} \subset 2^\Omega$ with $\emptyset \in \mathcal{E}$ then define $\mathcal{E}_\sigma$ as the class of all countable unions of sets in $\mathcal{E}$ and define $\mathcal{E}_{\sigma\delta}$ as the class of all countable intersections of sets in $\mathcal{E}_\sigma$. Verify that $(1)$ $\mathcal{E}_\sigma$ (or $\mathcal{E}_{\sigma\delta}$) is stable under the formation of
countable unions (or countable intersections). Prove that (2) if $\mathcal{E}$ is stable under the formation of finite intersections then so is $\mathcal{E}_{\sigma}$. Deduce (3) that if $\mathcal{E}$ is stable under the formation of finite unions and finite intersections then (a) for any $E$ in $\mathcal{E}_{\sigma}$ there exists a monotone increasing sequence $\{E_n\} \subset \mathcal{E}_{\sigma}$, $E_n \subset E_{n+1}$, such that $E = \bigcup_n E_n = \lim_n E_n$ and (b) for any $F$ in $\mathcal{E}_{\sigma^d}$ there exists a monotone decreasing sequence $\{F_n\} \subset \mathcal{E}_{\sigma}$, $E_n \supset F_{n+1}$, such that $F = \bigcap_n F_n = \lim_n F_n$.

**Exercise 1.7.** Let $\mathcal{E} \subset 2^X$ be a semi-ring such that $X$ is a countable union of elements in $\mathcal{E}$ and consider the class of subsets in $2^X$ defined by $\mathcal{F} = \{\sum_{k=1}^\infty E_k : E_k \in \mathcal{E}\}$, recall that $\sum$ means disjoint union of sets. (1) Modify the arguments of Exercise 1.1 to prove that $\mathcal{F}$ is stable under the formation of countable unions. (2) Show that $\mathcal{F}$ is stable under the formation of finite intersection. Why the distributive formula of Exercise 1.5 cannot be used to show that $\mathcal{F}$ is stable under countable intersections? (3) Show that if $A$ belongs to $\mathcal{F}$ and $B$ belongs to the ring generated by $\mathcal{E}$ then the difference $A \setminus B$ is $\mathcal{F}$.

**Exercise 1.8.** A (nonempty) class $\mathcal{L}$ of subsets of $2^X$ is called a lattice if it is stable under finite intersections and finite unions (which may or may not contain the empty set $\emptyset$). Verify that (a) any intersection of lattices is a lattice; (b) if $\mathcal{L}$ is a lattice then the complement class $\{A : A^c \in \mathcal{L}\}$ is also a lattice; (c) if $X = \mathbb{R}$ and $\mathcal{I}_c$ is the class of all bounded closed intervals $[a,b]$, $-\infty < a \leq b < \infty$, and the empty set $\emptyset$, then the smallest lattice class containing $\mathcal{I}_c$ is the class of all closed sets in $\mathbb{R}$; (d) if $X = \mathbb{R}$ and $\mathcal{I}_o$ is the class of all bounded open intervals $(a,b)$, $-\infty < a < b < \infty$, and the empty set $\emptyset$, then the smallest lattice class containing $\mathcal{I}_c$ is the class of all open sets in $\mathbb{R}$. (e) How about the equivalent of (c) and (d) for $X = \mathbb{R}^d$? Finally, note that a $\sigma$-lattice is a class stable under countable intersections and countable unions, show that a $\sigma$-lattice is a monotone class and give an example of a monotone class (with an infinite number of elements) which is not a lattice.

**Exercise 1.9.** A (nonempty) class $\mathcal{E}$ of subsets of $2^X$ is called an oval if for any elements $U, A, V$ in $\mathcal{E}$ we have $U | A | V := (U \cap A^c) \cup (V \cap A)$ in $\mathcal{E}$ (which may or may not contain the empty set $\emptyset$). Prove that $\mathcal{E}$ is a ring if and only if $\mathcal{E}$ is oval with $\emptyset \in \mathcal{E}$.

Actually, Exercises 1.8 and 1.9 are part of more advanced topics, the interested reader may consult Konig [72, Section 1.1, pp. 1-10] to understand the context.

Given a non empty set $\Omega$ (called space) with a $\sigma$-algebra $\mathcal{F}$, the couple $(\Omega, \mathcal{F})$ is called a measurable space and each element in $\mathcal{F}$ is called a measurable set. Moreover, the measurable space is said to be separable if $\mathcal{F}$ is countable generated, i.e., if there exists a countable class $\mathcal{K}$ such that $\sigma(\mathcal{K}) = \mathcal{F}$. An atom of a $\sigma$-algebra $\mathcal{F}$ is a set $F$ in $\mathcal{F}$ such that any other subset $E \subset F$ with $E$ in $\mathcal{F}$ is either the empty set, $E = \emptyset$, or the whole $F$, $E = F$. Thus, a $\sigma$-algebra separates points (i.e., for any $x \neq y$ in $\Omega$ there exist two sets $A$ and $B$ in $\mathcal{F}$ such that $x \in A$, $y \in B$ and $A \cap B = \emptyset$) if and only if the only atoms of $\mathcal{F}$ are the...
singletons (i.e., sets of just one point, \( \{x\} \) in \( \mathcal{F} \)). Sometimes, if \( \mathcal{K} \) is a class of subsets of \( \Omega \) and \( E \subset \Omega \) then the class \( \mathcal{K}_E \) of all sets of the form \( K \cap E \) with \( K \) in \( \mathcal{K} \) is referred to as the trace of \( \mathcal{K} \) on \( E \), and \( \mathcal{K}_E \) may be denoted by \( \mathcal{K} \cap E \).

### 1.2 Borel Sets and Topology

Recall that a topology on \( \Omega \) is a class \( \mathcal{T} \subset 2^\Omega \) with the following properties: (1) \( \emptyset, \Omega \in \mathcal{T} \), (2) if \( U, V \in \mathcal{T} \) then \( U \cap V \in \mathcal{T} \) (stable under finite intersections) and (3) if \( U_i \in \mathcal{T} \) for an arbitrary set of indexes \( i \in I \) then \( \bigcup_{i \in I} U_i \in \mathcal{T} \) (stable under arbitrary unions). Every element of \( \mathcal{T} \) is called open and the complement of an open set is called closed. A basis for a topology \( \mathcal{T} \) is a class \( \mathcal{B} \subset \mathcal{T} \) such that for any point \( x \in \Omega \) and any open set \( U \) containing \( x \) there exists an element \( V \in \mathcal{B} \) such that \( x \in V \subset U \), i.e., any open set can be written as a union of open sets in \( \mathcal{B} \). Clearly, if \( \mathcal{B} \) is known then also \( \mathcal{T} \) is known as the smallest class satisfying (1), (2), (3) and containing \( \mathcal{B} \). Moreover, a class \( \mathcal{B}_o \) containing \( \emptyset \) and such that \( \bigcup \{V \in \mathcal{B}_o\} = \Omega \) is called a sub-basis and the smallest class satisfying (1), (2), (3) and containing \( \mathcal{B}_o \) is called the weakest topology generated by \( \mathcal{B}_o \) (note that the class constructed as finite intersections of elements in a sub-basis forms a basis). A space \( \Omega \) with a topology \( \mathcal{T} \) having a countable basis \( \mathcal{B} \) is commonly used. If the topology \( \mathcal{T} \) is induced by a metric then the existence of a countable basis \( \mathcal{B} \) is obtained by assuming that the space \( \Omega \) is separable, i.e., there exists a countable dense set.

Given a family of spaces \( \Omega_i \) with a topology \( \mathcal{T}_i \) for \( i \) in some arbitrary family of indexes \( I \), the product topology \( \mathcal{T} = \prod_{i \in I} \mathcal{T}_i \) (also denoted by \( \mathcal{T}_i \)) on the Cartesian product space \( \Omega = \prod_{i \in I} \Omega_i \) is generated by the basis \( \mathcal{B} \) of open cylindrical sets, i.e., sets of the form \( \prod_{i \in I} U_i \), with \( U_i \in \mathcal{T}_i \) and \( U_i = \Omega_i \) except for a finite number of indexes \( i \). Certainly, it suffice to take \( U_i \) in some basis \( \mathcal{T}_i \) to get a basis \( \mathcal{B} \), and therefore, if the index \( I \) is countable and each space \( \Omega_i \) has a countable basis then so does the (countable!) product space \( \Omega \). Recall Tychonoff’s Theorem which states that any (Cartesian) product of compact (Hausdorff) topological spaces is again a compact (Hausdorff) topological space with the product topology.

On a topological space \((\Omega, \mathcal{T})\) we define the Borel \( \sigma \)-algebra \( \mathcal{B} = \mathcal{B}(\Omega) \) as the \( \sigma \)-algebra generated by the topology \( \mathcal{T} \). If the space \( \Omega \) has a countable basis \( \mathcal{B} \), then \( \mathcal{B} \) is also generated by \( \mathcal{B} \). However, if the topological space does not have a countable basis then we may have open sets which are not necessarily in the \( \sigma \)-algebra generated by a basis. The couple \((\Omega, \mathcal{B})\) is called a Borel space, and any element of \( \mathcal{B} \) is called a Borel set.

Similar to the product topology, if \{\((\Omega_i, \mathcal{F}_i) : i \in I\)\} is a family of measurable spaces then the product \( \sigma \)-algebra on the product space \( \Omega = \prod_{i \in I} \Omega_i \) is the \( \sigma \)-algebra \( \mathcal{F} = \prod_{i \in I} \mathcal{F}_i \) (also denoted by \( \mathcal{F}_i \)) generated by all sets of form \( \prod_{i \in I} A_i \), where \( A_i \in \mathcal{F}_i \), \( i \in I \) and \( A_i = \Omega_i \), \( i \notin J \) with \( J \subset I \), finite. However, only if \( I \) is finite or countable, we can ensure that the product \( \sigma \)-algebra \( \prod_{i \in I} \mathcal{F}_i \) is also generated by all sets of form \( \prod_{i \in I} A_i \), where \( A_i \in \mathcal{F}_i \), \( i \in I \). For a finite number of factors, we write \( \mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2 \times \cdots \times \mathcal{F}_n \). Sometimes, the notation
\( \mathcal{F} = \bigotimes_{i \in I} \mathcal{F}_i \) is used (i.e., with \( \otimes \) replacing \( \times \)), to distinguish from the Cartesian product (which is rarely used for classes of sets).

In an infinite product set \( \Omega = \prod_{i \in I} \Omega_i \), the cylindrical sets (or cylinder sets) are all sets of the form \( C_{J,A} \), where \( J \) is a finite subset of indexes (of \( I \)) and \( A \) is a (measurable) set in the (finite) product \( \sigma \)-algebra \( \mathcal{F}' = \prod_{i \in J} \mathcal{F}_i \) (of the finite product space \( \Omega' \)), and, by definition, \( (\omega_i : i \in I) \) belongs to \( C_{J,A} \) if and only if \( (\omega_i : i \in J) \) belongs to \( A \). Note that the representation of cylindrical sets in this form is not unique, e.g., \( C_{J,A} = C_{J',A'} \) for any \( A = \Omega_1 \times A' \) and \( J' = J \setminus \{1\} \). Thus, if \( \mathcal{C} \) denotes the class of all cylindrical subsets of \( \Omega \) then the identities (a) \( C_{J,A} = C_{J,A}^c \) (i.e., the complement of cylindrical set corresponding to \( J,A \) is indeed the cylindrical set corresponding to \( J,A^c \), with \( A^c \) being the complement of \( A \) in \( \Omega' \)) and (b) \( C_{J,A} \cup C_{J,B} = C_{J,A \cup B} \), prove that \( \mathcal{C} \) is an algebra in \( \Omega \).

**Exercise 1.10.** Let \( I \) be a family of indexes and \( \mathcal{E}_i \) be a class (of subsets of \( \Omega_i \)) generating a \( \sigma \)-algebra \( \mathcal{F}_i \). Verify that the product \( \sigma \)-algebra \( \prod_{i \in I} \mathcal{F}_i \) can be generated by cylinder sets of the form \( \prod_{i \in I} E_i \), where \( E_i \in \mathcal{E}_i \), \( i \in I \) and \( E_i = \Omega_i \), \( i \notin J \) with \( J \subset I \), finite. Now, for a finite product \( I = \{1, \ldots, n\} \) consider the \( \sigma \)-algebra \( \mathcal{A} \) generated by the hyper-rectangles of the form \( E_1 \times \cdots \times E_n \) with \( E_i \in \mathcal{E}_i \). Discuss why in proving that \( \mathcal{A} \) is actually the product \( \sigma \)-algebra \( \mathcal{F}_1 \times \cdots \times \mathcal{F}_n \), we may need the following assumption: for every \( i = 1, \ldots, n \) there exists a monotone increasing sequence \( \{E_{i,k} : k \geq 1\} \) such that \( \Omega_i = \bigcup_k E_{i,k} \).

\[ \square \]

**Proposition 1.8.** Let \( \Omega \) be a topological space such that every open set is a countable union of closed sets. Then the Borel \( \sigma \)-algebra \( \mathcal{B}(\Omega) \) is the smallest class stable under countable unions and countable intersections which contains all closed sets.

**Proof.** Let \( \mathcal{B}_0 \) be the smallest class stable under countable unions and intersections which contains all closed sets. Since every open set is a countable union of closed sets, we deduce that \( \mathcal{B}_0 \) contains all open sets. Define \( \Phi = \{B \in \mathcal{B}(\Omega) : B \in \mathcal{B}_0 \text{ and } B^c \in \mathcal{B}_0\} \). It is clear that \( \Phi \) is stable under countable unions and intersections, and it contains all closed sets. The minimal character of \( \mathcal{B}_0 \) implies that \( \Phi = \mathcal{B}_0 \), and because \( \Phi \) is also stable under the formation of complement, we deduce that \( \mathcal{B}_0 \) is a \( \sigma \)-algebra, i.e., \( \mathcal{B}_0 = \mathcal{B}(\Omega) \).

\[ \square \]

For instance, if \( d \) is a metric on \( \Omega \) then any closed \( C \) can be written as \( C = \bigcap_{n=1}^{\infty} \{x \in \Omega : d(x, C) < 1/n\} \), i.e., as a countable intersection of open sets, and by taking complement, any open set can be written as a countable union of closed sets. In this case, Proposition 1.8 proves that the Borel \( \sigma \)-algebra \( \mathcal{B}(\Omega) \) is the smallest class stable under countable unions and intersections which contains all closed (or open) sets.

The reader may want to take a look at the book Ask [5, Section 4.3-4, pp. 188–200], regarding measures on topological spaces.

On a topological space \( \Omega \) we define the classes \( \mathcal{F}_\sigma \) (and \( \mathcal{G}_\delta \)) as the countable unions of closed (intersections of open) sets. Thus, any countable unions of sets in \( \mathcal{F}_\sigma \) is again in \( \mathcal{F}_\sigma \) and any countable intersections of sets in \( \mathcal{G}_\delta \) is again
in $\mathcal{G}_\delta$. In particular, if the singletons (sets of only one point) are closed then any countable set is an $\mathcal{F}_\sigma$. However, we can show (with a so-called category argument) that the set of rational numbers is not a $\mathcal{G}_\delta$ in $\mathbb{R} = \Omega$.

In $\mathbb{R}$, we may argue directly that any open interval is a countable (disjoint) union of open intervals, and any open interval $(a, b)$ can be written as the countable union $\bigcup_{n=1}^{\infty} (a + 1/n, b - 1/n)$ of closed sets, an in particular, we show that any open set (in $\mathbb{R}$) is an $\mathcal{F}_\sigma$. In a metric space $(\Omega, d)$, a closed set $F$ can be written as $F = \bigcap_{n=1}^{\infty} F_n$, with $F_n = \{x \in \Omega : d(x, F) < 1/n\}$, which proves that any closed set is a $\mathcal{G}_\delta$, and by taking the complement, any open set in a metric space is a $\mathcal{F}_\sigma$.

Certainly, we can iterate these definitions to get the classes $\mathcal{F}_{\sigma\delta}$ (and $\mathcal{G}_{\delta\sigma}$) as countable intersections (unions) of sets in $\mathcal{F}_\sigma$ ($\mathcal{G}_\delta$), and further, $\mathcal{F}_{\sigma\delta\sigma}$, $\mathcal{G}_{\delta\sigma\delta}$, etc. Any of these classes are family of Borel sets, but in general, not every Borel set belongs necessarily to one of those classes.

**Exercise 1.11.** Show that the Borel $\sigma$-algebra $\mathcal{B}(\mathbb{R}^d)$ of the $d$-dimensional Euclidean space $\mathbb{R}^d$, which is generated by all open sets, can also be generated by (1) all closed sets, (2) all bounded open rectangles of the form $\{x \in \mathbb{R}^d : a_i < x_i < b_i, \forall i\}$, with $a_i, b_i \in \mathbb{R}$, (3) all unbounded open rectangles of the form $\{x \in \mathbb{R}^d : x_i < b_i, \forall i\}$, with $b_i \in \mathbb{R}$, (4) all unbounded open rectangles with rational extremes, i.e., $\{x \in \mathbb{R}^d : x_i < b_i, \forall i\}$, with $b_i \in \mathbb{Q}$, rational, (5) in the previous expressions we may replace the sign $<$ by any of the signs $\leq$ or $>$ or $\geq$ (i.e., replace open by closed or semi-closed) and the results remain true. Moreover, give an argument indicating that $\mathcal{B}(\mathbb{R}^d)$ has the cardinality of the continuum (e.g., see Rama [92, Section 4.5, pp. 110-112]). Finally, prove that $\mathcal{B}(\mathbb{R}^n \times \mathbb{R}^m) = \mathcal{B}(\mathbb{R}^n) \times \mathcal{B}(\mathbb{R}^m)$. \hfill \qed

**Exercise 1.12.** If $\{S_i : i \in I\}$ is a family of semi-algebras (semi-rings) then define $\mathcal{S}$ as the class of sets of the form $\prod_{i \in I} A_i$, with $A_i \in S_i$ and $A_i = \Omega_i$, $i \notin J$, for some finite non-empty index $J \subset I$. Prove that $\mathcal{S}$ is also a semi-algebra (semi-ring). *Hint:* For instance, make use of the equality 

$$(A_1 \times \cdots \times A_n \times \Omega_{n+1} \times \cdots)^c = (A_1 \times \cdots \times A_n)^c \times \Omega_{n+1} \times \cdots,$$

to reduce the question to the case when the index $I$ is finite. Next, for the case of $\Omega_1 \times \Omega_2$, we have the equalities $(A_1 \times A_2)^c = A_1^c \times \Omega_2 + A_1 \times A_2^c$, $(A_1 \times A_2) \cap (B_1 \times B_2) = (A_1 \cap B_1) \times (A_2 \cap B_2)$, $(A_1 \times A_2) \setminus (B_1 \times B_2) = (A_1 \setminus B_1) \times A_2 + (A_1 \cap B_1) \times (A_2 \setminus B_2)$. \hfill \qed

Besides using $\sigma$-algebras, sometimes one may be interested in what is called Souslin (and analytic) sets. The interested reader may take a quick look at Bogachev [16, Section 1.10, pp. 35-40], for some useful results.

As implicitly mentioned, a topology may not be defined on a measurable space. On the contrary, given a non empty space $\Omega$ with a (Hausdorff) topology, the couple $(\Omega, \mathcal{B})$ is called a *Borel space* whenever $\mathcal{B} = \mathcal{B}(\Omega)$ is the Borel $\sigma$-algebra (i.e., generated by the open sets), and any element in $\mathcal{B}$ is call a Borel set. Thus, a Borel space presuppose a given topology. Recall that a complete and separable metric (or metrizable) space is called a Polish space.
1.3 Measurable Functions

Let \((\Omega, \mathcal{F})\) and \((E, \mathcal{E})\) be two measurable spaces. A function \(f : \Omega \to E\) is called \textit{measurable} if \(f^{-1}(B) = \{\omega : f(\omega) \in B\}\) belong to \(\mathcal{F}\) for any \(B\) in \(\mathcal{E}\). Since \(\mathcal{A} = \{A \in \mathcal{E} : f^{-1}(A) \in \mathcal{F}\}\) is a \(\sigma\)-algebra, we deduce that if \(\mathcal{E} = \sigma(\mathcal{K})\) then for \(f\) to be measurable it suffices that \(K \in \mathcal{K}\) implies \(f^{-1}(K) \in \mathcal{F}\).

Usually, our interest is when \(E = \mathbb{R}^d\), but the particular case where \(E\) is a Lusin space (i.e., \(E\) is homeomorphic to a Borel subset of a compact metrizable space or equivalently, \(E\) is a one-to-one continuous image of a Polish space) and \(\mathcal{E} = \mathcal{B}(E)\) (its Borel \(\sigma\)-algebra) is sufficiently general to accommodate all situations of interest, for instance a complete metrizable space or a Borel set \(E \subset \mathbb{R}^d\) is a typical example. Recalling that a function \(f\) is continuous if and only if \(f^{-1}(U)\) is open in \(\Omega\) for any open set \(U\) in \(E\), we obtain that any continuous function is measurable (whenever any open set in \(\Omega\) belongs to \(\mathcal{F}\)).

**Exercise 1.13.** Verify that the function \(1_A\) is measurable if and only if the set \(A\) is measurable. Also show that if \(V\) is an additive group of measurable functions (i.e., any function in \(V\) is measurable, the function identically zero belongs to \(V\), and \(f \pm g\) is in \(V\) for every \(f, g\) in \(V\)) then the family of sets \(\mathcal{L} = \{A \subset \Omega : 1_A \in V\}\) is a \(\ell\)-class (or additive class) in \(2^\Omega\).

Suppose that \(E\) is a topological space where every open set \(O\) can be written as countable union of open sets with closure contained in \(O\), i.e., \(O = \bigcup_i O_i\), for a sequence of open set \(\{O_i : i = 1, 2, \ldots\}\) satisfying \(\overline{O_i} \subset O\), e.g., a metric space. If \(\{f_n\}\) is a sequence of measurable functions with values in \(E\) such that \(f_n(x) \to f(x)\) for every \(x \in \Omega\), then \(f\) is also measurable. Indeed, it suffices to write \(f^{-1}(O) = \bigcup_i \bigcap_{n \geq k} f_n^{-1}(F_i)\) for any open set \(O = \bigcup_i O_i\) in \(E\), with \(O_i\) open sets and \(F_i = \overline{O_i}\) closed sets. Similarly, if \(d\) is a complete metric on \(E\) and \(C\) is the subset of \(\Omega\) where \(\{f_n(x)\}\) converges then the expression \(C = \bigcap_k \bigcup_m \bigcap_{n \geq m} \{x \in \Omega : d(f_n(x), f_m(x)) < 1/k\}\) shows that \(C\) is a measurable set, and therefore, the limit \(f(x) = \lim_n f_n(x)\) for \(x \in C\) can be extended to a measurable function defined on the whole \(\Omega\).

The composition of measurable functions is clearly measurable and so, in particular, if \(E\) is a vector (algebra) topological space (i.e., the sum, scalar multiplication and product are continuous operations and \(E\) is endowed with its Borel \(\sigma\)-algebra) then \(cf + g\) (\(fg\)) is measurable for any scalar \(c\) and any measurable functions \(f\) and \(g\). Thus, the class of measurable functions \(\mathcal{L}^0 = \mathcal{L}^0(\Omega, \mathcal{F}; E)\) is a vector space if \(E\) is so. Note that if \(E\) is not separable then distinct notions of measurability may appear and a deeper analysis is necessary.

Sometimes we use measurable functions with values in either \((-\infty, +\infty]\) or \([-\infty, +\infty)\) or \(\overline{\mathbb{R}} = [-\infty, +\infty]\), i.e., extended real values. In this case, we have to specify how to handle the symbols \(-\infty\) and \(+\infty\). The corresponding Borel \(\sigma\)-algebra is obtained by simply adding the extra symbols, e.g., \(\bar{B} \in \mathcal{B}(\mathbb{R})\) if and only if \(\bar{B} \cap \overline{\mathbb{R}} \in \mathcal{B}(\mathbb{R})\). For a sequence \(\{f_n\}\) of functions taking values in \([-\infty, +\infty)\) or \(\overline{\mathbb{R}}\), the function \(f(x) = \inf_n f_n(x)\) is measurable if each \(f_n\) is so, and similarly with the sup, \(\liminf\) and \(\limsup\). Essentially, all countable
operation preserves measurability. However, if \( \{f_i : i \in I\} \) is a family of real-valued measurable functions with an infinite non countable index \( I \) such that \( f_i \leq C \) for some constant \( C \) and for every \( i \in I \) then the real-valued function \( f(x) = \sup \{f_i(x) : i \in I\} \) is not necessarily measurable.

**Exercise 1.14.** Let \( \{f_n\} \) be a sequence of measurable functions with extended real values and set \( \overline{f}(x) = \sup_n f_n \) and \( \underline{f}(x) = \inf_n f_n \). Discuss why the expressions \( \overline{f}^{-1}([-\infty, a]) = \bigcap_n f_n^{-1}([-\infty, a]) \) and \( \underline{f}^{-1}([a, +\infty]) = \bigcap_n f_n^{-1}([a, +\infty]) \) for every \( a \in \mathbb{R} \) show that \( \overline{f} \) and \( \underline{f} \) are measurable.

**Exercise 1.15.** Let \( \{f_n : n = 1, 2, \ldots\} \) be a sequence of measurable functions from a measurable space \((X, \mathcal{X})\) into another measurable space \((Y, \mathcal{Y})\). Prove that if \((Y, d)\) is also a complete metric space and \(Y\) is the corresponding Borel \(\sigma\)-algebra then the set \(\{x \in X : f_n(x) \text{ is convergent}\}\) is a measurable set. Hint: may need to use the fact that \(\{f_n(x)\}\) is convergent if and only if \(\sup_{n \geq k} d(f_n(x), f_k(x)) \to 0\) as \(k \to \infty\).

Let \( \{f_i : i \in I\} \) be a family of functions \( f_i : \Omega \to E_i \), where \( (E_i, \mathcal{E}_i) \) is measurable space. We denote by \( \sigma(\{f_i : i \in I\}) \) the \(\sigma\)-algebra generated by the class of sets \( \{f_i^{-1}(B_i) : B_i \in \mathcal{E}_i, i \in I\} \), which is the smallest \(\sigma\)-algebra in \(\Omega\) such that every \( f_i \) is measurable. It is clear that if \( f_i \) is \(\mathcal{F}\)-measurable for each \( i \) then \( \sigma(\{f_i : i \in I\}) \subset \mathcal{F} \). Moreover, if \( \mathcal{F}_i = \sigma(f_i) \) is the \(\sigma\)-algebra generated by \( \{f_i^{-1}(B_i) : B_i \in \mathcal{E}_i, i \in I\} \), a fixed \( f_i \), then \( \sigma(\bigcup_{i \in I} \mathcal{F}_i) = \sigma(f_i : i \in I) \), where \( \sigma(\bigcup_{i \in I} \mathcal{F}_i) \) is the smallest \(\sigma\)-algebra containing every \( \mathcal{F}_i \). A typical example of this construction is the case where \( \Omega = \prod_{i \in I} \Omega_i, \ E_i = \Omega_i, \ \mathcal{E} = \mathcal{F} \) and \( f_i = \pi_i \) are the projections, i.e., \( \pi_i : \Omega \to \Omega_i, \ \pi_i(\omega) = \omega_i \) for any \( \omega = (\omega_i : i \in I) \). It is easy to verify that the product \(\sigma\)-algebra \( \mathcal{F} = \prod_{i \in I} \mathcal{F}_i \) as defined in the previous section satisfies \( \mathcal{F} = \sigma(\{\pi_i : i \in I\}) \).

**Exercise 1.16.** Discuss the difference between the product Borel \(\sigma\)-algebras \( \mathcal{B}(\prod_{i \in I} \Omega_i) \) and \( \prod_{i \in I} \mathcal{B}(\Omega_i) \), when \( \Omega_i \) is a topological space and \( I \) is a countable or uncountable set of index.

**Exercise 1.17.** Let \( (\Omega, \mathcal{F}), (E_i, \mathcal{E}_i), i \in I \) be measurable spaces, and set \( E = \prod_i E_i, \ \mathcal{E} = \bigotimes_i \mathcal{E}_i \), and \( \pi_i : E \to E_i \) the projection. Prove that \( f : \Omega \to E \) is \((\mathcal{F}, \mathcal{E})\)-measurable if and only if \( \pi_i \circ f \) is \((\mathcal{F}, \mathcal{E}_i)\)-measurable for every \( i \in I \).

**Exercise 1.18.** Given an example of a non-measurable real-valued function \( f \) on a given measurable space \((\Omega, \mathcal{F})\) with \( \mathcal{F} \neq 2^\Omega \) such that \( f^2 \) is measurable. Prove that an (extended) real-valued \( f \) is measurable if and only if its positive part \( f^+ = \max\{f, 0\} \) and its negative part \( f^- = -\min\{f, 0\} \) are both measurable.

**Exercise 1.19.** Let \( f \) be a function between two measurable space \((X, \mathcal{X})\) and \((Y, \mathcal{Y})\). Prove that (1) \( f^{-1}(\mathcal{Y}) = \{f^{-1}(B) \in 2^X : B \in \mathcal{Y}\} \) is a \(\sigma\)-algebra in \(X\) and (2) \( \{B \in 2^Y : f^{-1}(B) \in \mathcal{X}\} \) is a \(\sigma\)-algebra in \(Y\).

For instance, for a product space \( \Omega = X \times Y \) with \((X, \mathcal{X})\) and \((Y, \mathcal{Y})\) two measurable spaces, and \( \mathcal{F} = \mathcal{X} \times \mathcal{Y} \) the product \(\sigma\)-algebra, we can consider
the sections of any set in $\mathcal{F}$. This is, for a fixed $y \in \mathcal{Y}$ and any $F \in \mathcal{F}$ define

$$F_y = \{x \in X : (x,y) \in F\}.$$  

If $F = A \times B$ is a rectangle with $A \in \mathcal{X}$ and $B \in \mathcal{Y}$ then $F_y = \emptyset$ for $y \not\in B$ or $F_y = A$ for $y \in B$. It is easy to check that for any $F \in \mathcal{F}$ the sections $F_y \in \mathcal{X}$, for every $y$ in $\mathcal{Y}$. Indeed, we verify that the family of sets $\Phi_y = \{F \in \mathcal{F} : F \in \mathcal{X}\}$ is a $\lambda$-class containing all measurable rectangle, which implies that $\Phi_y = \mathcal{F}$. Alternatively, we remark the properties $$(X \times Y \setminus F)_y = X \setminus F_y$$ and $$(\cup_n F_n)_y = \cup_n (F_n)_y,$$ for any sequence $\{F_n\}$ in $2^{X \times Y}$ and any $y$ in $\mathcal{Y}$, which proves directly that $\Phi_y$ is a $\sigma$-algebra. Certainly, sections can be defined for any product of measurable spaces and the above arguments remain valid.

### 1.4 Some Examples

It should be clear that our main example is the Borel line $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. The space $\mathbb{R}$ has a nice topology, in particular, it is a complete separable metric space (i.e., a Polish space). Even if the $\sigma$-algebra $\mathcal{B}(\mathbb{R})$ has the cardinality of the continuum, and so it is much smaller than $2^\mathbb{R}$, most of the sets (in $\mathbb{R}$) we encounter are Borel set and most functions are Borel function. This is to say $\mathcal{B}(\mathbb{R})$ has a reasonable size with respect to the space $\mathbb{R}$. Certainly, the same remarks apply to $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, which can be also viewed as a product space. We this in mind, let us consider the following examples:

[1] The space $\mathbb{R}^\infty$ or $\mathbb{R}^\mathbb{N}$ with $\mathbb{N} = \{1, 2, \ldots\}$ is the space of all sequences of real numbers. For instance, the family of (open) cylinder sets of the form

$$C = (a_1, b_1) \times \cdots \times (a_n, b_n) \times \mathbb{R} \times \mathbb{R} \times \cdots,$$

with $a_i < b_i$, is a basis of open sets for the product topology. Therefore a sequence in $\mathbb{R}^\infty$ (i.e., a double sequence of real numbers) converges if each coordinate (or component) converges. This space becomes a Polish space with the metric

$$d(x, y) = \sum_{i=1}^{\infty} \frac{2^{-i}|x_i - y_i|}{1 + |x_i - y_i|}, \quad \forall x = (x_i), y = (x_i) \in \mathbb{R}^\infty.$$  

The Borel $\sigma$-algebra $\mathcal{B}(\mathbb{R}^\infty)$ is equal to the product $\sigma$-algebra $\mathcal{B}^\infty(\mathbb{R})$, which is also generated by all sets of the form $B_1 \times \cdots \times B_n \times \cdots$, with $B_i \in \mathcal{B}(\mathbb{R})$ for any $i = 1, 2, \ldots$, i.e., we can impose any kind of Borel constraint on each coordinate and we get a Borel set. In this case, again, the size of the Borel $\sigma$-algebra $\mathcal{B}^\infty(\mathbb{R})$ is a reasonable, with respect to the space $\mathbb{R}^\infty$.

[2] The space $\mathbb{R}^T$, where $T$ is an infinite uncountable set (e.g., an interval in $\mathbb{R}$), is the space of all real-valued function defined on $T$. A basis for the product topology is the family of open cylinder sets of the form $C = \prod_{t \in T}(a_t, b_t)$, with $a_t < b_t$ for every $t \in T$, and $-a_t = b_t = +\infty$ for every $t$ except a finite number. Again, a sequence in $\mathbb{R}^T$ (i.e., a sequence of real-valued functions defined on $T$) converges if each coordinate converges, i.e., the pointwise convergence, and the topology becomes complicate. Moreover, the Borel $\sigma$-algebra $\mathcal{B}(\mathbb{R}^T)$ is not equal to the product $\sigma$-algebra $\mathcal{B}(\mathbb{R})^T$, which is generated by open (or Borel) cylinder sets as described in general early. It is not hard to show that elements...
in $\mathcal{B}^T(\mathbb{R})$ have the form $B \times \mathbb{R}^{T-S}$ (disregarding the order of indexes), with $B \in \mathcal{B}^S(\mathbb{R})$ for some countable subset $S \subset T$. This means that (product) Borel sets in $\mathbb{R}^T$ allow only a countable number of Borel constraint on each coordinate, and for instance, we deduce the unpleasant conclusion that the set of continuous functions is not a Borel set. In this sense, the product Borel $\sigma$-algebra $\mathcal{B}^T(\mathbb{R})$ is (too) small relative to the (too) big space $\mathbb{R}^T$. The Borel $\sigma$-algebra $\mathcal{B}(\mathbb{R}^T)$ is larger, but attached to the pointwise convergence, which create other serious complications.

[3] Usually, for a domain we mean a connected set which is the closure of its interior. Thus, the set $C(D)$ of all real-valued bounded continuous functions defined on a domain $D \subset \mathbb{R}^d$ with the uniform convergence is a good example of a Banach (complete normed) space. If $D$ is bounded then the space $C(D)$ is separable, a very important property for the construction of the Borel $\sigma$-algebra. When $D$ is unbounded (e.g., $D = \mathbb{R}^d$), we prefer to use the locally uniform convergence. Actually, this is also the case when considering continuous functions on an open set $O \subset \mathbb{R}^d$. As discussed later Chapters, this space has a nice topology, referred to as locally convex topological vector spaces. For instance, as in the case of $\mathbb{R}^\infty$, we may choose a increasing sequence $\{K_i\}$ of compact subsets of $\mathbb{R}^d$ such that either $\mathbb{R}^d = \bigcup_i K_i$ or $O = \bigcup_i K_i$ to define a metric

$$d(f, g) = \sum_{i=1}^{\infty} \frac{2^{-i} \|f - g\|_n}{1 + \|f - g\|_n}, \quad \forall f, g \in C(\mathbb{R}^d) \text{ or } C(O),$$

where $\|\cdot\|_n$ is the supremum norm within $K_n$. Thus, $C(\mathbb{R}^d)$ and $C(O)$ are complete separable metric spaces under the locally uniform convergence topology. Actually, if $X$ is a locally compact space, we may consider the space $C_0(X)$ of real-valued continuous functions with compact support. Then, besides the Borel $\sigma$-algebra on $X$, we may consider smaller $\sigma$-algebra which make all functions in $C_0(X)$ measurable, i.e., the Baire $\sigma$-algebra on $X$. If $X$ is a locally compact Polish space (e.g., $X$ is a domain or an open set in $\mathbb{R}^d$) both $\sigma$-algebra coincide, but this is not the case in general.

[4] As mentioned early, a Polish space $\Omega$ is a complete separable metric space, i.e., the topology of the space $\Omega$ is also generated by a basis composed of open balls $B(x, r) = \{y \in \Omega : d(y, x) < r\}$ for $x$ in some countable dense set of $\Omega$, $r$ any positive rational and there exists some metric (equivalent to $d$) which makes $\Omega$ complete. For instance, $\Omega$ is a closed subset of $\mathbb{R}$ with the induced or relative topology; or a more elaborated example $\Omega$ is the space of real-valued continuous functions defined on some locally compact space with the locally uniform convergence. Since, for any closed set $F \subset \Omega$ the function $d(x, F) = \inf\{d(x, y) : y \in F\}$ is continuous, we deduce that the Borel $\sigma$-algebra $\mathcal{B}(\Omega)$ in a Polish space is the smallest $\sigma$-algebra for which every real-valued continuous function defined on $\Omega$ is measurable. This fact is not granted for a general topological space and give rise to the Baire $\sigma$-algebra. To study stochastic processes we use the so-called canonical sample space $D([0, \infty])$ of cad-lag real-valued functions, i.e., functions $\omega : [0, \infty] \to \mathbb{R}$ which are right-
1.5. Various Tools

Let us consider real-valued measurable functions defined on \((\Omega, \mathcal{F})\). A measurable function \(\varphi\) taking a finite number of values is called a simple function (or a measurable simple function), i.e., if \(\varphi\) takes only the values \(a_1, \ldots, a_n\) then \(\varphi(x) = \sum_{i=1}^{n} a_i 1_{A_i}(x)\), where \(A_i = f^{-1}(\{a_i\})\) and \(1_A(x) = 1_{\{x \in A\}}\) is the characteristic function of the set \(A\) (or indicator of the condition \(x \in A\)). Thus \(\varphi\) is a simple function if there exist a finite number of measurable sets \(B_1, \ldots, B_n\) and values \(b_1, \ldots, b_n\) such that \(\varphi(x) = \sum_{i=1}^{n} b_i 1_{B_i}(x)\), for every \(x \in \Omega\); and this representation is by no means unique. It is not so hard to show that \(f\) is a simple function if and only if \(f^{-1}(\mathcal{B}(\mathbb{R}))\) is a finite sub \(\sigma\)-algebra of \(\mathcal{F}\).

The set of simple functions form an algebra and a lattice, i.e., if \(\varphi\) and \(\psi\) are simple functions so are the their sum \(\varphi + \psi\), their product \(\varphi \psi\), their max \(\varphi \lor \psi\), and their min \(\varphi \land \psi\). A key point used later is the following approximation result.

**Proposition 1.9.** If \((\Omega, \mathcal{F})\) is a measurable space and \(f: \Omega \to [0, \infty]\) is measurable, then there exists a sequence of measurable simple functions \(\{f_n\}\) such that \(0 \leq f_1 \leq \ldots \leq f_n \leq \ldots \leq f\), with \(f_n \to f\) pointwise in \(\Omega\), and \(f_n \to f\) uniformly on every set where \(f\) is bounded.

**Proof.** Take \(n\) and define \(F_n^k = f^{-1}([k2^{-n}, (k+1)2^{-n}])\) and \(F_n = f^{-1}([2^n, \infty])\), for every \(k\) between 1 and \(2^{2n} - 1\).

Because \(f\) is measurable \(F_n^k, F_n \in \mathcal{F}\). Now, set

\[
f_n(x) = 2^n 1_{F_n}(x) + \sum_{k=1}^{2^{2n} - 1} k2^{-n} 1_{F_n^k}(x), \quad \forall x \in \Omega.
\]

By construction we have \(f_n \leq f_{n+1}\) for any \(n\), and \(0 \leq f - f_n \leq 2^{-n}\) on the set where \(f \leq 2^n\). Hence, conclusion follows. \(\Box\)

If we apply the above arguments by components or coordinates then the previous approximation result remains true for a measurable function with values in \([0, \infty]^d\).

**Corollary 1.10.** Let \((\Omega, \mathcal{F})\) be a measurable space and \((E, d)\) be a separable metric space. If \(f: \Omega \to E\) is measurable then there exists a sequence of measurable simple functions \(\{f_n\}\) such that \(d(f_n, f) \geq d(f_{n+1}, f) \to 0\), pointwise in \(\Omega\).
(1) Verify that if a characterization of a function in $S$ besides having a finite number of values, what other property is needed to give functions in $S$ can be expressed as the pointwise limit of a monotone increasing sequence of $\bar{S}$.

(2) Now, define $\bar{S}$ by $x \in \bar{S}$ if and there exists a sequence $\{v_n\}$ is a monotone increasing convergent sequence of functions in $V$, i.e., $v_n \leq v_{n+1}$ for every $n$ and $v_n(x) \to v(x)$, finite $\forall x \in \Omega$, then $v \in V$. Then $V$ contains all $\sigma(G)$ measurable functions.

Proof. Let $A$ be the class of $A \subset \Omega$ such that $1_A \in V$. Since $1_{A \Delta B} = 1_A - 1_B$ if $A \supset B$ and $V$ is a vector space, the class $A$ is stable under monotone differences. Moreover, $A$ is stable under monotone countable unions because $V$ is stable under the monotone increasing pointwise convergence. Hence $A$ is a $\lambda$-class containing $G$, and invoking Proposition 1.9, we deduce $\sigma(G) \subset A$. Now, writing any measurable function $f = f^+ - f^-$ and applying Proposition 1.9, we conclude.

- Remark 1.12. There are other forms of Corollary 1.11, for instance the following one. Let $H$ is a monotone vector space of bounded real-valued functions (i.e., a lattice) and $\bar{V}$ is the limit of any monotone pointwise-convergence sequence in $H$ belongs to $H$ whenever it is bounded) containing a multiplicative $M$ subspace (i.e., if $f$ and $g$ belong to $M$ then $fg$ also belongs to $M$). Then $H$ contains all bounded and measurable functions with respect to the $\sigma$-algebra generated by the functions in $M$, e.g., see Dellacherie and Meyer [31, Chapter 1, Theorem 21, pp. 20–21] or Sharpe [105, Appendix A0, pp. 264–266].

Exercise 1.21. Let $S$ be a semi-ring of measurable sets in $(\Omega, F)$ such that $\sigma(S) = F$ and there exists a sequence $S_i$ in $S$ satisfying $\Omega = \bigcup_i S_i$. Denote by $\mathbb{S}$ the vector space generated by all functions of the form $1_A$ with $A$ in $S$. Besides having a finite number of values, what other property is needed to give a characterization of a function in $\mathbb{S}$? Consider the following questions:

(1) Verify that if $\varphi, \phi \in \mathbb{S}$ then $\max\{\varphi, \phi\} \in \mathbb{S}$ (i.e., $\mathbb{S}$ a lattice) and $\varphi \phi \in \mathbb{S}$.

(2) Now, define $\bar{\mathbb{S}}$ as the semi-space of extended real-valued functions which can be expressed as the pointwise limit of a monotone increasing sequence of functions in $\mathbb{S}$. Show that if $f(x) = \lim_n f_n(x)$, with $f_n(x) \leq f_{n+1}(x)$, for every $x$ in $\Omega$ and $f_n$ in $\bar{\mathbb{S}}$, then $f$ belongs to $\bar{\mathbb{S}}$. Verify that the function $1 = 1_\Omega$ may not belongs to $\mathbb{S}$, but $1$ belongs to $\bar{\mathbb{S}}$. Moreover, if $u : \mathbb{R}^2 \to \mathbb{R}$ is a continuous function, nondecreasing in each variable and with $u(0,0) = 0$, then show that $x \mapsto u(f(x), g(x))$ belongs to $\bar{\mathbb{S}}$ for every $f$ and $g$ in $\bar{\mathbb{S}}$. Therefore, $\bar{\mathbb{S}}$ is a lattice but not necessarily a vector space.

Exercise 1.22. Let \((X, \mathcal{X})\) and \((Y, \mathcal{Y})\) be two measurable spaces and let \(f : X \times Y \to [0, +\infty]\) be a measurable function. Consider the functions \(f_*(x) = \inf_{y \in Y} f(x, y)\) and \(f^*(x) = \sup_{y \in Y} f(x, y)\) and assume that for every simple (measurable) function \(f\) the functions \(f_*\) and \(f^*\) are also measurable. By means of the approximation by simple functions given in Proposition 1.9:

(1) Prove that \(f_*\) and \(f^*\) are measurable functions.

(2) Extend this result to a real-valued function \(f\).

(3) Now, if \(g : \mathbb{R}^d \to \mathbb{R}\) is a locally bounded (Borel) measurable function, then show that \(\overline{g}(x) = \lim_{r \to 0} \sup_{0 < |y - x| < r} g(y)\) and \(g(x) = \lim_{r \to 0} \inf_{0 < |y - x| < r} g(y)\) are measurable functions.

(4) Finally, give some comments on the assumption about the measurability of \(f_*\) and \(f^*\) for a simple function \(f\).

Hint: apply the transformation \(\arctan\) to obtain a bounded function, i.e., \(f_*(x) = \tan\left(\inf_{y \in Y} \arctan f(x, y)\right)\) and \(f^*(x) = \tan\left(\sup_{y \in Y} \arctan f(x, y)\right)\). Regarding (4), the brief comments in next Chapter about analytic sets and universal completion hold an answer, see Proposition 2.3.

\[\square\]

Proposition 1.13. Let \((\Omega, \mathcal{F})\) and \((E, \mathcal{E})\) be two measurable spaces, \(g : \Omega \to E\) be a measurable function, and \(\mathbb{R} = [-\infty, +\infty]\) the extended real numbers. Denote by \(\mathcal{G} = \sigma(g) = g^{-1}(\mathcal{E})\) the \(\sigma\)-algebra generated by \(g\). Then a function \(h\) is measurable from \((\Omega, \mathcal{G})\) into \((\mathbb{R}, \mathcal{B}(\mathbb{R}))\) (or into \(\mathbb{R}\)) if and only if there exists a measurable function \(k\) from \((E, \mathcal{E})\) into \((\mathbb{R}, \mathcal{B}(\mathbb{R}))\) (or into \(\mathbb{R}\)) such that \(h = k \circ g\).

Proof. If \(h(x) = k(g(x))\) for a Borel function \(k\), then for every Borel set \(B\), the pre-image \(k^{-1}(B)\) is in \(\mathcal{E}\) and \(h^{-1}(B) = g^{-1}(k^{-1}(B))\) belongs to \(\mathcal{G}\), i.e., \(h\) is \(\mathcal{G}\)-measurable.

Now, suppose that \(h : \Omega \to \mathbb{R}\) is \(\mathcal{G}\)-measurable, i.e., \(h^{-1}(B) \in \mathcal{G}\) for any set \(B\) in the Borel \(\sigma\)-algebra \(\mathcal{B}\) of \(\mathbb{R}\). Since \(\mathcal{G} = g^{-1}(\mathcal{E})\), if \(h = \mathbb{1}_A\), there exists \(B \in \mathcal{E}\) such that \(A = g^{-1}(B)\), so we can take \(k = \mathbb{1}_B\) to have \(h = k \circ g\). A little more general, if \(h\) is a simple function, i.e., \(h = \sum_{i=1}^n a_i \mathbb{1}_{A_i}\) with \(\{A_i\}\) disjoint and \(a_i \neq a_j\) if \(i \neq j\), then we take \(k = \sum_{i=1}^n \mathbb{1}_{B_i}\) with \(A_i = g^{-1}(B_i)\).

In general, we write \(h = h^+ - h^-\) and we consider only the case \(h \geq 0\). Approximating \(h\) with an increasing sequence \(\{h_n\}\) of simple functions as in Proposition 1.9, we have \(h = \lim_n h_n\) and \(h_n = k_n \circ g\). By using the pointwise definition \(k(y) = \limsup_n k(y)\) (or with \(\liminf\)) for any \(y\) in \(E\), the following equality \(k(g(x)) = \limsup_n h_n(f(x)) = \lim_n h_n(f(x)) = g(x)\) holds for every \(x \in \Omega\). Finally, if \(h\) has only finite values then we desire finite values for the function \(k\), for instance, we can modify the definition of \(k\) into \(\bar{k}\), namely, \(\bar{k}(y) = k(y)\) if \(|k(y)| < \infty\) and \(\bar{k}(y) = 0\) if \(|k(y)| = \infty\), which is a measurable with values in \(\mathbb{R}\).

\[\square\]

Note that the measurable space \((E, \mathcal{E})\) may take a product form. This means that if \((\Omega, \mathcal{F})\) is a measurable space and \(g = \{g_i : i \in I\}\) is a family of measurable functions with \(f_i\) taking values in some measurable space \((E_i, \mathcal{E}_i)\), and \(\mathcal{G}\) is the \(\sigma\)-algebra generated by \(\{g_i : i \in I\}\), i.e., the minimal \(\sigma\)-algebra containing...
Let \( g_3 \) be the following statement: the sets \( \{ E \} \) are equal, for every closed set \( C \) and in general, the family (possible uncountable) of all atoms (of \( F \)) may not contain any uncountable Borel set with an uncountable complement does not belong to \( F \) and in general, the family (possible uncountable) of all atoms (of \( F \)) may not contain any uncountable Borel set with an uncountable complement does not belong to \( F \).

Also, by means of Exercise 1.17, we can extend the arguments of the previous result to the case where the spaces \( \left( \mathbb{R}, \mathcal{B}(\mathbb{R}) \right) \) or \( \left( \mathbb{R}, \mathcal{B}(\mathbb{R}) \right) \) are replaced by the product spaces \( \left( \mathbb{R}^I, \mathcal{B}^I(\mathbb{R}) \right) \) or \( \left( \mathbb{R}^I, \mathcal{B}^I(\mathbb{R}) \right) \), for any index \( I \).

- **Remark 1.14.** If \( \{ g_i : i \in I \} \) is a family of measurable functions then the \( \sigma \)-algebra \( G = \sigma(g_i : i \in I) \) generated by this family is **countable independent** in the following sense: For any set \( A \) in \( G \) there exists a countable subset of indexes \( J \) of \( I \) such that \( A \) is also measurable with respect to \( \sigma(g_i : i \in J) \). Indeed, to check this, observe that the class of sets having the above property forms a \( \sigma \)-algebra. Thus, if \( h \) is a measurable function on \( (\Omega, G) \) assuming only a finite number (or countable) of values (i.e., a simple function) then there exist a measurable function \( k \) and a countable subset \( J \) of \( I \) such that \( h = k(g_i : i \in J) \), i.e., \( k \) is independent of the coordinates \( i \) in \( I \backslash J \). Indeed, such a function \( h \) has the form
  \[
  h = \sum_n a_n 1_{A_n}
  \]
  for some sequence \( \{ A_n \} \) of disjoint measurable sets and some sequence \( \{ a_n \} \) of values. Each \( A_n \) is measurable with respect to \( \sigma(g_i : i \in J_n) \) for some countable subset of indexes \( J_n \) of \( I \), and so, \( h \) is measurable with respect to \( \sigma(g_i : i \in J) \) for the measurable subset \( J = \bigcup_n J_n \) of \( I \). Therefore, the function \( k \) can be taken measurable with respect to \( \sigma(g_i : i \in J) \). \( \Box \)

**Exercise 1.23.** Let \( \{ f_t : t \in T \} \) be a family of measurable functions from \( (\Omega, F) \) into a Borel space \( (E, \mathcal{E}) \). Assume that the set of indexes \( T \) has a (sequential) topology and there exists a countable subset of indexes \( Q \subset T \) such that for every \( x \in \Omega \) and any \( t \in T \) there is a sequence \( \{ t_n \} \) of indexes in \( Q \) such that \( t_n \to t \) and \( f_{t_n}(x) \to f_t(x) \).

(a) Prove that for metric spaces, the above condition is equivalent to the following statement: the sets \( \{ x \in \Omega : f_t(x) \in C, \forall t \in O \} \) and \( \{ x \in \Omega : f_t(x) \in C, \forall t \in O \cap Q \} \) are equal, for every closed set \( C \) in \( E \) and any open set \( O \) in \( T \).

(b) If the mapping \( t \mapsto f_t(x) \) is continuous for every fixed \( x \in \Omega \), then prove that any countable dense set \( Q \subset T \) satisfies the above condition.

(c) If \( E = [\infty, +\infty] \) then for the family \( \{ f_t : t \in T \} \) of extended real-valued functions, we have
  \[
  f^*(x) = \sup_{t \in T} f_t(x) = \sup_{t \in Q} f_t(x), \quad f_*(x) = \inf_{t \in T} f_t(x) = \inf_{t \in Q} f_t(x),
  \]
  which prove that \( f^* \) and \( f_* \) are measurable functions. \( \Box \)

Given a measurable space \( (\Omega, F) \), we may not necessarily know if a singleton is measurable, i.e., \( \{ \omega \} \in F \). However, we define the **atoms of** \( F \) as elements \( A \in F \) such that \( A \neq \emptyset \), and any \( B \subset A \) with \( B \in F \) results \( B = \emptyset \) or \( B = A \). We can show that any measurable function must be constant on every atom, and in general, the family (possible uncountable) of all atoms (of \( F \)) may not generate \( F \). For instance, the Borel \( \sigma \)-algebra \( \mathcal{B}(\mathbb{R}^d) \) contains all singletons, but, any uncountable Borel set with an uncountable complement does not belong to the \( \sigma \)-algebra generated by \( \{ x \} \) with \( x \in \mathbb{R}^d \).
Exercise 1.24. Let $\mathcal{P} = \{\Omega_1, \ldots, \Omega_n\}$ be a finite partition of $\Omega$, i.e., $\Omega = \sum_{i=1}^{n} \Omega_i$, with $\Omega_i \neq \emptyset$. Describe the algebra $\mathcal{A}$ generated by the finite partition $\mathcal{P}$ and prove that each $\Omega_i$ is an atom of $\mathcal{A}$. How about if $\mathcal{P}$ is a countable or uncountable partition?

Perhaps, some readers could benefit from taking a quick look at Al-Gwaiz and Elsanousi [1, Chapter 10, pp. 349–392], for a discussion on preliminaries of the Lebesgue measure in $\mathbb{R}$. 
Chapter 2

Measure Theory

First the definition and initial properties of a measure is given in general (as a set-function with desired properties), which are discussed in some detail. Next, the real question is addresses, namely, how to construct a measure or in other words, how to extend a suitable definition of a set-function to become a measure. Finally, the Lebesgue measure is defined and studied.

2.1 Abstract Measures

Let us begin with the key concepts of a set-function (i.e., a function defined on a class of sets) being finitely additive and countably additive. A (set) function \( \mu : \mathcal{K} \to [0, \infty] \), with \( \emptyset \in \mathcal{K} \subseteq 2^\Omega \) and \( \mu(\emptyset) = 0 \) is called additive or finitely additive if \( A, B \in \mathcal{K} \), with \( A \cup B \in \mathcal{K} \) and \( A \cap B = \emptyset \) imply \( \mu(A \cup B) = \mu(A) + \mu(B) \). Similarly, \( \mu \) is called \( \sigma \)-additive or countably additive if \( A_i \in \mathcal{K} \), with \( \bigcup_{i=1}^\infty A_i = A \in \mathcal{K} \) and \( A_i \cap A_j = \emptyset \) for \( i \neq j \) imply \( \mu(A) = \sum_{i=1}^\infty \mu(A_i) \).

It is clear that if \( \mu \) is \( \sigma \)-additive then \( \mu \) is also additive, but the converse is false; e.g., take \( \mathcal{K} = 2^\Omega \), with \( \Omega \) an infinite set and \( \mu(A) = 0 \) if \( A \) is finite and \( \mu(A) = \infty \) otherwise. Naturally, if \( \mu \) is redefined as \( \mu(A) = 0 \) if \( A \) is a countable set (finite or infinite) and \( \mu(A) = \infty \) otherwise, then \( \mu \) is \( \sigma \)-additive. Certainly, if \( \mathcal{K} \) is a \( (\sigma-) \)ring then \( (\sigma-) \)additivity is neat and written as: \( A, B \in \mathcal{K} \) implies \( \mu(A + B) = \mu(A) + \mu(B) \) or \( A_i \in \mathcal{K} \) implies \( \mu(\sum_{i=1}^\infty A_i) = \sum_{i=1}^\infty \mu(A_i) \), plus the implicit condition \( \mu(\emptyset) = 0 \). The above properties will be referred to as \( \mu \) being \( (\sigma-) \)additivity on \( \mathcal{K} \), and in this context, the condition \( \mu(\emptyset) = 0 \) will be always implicitly assumed.

In view of Exercise 1.1, it is simple (but perhaps tedious) to verify that any additive (or \( \sigma \)-additive) set-function defined on a semi-ring (or semi-algebra) \( \mathcal{K} \) has a unique extension to the ring (or algebra) generated by \( \mathcal{K} \). Thus, a set-function being additive or \( \sigma \)-additive is meaningful. However, if \( \mathcal{K} \) is only a \( \pi \)-class (e.g., the class of all closed intervals in \( \mathbb{R} \)) then it may not be any sets \( A \neq B \) in \( \mathcal{K} \) with \( A \cup B \in \mathcal{K} \) and \( A \cap B = \emptyset \) to test the \( (\sigma-) \)additivity property, i.e., \( (\sigma-) \)additivity as defined above is of no use when the class \( \mathcal{K} \) is
only a $\pi$-class $\mathcal{K}$. In the same way that a semi-ring generated a ring, a $\pi$-class $\mathcal{K}$ generated a lattice $\bar{\mathcal{K}}$ (i.e., the class of finite unions of sets in $\mathcal{K}$), and as seen later, the concept of additivity is translated as $\mathcal{K}$-tightness for a $\pi$-class, while the concept of $\sigma$-additivity becomes the $\sigma$-smoothness property at $\emptyset$ (or monotone continuity at $\emptyset$) for the lattice $\bar{\mathcal{K}}$.

To discuss set functions with possible infinite values, besides the usual order and sign rules with the symbols $+\infty$ and $-\infty$, the convention $0 \cdot (\pm \infty) = 0$ is used (unless otherwise stated) in all what follows. The extended real numbers (i.e., $\mathbb{R}$ and the two infinite symbols with the above conventions) is usually denoted by $\bar{\mathbb{R}} = [-\infty, +\infty]$.

**Definition 2.1.** A set-function $\mu : \mathcal{A} \to [0, \infty]$ is called a measure (on a $\sigma$-ring or algebra or ring) if $\mu$ is $\sigma$-additive and $\mathcal{A}$ is a $\sigma$-algebra (or a $\sigma$-ring or algebra or ring). If $\mu$ also satisfies $\mu(\Omega) = 1$ then $\mu$ is called a probability measure or in short a probability. Thus a tern $(\Omega, \mathcal{A}, \mu)$ is called measure space if $\mu$ is a measure on the measurable space $(\Omega, \mathcal{A})$.

Similarly, $(\Omega, \mathcal{F}, P)$ is called a probability space when $P$ is a probability on the measurable space $(\Omega, \mathcal{F})$. Sometimes we use the name additive measure (or additive probability) to say that $\mu$ is finitely additive on an algebra $\mathcal{A}$. If $\mu$ is an additive measure and $A, B \in \mathcal{A}$ with $A \subseteq B$ then by writing $A \cup B = A + (B \setminus A)$ we deduce $\mu(A) \leq \mu(B)$ (monotony) and $\mu(B \setminus A) = \mu(B) - \mu(A)$ if $\mu(A) < \infty$. Moreover, if $A_i \in \mathcal{A}$, for $i = 1, \ldots, n$ then $\mu(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} \mu(A_i)$ (subadditivity); and similarly with the sub $\sigma$-additivity if $\mu$ is a measure. Note that occasionally, we have to study measures defined on $\sigma$-rings instead of $\sigma$-algebras, e.g., see Halmos [57].

Perhaps the simplest example is the Dirac measure, namely, take a fix element $x_0$ in $\Omega$ to define

$$
\delta : 2^\Omega \to [0, 1], \quad \delta(A) = 1 A(x_0),
$$

i.e., $\delta(A)$ is equal to 1 if $x_0 \in A$ and is equal to 0 otherwise. This gives rise to the discrete measures after using the fact that $\mu(A) = \sum_{i=1}^{\infty} a_i m_i(A)$ is a measure if each $m_i$ is so, provided that $a_i$ are nonnegative real numbers.

To clarify the difference between additivity and $\sigma$-additivity consider the following properties on $\mu$:

1. monotone continuity from below, i.e., if $\bigcup_{i=1}^{\infty} A_i = A$ with $A_i \subset A_{i+1}$, and $A_i,A \in \mathcal{A}$ imply $\lim_{i\to\infty} \mu(A_i) = \mu(A)$;
2. monotone continuity from above, i.e., if $\bigcap_{i=1}^{\infty} A_i = A$ with $A_i \supset A_{i+1}$, and $A_i,A \in \mathcal{A}$ then $\lim_{i\to\infty} \mu(A_i) = \mu(A)$;
3. monotone continuity at $\emptyset$, i.e., if $\bigcap_{i=1}^{\infty} A_i = \emptyset$ with $A_i \supset A_{i+1}$, and $A_i,A \in \mathcal{A}$ then $\lim_{i\to\infty} \mu(A_i) = 0$.

**Proposition 2.2.** Let $\mu : \mathcal{A} \to [0, \infty]$ be an additive set function on the algebra $\mathcal{A} \subset 2^\Omega$. Then $\mu$ is $\sigma$-additive on $\mathcal{A}$ if and only if $\mu$ is monotone continuous from below. Moreover, assuming $\mu(\Omega) < \infty$ the $\sigma$-additivity of $\mu$ on $\mathcal{A}$ is equivalent a either (a) monotone continuity from above or (b) monotone continuity at $\emptyset$. 


Proof. Only the case of a finite measure $\mu$ is considered, since we may proceed similarly for the general case.

If $\mu$ is $\sigma$-additive and $\{A_i\}$ is a monotone increasing sequence with $A = \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$, then define $B_i = A_i \setminus A_{i-1}$, with $A_0 = \emptyset$ to get $B_i$ disjoint, $\bigcup_{i=1}^{n} B_i = A_n$ and $\bigcup_{i=1}^{\infty} B_i = A$. Hence, the $\sigma$-additivity property implies $\mu(A) = \lim_{n} \sum_{i=1}^{n} \mu(B_i) = \lim_{n} \mu(A_n)$, i.e., $\mu$ is monotone continuous from below.

If $\mu$ is monotone continuous from below and $\{A_i\}$ is a monotone decreasing $A = \bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$, then define $B_i = A \setminus A_i$ to get a monotone increasing sequence $\{B_i\}$ with $\bigcup_{i=1}^{\infty} B_i = A_1 \setminus A$ and $\mu(B_i) = \mu(A_1) - \mu(A_i)$. Then monotone continuity from below applied to $\{B_i\}$ implies $\mu(A_1) - \mu(A) = \lim_{n} \mu(B_i) = \mu(A_1) - \lim_{n} \mu(A_i)$, i.e., $\mu$ is monotone continuity from above.

Trivially, if $\mu$ is monotone continuous from above then $\mu$ is monotone continuous in $\emptyset$.

If $\mu$ is monotone continuous in $\emptyset$ and $\{A_i\}$ is a sequence of disjointed sets with $A = \bigcup_{i=1}^{n} A_i \in \mathcal{A}$, then define $B_n = A \setminus \bigcup_{i=1}^{n} A_i$ to get a decreasing sequence $\{B_n\}$ to $\emptyset$ and $\mu(B_n) = \mu(A) - \sum_{i=1}^{n} \mu(A_i)$. Then the monotone continuity in $\emptyset$ applied to $\{B_i\}$ implies $\lim_{n} \mu(B_n) = 0$, i.e., $\mu$ is $\sigma$-additive.

Note that in particular, for a finitely additive probability measure the $\sigma$-additivity can be replaced by the monotone continuity in $\emptyset$, which is much simpler to prove.

**Exercise 2.1.** Let $(\Omega, \mathcal{A}, \mu)$ be a measure space and $\{A_i : i \geq 1\}$ be a sequence in $\mathcal{A}$. We write $\liminf_{k} A_k = \bigcap_{n} \bigcup_{i \geq n} A_i$ and $\limsup_{k} A_k = \bigcap_{n} \bigcup_{i \geq n} A_i$. Prove (a) $\mu(\liminf_{k} A_k) \leq \liminf_{k} \mu(A_k)$ and (b) if $\mu(\bigcup_{i \geq n} A_i) < \infty$ for some $n$ then $\mu(\limsup_{k} A_k) \geq \limsup_{k} \mu(A_k)$.

**Exercise 2.2.** Let $(\Omega, \mathcal{A}, \mu)$ be a measure space. Verify that (a) if $A$ and $B$ belong to $\mathcal{A}$ then $\mu(A) + \mu(B) = \mu(A \cup B) + \mu(A \cap B)$. Also, prove that (a) if $\{A_i : i \geq 1\}$ is a sequence of sets in $\mathcal{A}$ satisfying $\mu(A_i \cap A_j) = 0$ for $i \neq j$ then $\mu(\bigcup_{i} A_i ) = \sum_{i} \mu(A_i)$.

**Exercise 2.3.** Let $\{\mu_n\}$ be a sequence of measures. Prove (a) if $\{c_n\}$ is a sequence of nonnegative numbers then $\mu = \sum_n c_n \mu_n$ is a measure; (b) if the sequence $\{\mu_n\}$ is increasing then $\mu = \lim\mu_n$ is a measure; (c) if the sequence $\{\mu_n\}$ is decreasing and $\mu_1$ finite then $\mu = \lim\mu_n$ is a measure.

A set $N \in \mathcal{A}$ with $\mu(N) = 0$ is called a negligible set or null set or set of measure zero. In view of the $\sigma$-additivity, we note that a countable union of negligible sets is again a negligible set.

If a property relative (to elements in $\Omega$, e.g., pointwise equality) is satisfied everywhere except on a set of measure zero then we say that the property is satisfied almost everywhere (in short a.e.), or almost surely (in short a.s) when dealing with a probability measure. For instance, a function $f : \Omega \to E$ is called almost everywhere measurable if there exists a null set $N$ such that $f : \Omega \setminus N \to E$ is measurable. For instance, the interest reader may check the book Wang [Preliminary] Menaldi November 11, 2016
and Klr [118] on generalized measure theory, where besides almost everywhere, the concept of pseudo-almost everywhere (among many other properties) is also discussed.

A measure \( \mu \) or properly saying, a measure space \( (\Omega, A, \mu) \) is called complete if the \( \sigma \)-algebra \( A \) contains all subsets of negligible sets of \( \mu \), i.e., if \( N \in A \) with \( \mu(N) = 0 \) then any \( N_1 \subset N \) belongs to \( A \) and necessarily \( \mu(N_1) = 0 \). If a measure space \( (\Omega, A, \mu) \) is not complete then we can always completed (or extended to a complete measure space) in a natural way as follows: define

\[
\tilde{A} = \{ \tilde{A} \subset \Omega : \tilde{A} = A \cup N_1, \ N_1 \subset N, \ A, N \in A, \ \mu(N) = 0 \}
\]

and \( \tilde{\mu}(\tilde{A}) = \mu(A) \) to deduce that \( A \subset \tilde{A} \) is a \( \sigma \)-algebra and that indeed \( (\Omega, \tilde{A}, \tilde{\mu}) \) is a complete measure space with \( \tilde{\mu} = \mu \) on \( A \). This is called the completion of \( (\Omega, A, \mu) \).

Exercise 2.4. Let \( (\tilde{A}, \tilde{\mu}) \) be the completion of \( (A, \mu) \) as above. Verify that \( \tilde{A} \) is a \( \sigma \)-algebra and that \( \tilde{\mu} \) is indeed well defined. Moreover, show that if \( \tilde{A} \in \tilde{A} \) then there exist \( A, B \in A \) such that \( A \subset \tilde{A} \subset B \) and \( \mu(A) = \mu(B) \). Furthermore, if \( \tilde{A} \subset \Omega \) and there exist \( A, B \in A \) such that \( A \subset \tilde{A} \subset B \) with \( \mu(A) = \mu(B) < \infty \) then \( \tilde{A} \in \tilde{A} \).

Given a family \( \Upsilon \) of measures on a measurable space \( (\Omega, F) \) we may define the \( \Upsilon \)-universal completion of \( F \) as \( F^\Upsilon = \bigcap_{\mu \in \Upsilon} F^\mu \), where \( F^\mu \) is the completion of \( F \) relative to \( \mu \). This is commonly used in probability when dealing with Markov processes, where \( \Upsilon \) is the family of all probability measures \( P \) on \( F \). This concept is related to the absolute measurable spaces, e.g., see Nishiura [88]. Note that \( A \in F^\Upsilon \) if and only if for every \( \mu \in \Upsilon \) there exist \( B \) and \( N \) in \( F \) such that \( B \setminus N \subset A \subset B \cup N \) and \( \mu(N) = 0 \) (since \( B \) and \( N \) may depend on \( \mu \), clearly this does not necessarily imply that \( \mu(N) = 0 \) for every \( \mu \)). Thus, a \( \Upsilon \)-universally complete measurable space satisfies \( F = F^\Upsilon \). The concept of universally measurable is particularly interesting when dealing with measures in a Polish space \( \Omega \), where \( F = B(\Omega) \) is its Borel \( \sigma \)-algebra, and then any subset of \( \Omega \) belonging to \( F^\Upsilon \) is called universally measurable if \( \Upsilon \) is the family of all probability measures over \( (\Omega, B(\Omega)) \). In this context, it is clear that a Borel set is universally measurable. Also, it can be proved that any analytic set (i.e., a continuous or Borel images of Borel sets in a Polish space) is universally measurable, and on any uncountable Polish space there exists a analytic set (with not analytic complement) which is not a Borel set, e.g., see Dudley [37, Section 13.2]. For further reference we state the following particular case:

Proposition 2.3. Let \( X \) and \( Y \) be two Polish spaces and \( \psi : X \to Y \) be a measurable function. If \( A \) be a Borel set in \( X \) and \((\tilde{\mu}, \tilde{B})\) is the completion of a \( \sigma \)-finite measure \( \mu \) on the Borel \( \sigma \)-algebra \( B = B(Y) \) then \( \psi(A) \) belongs to \( \tilde{B} \). Moreover, if \( A \) is a Borel subset of \( X \times Y \) and \( C \) is its projection on \( Y \), i.e., \( C = \{ y \in Y : (x, y) \in A \} \), then there exists a measurable function from the \( \mu \)-completion \( (Y, \tilde{B}(Y)) \) into the Borel space \( (X, B(X)) \) such that \((g(y), y)\) belongs to \( A \) for every \( y \) in \( C \).
Summing up, on a Polish space \( \Omega \) we can define the family (or class of subsets) \( \mathcal{A} = \mathcal{A}(\Omega) \) of analytic sets in \( \Omega \) (i.e., \( A \subset \Omega \) is analytic if \( A = f(X) \) for some Polish space \( X \) and some continuous function \( f : X \to \Omega \)). This class \( \mathcal{A} \) is not necessarily a ring, but it is closed under the formation of countable unions and intersections. Any Borel set is analytic, i.e., \( \mathcal{B}(\Omega) \subset \mathcal{A}(\Omega) \), and if \( \mu^* \) is a finite outer measure (see next Definition 2.4) such that its restriction \( \mu \) to the Borel \( \sigma \)-algebra is a measure, then for any analytic set \( A \) there is a Borel set \( B \) such that \( \mu^*((A \setminus B) \cup (B \setminus A)) = 0 \), i.e., any analytic set is \( \mu^* \)-measurable (in other words, any analytic set belongs to the the \( \mu \)-completion of the \( \sigma \)-algebra of Borel sets). Furthermore, a continuous function between Polish spaces preserves analytic sets, but does not necessarily preserve \( \mu^* \)-measurable sets. For instance, the reader may check Cohn [28, Chapter 8, pp. 251-296] for a detailed discussion on Polish spaces and analytic sets or Bogachev [16, Chapter 6, pp. 1-66] for a carefully presentation on Borel, Baire and Souslin sets.

It is rather simple to define a finitely additive measure. For instance the Jordan-Riemann measure \( m \) in \( \mathbb{R} \), namely, \( \mathcal{A} \subset \mathbb{R}^2 \) is the algebra of sets that can be written as a disjoint finite union of intervals (closed, open, semi-open, bounded, unbounded), say a generic interval different from \( \mathbb{R} \) is denoted by \( I \) and has the form \( (a, b), [a, b], (a, b] \) or \( (a, b] \) with \( a \leq b, a, b \in [−\infty, +\infty] \), and \( m(I) = b - a \), \( m(\mathbb{R}) = \infty \) and finally for \( A = \bigcup_{i=1}^n I_i \) with \( I_i \cap I_j = \emptyset \) if \( i \neq j \) where we define \( m(A) = \sum_{i=1}^n m(I_i) \). A more difficult step is to show the \( \sigma \)-additivity and to extend the definition of \( m \) to a \( \sigma \)-algebra (the Borel \( \sigma \)-algebra \( \mathcal{B}(\mathbb{R}) \) in the case of the Jordan-Riemann measure).

Thus a millstone to overcome is the construction of measures. It is our intention to discuss three methods (even if only one is usually sufficient) used to obtain a measure from a given set-function naturally defined (a priori) in a small class of sets, which is not necessarily a \( \sigma \)-ring.

2.2 Caratheodory’s Arguments

This method is also called the outer-measure construction, and it begins by considering set-functions defined for every possible subsets of a given reference abstract set \( \Omega \).

Definition 2.4. A function \( \mu^* : 2^{\Omega} \rightarrow [0, \infty] \) is called an outer measure (or exterior measure) on \( \Omega \) if (1) \( \mu^*(\emptyset) = 0 \), (2) \( A \subset B \) implies \( \mu^*(A) \leq \mu^*(B) \) (monotone or isotone), and (3) \( \mu^*(\bigcup_{i=1}^n A_i) \leq \sum_{i=1}^n \mu^*(A_i) \) (sub \( \sigma \)-additive). Next a subset \( A \subset \Omega \) is said to be \( \mu^* \)-measurable if \( \mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c) \), for every \( E \subset \Omega \), i.e., \( \mu^*(E) \geq \mu^*(E \cap A) + \mu^*(E \cap A^c) \), in view of the sub-additivity.

In some modern books (e.g., Evans and Gariepy [42], Mattila [80]), the definition of measure is actually the above definition of outer measure. This is actually not a problem since the results in this section show how to construct a measure from a given outer measure and conversely. Indeed, based on the additivity of the inf-operation, it is simple to show that if \( (\Omega, \mu, \mathcal{A}) \) is a measure
space then

\[ \mu^*(A) = \inf \{ \mu(B) : A \subset B \in \mathcal{A} \}, \quad \forall A \in 2^\Omega, \]

is an extension (non-necessarily unique) of the measure \( \mu \) to an outer measure. However, the delicate point (considered in section) is the converse, i.e., for a given outer measure, how to induce a measure.

In this section, a measure (on the space \( \Omega \)) is a non-negative \( \sigma \)-additive set-function \( \mu \) defined on a \( \sigma \)-algebra \( \mathcal{F} \subset 2^\Omega \), and the pair \((\mu, \mathcal{F})\) is called a complete measure if \( \mathcal{F} \) contains all the subsets of any negligible set, i.e., if \( N \) belongs to \( \mathcal{F} \) and \( \mu(N) = 0 \) then any subset \( N' \subset N \) belongs also to \( \mathcal{F} \) (and \( \mu(N') = 0 \)).

**Theorem 2.5.** If \( \mu^* \) is an outer measure on \( \Omega \) and \( \mathcal{F} \) is the class of all \( \mu^* \)-measurable sets then \( \mathcal{F} \) is a \( \sigma \)-algebra and the restriction \( \mu \) of \( \mu^* \) to \( \mathcal{F} \) is a complete measure.

**Proof.** First, because the definition of \( \mu^* \)-measurability is symmetric in \( A \) and \( A^c \), the class \( \mathcal{F} \) is stable under the formation of complement. Next, if \( A, B \in \mathcal{F} \) and \( E \subset \Omega \), by the subadditivity we have

\[ \mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c) = \mu^*(E \cap A \cap B) + \mu^*(E \cap A \cap B^c) + \mu^*(E \cap A^c \cap B) + \mu^*(E \cap A^c \cap B^c) \geq \mu^*(E \cap (A \cup B)) + \mu^*(E \cap (A \cup B)^c). \]

Hence \( A \cup B \in \mathcal{F} \), i.e., the class \( \mathcal{F} \) is an algebra. Moreover, if \( A, B \in \mathcal{F} \) and \( A \cap B = \emptyset \) then

\[ \mu^*(A \cup B) = \mu^*((A \cup B) \cap A) + \mu^*((A \cup B) \cap A^c) = \mu^*(A) + \mu^*(B), \]

i.e., \( \mu^* \) is finitely additive on \( \mathcal{F} \).

To show that \( \mathcal{F} \) is a \( \sigma \)-algebra we have to prove only that \( \mathcal{F} \) is stable under countably disjoint unions. Thus, for any sequence \( \{A_j\} \) of disjoint sets in \( \mathcal{F} \), define \( B_n = \bigcup_{j=1}^n A_j \) and \( B = \bigcup_{j=1}^\infty A_j \) to get

\[ \mu^*(E \cap B_n) = \mu^*(E \cap B_n \cap A_n) + \mu^*(E \cap B_n \cap A_n^c) = \mu^*(E \cap A_n) + \mu^*(E \cap B_{n-1}), \quad \forall E \subset \Omega, \]

and by induction, this yields \( \mu^*(E \cap B_n) = \sum_{j=1}^n \mu^*(E \cap A_j) \). Therefore

\[ \mu^*(E) = \mu^*(E \cap B_n) + \mu^*(E \cap B_n^c) \geq \sum_{j=1}^n \mu^*(E \cap A_j) + \mu^*(E \cap B^c), \]

and as \( n \to \infty \) we obtain

\[ \mu^*(E) \geq \sum_{j=1}^\infty \mu^*(E \cap A_j) + \mu^*(E \cap B^c) \geq \mu^* \left( \bigcup_{j=1}^\infty (E \cap A_j) \right) + \mu^*(E \cap B^c) = \mu^*(E \cap B) + \mu^*(E \cap B^c) \geq \mu^*(E), \]
implies \( \bigcup F \) satisfied). Pick any set \( E \) every
Hence \( \varepsilon > \infinium \) ensures that for every \( A \) elements in
Since \( \varepsilon \)
Proof. Let \( \\Omega = \bigcup \infinium \) and by taking the infimum over all covers we deduce
Finally, if \( \mu^*(A) = 0 \) then we have
\[
\mu^*(E) \leq \mu^*(E \cap A) + \mu^*(E \cap A^c) = \mu^*(E \cap A^c) \leq \mu^*(E), \quad \forall E \subset \Omega,
\]
i.e., \( A \in \mathcal{F} \), and \( \mu = \mu^*|_\mathcal{F} \) is a complete measure. \( \square \)
At this point, we need to discuss how to construct an outer measure.
\[\textbf{Proposition 2.6.} \text{Let } \mathcal{E} \subset 2^\Omega \text{ and } \mu: \mathcal{E} \to [0, +\infty] \text{ be such that } \emptyset \in \mathcal{E}, \mu(\emptyset) = 0 \text{ and } \Omega = \bigcup_n \Omega_n, \text{ for some sequence } \{\Omega_n\} \text{ in } \mathcal{E}. \text{ Define}
\]
\[
\mu^*(A) = \inf \left\{ \sum_{n=1}^\infinium \mu(E_n) : E_n \in \mathcal{E}, A \subset \bigcup_{n=1}^\infinium E_n \right\}, \quad \forall A \subset \Omega. \tag{2.1}
\]
Then \( \mu^* \) is an outer measure on \( \Omega \). Moreover, if a set \( A \subset \Omega \) satisfies \( \mu(E) \geq \mu^*(E \cap A) + \mu^*(E \cap A^c) \), for every \( E \in \mathcal{E} \) with \( \mu(E) < \infinium \) then \( A \) is \( \mu^* \)-measurable.

Proof. Since \( \emptyset \in \mathcal{E} \) and \( \Omega = \bigcup_n \Omega_n \), with \( \Omega_n \in \mathcal{E} \), the set function \( \mu^* \) is defined for every \( A \in 2^\Omega \) and \( \mu^*(\emptyset) = 0 \). If \( A \subset B \) then any time we cover \( B \) with elements in \( \mathcal{E} \) also we cover \( A \), and so the infimum satisfies \( \mu^*(A) \leq \mu^*(B) \).

To check the sub \( \sigma \)-additivity, let \( \{A_n\} \) a sequence in \( 2^\Omega \). The definition of infimum ensures that for every \( \varepsilon > 0 \) and \( n \) there is a sequence \( \{E^n_j\} \) such that
\[
A_n \subset \bigcup_{j=1}^\infinium E^n_j \quad \text{and} \quad \sum_{j=1}^\infinium \mu(E^n_j) \leq \mu^*(A_n) + 2^{-n}\varepsilon.
\]
Hence
\[
\bigcup_{n=1}^\infinium A_n \subset \bigcup_{j,n=1}^\infinium E^n_j \quad \text{and} \quad \mu^* \left( \bigcup_{n=1}^\infinium A_n \right) \leq \sum_{j,n=1}^\infinium \mu(E^n_j) \leq \varepsilon + \sum_{n=1}^\infinium \mu^*(A^n).
\]
Since \( \varepsilon \) is arbitrary, definition (2.1) yields a \( \mu^* \) sub \( \sigma \)-additivity, i.e., \( \mu^* \) is an outer measure.

Finally, pick a set \( A \subset \Omega \) satisfying \( \mu(E) \geq \mu^*(E \cap A) + \mu^*(E \cap A^c) \), for every \( E \in \mathcal{E} \) with \( \mu(E) < \infinium \) (note that for \( \mu(E) = \infinium \) the inequality is trivially satisfied). Pick any set \( F \subset \Omega \) and a sequence \( \{E_n\} \subset \mathcal{E} \) covering \( F \). Since \( \bigcup_n (E_n \cap A) \supset F \cap A \) and \( \bigcup_n (E_n \cap A^c) \supset F \cap A^c \), the sub \( \sigma \)-additivity of \( \mu^* \) implies
\[
\sum_n \mu(E_n) \geq \sum_n \mu^*(E_n \cap A) + \sum_n \mu^*(E_n \cap A^c) \geq \\
\geq \mu^*(F \cap A) + \sum_n \mu^*(F \cap A^c),
\]
and by taking the infimum over all covers we deduce \( \mu(F) \geq \mu^*(F \cap A) + \mu^*(F \cap A^c) \), which means that \( A \) is \( \mu^* \)-measurable. \( \square \)
Remark 2.7. It is clear that $\mu^*(E) \leq \mu(E)$, for every $E$ in $\mathcal{E}$. Moreover, iterating the inf, i.e.,

$$
\mu^{**}(A) = \inf \left\{ \sum_{n=1}^{\infty} \mu^*(E_n) : E_n \in \mathcal{E}, A \subset \bigcup_{n=1}^{\infty} E_n \right\}, \quad \forall A \subset \Omega,
$$

we obtain the same outer measure, i.e., $\mu^{**} = \mu^*$. Indeed, we need only to show that $\mu^{**}(A) \geq \mu^*(A)$ for every $A \subset \Omega$ with $\mu^{**}(A) < \infty$. For every $\varepsilon > 0$ there exists a sequence $\{E_n\}$ in $\mathcal{E}$ such that $A \subset \bigcup_n E_n$ and $\varepsilon + \mu^*(A) \geq \sum_n \mu^*(E_n)$, and therefore, for each $n \geq 1$ there is a sequence $\{E_{n,k}\}$ such that $E_n \subset \bigcup_k E_{n,k}$ and $2^{-n}\varepsilon + \mu^*(E_n) \geq \sum_k \mu(E_{n,k})$. Hence

$$
2\varepsilon + \mu^{**}(A) \geq \sum_{n,k} \mu(E_{n,k}) \quad \text{and} \quad A \subset \bigcup_{n,k} E_{n,k},
$$

which shows that $\mu^{**} \geq \mu^*$.

Exercise 2.5. With the notation of Proposition 2.6, if the initial set measure $\mu(E) = \sum_{\omega_i \in E} r_i$, for some sequences $\{\omega_i\} \subset \Omega$ and $\{r_i\} \subset [0, \infty]$ then the same weighted counting expression holds $\mu^*(E)$, for any $E$ in $\mathcal{E}$, i.e.,

$$
\mu^*(E) = \sum_{\omega_i \in E} r_i = \sum_{i=1}^{\infty} r_i \mathbb{1}_{\omega_i \in E}, \quad \forall E \in \mathcal{E},
$$

where $\mathbb{1}_{\omega_i \in E} = 1$ if $\omega_i$ is in $E$ and vanishes otherwise. What can be said about $\mu^*(A)$, for any $A \subset 2^\Omega$?

If there is not a sequence $\{\Omega_n : n \geq 1\}$ such that $\Omega = \bigcup_n \Omega_n$, we may include the whole space $\Omega$ into the class $\mathcal{E}$ and define $\mu(\Omega) = \infty$. Thus, ensuring that $\mu^*$ given by (2.1) is defined for every subset $A$ of $\Omega$. Alternatively, we may consider the $\sigma$-ring of all sets covered by some sequence of sets in $\mathcal{E}$ or equivalently, the hereditary $\sigma$-ring $\mathcal{R}$ generated by the class $\{A \subset E : E \in \mathcal{E}\}$, further details are given in Exercise 2.12 below.

Remark 2.8. Recall the notation $\sum_n E_n$ to indicate a disjoint union, i.e., $\sum_n E_n = \bigcup_n E_n$ with $E_n \cap E_m = \emptyset$ if $n \neq m$. Assume that the class $\mathcal{E}$ is a semi-ring and $\mu$ is additive on $\mathcal{E}$, i.e., $E = \sum_{i=1}^{n} E_i$, $E$ and $E_i$ belong to $\mathcal{E}$ yield $\mu(E) = \sum_{i=1}^{n} \mu(E_i)$. Then the outer measure $\mu^*$ induced by $\mu$ by means of (2.1) satisfies

$$
\mu^*(A) = \inf \left\{ \sum_{n=1}^{\infty} \mu(E_n) : E_n \in \mathcal{E}, A \subset \sum_{n=1}^{\infty} E_n \right\}, \quad \forall A \subset \Omega.
$$

Indeed, if $\{E_n : n \geq 1\} \subset \mathcal{E}$ is a covering of $A$ then define $E'_1 = E_1$, $E'_2 = E_2 \setminus E_1$, and by induction

$$
E'_n = (E_n \setminus E_{n-1}) \cup (E_n \setminus E_{n-2}) \cup \cdots (E_n \setminus E_1).
$$
Because the class $\mathcal{E}$ is a semi-ring, we can write each $E'_n$ as a disjoint union of sets in $\mathcal{E}$, i.e., $E'_n = \sum_{i=1}^{k_n} E'_{n,i}$. The additivity of $\mu$ implies that $\mu(E_n) \geq \sum_{i=1}^{k_n} \mu(E'_{n,i})$. Hence, $\{E_{n,i} : i = 1, \ldots, k_n, n \geq 1\} \subseteq \mathcal{E}$ is a countable cover of $A$ satisfying

$$A \subseteq \sum_{n} \sum_{i=1}^{k_n} E'_{n,i} \quad \text{and} \quad \sum_{n} \mu(E_n) \geq \sum_{n} \sum_{i=1}^{k_n} \mu(E'_{n,i}),$$

which complete the proof. \qed

From the inf definition (2.1) it is clear that if $\mu_i$, $i \geq 1$, are set measures, $\mu_i : \mathcal{E} \rightarrow [0, +\infty]$ as in Proposition 2.6, then $\sum_i (\mu_i)^* \geq \sum_i \mu_i^*$. 

**Exercise 2.6.** Denote by $\mu^*(\cdot, \mathcal{E})$ the outer measure (2.1). Assume that $\mu$ is initially defined on $\mathcal{E}' \supset \mathcal{E}$. Verify (1) that $\mu^*(\cdot, \mathcal{E}') \leq \mu^*(\cdot, \mathcal{E})$. Next, prove (2) that if for every $\varepsilon > 0$ and $E'$ in $\mathcal{E}'$ there exists $E$ in $\mathcal{E}$ such that $E \supset E'$ and $\mu(E') \leq \mu(E) + \varepsilon$ then $\mu^*(\cdot, \mathcal{E}') = \mu^*(\cdot, \mathcal{E})$. \qed

Now, if we require that the initial $\mu$ is a $\sigma$-additive on some algebra $\mathcal{E}$ then we close the circle, i.e., we are able to extend a measure (initially defined on an algebra) to a $\sigma$-algebra.

**Theorem 2.9.** If $\mu$ is a measure on an algebra $\mathcal{E}$ and $\mu^*$ is defined by (2.1) then (a) $\mu^*|_E = \mu$ and (b) every set in $A = \sigma(\mathcal{E})$ is $\mu^*$-measurable and $\bar{\mu} = \mu^*|_A$ is a measure. Moreover, if $\bar{\mu}$ is $\sigma$-finite (i.e., there exists $\{A_n\} \subset A$ such that $\bigcup_{n=1}^{\infty} A_n = \Omega$ with $\bar{\mu}(A_n) < \infty$) then $\bar{\mu}$ is uniquely determinate on $A$, i.e., if $\nu$ is another measure on $A$ such that $\nu|_E = \mu$ then $\nu = \bar{\mu}$.

**Proof.** To show (a), take a generic element $E \in \mathcal{E}$ and for any countable cover $\{E_n\} \subset \mathcal{E}$ define $F_n = E \cap (E_n \setminus \bigcup_{i=1}^{n-1} E_i)$ to satisfy $F_n \in \mathcal{E}$, $E = E \bigcap_{n=1}^{\infty} F_n$, $F_n \cap F_m = \emptyset$ for $n \neq m$ and $F_n \subset E_n$. Hence $\mu(E) = \sum_{n=1}^{\infty} \mu(F_n) \leq \sum_{n=1}^{\infty} \mu(E_n)$, and since the cover is arbitrary, we deduce $\mu(E) \leq \mu^*(E)$. On the other hand, choosing $E_1 = E$ and $E_i = \emptyset$ for $i \geq 2$ we get $\mu^*(E) \leq \mu(E) + 0$, i.e., $\mu(E) = \mu^*(E)$ for every $E \in \mathcal{E}$.

To establish (b), we need to show that every set $E \in \mathcal{E}$ is $\mu^*$-measurable. Thus, take any $F \subset \Omega$ and $\varepsilon > 0$ and by definition of $\mu^*(F)$, there exists a countable cover $\{F_n\} \subset \mathcal{E}$ of $F$ such that $\mu^*(F) + \varepsilon \geq \sum_{n=1}^{\infty} \mu(F_n)$. Since $\{F_n \cap E\}$ and $\{F_n \cap E^c\}$ cover $F \cap E$ and $F \cap E^c$, the additivity of $\mu$ on $\mathcal{E}$ implies

$$\mu^*(F) + \varepsilon \geq \sum_{n=1}^{\infty} (\mu(F_n \cap E) + \mu(F_n \cap E^c)) \geq \mu^*(F \cap E) + \mu^*(F \cap E^c),$$

and because $\varepsilon$ is arbitrary, the set $E$ is $\mu^*$-measurable. Next, by means of Theorem 2.5, $\mu$ induces an outer measure $\mu^*$. In turn, $\mu^*$ yields a measure $\bar{\mu}$ on the $\sigma$-algebra $\mathcal{A}^*$ of $\mu^*$-measurable sets. Since $\mathcal{E} \subset \mathcal{A}^*$ we deduce that $\sigma(\mathcal{E}) = \mathcal{A} \subset \mathcal{A}^*$. Moreover, by (a), $\mu^*|_E = \mu$.
Let us prove that the extension to $\mathcal{A}$ is unique. Suppose that $\nu$ is another measure such that $\nu|_\mathcal{E} = \mu$. For any $A \in \mathcal{A}$ and any sequence $\{E_i\} \subset \mathcal{E}$ with $A \subset \bigcup_{i=1}^{\infty} E_i$ we have $\nu(A) \leq \sum_{i=1}^{\infty} \nu(E_i) = \sum_{i=1}^{\infty} \mu(E_i)$, which yields $\nu(A) \leq \bar{\mu}(A)$. Setting $E = \bigcup_{i=1}^{\infty} E_i$, we get $\nu(E \setminus A) \leq \bar{\mu}(E \setminus A)$ and

$$\nu(E) = \lim_{n \to \infty} \nu \left( \bigcup_{i=1}^{n} E_i \right) = \lim_{n \to \infty} \mu \left( \bigcup_{i=1}^{n} E_i \right) = \bar{\mu}(E).$$

If $\bar{\mu}(A) < +\infty$, for any $\varepsilon > 0$ we can choose a cover $\{E_i\}$ such that $\bar{\mu}(E) < \bar{\mu}(A) + \varepsilon$, i.e., $\bar{\mu}(E \setminus A) < \varepsilon$. Then

$$\bar{\mu}(A) \leq \bar{\mu}(E) = \nu(E) = \nu(A) + \nu(E \setminus A) \leq \nu(A) + \bar{\mu}(E \setminus A) \leq \nu(A) + \varepsilon,$$

and because $\varepsilon$ is arbitrary, we have $\bar{\mu}(A) = \nu(A)$. Finally, if $\bar{\mu}$ is $\sigma$-finite then $\Omega = \bigcup_{n=1}^{\infty} A_n$, with $\bar{\mu}(A_n) < +\infty$, and we may assume that $A_n \cap A_m = \emptyset$ for $n \neq m$. Hence for any $A \in \mathcal{A}$ we have

$$\bar{\mu}(A) = \sum_{n=1}^{\infty} \bar{\mu}(A \cap A_n) = \sum_{n=1}^{\infty} \nu(A \cap A_n) = \nu(A),$$

i.e., $\nu = \bar{\mu}$.

\[ \square \]

- **Remark 2.10.** If $\mathcal{E}$ is a $\pi$-class (i.e., closed under finite intersections and contains the empty set $\emptyset$) and $\mu$ is a (nonnegative) set function defined on $\mathcal{E}$ then we say that $\mu$ is additive on $\mathcal{E}$ if for every $\varepsilon > 0$ and every $F$ and $E$ in $\mathcal{E}$ there exists a sequence (possible finite) $\{E_n\} \subset \mathcal{E}$ such that $F \setminus E \subset \bigcup_n E_n$ and $\mu(F) + \varepsilon > \mu(F \setminus E) + \sum_n \mu(E_n)$. Similarly, we say that $\mu$ is a pre-outer measure if (a) $\mu(\emptyset) = 0$, (b) $E \subset F$, $E$ and $F$ in $\mathcal{E}$ implies $\mu(E) \leq \mu(F)$ (i.e., monotone on $\mathcal{E}$), (c) $E \subset \bigcup_n E_n$, $E$ and $E_n$ in $\mathcal{E}$ implies $\mu(E) \leq \sum_n \mu(E_n)$ (i.e., sub $\sigma$-additive on $\mathcal{E}$). Now, remark that in the proof of the precedent Theorem 2.9, we have also proved that (1) if $\mathcal{E}$ is a $\pi$-class and the initial set function $\mu$ is additive then any set in the $\sigma$-algebra generated by $\mathcal{E}$ is $\mu^*$-measurable; and (2) if the initial set function $\mu$ is a pre-outer measure then $\mu^* = \mu$ on $\mathcal{E}$. In particular, if the initial set function $\mu$ can be extended to a measure on the $\sigma$-algebra $\mathcal{A} = \sigma(\mathcal{E})$ generated by a class $\mathcal{E}$ (satisfying the assumptions of Proposition 2.6) then $\mu = \mu^*$ on $\mathcal{E}$ (but not necessarily on $\mathcal{A}$); and moreover, if $\mathcal{E}$ is a $\pi$-class then any set in $\mathcal{A}$ is $\mu^*$-measurable. \[ \square \]

**Exercise 2.7.** Let $\mu^*$ be the outer measures induced by a set function $\mu : \mathcal{E} \to [0, +\infty]$, as in Proposition 2.6. Denote by $\mathcal{E}_\sigma$ the class of countable unions of sets in $\mathcal{E}$, and by $\mathcal{E}_{\sigma\delta}$ the class of countable interections of sets in $\mathcal{E}_\sigma$. Prove that (1) for every $A \subset \Omega$ and any $\varepsilon > 0$ there exists a set $E$ in $\mathcal{E}_\sigma$ such that $A \subset E$ and $\mu^*(E) \leq \mu(A) + \varepsilon$. Deduce that (2) for every $A \subset \Omega$ there exists a set $F$ in $\mathcal{E}_{\sigma\delta}$ such that $A \subset F$ and $\mu^*(F) = \mu^*(A)$. \[ \square \]

**Exercise 2.8.** Let $\mu_i^* (i = 1, 2)$ be the outer measures induced by the initial set functions $\mu_i : \mathcal{E} \to [0, +\infty]$, as in Proposition 2.6, and assume that the class $\mathcal{E}$
is stable under the formation of finite unions and finite intersections, and that every set in $\mathcal{E}$ is $\mu^*_i$-measurable (for $i = 1, 2$). Show (1) if $A$ and $B$ belong to $\mathcal{E}_{\sigma \delta}$ then $A \cap B$ also belongs to $\mathcal{E}_{\sigma \delta}$. Prove that (2) if $\mu^*_1(E) \leq \mu^*_2(E)$ for every $E$ in $\mathcal{E}$ then $\mu^*_1(A) \leq \mu^*_2(A)$, for every $A \subset \Omega$. \hfill $\square$

**Exercise 2.9.** Let $\mu_i$ ($i \geq 1$) be initial set function $\mu_i : \mathcal{E} \to [0, +\infty]$ as in Proposition 2.6, where $\mathcal{E}$ is now a ring. Assume that all set initial functions are additive, and all but one are $\sigma$-additive, i.e., $\mu_1$ is additive and $\mu_i$ ($i \geq 2$) are $\sigma$-additive. Prove that $(\sum_i \mu_i)^* = \sum_i \mu_i^*$.

**Exercise 2.10.** Let $\mu^*$ be the outer measure on $\Omega$ induced by an additive set function $\mu$ defined on an algebra (actually, a semi-ring suffices) $\mathcal{A} = \mathcal{E}$, given by (2.1). Denote by $\mathcal{A}_\sigma$ the class of countable unions of sets in $\mathcal{A}$, and by $\mathcal{A}_{\sigma \delta}$ the class of countable intersections of sets in $\mathcal{A}_\sigma$. (1) Prove that for every subset $E \subset \Omega$ and every $\epsilon > 0$ there exists a set $A$ in $\mathcal{A}_\sigma$ such that $E \subset A$ and $\mu^*(A) \leq \mu^*(E) + \epsilon$, and (2) deduce that for some $B$ in $\mathcal{A}_{\sigma \delta}$ such that $E \subset B$ and $\mu^*(E) = \mu^*(B)$. Now, a subset $E$ of $\Omega$ is called $\sigma$-finite (relative to $\mu^*$) if there exists a sequence $\{F_i : i \geq 1\}$ in $2^\Omega$ such that $E \subset \bigcup_i F_i$ and $\mu^*(F_i) < \infty$, for every $i$. (3) Show that a $\sigma$-finite set $E$ is $\mu^*$-measurable if and only if there exists $B$ in $\mathcal{A}_{\sigma \delta}$ such that $E \subset B$ and $\mu^*(B \setminus E) = 0$. Finally, if $\mu^*(\Omega) < \infty$ then (4) prove that $E$ is $\mu^*$-measurable if and only if $\mu^*(E) = \mu(\Omega) - \mu^*(\Omega \setminus E)$. \hfill $\square$

**Exercise 2.11.** Let $\mu^*$ be an outer measure on $\Omega$ and $\{\Omega_i : i \geq 1\}$ be a sequence of disjoint $\mu^*$-measurable sets. Prove that $\mu^*(E \cap (\bigcup_i \Omega_i)) = \sum_i \mu^*(E \cap \Omega_i)$, for every $E$ in $2^\Omega$.

A class $\mathcal{R} \subset 2^\Omega$ is called hereditary if $A \in \mathcal{R}$ and $B \subset A$ implies $B \in \mathcal{R}$. The concept of outer measure can be consider on any hereditary $\sigma$-field $\mathcal{R}$ instead of $2^\Omega$. Similarly, a measure can be defined only on a $\sigma$-ring, instead of a $\sigma$-algebra.

**Exercise 2.12.** Revise the definitions and proofs of this section to use outer measures defined on hereditary $\sigma$-rings. The class of $\mu^*$-measurable set becomes a hereditary $\sigma$-ring in Theorem 2.5. The class $\mathcal{E}$ needs not to cover the whole space $\Omega$ in Propositions 2.6. Theorem 2.9 is practically undisturbed if $\sigma$-algebra is replaced by $\sigma$-ring, e.g., see Halmos [57, section 10, pp 41-48]. Referring to Propositions 2.6, as a typical application, consider the hereditary $\sigma$-ring of all set covered by a sequence of sets in $\mathcal{E}$ with $\mu$-finite value (i.e., $\sigma$-finite sets relative to $\mathcal{E}$), to construct $\sigma$-finite measures defined on $\sigma$-rings. Consider the alternative way of beginning with $\mu$ define on a class $\mathcal{E}$ such that $\Omega$ belongs to $\mathcal{E}$ and $\mu(E) < \infty$ for every $E \neq \Omega$. \hfill $\square$

**From a Semi-Ring**

Now, we want to extend the notion of length (or area or volume) naturally defined for intervals (or rectangles or cuboid) to other general sets. Since the class of all intervals form a semi-ring, we need to be able to redo the previous constructions starting from a semi-ring, instead of an algebra.
Thus, if the initial set function $\mu$ is a finitely additive measure on a ring $\mathcal{E}$ then we can define the outer measure $\mu^*$, for any $A \subset \Omega$, by

\[
\text{either } \mu^*(A) = \inf \left\{ \lim_{n \to \infty} \mu(E_n) : E_n \in \mathcal{E}, A \subset \bigcup_{n=1}^{\infty} E_n, E_n \subset E_{n+1} \right\},
\]

\[
\text{or } \mu^*(A) = \inf \left\{ \sum_{n=1}^{\infty} \mu(E_n) : E_n \in \mathcal{E}, A \subset \sum_{n=1}^{\infty} E_n \right\},
\]

instead of using (2.1). Actually, the last expression with coverings in the form of disjoint unions remains valid for a semi-ring $\mathcal{E}$. Similarly, if $\mu$ is a measure on a $\sigma$-algebra $\mathcal{A}$ then

\[
\mu^*(A) = \inf \left\{ \mu(E) : E \in \mathcal{E}, A \subset E, \right\}, \quad \forall A \subset \Omega,
\]

yields an outer measure. Denoting by $\mathcal{A}^*$ the $\sigma$-algebra of all $\mu^*$-measurable sets, we have a complete measure $(\overline{\mu}, \mathcal{A}^*)$ by taking $\overline{\mu} = \mu^*|_{\mathcal{A}^*}$, which is an extension of $(\mu, \mathcal{A})$, and a set $N \subset \Omega$ is negligible if and only if $\mu^*(N) = 0$.

Recall Exercise 1.1, where we verify that the algebra $\mathcal{A}$ (ring) generated by a $\mathcal{S}$ semi-algebra (semi-ring) is the class of finite disjoint unions, i.e., $A \in \mathcal{A}$ if and only if $A = \sum_{i=1}^{n} A_i$ for some $A_i \in \mathcal{S}$.

**Proposition 2.11.** Let $\mathcal{E}$ be a semi-ring and $\mu : \mathcal{E} \to [0, \infty)$ be a $\sigma$-additive finite-valued set function. Then $\mu$ can be uniquely extended to $\sigma$-additive set function on the $\sigma$-ring $\mathcal{R}$ generated by $\mathcal{E}$. Moreover, a further unique extension of the measure $\mu$ to the $\sigma$-ring $\widehat{\mathcal{R}}$ of all ($\sigma$-finite) $\mu^*$-measurable sets is also possible. In particular, if there exists sequence $\{E_n\} \subset \mathcal{E}$ such that $\Omega = \bigcup_{n=1}^{\infty} E_n$ and $\mu(E_n) < \infty$, then $\mu$ can be uniquely extended to a measure on the $\sigma$-algebra $\mathcal{A}$ generated by $\mathcal{E}$. Furthermore, a set $A \subset \Omega$ is $\mu^*$-measurable if and only if $\mu(E) \geq \mu^*(E \cap A) + \mu^*(E \cap A^c)$, for every $E \in \mathcal{E}$.

**Proof.** If $\mathcal{R}_0$ is the ring generated by $\mathcal{E}$ then, recalling that any set in $\mathcal{R}_0$ can be written as a finite disjoint union of elements in $\mathcal{E}$, we extend the definition of $\mu$ to $\mathcal{R}_0$,

\[
\mu(A) = \sum_{i=1}^{n} \mu(E_i), \quad A = \bigcup_{i=1}^{n} E_i, \quad E_i \cap E_j = \emptyset \text{ if } i \neq j.
\]

Because there is only a finite sum (or disjoint union), we deduce that $\mu$ remains $\sigma$-additive on the ring $\mathcal{R}_0$. At this point, we revise the proof of Theorem 2.9 (remarking that $F_n \cap E^c = F_n \setminus E \in \mathcal{R}_0$, for every $F_n, E \in \mathcal{R}_0$) to check that the algebra generated by $\mathcal{E}$ can be replaced by the ring $\mathcal{R}_0$ and the results remain valid. Hence, $\mu$ has a unique extension to $\sigma$-ring $\mathcal{R}$ generated by $\mathcal{E}$.

It is clear that if $\Omega = \bigcup_{n=1}^{\infty} E_n$ and $\mu(E_n) < \infty$, for every $n$, then the $\sigma$-ring $\mathcal{R}$ is indeed the $\sigma$-algebra $\mathcal{A} = \sigma(\mathcal{E})$.

Finally, Theorem 2.9 also ensure a unique extension to the $\sigma$-ring $\widehat{\mathcal{R}}$ of all $\mu^*$-measurable sets. Because the initial set function $\mu$ assume only finite values,
all set in σ-ring \( \mathcal{R} \) are σ-finite. In any case, the uniqueness of the extension is only warranty on the σ-ring \( \mathcal{R} \) of all σ-finite \( \mu^* \)-measurable sets.

It is also clear that because \( \mu^* = \mu \) on \( \mathcal{E} \), Proposition 2.6 yields the stated characterization of a \( \mu^* \)-measurable set in term of sets in the semi-ring \( \mathcal{E} \).

**Exercise 2.13.** Complete the proof of Proposition 2.11, i.e., verify that the extension of \( \mu \) to \( \mathcal{R}_0 \) is well defined and indeed is σ-additive.

- **Remark 2.12.** In the statement of Proposition 2.11, we may initially assume \( \mu : \mathcal{E} \to [0, \infty] \) and define \( \mathcal{E}_0 = \{ E \in \mathcal{E} : \mu(E) < \infty \} \), which is again a semi-ring, i.e., \( \mathcal{R} \) is the σ-ring generated by \( \mathcal{E}_0 \). In general, a subset \( A \) of \( \Omega \) is called σ-finite relative to a set function \( \mu \) defined on a class \( \mathcal{E} \subset 2^\Omega \) if there exists a sequence \( \{ E_n \} \) in \( \mathcal{E} \) such that \( A \subset \bigcup_n E_n \) and \( \mu(E_n) < \infty \), for every \( n \). Thus \( \mathcal{R} \) is the σ-ring generated by the σ-finite sets in \( \mathcal{E} \) relative to \( \mu \). Therefore, if the initial class \( \mathcal{E} \) is a semi-algebra then we may be forced to define the semi-ring \( \mathcal{E}_0 \) as above, which may not be a semi-algebra.

- **Remark 2.13.** The reader can verify that only the finitely additive character (instead of the σ-additivity) of the set function \( \mu \) is used to prove that any set in \( \mathcal{E} \) is \( \mu^* \)-measurable, that \( \mu \leq \mu^* \) on \( \mathcal{E} \) and that a set \( A \subset \Omega \) is \( \mu^* \)-measurable if and only if \( \mu(E) \geq \mu^*(E \cap A) + \mu^*(E \cap A^c) \), for every \( E \) in \( \mathcal{E} \). However, to check that \( \mu = \mu^* \) on \( \mathcal{E} \) the σ-additivity is involved. Sometimes, a function defined on a (semi-)ring is called content or additive measure if it is additive and pre-measure if it is σ-additive. In this context, finitely additive on a semi-ring \( \mathcal{E} \) means that \( \mu(E) = \sum_{i<n} \mu(E_i) \) whenever \( E = \sum_{i<n} E_i \) with all sets in \( \mathcal{E} \), just the case of two sets may not be sufficient.

- **Remark 2.14.** Recall that if \( \mathcal{S}_i \) is a semi-ring (semi-algebra) in a measure space \((\Omega_i, \mathcal{F}_i, \mu_i)\), for \( i = 1, 2 \), then \( \mathcal{S} = \{ S_1 \times S_2 : S_i \in \mathcal{S}_i, \ i = 1, 2 \} \), is a semi-ring (semi-algebra), see Exercise 1.12. Thus the product expression \( \mu(S_1 \times S_2) = \mu_1(S_1)\mu_2(S_2) \) defines an additive measure on \( \mathcal{S} \) (or in Cartesian product \( \mathcal{F}_1 \times \mathcal{F}_2 \)), which can be extended to the product σ-algebra \( \mathcal{F} = \sigma(S_1) \otimes \sigma(S_1) \), by the Caratheodory’s extension Theorem 2.9. However, to verify that \( \mu^* = \mu \) on the semi-ring \( \mathcal{S} \), we need to check that \( \mu = \mu_1 \times \mu_2 \) is indeed σ-additive on \( \mathcal{S} \). Actually, this will be address later by either the construction of the integral or a discussion on inner measures (more tools are needed to prove this fact).

Summing up, the construction of a (σ-finite) measure on a σ-algebra \( \mathcal{A} \) begins with a σ-additive (set) function defined on a semi-ring \( \mathcal{E} \), which generates \( \mathcal{A} \). Actually, the σ-algebra \( \mathcal{A}^* \) of all \( \mu^* \)-measurable sets is usually strictly larger than \( \mathcal{A} = \sigma(\mathcal{E}) \). Usually, the passage of a finitely additive measure defined on an algebra to a σ-additive measure on the generated σ-algebra is called Hopf’s extension theorem, e.g., see Richardson [93, Section 2.4, pp. 24–30].

We can refine a little the argument on the uniqueness.

**Proposition 2.15.** Let \( \mathcal{E} \subset 2^\Omega \) be a \( \pi \)-class. Suppose that \( \mu \) and \( \nu \) are two measures on \( \mathcal{A} = \sigma(\mathcal{E}) \) such that (1) \( \mu = \nu \) on \( \mathcal{E} \) and (2) there exists a monotone increasing sequence \( \{ E_n \} \) of elements in \( \mathcal{E} \) satisfying \( \Omega = \bigcup_n E_n \) and \( \mu(E_n) = \nu(E_n) < \infty \) for every \( n \). Then \( \mu = \nu \) on \( \mathcal{A} \).
Proof. For any fixed $E \in \mathcal{E}$ with $\mu(E) = \nu(E) < \infty$ consider the class $\Phi_E = \{A \in \mathcal{A} : \mu(E \cap A) = \nu(E \cap A)\}$. It is clear that $\Omega \in \Phi_E$ and that the class $\Phi_E$ is stable under monotone differences, and by means of Proposition 2.2, the class $\Phi_E$ is stable also by countably monotone unions and intersections, i.e., $\Phi_E$ is a $\lambda$-class. Because $\mathcal{E} \subset \Phi_E$, Proposition 1.6 implies that $\sigma(\mathcal{E}) \subset \Phi_E$, i.e., $\Phi_E = \mathcal{A}$.

In particular, for $E = E_n$ we deduce that $\mu(E_n \cap A) = \nu(E_n \cap A)$, for every $A \in \mathcal{A}$. Because $E_n \subset E_{n+1}$ (i.e., we need to know that the sequence is monotone increasing) we can use again Proposition 2.2 to get

$$
\mu(A) = \lim_n \mu(E_n \cap A) = \lim_n \nu(E_n \cap A) = \nu(A), \quad \forall A \in \mathcal{A},
$$

i.e., $\mu = \nu$. \hfill \Box

In the case of two probability measures $P$ and $Q$, we may take $E_n = \Omega$ for every $n$ and conclude that: if $P = Q$ on a $\pi$-class $\mathcal{E}$ then $P = Q$ on a $\mathcal{A} = \sigma(\mathcal{E})$.

- **Remark 2.16.** In Proposition 2.15, as well as in previous statements, the unique extension of a measure $\mu$ initially defined on a $\pi$-class $\mathcal{E} \subset 2^\Omega$ requires the $\sigma$-finite property of $\mu$ with respect to $\mathcal{E}$. In general, the assumption $\mu(E) < \infty$ (for every $E \in \mathcal{E}$) yields a unique measure on the $\sigma$-ring $\mathcal{R}$ generated by $\mathcal{E}$, see also Remark 2.12. Indeed, as in the above proof, a monotone argument shows that the class of sets in $\mathcal{R}$ which are included in a countable union of sets in $\mathcal{E}$ is indeed the whole $\sigma$-ring $\mathcal{R}$. Hence, we deduce that $\mu = \nu$ on $\mathcal{R}$. \hfill \Box

Recalling the notation of the symmetric difference $A \Delta B = (A \setminus B) \cup (B \setminus A)$, we have

**Proposition 2.17.** Let $(\Omega, \mathcal{A}, \mu)$ be a finite measure space, i.e., $\mu(\Omega) < \infty$, with $\mathcal{A} = \sigma(\mathcal{E})$, the $\sigma$-algebra generated by some ring $\mathcal{E}$. Then, for every $A$ in $\mathcal{A}$, there exists a sequence $\{A_n\} \subset \mathcal{E}$ such that $\mu(A \Delta A_n) \to 0$.

**Proof.** If $\mathcal{F}$ denotes the class of sets in $\mathcal{A}$ having the required property, then by means of Proposition 1.4, we need only to show that $\mathcal{F}$ is a monotone class, i.e., that $\mathcal{F}$ is stable under countable monotone unions and intersections.

To this purpose, first we check that $\mathcal{F}$ is stable under monotone differences. Indeed, if $E \subset F$ are two sets in $\mathcal{F}$ then for any $E_n$ and $F_n$ in $\mathcal{R}$ such that $\mu(F \Delta F_n) \to 0$ and $\mu(E \Delta E_n) \to 0$, and the inequality

$$
|(|1_F - 1_E| - (1_{F_n} - 1_{E_n}))| \leq |1_F - 1_{F_n}| + |1_E - 1_{E_n}|,
$$

implies

$$
\mu((F \setminus E) \Delta (F_n \setminus E_n)) \leq \mu(F \Delta F_n) + \mu(E \Delta E_n),
$$

i.e., we deduce that $F \setminus E$ also belongs to $\mathcal{F}$.

Next, given a monotone increasing sequence $\{F_n\} \subset \mathcal{F}$, we have to show that $\bigcup_n F_n = F$ belongs to $\mathcal{F}$. Indeed, because $\mathcal{F}$ is stable under monotone differences, we can write $F = \sum_n E_n$, with $E_n = F_{n+1} \setminus F_n$, $n \geq 1$, $F_0 = 0$, and $\{E_n\} \subset \mathcal{F}$. Now, for any $\varepsilon > 0$ we select $m$ sufficiently large to have
\( \mu(F \setminus F_m) = \sum_{k>m} \mu(E_k) < \varepsilon/2; \) and for each \( E_n \), with \( n = 1, \ldots, m \), there exists \( G_n \) in \( \mathcal{R} \) such that \( \mu(E_n \Delta G_n) < 2^{-n} \), for every \( n = 1, \ldots, m \). Now, again we have

\[
\left| \mathbb{1}_F - \sum_{n=1}^{m} \mathbb{1}_{G_n} \right| \leq \sum_{n>m} \mathbb{1}_{E_n} + \sum_{n=1}^{m} \left| \mathbb{1}_{E_n} - \mathbb{1}_{G_n} \right|,
\]

which yields

\[
\mu(F \Delta R_m) \leq \sum_{n>m} \mu(E_n) + \sum_{n=1}^{m} \mu(E_n \Delta G_n) \leq \varepsilon,
\]

for \( R_m = \bigcup_{n=1}^{m} G_n, \ m = m(\varepsilon) \), i.e., the limiting set \( F \) also belongs to \( \mathcal{F} \).

Finally, if \( \{F_n\} \subset \mathcal{F} \) is a monotone decreasing sequence with limit \( F = \bigcap_n F_n \) then by considering the monotone increasing sequence \( G_n = F_1 \setminus F_n \) with limit \( G = F_1 \setminus F \), we complete the proof.

\( \square \)

The reader may check the book by Kemeny et al. [68, Section 1.2, pp, 10-18] for more details on Proposition 2.17.

\textbf{Remark 2.18.} Let \((\Omega, \mathcal{A}, \mu)\) be a measure space (non necessarily finite) and consider the \( \sigma \)-ring \( \mathcal{R} \) of \( \sigma \)-finite measurable sets, i.e., a set \( A \) belongs to \( \mathcal{R} \) if and only if \( A = \bigcup_i A_i \) for some sequence \( \{A_i\} \) in \( \mathcal{A} \) with \( \mu(A_i) < \infty \), for every \( i \). Suppose that \( \mathcal{R} \) is the \( \sigma \)-ring generated by some ring \( \mathcal{K} \) satisfying \( \mu(K) < \infty \), for every \( K \) in \( \mathcal{K} \), i.e., \( \mathcal{K} \) is contained in \( \mathcal{R}_0 = \{R \in \mathcal{R} : \mu(R) < \infty\} \). Now, for a given measurable set \( R \) with finite measure (i.e., in \( \mathcal{R}_0 \)), we may consider the finite measure space \((R, \mathcal{A}|_R, \mu|_R)\), the restriction of \((\Omega, \mathcal{A}, \mu)\) to \( R \), namely, \( \mathcal{A}|_R = \{A \cap R : A \in \mathcal{A}\} \) and \( \mu|_R = \mu \) on \( \mathcal{A}|_R \). Thus, in view of Proposition 2.17, there exists a sequence \( \{K_n\} \) in \( \mathcal{K} \) such that \( \mu(RK_n) \to 0 \).

\( \square \)

\textbf{Exercise 2.14.} Let \((X, \mathcal{X}, \mu)\) be a \( \sigma \)-finite measure space and \((Y, \mathcal{Y})\) be a measurable space. By means of a monotone argument prove that the function \( y \mapsto \mu(A_y) \) is measurable from \((Y, \mathcal{Y})\) into the Borel space \([0, \infty)\), for every \( A \) in the product \( \sigma \)-algebra \( \mathcal{X} \times \mathcal{Y} \) with section \( A_y = \{x \in X : (x, y) \in A\} \). Hint: the equality \( (A \cup B)_x = A_x \cup B_x \) could be of some use here.

\( \square \)

A measure \( \mu \) on \((\sigma \text{-algebra}) \mathcal{A}\) is called \textit{semi-finite} if for every \( A \) in \( \mathcal{A} \) with \( \mu(A) = \infty \) we can find \( F \) in \( \mathcal{A} \) satisfying \( F \subset A \) and \( 0 < \mu(F) < \infty \).

\textbf{Exercise 2.15.} First (1) show that any \( \sigma \)-finite measure is semi-finite, and give an example of a semi-finite measure which is not \( \sigma \)-finite. Next, (2) prove that if \( \mu \) is a semi-finite measure then for every measurable set \( A \) with \( \mu(A) = \infty \) and any real number \( c > 0 \) there exists a measurable set \( F \) such that \( F \subset A \) and \( c < \mu(F) < \infty \). Now, if \( \mu \) is a measure (non necessarily semi-finite) then the finite part \( \mu_f \) is defined by the expression

\[
\mu_f(A) = \sup\{\mu(F) : F \in \mathcal{A}, \ F \subset A, \ \mu(F) < \infty\}.
\]

Finally, (3) prove that \( \mu_f \) is a semi-finite measure, and that if \( \mu \) is semi-finite then \( \mu = \mu_f \). Moreover, (4) show that there exists a measure \( \nu \) (non necessarily unique) which assumes only the values 0 and \( \infty \) such that \( \mu = \mu_f + \nu \).

\( \square \)
The reader may take a look at Taylor [114, Chapter 4, 177–225] and Yeh [120, Chapter 5, pp. 481–596].

2.3 Inner Approach

The second method to construct is a kind of dual technique, instead of using an external approximation, we use an internal approximation, i.e., the key notion is the concept of inner measure. Thus, in a way analogous to the outer measure in Section 2.2 (using the Caratheodory splitting method), we develop the inner measure construction. However, this section is not referred to for the typical Lebesgue measure defined in the next chapter, it could be only used later, when topology is involved. Begin with

**Definition 2.19.** A function $\mu_* : 2^\Omega \to [0, \infty]$ is called an inner measure (or interior measure) on $\Omega$ if (1) $\mu_*(\emptyset) = 0$, (2) $A \subset B$ implies $\mu_*(A) \leq \mu_*(B)$ (monotone or isotone), and (3) $A \cap B = \emptyset$ implies $\mu_*(A \cup B) \geq \mu_*(A) + \mu_*(B)$ (super-additive). Next a subset $A \subset \Omega$ is said to be $\mu_*$-measurable if $\mu_* (E) = \mu_*(E \cap A) + \mu_*(E \cap A^c)$, for every $E \subset \Omega$, i.e., $\mu_*(E) \leq \mu_*(E \cap A) + \mu_*(E \cap A^c)$, in view of the super-additivity.

Note that by induction, the monotony and super-additivity of $\mu_*$ implies $\mu_* \left( \sum_{i=1}^{\infty} A_i \right) \geq \mu_*(\sum_{i=1}^{n} A_i) \geq \sum_{i=1}^{n} \mu_*(A_i)$, and as $n \to \infty$, we deduce a property that could be called super $\sigma$-additivity. It is also clear that the sets $\emptyset$ and $\Omega$ are $\mu_*$-measurable.

In the same way that a measure $(\mu, \mathcal{F})$ induces an outer measure, namely,

$$\mu^*(A) = \inf \{ \mu(F) : A \subset F \in \mathcal{F} \}, \quad \forall A \in 2^\Omega,$$

an inner measure is simply induced by the expression

$$\mu_*(A) = \sup \{ \mu(F) : A \supset F \in \mathcal{F} \}, \quad \forall A \in 2^\Omega.$$

However, the actual problem is to begin with a set-function $\mu$ defined on a small class of sets $\mathcal{K}$, and then construct an inner measure to be able to obtain a measure by restricting the definition of the inner measure $\mu_*$ to the sets that are $\mu_*$-measurable. The interested reader may take a look at Halmos [58, Section 14, pp. 58–62], but what follows is more related to Pollard [90, Appendix A, pp. 289–300].

**Proposition 2.20.** If $\mu_*$ is an inner measure on $\Omega$ and $\mathcal{A}$ is the class of all $\mu_*$-measurable sets then $\mathcal{A}$ is an algebra and the restriction $\mu$ of $\mu_*$ to $\mathcal{A}$ is a complete finitely additive measure.

**Proof.** First, because the definition of $\mu_*$-measurability is symmetric in $A$ and $A^c$, the class $\mathcal{A}$ is stable under the formation of complement. Next, for any $A, B \in \mathcal{A}$ and $E \subset \Omega$, the equality

$$(E \cap A^c \cap B) \cup (E \cap A \cap B^c) \cup (E \cap A^c \cap B^c) = E \cap (A \cap B)^c$$
and the super-additivity of $\mu_*$ imply

$$
\mu_*(E) = \mu_*(E \cap A) + \mu_*(E \cap A^c) = \mu_*(E \cap A \cap B) + \\
+ \mu_*(E \cap A \cap B^c) + \mu_*(E \cap A^c \cap B) + \mu_*(E \cap A^c \cap B^c) \leq \\
\leq \mu_*(E \cap (A \cap B)) + \mu_*(E \cap (A \cap B)^c).
$$

Hence $A \cap B \in \mathcal{A}$, i.e., the class $\mathcal{A}$ is an algebra. Moreover, if $A, B \in \mathcal{A}$ and $A \cap B = \emptyset$ then

$$
\mu_*(A \cup B) = \mu_*((A \cup B) \cap A) + \mu_*((A \cup B) \cap A^c) = \mu_*(A) + \mu_*(B),
$$
i.e., $\mu_*$ is finitely additive on $\mathcal{A}$.

Finally, if $\mu_*(A) = 0$ and $B \subset A$ with $A \in \mathcal{A}$ then the monotony of $\mu_*$ implies

$$
\mu_*(E) \leq \mu_*(E \cap A) + \mu_*(E \cap A^c) = \\
= \mu_*(E \cap A^c) \leq \mu_*(E \cap B) + \mu_*(E \cap B^c), \quad \forall E \subset \Omega,
$$
i.e., $B \in \mathcal{A}$, and $\mu = \mu_*|_{\mathcal{A}}$ is a complete finitely additive measure. \hfill \Box

The essential properties of an inner measure are captured by the expression

$$
\mu_*(A) = \sup \{ \mu_*(B) : B \subset A, \, \mu_*(B) < \infty \}, \quad \forall A \in 2^\Omega. \tag{2.2}
$$

Indeed, any set function $\mu_*$ with $\mu_*(\emptyset) = 0$ satisfying the sup representation (2.2) is monotone, super-additive, and semi-finite (i.e., for every set $A$ with $\mu_*(A) = \infty$ there is a sequence $\{A_n\}$ such that $A_n \subset A$ and $\mu_*(A_n) \to \infty$). Conversely, any semi-finite inner measure $\mu_*$ satisfies (2.2).

Similarly to the previous sections, our intension is to construct an inner measure $\mu_*$ (such that its restriction to the $\mu_*$-measurable sets is a measure) out of a finite-valued set $\mu : \mathcal{K} \to [0, \infty)$ defined on a $\pi$-class $\mathcal{K}$ with $\mu(\emptyset) = 0$. A good candidate is the following sup expression

$$
\mu_*(A) = \sup \{ \sum_{i=1}^n \mu(K_i) : \sum_{i=1}^n K_i \subset A, \, K_i \in \mathcal{K} \}, \quad \forall A \in 2^\Omega. \tag{2.3}
$$

Due to the supremum, there is not need to allow infinite series of sets inside $A$, but because $\mathcal{K}$ is only a $\pi$-class, a finite union is needed.

**Definition 2.21.** A set-function $\mu$ defined on a $\pi$-class $\mathcal{K}$ is called $\mathcal{K}$-tightness if for every $K$ and $K'$ in $\mathcal{K}$ with $K' \subset K$ we have $\mu(K) = \mu(K') + \mu_*(K \setminus K')$. Moreover, a finite-valued set-function $\mu$ is called $\sigma$-smooth on $\mathcal{K}$ at $\emptyset$ if for any decreasing sequence $\{\tilde{K}_n\}$ of finite disjoint unions of sets in $\mathcal{K}$, $\tilde{K}_n = \sum_{i<n} K_{n,i}$, with $\bigcap_n \tilde{K}_n = \emptyset$ then $\sum_{i<n} \mu(K_{n,i}) \to 0$ as $n \to \infty$. If $\mathcal{K} \subset 2^\Omega$ is a $\pi$-class then $\mathcal{K}$ is the lattice generated by $\mathcal{K}$, i.e., the class of finite unions of sets in $\mathcal{K}$. The class of all sets $F \subset \Omega$ satisfying $F \cap K \in \mathcal{K}$, for every $K \in \mathcal{K}$, is denoted by $\mathcal{K}_*$, and clearly, $\mathcal{K} \subset \mathcal{K}_*$. \hfill \Box
Contrary to the case of a semi-ring, additivity on a $\pi$-class is almost meaningless and replaced with the so-called $\mathcal{K}$-tightness as above, i.e., two conditions, (a) $\mu$ is monotone (i.e., $\mu(K') \leq \mu(K)$ if $K$ and $K'$ in $\mathcal{K}$ with $K' \subset K$) and (b) for every $\varepsilon > 0$ there exists a finite sequence of disjoint sets $\{K_i : i < n\} \subset \mathcal{K}$ such that $\sum_{i<n} K_i \subset K \setminus K'$ and $\mu(K) \leq \mu(K') + \varepsilon + \sum_{i<n} \mu(K_i)$. In this context, an important role is played by the lattice $\bar{\mathcal{K}}$ generated by $\mathcal{K}$ and the larger class $\bar{\mathcal{K}}_0 \supset \bar{\mathcal{K}}$. Moreover, it is clear that the property of being either $\mathcal{K}$-tightness or $\bar{\mathcal{K}}$-tightness are actually the same. Furthermore, the property of $\mu$ being $\sigma$-smooth on $\mathcal{K}$ at $\emptyset$ combined with $\mathcal{K}$-tightness could be call monotone continuity in $\emptyset$ on the lattice $\bar{\mathcal{K}}$, and clearly, it is only usable when $\mu$ is finite.

We are ready to state the main result

**Theorem 2.22.** Let $\mu$ be a finite-valued set defined on a $\pi$-class $\mathcal{K}$ with $\mu(\emptyset) = 0$. Then $\mu_*$ defined by (2.3) is an inner measure. Now, denote by $\mathcal{A}$ the algebra of $\mu_*$-measurable sets and assume that $\mu$ is $\mathcal{K}$-tight. Then

$$A \in \mathcal{A} \iff \mu(K) \leq \mu_*(K \cap A) + \mu_*(K \setminus A) \quad \forall K \in \mathcal{K},$$

(2.4)

the algebra $\mathcal{A}$ contains the class $\bar{\mathcal{K}}_*$ defined in Definition 2.21, and $\mu_*|_{\mathcal{K}} = \mu$. Moreover, if $\mu$ is $\sigma$-smooth on $\mathcal{K}$ at $\emptyset$ then $\mathcal{A}$ is a $\sigma$-algebra and $\mu_*$ is a semi-finite complete measure on $\mathcal{A}$, uniquely determined by $\mu$ on the $\mathcal{K}$, i.e., if $\nu$ is another semi-finite measure on a $\sigma$-algebra $\mathcal{F}$ with $\mathcal{K} \subset \mathcal{F} \subset \mathcal{A}$ such that $\nu|_{\mathcal{K}} = \mu$ then $\nu = \mu_*$ on $\mathcal{F}$.

**Proof.** If $E \subset F$ then the supremum defining $\mu_*(F)$ is taken over a larger family, so $\mu_*(E) \leq \mu_*(F)$. When $E \cap F = \emptyset$, each finite disjoint sequences $\{K_i\}$ and $\{K'_i\}$ with $\sum_i K_i \subset E$ and $\sum_i K'_i \subset F$ we can construct another finite disjoint sequence $\{K''_i\}$ with $\sum_i K''_i \subset E \cup F$ and $\sum_i \mu(K''_i) = \sum_i \mu(K_i) + \sum_i \mu(K'_i)$, which means that $\mu_*(E) + \mu_*(F) \leq \mu_*(E \cup F)$. This shows that $\mu_*$ is monotone and super-additive on $2^\Omega$, and thus (2.3) defines an inner measure $\mu_*$. Therefore, Proposition 2.20 implies that $\mu_*$ is an additive set function (i.e., a finite additive measure) on algebra $\mathcal{A}$ of all $\mu_*$-measurable sets.

Let $\mathcal{A}$ be a set satisfying $\mu(K) \leq \mu_*(K \cap A) + \mu_*(K \setminus A)$ for any $K$ in $\mathcal{K}$. Since $\mu_*$ is super-additive and monotone, if $\sum_{i=1}^n K_i = K \subset E$ with $K_i$ in $\mathcal{K}$ then

$$\sum_{i=1}^n \mu(K_i) \leq \sum_{i=1}^n \mu_*(K_i \cap A) + \sum_{i=1}^n \mu_*(K_i \setminus A) \leq \mu_*(K \cap A) + \mu_*(K \setminus A) \leq \mu_*(E \cap A) + \mu_*(E \setminus A),$$

and taking the supremum over all finite disjoint sequences $\{K_i\}$ we deduce $\mu_*(E) \leq \mu_*(E \cap A) + \mu_*(E \setminus A)$. The reverse inequality follows from the super-additivity, and therefore, $A$ belongs to $\mathcal{A}$. This shows (2.4) as desired.

The fact that $\mathcal{K}$ is stable under finite intersections was not used in the current (or the previous) paragraph, but it is needed for later arguments.

**Step 1 (with tightness)** From the definition of $\mu_*$ follows that $\mu(K) \leq \mu_*(K)$ for every $K$ in $\mathcal{K}$. Now, if $K''$ belongs to $\mathcal{K}$ then apply the tightness
property to any set \( K \) in \( \mathcal{K} \) and \( K' = K \cap K'' \) to get

\[
\mu(K) = \mu(K \cap K'') + \mu_*(K \setminus K'') \leq \mu_*(K \cap K'') + \mu_*(K \setminus K''),
\]

which implies, after invoking (2.4), that \( K'' \) belongs to the algebra \( \mathcal{A} \), i.e., \( \mathcal{K} \subset \mathcal{A} \). Moreover, if \( F \) belongs to \( \tilde{\mathcal{K}}_* \) then for any set \( K \) in \( \mathcal{K} \), the intersection \( K \cap F \) is a finite union of sets in \( \mathcal{K} \subset \mathcal{A} \). Thus \( K \setminus F = K \setminus (K \cap F) \) and \( K \cap F \) belong to the algebra \( \mathcal{A} \), and hence, the additivity of \( \mu_* \) yields

\[
\mu(K) \leq \mu_*(K) = \mu_*(K \cap F) + \mu_*(K \setminus F)
\]

which implies that \( F \) belongs to the algebra \( \mathcal{A} \), i.e., \( \tilde{\mathcal{K}}_* \subset \mathcal{A} \).

To show that \( \mu = \mu_* \) on \( \mathcal{K} \), pick \( K = \sum_{i=1}^n K_i \) with all sets in \( \mathcal{K} \) and use the tightness condition with \( K \) and \( K' = K_1 \) to obtain \( \mu(K) = \mu(K_1) + \mu_*(K \setminus K_1) \). Since \( \mu \leq \mu_* \) on \( \mathcal{K} \), \( K \setminus K_1 = \sum_{i=2}^n K_i \) and \( \mu_* \) is additive on \( \mathcal{A} \supset \mathcal{K} \) we have \( \mu_*(K \setminus K_1) \geq \sum_{i=2}^n \mu(K_i) \), which yields \( \mu(K) \geq \sum_{i=1}^n \mu(K_i) \), the super-additivity of \( \mu \). Therefore, the sup defining \( \mu_*(K) \) is achieved for \( K \) and \( \mu(K) = \mu_*(K) \) for every \( K \) in \( \mathcal{K} \), which means that \( \mu = \mu_* \) is additive on \( \mathcal{K} \).

Step 2 (\( \sigma \)-smooth) Even if we suppose that \( \mu_* \) is monotone continuous from above on \( \mathcal{K} \) at \( \emptyset \) (i.e., \( \sigma \)-smooth), then \( \mu_* \) is \( \sigma \)-additive on the algebra \( \mathcal{A} \), and therefore, Caratheodory extension Theorem 2.9 ensures that \( \mu_* \) can be extended to a measure on the \( \sigma \)-algebra generated by \( \mathcal{A} \), but a priori, the extension needs not to be preserve the sup representation (2.3).

The next point is to show that \( \mathcal{A} \) is a \( \mu_* \)-complete \( \sigma \)-algebra, independent of the fact that Caratheodory extension of \( (\mu_*, \mathcal{A}) \) yields a complete measure \((\tilde{\mu}_*, \mathcal{A})\). Actually, the completeness of \( \mu_* \) comes from Proposition 2.20.

Let us prove that \( \mu_* \) is \( \sigma \)-smooth on \( \mathcal{A} \) at \( \emptyset \), i.e., if \( \{A_n\} \subset \mathcal{A} \) is a decreasing sequence with \( \bigcap_n A_n = \emptyset \) and \( \mu_*(A_1) < \infty \) then \( \mu_*(A_n) \to 0 \). Indeed, the sup definition (2.3) of \( \mu_* \) ensures that for any \( \varepsilon > 0 \) and for any \( n \geq 1 \) there exist a finite disjoint union \( \tilde{K}_n \) of sets in \( \mathcal{K} \) such that \( \tilde{K}_n \subset A_n \) and \( \mu_*(A_n) - 2^{-n} < \mu_*(\tilde{K}_n) \). Define the decreasing sequence \( \{\tilde{K}_n\} \) with \( \tilde{K}_n = \cap_{i \leq n} \tilde{K}_i \) (which can be written as a finite disjoint union of sets in \( \mathcal{K} \)) and use the \( \sigma \)-smoothness property of \( \mu \) to obtain \( \mu_*(\tilde{K}_n) \to 0 \). Since the inclusion \( A_n \setminus \tilde{K}_n \subset \bigcup_{i \leq n} (A_i \setminus \tilde{K}_i) \) yields

\[
\mu_*(A_n \setminus \tilde{K}_n) \leq \sum_{i \leq n} \mu_*(A_i \setminus \tilde{K}_i) \leq \sum_{i \leq n} \varepsilon 2^{-i} \leq \varepsilon,
\]

and \( \mu_*(A_n) = \mu_*(\tilde{K}_n) + \mu_*(A_n \setminus \tilde{K}_n) \), we deduce that \( \mu_*(A_n) \to 0 \), i.e., \( \mu_* \) is \( \sigma \)-smooth on \( \mathcal{A} \) at \( \emptyset \).

Step 3 (finishing) Now, to check that \( \mathcal{A} \) is a \( \sigma \)-algebra, we have to show only that \( \mathcal{A} \) is stable under the formation of countable intersections, i.e., if \( \{A_i, i \geq 1\} \) is a sequence of sets in \( \mathcal{A} \) then we should show that \( \mathcal{A} = \cap_i A_i \) also belongs to \( \mathcal{A} \). For this purpose, from the sup definition (2.3) of \( \mu_* \) and because \( \mathcal{A} \) contains any finite union of sets in \( \mathcal{K} \), for any \( \varepsilon > 0 \) and for any set \( K \) in \( \mathcal{K} \) there exist a set \( A' \subset K \cap A \) in \( \mathcal{A} \) such that \( \mu_*(K \cap A) - \varepsilon < \mu_*(A') \). Thus, define the decreasing sequence \( \{B_n\} \) with \( B_n = \cap_{i \leq n} A_i \) to have \( \bigcap_n (K \cap B_n \cap A') = A' \).
and to use the $\sigma$-smoothness of $\mu_*$ on $\mathcal{A}$ with the sequence $(K \cap B_n \cap A') \setminus A$. Hence $\lim_n \mu_*(K \cap B_n \cap A') = \mu_*(A')$, which yields

$$\lim_n \mu_*(K \cap B_n) \geq \lim_n \mu_*(K \cap B_n \cap A') = \mu_*(A') > \mu_*(K \cap A) - \epsilon$$

and proves that $\lim_n \mu_*(K \cap B_n) = \mu_*(K \cap A)$. Recall that $B_n$ is in $\mathcal{A}$ to have

$$\mu(K) \leq \mu_*(K \cap B_n) + \mu_*(K \cap B_n) \leq \mu_*(K \cap B_n) + \mu_*(K \cap A),$$

and, after taking $n \to \infty$ and invoking the condition (2.4), to deduce that $A$ belongs to $\mathcal{A}$, i.e., $\mathcal{A}$ is a $\sigma$-algebra.

The final argument is to show that $\mu_*$ is $\sigma$-additive. Indeed, pick a sequence $\{A_n\} \subset \mathcal{A}$ with $A = \sum_n A_n$. If $A$ is a set in $\mathcal{A}$ with finite measure $\mu_*(A) < \infty$ then the $\sigma$-smoothness property of $\mu_*$ on $\mathcal{A}$ implies that $\mu_*(A \setminus \bigcup_{i<n} A_i) \to 0$, i.e., $\mu_*(A) = \sum_n \mu_*(A_n)$. If $\mu_*(A) = \infty$, the sup definition (2.3) ensures that there exists a sequence $\{A'_k\} \subset \mathcal{A}$ such that $A'_k \subset A$, $\mu_*(A'_k) < \infty$ and $\mu_*(A'_k) \to \infty$. Hence $\mu_*(A'_k) = \sum_n \mu_*(A'_k \cap A_n) \leq \sum_n \mu_*(A_n)$, and as $k \to \infty$ we deduce $\infty = \sum_n \mu_*(A_n)$, i.e., $\mu_*$ is $\sigma$-additive on the $\sigma$-algebra $\mathcal{A}$.

The uniqueness of $\mu_*$ is not really an issue, we have to show that if another semi-finite measure $\nu$ on a $\sigma$-algebra $\mathcal{F} \subset \mathcal{A}$ containing the class $\mathcal{K}$ and such $\nu = \mu$ on $\mathcal{K}$ then $\nu = \mu_*$ on $\mathcal{F}$. Indeed, they both agree on any set of finite measure (e.g., see Proposition 2.15), and for any set $F$ in $\mathcal{F}$ with infinite measure there exists a sequence $\{F_n\} \subset \mathcal{F}$ with $\nu(F_n) < \infty$, $F_n \subset F$ and $\lim_n \nu(F_n) = \nu(F)$, i.e., $\nu(F) = \mu_*(F)$ too.

Note that the $\sigma$-smoothness on $\mathcal{K}$ at $\emptyset$ and the $\mathcal{K}$-tightness assumptions are really conditions on the $\pi$-class $\tilde{\mathcal{K}}$ of all disjoint unions of sets in $\mathcal{K}$. Indeed, it is clear that if (a) $\mu$ is monotone on $\mathcal{K}$ and (b) $\mu$ is additive on $\mathcal{K}$ (i.e., $\mu(K) = \sum_{i<n} \mu(K_i)$ whenever $K = \sum_{i<n} K_i$ are sets in $\mathcal{K}$) then $\mu$ can be extended (in a unique way) to the $\pi$-class $\tilde{\mathcal{K}}$ preserving (a) and (b) by setting $\mu(\sum_{i<n} K_i) = \sum_{i<n} \mu(K_i)$. Therefore, $\mathcal{K}$-tightness translates into three properties: (a), (b) and (c) for every $K \supset K'$ sets in $\mathcal{K}$ (could be in $\tilde{\mathcal{K}}$) and every $\epsilon > 0$ there exists $\tilde{K} \subset K \setminus K'$ in $\tilde{\mathcal{K}}$ such that $\mu(K) \leq \mu(K') + \epsilon + \mu(\tilde{K})$. Similarly, $\sigma$-smoothness on $\mathcal{K}$ at $\emptyset$ translates into one condition: any decreasing sequence $\{\tilde{K}_n\}$ of sets in $\tilde{\mathcal{K}}$ such that $\bigcap \tilde{K}_n = \emptyset$ satisfies $\mu(\tilde{K}_n) \to 0$. With this in mind, there is not loss of generality if in Theorem 2.22 we assume that the $\pi$-class $\mathcal{K}$ is also stable under the formation of finite disjoint unions, see also Exercise 2.17.

- **Remark 2.23.** If the class $\mathcal{K}$ contains the empty set $\emptyset$, but it is not necessarily stable under finite intersections, then the sup-expression (2.3) defines an inner measure $\mu_*$. Hence, Proposition 2.20 proves that $\mu_*$ is a finitely additive set function on the algebra $\mathcal{A}$ of $\mu_*$-measurable sets. Moreover, if a subset $A$ of $\Omega$ satisfies $\mu(K) \leq \mu_*(K \cap A) + \mu_*(K \setminus A)$ for any $K$ in $\mathcal{K}$ then $A$ belongs to $\mathcal{A}$. However, it is not affirmed that $\tilde{\mathcal{K}} \subset \mathcal{A}$.

- **Remark 2.24.** If the finite-valued set function $\mu$ defined on the $\pi$-class $\mathcal{K}$ can be extended to a (finitely) additive set function $\bar{\mu}$ define on the semi-ring $\mathcal{S}$ generated by $\mathcal{K}$ then $\mu$ is necessarily $\mathcal{K}$-tight. Indeed, first recall that $\bar{\mu}$
is additive on the semi-ring $S$ if (by definition) $ar{\mu}(S) = \sum_{i<n} \bar{\mu}(S_i)$, for any finite sequence \{\(S_i, i < n\)\} of disjoint sets in $S$ with $S = \sum_{i<n} S_i$ also in $S$. Thus, if $K \supset K'$ are sets in $K$ then $K \setminus K'$ is a finite disjoint unions of sets in $S$, i.e., $K \setminus K' = \sum_{i<n} S_i$, and the additivity of $\bar{\mu}$ implies $\mu(K) = \mu(K') + \sum_{i<n} \bar{\mu}(S_i)$. Hence, this yields the following monotone property: if $K \supset K'$ are sets in $\mathcal{K}$ then $K \setminus K'$ is a finite disjoint union of sets in $S$, i.e., $K \setminus K' = \sum_{i<n} S_i$, and the additivity of $\bar{\mu}$ implies $\mu(K) = \mu(K') + \sum_{i<n} \mu(S_i)$, i.e., $\mu$ is $\mathcal{K}$-tight. Therefore, Proposition 2.11 on Caratheodory extension from a semi-ring and the previous Theorem 2.22 can be combined to show a $\sigma$-additive set function $\mu$ defined on a semi-ring $S$ can be extended to an inner measure by means of the sup expression (2.3) with $\mathcal{K}$ replaced by $S$. In this case $\mu_* \leq \mu^*$ in $2^\Omega$, and $\mu_* = \mu^*$ on the completion of the $\sigma$-algebra generated by $S$.

**Exercise 2.16.** With the notation of in Theorem 2.22, verify that the class $\mathcal{K}$ of all finite unions of sets in $\mathcal{K}$ is a lattice (i.e., a $\pi$-class stable under the formation of finite unions), and that the expression $\mu(K \cup K') = \mu(K) + \mu(K') - \mu(K \cap K')$ defines, by induction, a unique extension of $\mu$ to the class $\mathcal{K}$, which results $\mathcal{K}$-tight. At this point, instead of (2.3), we could redefine $\mu_*$ as

$$
\mu_*(A) = \sup \{\mu(K) : K \subset A, K \in \mathcal{K}\}, \quad \forall A \subset \Omega,
$$

where $\mu$ is a $\mathcal{K}$-tight finite-valued set function with $\mu(\emptyset) = 0$ defined on the lattice $\mathcal{K}$, i.e., without any loss of generality we could assume, initially, that the class $\mathcal{K}$ is a lattice.

**Exercise 2.17.** Let $\mathcal{K} \subset 2^\Omega$ be a lattice (i.e., a class containing the empty set and stable under finite unions and finite intersections) and $\mu : \mathcal{K} \to [0, \infty)$ be a finite-valued set function with $\mu(\emptyset) = 0$. Consider

$$
\mu_*(A) = \sup \{\mu(K) : K \subset A, K \in \mathcal{K}\}, \quad \forall A \subset \Omega,
$$

and assume that $\mu$ is $\mathcal{K}$-tight. Show that (1) $\mu$ can be uniquely extended to an (finitely) additive finite-valued set function $\bar{\mu}$ defined on the ring $\mathcal{R}$ generated by the lattice $\mathcal{K}$. Next, prove that (2) if $\mu$ is $\sigma$-smooth on $\mathcal{K}$ (i.e., $\lim_n \mu(K_n) = 0$ whenever $\{K_n\} \subset \mathcal{K}$ is a decreasing sequence with $\bigcap_n K_n = \emptyset$) then the extension $\bar{\mu}$ is $\sigma$-additive on $\mathcal{R}$.

The interested reader may check the books by Pollard [90, Appendix A, pp. 289–300] and Halmos [57, Section III.14, pp. 58–62]. For instance, the book by Cohn [28] could be used for even further details.

### 2.4 Geometric Construction

This is the third method for constructing a suitable measure in a more geometric way, and without any emphasis in extending a previously defined set-function on a small class. However, the space abstract $X$ need to have a topology, actually a metric $d$ so that $(X, d)$ is a metric space, usually separable. Actually, this
section is not a fully independent construction and proofs are not completely self-contained, a reference to a later chapter is necessary. For instance, the reader may take a look at Morgan [83] for a beginner’s guide.

As mentioned early, once a topology is given, the $\sigma$-algebras $B = B(X)$ generated by the open sets is called the Borel $\sigma$-algebra. A Borel measure is a measure define on the Borel $\sigma$-algebra $B$ (of a topological space $X$), which induces the outer measure

$$
\mu^*(A) = \inf \left\{ \mu(B) : A \supset B \in B \right\},
$$

which may not be a unique extension to an outer measure. If we begin with an outer measure $\mu^*$ then the restriction $\mu$ to the $\sigma$-algebra of $\mu^*$-measurable sets is a measure, but to recover these properties we need to impose two conditions: (a) $\mu^*$ is called a Borel outer measure if any Borel set is $\mu^*$-measurable and (b) $\mu^*$ is called a Borel regular outer measure if, beside being a Borel measure, for any subset $A \subset \Omega$ there exists a Borel set $B \supset A$ such that $\mu^*(A) = \mu(B)$. With this understanding, to insist in the outer measure character $\mu^*$ we refer to a Borel regular outer measure, and to insist in the measure character $\mu$ we refer to a Borel measure, but actually, they are the same concepts.

Let us go back to the construction in Proposition 2.6 of an outer measure. Suppose that $(X, d)$ is a separable metric space, $E$ is a family of subsets of $X$ and $b : E \to [0, \infty]$ satisfying:

(a) For every $\delta > 0$ there exist a sequence $\{E_n\}$ in $E$ such that $X = \bigcup_{n=1}^{\infty} E_n$ and $d(E_n) < \delta$,

(b) For every $\delta > 0$ there exist $E \in E$ such that $b(E) < \delta$ and $d(E) < \delta$,

where $d(E)$ means the diameter of $E$, i.e., $d(E) = \sup\{d(x, y) : x, y \in E\}$.

Therefore, for $\delta > 0$ and $A \subset X$ define

$$
\mu^*_\delta(A) = \inf \left\{ \sum_{n=1}^{\infty} b(E_n) : E_n \in E, A \subset \bigcup_{n=1}^{\infty} E_n, d(E_n) \leq \delta \right\},
$$

and we may replace the condition $d(E_n) \leq \delta$ by the strictly inequality $d(E_n) < \delta$, without any other changes. Condition (a) ensures that $\mu^*_\delta$ is properly defined for every $A \subset X$ and (b) implies that $\mu^*_\delta(\emptyset) = 0$. Thus, the set-function $\mu^*_\delta$ is an outer measure, but not necessarily an Borel outer measure. This construction is different from the one used in Propositions 2.6 and 2.11, in a way, the local geometry of the metric space $(X, d)$ is involved.

**Theorem 2.25.** Under the previous assumptions, the set-functions $\mu^*_\delta$ are non-increasing in $\delta$ and

$$
\mu^*(A) = \lim_{\delta \to 0} \mu^*_\delta(A) = \sup_{\delta > 0} \mu^*_\delta(A), \quad \forall A \subset X
$$

is a Borel outer measure. Moreover, if any set in $E$ is a Borel set, then $\mu^*$ is Borel regular outer measure.
Proof. First, by Proposition 2.6 each \( \mu_\delta^* \) is an outer measure. Thus the limit (or sup) \( \mu^* \) is monotone and \( \mu^*(\emptyset) = 0 \). To see that \( \mu^* \) is also sub-\( \sigma \)-additive, note that \( \sum_n \mu^*(A_n) \geq \sum_n \mu^*_\delta(A_n) \geq \mu^*_\delta(\bigcup_i A_i) \), for any \( \delta > 0 \). Therefore \( \mu^* \) is an outer measure.

To show that \( \mu^* \) is a Borel outer measure we need to invoke the so-called Caratheodory’s Criterion (of measurability), which is proved in a later chapter (see Proposition 3.7) and is more topological involved. Caratheodory’s Criterion affirms that \( \mu^* \) is a Borel measure if \( \mu^*(A \cup B) = \mu^*(A) + \mu^*(B) \) for every sets \( A, B \subset X \) with \( d(A, B) > 0 \). Thus, to establish this condition, for any sets \( A, B \subset X \) with \( d(A, B) > 0 \) choose \( \delta > 0 \) such that \( 2\delta < d(A, B) \). Hence, if the sequence \( \{ E_n \} \subset \mathcal{E} \) covers \( A \cup B \) and \( d(E_n) < \delta \) then none of them can meet \( A \) and \( B \) and so

\[
\sum_n b(E_n) \geq \sum_{A \cap E_n \neq \emptyset} b(E_n) + \sum_{B \cap E_n \neq \emptyset} b(E_n) \geq \mu^*_\delta(A) + \mu^*_\delta(B).
\]

This implies \( \mu^*_\delta(A \cup B) = \mu^*_\delta(A) + \mu^*_\delta(B) \) and this equality is preserved as \( \delta \to 0 \).

To check that \( \mu^* \) is regular, for any \( A \subset X \) and \( i = 1, 2, \ldots \), let us choose sets \( E_{i,n} \in \mathcal{E} \), for \( n = 1, 2, \ldots \), such that

\[
A \subset \bigcup_n E_{i,n}, \quad d(E_{i,n}) \leq 1/i, \quad \sum_n b(E_{i,n}) \leq \mu^*_1(A) + 1/i.
\]

Then \( B = \bigcap_i \bigcup_n E_{i,n} \) is a Borel set such that \( A \subset B \) and \( \mu^*(A) = \mu^*(B) \). \( \square \)

Sometimes we use the notation \( \mu^*(A) = \mu^*(A, \mathcal{E}) \) to emphasize the dependency on the cover class \( \mathcal{E} \). Note that if every \( E \in \mathcal{E} \) can be written as a countable disjoint union \( \bigcup_n E_n \) of elements in \( \mathcal{E} \) with \( d(E_n) < \delta \), and \( b \) is \( \sigma \)-additive on \( \mathcal{E} \), then both (Caratheodory and Hausdorff) constructions agree, i.e,

\[
\mu^*_\delta(A) = \inf \left\{ \sum_{n=1}^{\infty} b(E_n) : E_n \in \mathcal{E}, A \subset \bigcup_{n=1}^{\infty} E_n, d(E_n) \leq \delta \right\} = \mu^*(A),
\]

for every \( \delta > 0 \). Therefore, the interest situation for Hausdorff construction should occurs when \( b \) is not \( \sigma \)-additive.

Typically, we may take \( b = d^s \), where \( s \geq 0 \) represent the dimension. In this case, if the class \( \mathcal{K} \) of closed d-balls has the property: for every set \( E \) there exists a closed ball \( K \) such that \( E \subset K \) and \( d(E) = d(K) \) then \( \mu^*_\delta(A, \mathcal{K}) = \mu^*_\delta(A, 2^X) \), for every \( \delta > 0 \). Note that this property holds in \( \mathbb{R} \), but fails in higher dimension for the Euclidean norm, e.g., a equilateral triangle \( E \) in \( \mathbb{R}^2 \) can be contained in a circle of radius \( d(E)/2 \). Thus, the class \( \mathcal{E} \) used to take the infimum really could matter.

An outer measure \( \mu^* \) on \( X \) is called semi-finite (compare with Example 2.15) if for every \( A \subset X \) with \( \mu(A) = \infty \) we can find a sequence \( \{ A_n \} \) satisfying \( A_n \subset A \), \( \mu(A_n) < \infty \) and \( \mu(A_n) \to \infty \). From the definition \( \mu^* = \lim_{\delta \to 0} \mu^*_\delta = \sup_{\delta > 0} \mu^*_\delta \), it is clear that \( \mu^* \) is semi-finite if each \( \mu^*_\delta \) is so. However, even if each \( \mu^*_\delta \) is \( \sigma \)-finite, the Borel outer measure \( \mu^* \) is not necessarily \( \sigma \)-finite.

For instance, the book by Rogers [96] is dedicated to the study of the Hausdorff measures, in particular in \( \mathbb{R}^d \).
Exercise 2.18. Consider the space $\mathbb{R}^d$ with the non-Euclidean distance $d$ derived from norm $|x| = \max\{|x_i| : i = 1, \ldots, d\}$. Let $E$ be the semi-ring of all $d$-intervals of the form $[a, b]$ with $a$ and $b$ in $\mathbb{R}^d$, and for a fixed $r = 1, \ldots, d$, let $b_r$ be the hyper-volume of its $r$-projection, i.e.,

$$b_r(I) = (b_1 - a_i) \ldots (b_r - a_r), \quad \forall I = [a, b] \in E.$$ 

Denote by $\mu_r^*$ and $\mu_r^*$ the Hausdorff measure constructed from these $E$ and $b_r$. Prove (1) that $\mu_r^*$ with $r = d$ is the Lebesgue outer measure. Next, consider the injection and projection mappings $i_r : x' \mapsto x = (x', 0)$ from $\mathbb{R}^r$ into $\mathbb{R}^d$ and $\pi_r : x \mapsto x'$ from $\mathbb{R}^d$ into $\mathbb{R}^r$, when $r = 1, \ldots, d - 1$. Show (2) that $\mu_r^*(A) = \ell_r(i^{-1}(A))$, for every $A \subset i_r(\mathbb{R}^r)$, where $\ell_r$ is the Lebesgue outer measure. Also, show (3) that $\mu_r^*(A) = \infty$ if the projection $\pi_{r+1}(A)$ contains an open $(r+1)$-interval in $\mathbb{R}^{r+1}$, and thus, $\mu_r^*$ is not $\sigma$-finite in $\mathbb{R}^d$ for $r = 1, \ldots, d-1$, but only semi-finite. On the other hand, let $\mathcal{R}$ be the ring of all finite unions of semi-open cubes with edges parallel to the axis and with rational endpoints (i.e., $d$-intervals of the form $[a, b] \subset \mathbb{R}^d$ with $b_i, a_i$ rational numbers and $b_i - a_i = h$). Show (4) that $\mu_{r,\delta}^*(A, \mathcal{R}) = \inf \left\{ \sum_{n=1}^{\infty} d^r(E_n) : E_n \in \mathcal{R}, A \subset \bigcup_{n=1}^{\infty} E_n, d(E_n) \leq \delta \right\}$, and $\mu_r^*(\cdot, \mathcal{R}) = \lim_{\delta \to 0} \mu_{r,\delta}^*(\cdot, \mathcal{R})$ satisfy also (1), (2) and (3) above.

Three approaches for the construction of measures have been described, first the outer measure, which begins with almost no assumptions (Caratheodory’s construction Theorem 2.5 and Proposition 2.6), but they are really useful under the semi-ring condition of Proposition 2.11. Next, the inner measure, which begins from a $\pi$-class (Proposition 2.20 and Theorem 2.22), but it is mainly used in conjunction with topological spaces. Finally, this last approach (Hausdorff Construction Theorem 2.25), which has a geometric character and it can be viewed as a particular case of the outer measure construction. The key point is to force covers with sets of small diameters. Exercise 2.18 gives an idea of this situation, but it not adequate to handle “curves” sets, a more specific analysis is needed. The reader may find a deeper study in the book Lin and Yang [78] or Mattila [80]. Also, a quick look at the books Falconer [43] and Munroe [85] may be really beneficial.

## 2.5 Lebesgue Measures

Any of the previous three methods (outer, inner and geometric) for the construction of abstract measures can be used to define the so-called Lebesgue measure in $\mathbb{R}^d$, denoted by either $m$ or $\ell$. Several books are entirely dedicated to this purpose, with a great simplification of the previous section (but dealing with more details on other aspect of the theory), e.g., Burk [23], Jain and Gupta [64], Jones [65], among others.
As seen early, the Borel σ-algebra $\mathcal{B}(\mathbb{R}^d)$ can be generated by the class $\mathcal{I}_d$ of all $d$-dimensional intervals as
\[
[a, b] = \{ x \in \mathbb{R}^d : a_i < x_i \leq b_i, \ i = 1, \ldots, d \}, \quad \forall a, b \in \mathbb{R}^d, \ a \leq b,
\]
in the sense that $a_i \leq b_i$ for every $i$. The class $\mathcal{I}_d$ is a semi-ring in $\mathbb{R}^d$ and clearly, we can cover the whole space with an increasing sequence of intervals in $\mathcal{I}_d$. Sometimes, we prefer to use a semi-algebra of $d$-intervals, e.g., adding the cases $]-\infty, b_i]$ or $]a_i, +\infty[$ for $a_i, b_i \in \mathbb{R}$, among others, under the convention that $0\infty = 0$ in the product formula below. Therefore, to define a σ-finite measure on $\mathcal{B}(\mathbb{R}^d)$ (via the Caratheodory’s construction Proposition 2.11) we need only to know its (nonnegative real) values and to show that it is σ-additive on the class $\mathcal{I}_d$ of all $d$-dimensional intervals.

**Proposition 2.26.** The Lebesgue measure $m$, defined by
\[
m([a, b]) = \prod_{i=1}^{d} (b_i - a_i), \quad \forall a, b \in \mathbb{R}^d,
\]is σ-additive on $\mathcal{I}_d$.

**Proof.** Using the fact that for any two intervals $[a, b]$ and $[c, d]$ in $\mathcal{I}_d$ such that $[a, b] \cap [c, d] = \emptyset$ and $[a, b] \cup [c, d]$ belongs to $\mathcal{I}_d$ there exists exactly one coordinate $j$ such that $[a_j, b_j] \cap [c_j, d_j] = [a_j \wedge c_j, b_j \vee d_j]$ and $[a_i, b_i] = [c_i, d_i]$ for any $i \neq j$, it is relatively simple to check that the above definition produces an additive measure, and to show the σ-additivity, we use the character locally compact of $\mathbb{R}^d$. Indeed, let $I, I_n \in \mathcal{I}_d$ be such that $I = \sum_{n=1}^{\infty} I_n$, and for any $\varepsilon > 0$ define
\[
J_n = J_n(\varepsilon) = \{ x \in \mathbb{R}^d : a_{n,i} < x_i \leq b_{n,i} + 2^{-n}\varepsilon \},
\]
for $I_n = [a_n, b_n]$. It is clear that there is a constant $c > 0$ such that $b_{n,i} - a_{n,i} \leq c$, for every $n, i$, which yields the estimate
\[
0 \leq \sum_{n=1}^{\infty} (m(J_n) - m(I_n)) \leq C\varepsilon, \quad \forall \varepsilon \in (0, 1],
\]
for a suitable constant $C = C(c, d)$ depending only on $c$ and the dimension $d$. Similarly, if $I = [a, b]$ and $I_{\varepsilon} = \{ x \in \mathbb{R}^d : a_i + \varepsilon < x_i \leq b_i \}$, then $m(I_{\varepsilon}) \to m(I)$, as $\varepsilon$ decreases to 0.

Now, the interiors $\{ J_{n_i}^{\circ}(\varepsilon) \}$ constitute a sequence of open sets which cover the (compact) closure $\bar{I}_{\varepsilon}$, and therefore, there exists a finite subcover, namely $J_{n_1}^{\circ}(\varepsilon), \ldots, J_{n_k}^{\circ}(\varepsilon)$. Hence, $J_{n_1}(\varepsilon), \ldots, J_{n_k}(\varepsilon)$ will cover $I_{\varepsilon}$, and in view of the sub-additivity we deduce
\[
m(I_{\varepsilon}) \leq \sum_{i=1}^{k} m(J_{n_i}) \leq \sum_{n=1}^{\infty} m(J_n) \leq C\varepsilon + \sum_{n=1}^{\infty} m(I_n).
\]
Because $\varepsilon > 0$ is arbitrary, we get $m(I) \leq \sum_{n=1}^{\infty} m(I_n)$.

Finally, since $I \supset \sum_{n=1}^{k} I_n$, the additivity implies $m(I) \geq \sum_{n=1}^{k} m(I_n)$, and as $k \to \infty$ we conclude.

[2.5. Lebesgue Measures]
Usually, the measure \( m \) (or sometimes denotes by \( \ell \) or \( \ell_d \) to make explicit the dimension \( d \)) considered on the Borel \( \sigma \)-algebra is called Lebesgue-Borel measure and its extension (or completion) to the \( \sigma \)-algebra \( \mathcal{L} \) of all \( m^* \)-measurable sets is called the Lebesgue measure.

**Exercise 2.19.** Consider the outer Lebesgue measure \( \ell^* \) on \( (\mathbb{R}^d, \mathcal{L}) \). First, (1) verify that any Borel set is measurable and that the boundary \( \partial I \) of any semi-open (semi-close) \( d \)-interval \( I \) in the semi-ring \( \mathcal{I}_d \) has Lebesgue measure zero. Second, (2) show that for any subset \( A \) of \( \mathbb{R}^d \) and any \( \varepsilon > 0 \) there is an open set \( O \) containing \( A \) such that \( \ell^*(A) + \varepsilon \geq \ell(O) \). Deduce that also there is a countable intersection of open sets \( G \) containing \( A \) such that \( \ell^*(A) = \ell(G) \). \( \square \)

**Exercise 2.20.** Consider the Lebesgue measure \( \ell \) on \( (\mathbb{R}^d, \mathcal{L}) \). First, (1) show that for any measurable set \( A \) with \( \ell(A) < \infty \) and any \( \varepsilon > 0 \) there exits an open set \( O \) with \( \ell(O) < \infty \) and a compact set \( K \) such that \( K \subset A \subset O \) and \( \ell(O \setminus K) < \varepsilon \). Next, (2) prove that for every measurable set \( A \subset \mathbb{R}^d \) and any \( \varepsilon > 0 \) there exits a closed set \( C \) and an open set \( O \) such that \( C \subset A \subset O \) and \( \ell(O \setminus C) < \varepsilon \). Finally, if \( \mathcal{F}_\sigma \) denotes the class of countable unions of closed sets in \( \mathbb{R}^d \) and \( \mathcal{G}_\delta \) denotes the class of countable intersections of open sets in \( \mathbb{R}^d \) then (3) prove that for any measurable set \( A \) there exits a set \( G \) in \( \mathcal{G}_\delta \) and a set \( F \) in \( \mathcal{F}_\sigma \) such that \( F \subset A \subset G \) and \( \ell(G \setminus F) = 0 \). \( \square \)

**Exercise 2.21.** Consider the class \( \mathcal{I}_d \) of open bounded \( d \)-intervals in \( \mathbb{R}^d \) and the hyper-volume set function \( m \), i.e., of the form \( I = (a_1, b_1) \times \cdots \times (a_d, b_d) \), with \( a_i \leq b_i \) in \( \mathbb{R} \), \( i = 1, \ldots, d \), and \( m(I) = (b_1 - a_1) \cdots (b_d - a_d) \). Even if \( \mathcal{I}_d \) is not a semi-ring, we can define the outer measure

\[
m^*(A) = \inf \left\{ \sum_{n=1}^{\infty} m(I_n) : I_n \in \mathcal{I}_d, A \subset \bigcup_{n=1}^{\infty} I_n \right\}, \quad \forall A \subset \mathbb{R}^d,
\]

as in Caratheodory’s construction Proposition 2.6. Compare with the construction of the Lebesgue measure given in Proposition 2.26 and show that both definition are equivalent. \( \square \)

**Exercise 2.22.** Let \( \mathcal{J}_d \) be the class of all \( d \)-intervals in \( \mathbb{R}^d \) (which includes any open, non-open, closed, non-closed, bounded or unbounded intervals) with boundary points \( a = (a_i) \) and \( b = (b_i) \), where \( a_i \) and \( b_i \) belong to \([-\infty, +\infty]\). By considering in some detail the case \( d = 1 \) (with comments to the general case \( d \geq 2 \), do as follow:

1. Prove that \( \mathcal{J}_d \) is a semi-algebra of subsets of \( \mathbb{R}^d \).
2. Define an additive set function \( \bar{m} \) on \( \mathcal{J}_d \) such that \( \bar{m} \) restricted to the semi-ring \( \mathcal{I}_d \) of (left-open and right-closed) \( d \)-intervals used in Proposition 2.26 agrees with the expression (2.5). Extend the definition of \( \bar{m} \) to an additive set function on the algebra \( \mathcal{J}_d \) generated by the semi-algebra \( \mathcal{J}_d \).
3. Show that

\[
\bar{m}(J) = \sup \{ \bar{m}(K) : J \supseteq K, \text{ compact } K \in \mathcal{J}_d, \}
\]

for every \( J \) in \( \mathcal{J}_d \).
(4) Deduce from (3) that \( \bar{m} \) is \( \sigma \)-additive on \( \mathcal{J}_d \) and show that the extension of \( \bar{m} \) is also the Lebesgue measure.

**Exercise 2.23.** Consider the Lebesgue measures \( \ell_d, \ell_h \) and \( \ell_{d+h} \) on the spaces \( \mathbb{R}^d, \mathbb{R}^h \) and \( \mathbb{R}^{d+h} \). Discuss the additive product measure \( \ell_d \times \ell_h \) (see Remark 2.14), its outer measure extension \( \ell^* \) in \( \mathbb{R}^{d+h} \) and the \((d+h)\)-dimensional Lebesgue measure \( \ell_{d+h} \). Actually, verify that \( \ell_{d+h} \) is the completion of the product Lebesgue measure \( \ell_d \times \ell_h \).

Similarly to the construction of the Lebesgue measure in Proposition 2.26, we can check that if \( F_i: \mathbb{R} \to \mathbb{R} \), \( i = 1, \ldots, d \) then

\[
m_F([a,b]) = \prod_{i=1}^{d} (F_i(b_i) - F_i(a_i)), \quad \forall a, b \in \mathbb{R}^d
\]

defines a finitely additive measure \( \mathcal{I}_d \). To show the \( \sigma \)-additive on \( \mathcal{I}_d \) we proceed as above. The right-continuity is used to build the intervals \( J_n(\varepsilon) \) as follows

\[
J_n = J_n(\varepsilon) = \{ x \in \mathbb{R}^d : a_{n,i} < x_i \leq b_{n,i} + \delta_n \},
\]

where \( \delta_n > 0 \) is such that \( F_i(b_{n,i} + \delta_n) - F_i(b_{n,i}) < 2^{-n}\varepsilon \). Now, if \( C \) satisfies \( F_i(b_{n,i}) - F_i(a_{n,i}) \leq C \), for any \( n, i \), then estimate (2.6) remains true. This yields the Lebesgue-Stieltjes measure \( \mathcal{B}(\mathbb{R}^d) \), for instance, the reader may check the classic book by Munroe [85, Chapter III, Section 14, pp. 115–127].

**Exercise 2.24.** Let \( F_i: \mathbb{R} \to \mathbb{R}, i = 1, \ldots, d \), be non-decreasing functions. Verify that the expression (2.7) defines an additive set function \( m_F \) on the semi-ring \( \mathcal{I}_d \), which is \( \sigma \)-additive if each \( F_i \) is right-continuous. Give some details on how the alternative construction of the Lebesgue measure presented in the previous Exercise 2.22 can be used for the Lebesgue-Stieltjes measure.

Instead of having \( F = (F_1, \ldots, F_d) \) with \( F_i \) depending only on the single variable \( x_i \), we may have \( F: \mathbb{R}^d \to \mathbb{R} \). However, it is rather complicate to characterize the functions \( F \) of \( d \)-variables such that a product formula similar to the above produces a \( \sigma \)-additive measure. Hence, the Lebesgue-Stieltjes measure is mainly used on \( \mathbb{R} \) and \( F \) is referred to (in probability) as the cumulative distribution of \( m_F \).

Actually, the Lebesgue measure \( m \) (in general \( m_F \)) is complete and defined on \( \mathcal{L}(\mathbb{R}^d) \) (Lebesgue-measurable sets), which is a strictly larger \( \sigma \)-algebra than the (countable generated) Borel \( \sigma \)-algebra \( \mathcal{B}(\mathbb{R}^d) \). One way of seeing this fact is to construct a set \( C \) of measure zero with the cardinality of the continuum (e.g., the Cantor set) and then, any subset of \( C \) is measurable, i.e., \( \mathcal{L}(\mathbb{R}^d) \) has the cardinality of the \( 2^\mathbb{R} \), while \( \mathcal{B}(\mathbb{R}^d) \) has only the cardinality of the continuum. On the other hand, using the axiom of choice, we can construct a non measurable set. Thus \( \mathcal{B}(\mathbb{R}^d) \subset \mathcal{L}(\mathbb{R}^d) \subset 2^{\mathbb{R}^d} \) and all inclusions are strict (see Exercises below).
Exercise 2.25. Recall that the Cantor set $C$ as the set of all real numbers $x$ in $[0,1]$ expressed in the ternary system $x = \sum_n a_n 3^{-n}$ with $a_n \in \{0, 2\}$; and consider the Cantor function $f$ initially defined by $f(x) = \sum_n b_n 2^{-n}$, with $b_n = a_n/2$. Give a quick argument justifying that the Cantor set is uncountable and compact, with empty interior and no isolated points. Show that the Cantor set has Lebesgue measure zero, i.e., $m(C) = 0$. Extend $f$ to the function $f : [0,1] \to [0,1]$, which is constant on the complement $[0,1] \setminus C$ and strictly increasing on $C$ (except at the two endpoints of each interval removed). Again, verify that $f$ is a continuous function, e.g., see Folland [45, Proposition 1.22, pp. 38–39].

For the Lebesgue measure, any hyperplane perpendicular to any axis, e.g., $\pi = \{x \in \mathbb{R}^d : x_1 = 0\}$ has measure zero. Indeed, for any $\varepsilon > 0$ we use intervals of the form

$$I_n = \{x \in \mathbb{R}^d : -2^{-n} \varepsilon < x_1 \leq 2^{-n} \varepsilon, -n < x_i \leq n, i \geq 2\},$$

which cover the hyperplane $\pi$ and satisfy $m(I_n) = 2^{d-n}n^{d-1}\varepsilon$. Since

$$m(\pi) \leq \sum_n m(I_n) \leq \varepsilon \sum_n 2^{d-n}n^{d-1} < \infty,$$

we deduce $m(\pi) = 0$. Therefore, any countable union of hyperplane (as above) is a negligible set. Actually, any hyperplane has measure zero, as seen later (using the invariance under rotation of $m$).

Exercise 2.26. The translation invariance of the Lebesgue measure can be used to show the existence of a non-measurable set in $\mathbb{R}$. Indeed, consider the addition modulo 1 acting on $[0,1] \times [0,1]$, i.e., for any $a$ and $b$ in $[0,1)$ we set $a + b = c \mod 1$ with $c = a + b$ if $a + b < 1$ or $c = a + b - 1$ if $a + b \geq 1$. Verify that for any Lebesgue measurable set $E$ and any $b$ in $[0,1)$ the set $A + b \mod 1 = \{a + b \mod 1 : a \in A\}$ is Lebesgue measurable and $m(A + b \mod 1) = m(A)$. Next, define the equivalence relation $x \sim y$ if and only if $x - y$ is rational, and consider the family $\bar{x}$ of equivalence classes in $[0,1)$, i.e., $\bar{x} = \{y \in [0,1) : y \sim x\}$. Certainly, all rational numbers belong to the same equivalence class, and by means of the axiom of choice, we can select one (and only one) element of each equivalence class to form a subset $E'$ of $[0,1)$ such that (1) any two distinct element $x$ and $y$ in $E$ does not belong to the same equivalence class, and (2) $\bar{x}$ with $x$ in $E$ yield all possible equivalence classes. Let $\{r_n\}$ be an enumeration of the rational numbers in $[0,1)$ with $r_0 = 0$. Prove that $E_n = E + r_n \mod 1$ defines a sequence of disjoint sets in $[0,1)$, with $m(E_n) = m(E)$. Finally, show that $[0,1) = \bigcup_n E_n$ and deduce that $E$ cannot be a Lebesgue measurable set. Actually, this assertion can be rephrased as: Every Lebesgue measurable set of positive measure contains a set that is not Lebesgue measurable. For instance, the reader may check the book Burk [23, Appendix B, C] or Kharazishvili [69] to find a comprehensive discussion on non-measurable sets.

Exercise 2.27. Verify that if $I$ is a bounded $d$-intervals in $\mathbb{R}^d$ (which includes any open, non-open, closed, non-closed) with endpoints $a$ and $b$ then the Lebesgue measure $m(I)$ is equal the product $\prod_{i=1}^d (b_i - a_i)$. 

---

(1) Let $\mathcal{E}$ be the class of all open bounded intervals with rational endpoints, i.e., $(a, b)$ with $a$ and $b$ in $\mathbb{Q}^d$. Denote by $m_1^*$ and $m_1$ the outer measure and measure induced by the Caratheodory construction relative to the Lebesgue measure restricted to the (countable) class $\mathcal{E}$. Check that a subset $A$ of $\mathbb{R}^d$ is $m_1^*$-measurable if and only if $A$ is Lebesgue measurable, and prove that $m_1 = m$. Can we show that $m_1^* = m^*$?

(2) Similarly, let $\mathcal{C}$ be the class of all open cubes with edges parallel to the axis with rational endpoints (i.e., $(a, b)$ with $a, b$ in $\mathbb{Q}^d$ and $b_i - a_i = r$, for every $i$), and let $(m_2^*)$ $m_2$ be the corresponding (outer) measure generated as above, with $\mathcal{C}$ replacing $\mathcal{E}$. Again, prove results similar to item (1). What if $\mathcal{E}$ is the class of semi-open dyadic cubes $[(i - 1)2^{-n}, i2^{-n}]^d$ for $i = 0, \pm 1, \ldots \pm 4^n$?

(3) How can we extend all these arguments to the Lebesgue-Stieltjes measure. State precise assertions with some details on their proof.

(4) Consider the class $\mathcal{D}$ of all open balls with rational centers and radii. Repeat the above arguments and let $(m_3^*)$ $m_3$ (outer) measure associated with the class $\mathcal{D}$. How can we easily verify the validity of the previous results for this setup (see later Corollary 2.35).

\begin{exercise}
Let $F: \mathbb{R} \to \mathbb{R}$ be a nondecreasing right-continuous functions and $m_F$ be its Lebesgue-Stieltjes measure associated. Verify that (1) $m_F([a, b]) = F(b) - F(a)$. Prove that (2) there is a bijection between the Lebesgue-Stieltjes measures in $\mathbb{R}$ and the semi-space of all nondecreasing right-continuous functions $F$ from $\mathbb{R}$ into itself satisfying $F(0) = 0$. Finally, (3) can we do the same for $\mathbb{R}^d$?
\end{exercise}

\begin{exercise}
Let $F: \mathbb{R} \to \mathbb{R}$ be a nondecreasing right-continuous functions and $m_F$ be its corresponding Lebesgue-Stieltjes measure, i.e., $m_F([a, b]) = F(b) - F(a)$. Now consider the measure $m_F$ restricted to the interval $[0, 1]$, where $F = f$ is now the Cantor function as in Exercise 2.25. Verify that $m_F(C) = 1$ and $m_F([0, 1] \setminus C) = 0$, where $C$ is Cantor set.
\end{exercise}

\begin{exercise}
Let $F : \mathbb{R} \to \mathbb{R}$ be a nondecreasing function, define $G(x) = F(x^+) = \lim_{y \to x^+} F(y)$ and consider the (Lebesgue-Stieltjes type) measures $m_F$ and $m_G$ induced by $F$ and $G$, via the expressions $\tilde{m}_F([a, b]) = F(b) - F(a)$ and $\tilde{m}_G([a, b]) = G(b) - G(a)$, respectively, and the semi-ring $\mathcal{I}$ of intervals $I = (a, b)$, with $a$ and $b$ in $\mathbb{R}$.

(1) Show that $G$ is right-continuous, and that

$$G(x-) = \lim_{y \to x^-} G(x) = \lim_{y \to x^-} F(x) = F(x^-).$$

Also give some details on the construction of the measures $m_F$ and $m_G$. Verify that the expressions of $m_F$ and $m_G$ does not change if we assume that $F(0^+) = G(0) = 0$.

(2) Prove that any Borel set is measurable relative to either $m_F^*$ or $m_G^*$. Check that $m_G((a, b]) = G(b) - G(a)$ and $m_F((a, b]) \leq F(b) - F(a)$.
(3) Verify that for every singleton (set of only one point) \( \{x\} \) we have
\[
m_F(\{x\}) = F(x) - F(x-) \leq G(x) - G(x-) = m_G(\{x\}),
\]
and deduce (a) if \( F \) is left-continuous then \( m_F \) has no atoms, (b) any atom of \( m_F \) is also an atom of \( m_G \) and (c) a point \( \alpha \) is an atom for \( m_G \) if and only if \( \alpha \) is a point of discontinuity for \( G \) (or equivalently, if and only if \( \alpha \) is a point of discontinuity for \( F \)). Moreover, calculate \( m_F(I) \) and \( m_G(I) \), for any interval \( I \) (non necessarily of the form \( [a,b] \)) with endpoints \( a \) and \( b \), none of them being atoms. Can you find \( m_F(I) \)?.

(4) Deduce that for every bounded interval \( [a,b] \) and \( c > 0 \), there are only finite many atoms \( \{a_n\} \subset [a,b] \) (possible none) with \( m_F(\{a_n\}) \geq c > 0 \). Thus, conclude that \( F \) and \( G \) can have only countable many points of discontinuities, and \( \sum_n b_n \mathbf{1}_{b_n<\varepsilon} \to 0 \) as \( \varepsilon \to 0 \), for either \( b_n = m_F(\{a_n\}) \) or \( b_n = m_G(\{a_n\}) \), with \( \{\alpha_k\} \) the sequence of all atoms in \( [a,b] \).

(5) Assume that \( F \) is a nondecreasing purely jump function, i.e., for some sequence \( \{\alpha_i\} \) of points and some sequence \( \{f_i\} \) of positive numbers we have \( H_l \leq F \leq H_r \), where
\[
H_l(x) = \sum_{0 < \alpha_i < x} f_i, \quad \forall x > 0 \quad \text{and} \quad H_l(x) = \sum_{x \leq \alpha_i \leq 0} f_i, \quad \forall x \leq 0,
\]
\[
H_r(x) = \sum_{0 < \alpha_i \leq x} f_i, \quad \forall x > 0 \quad \text{and} \quad H_l(x) = \sum_{x < \alpha_i \leq 0} f_i, \quad \forall x \leq 0.
\]
Verify that \( H_l \) is a left-continuous and \( H_r \) is right-continuous, and if \( H_l^\varepsilon \) and \( H_r^\varepsilon \) are as above with \( \{f_i\} \) replaced with \( \{f_i^\varepsilon\} \), \( f_i = f_i \mathbf{1}_{f_i \geq \varepsilon} \), then \( H_l^\varepsilon \) and \( H_r^\varepsilon \) have only a finite number of jumps, and \( H_l^\varepsilon \to H_l \) and \( H_r^\varepsilon \to H_r \), uniformly on any bounded interval \( [a,b] \).

(6) Prove that if \( F = H_r \) then \( m_F^* \) is a purely atomic measure, i.e., \( \{\alpha_i\} \) are the atoms of \( m_F^* \) and \( m_F^*(A) = \sum_{\alpha_i \in A} f_i \), for every \( m_F^* \)-measurable set \( A \subset \mathbb{R} \). Similarly, if \( F = H_l \) then \( m_F^* = 0 \).

(7) If \( F(x-) \) denotes the left-hand limit of \( F \) at the point \( x \) then define the function
\[
\bar{F}_r(x) = \sum_{0 < y \leq x} [F(y) - F(y-)] \quad \text{or} \quad \bar{F}_r(x) = \sum_{x < y \leq 0} [F(y-) - F(y)],
\]
depending on the sign of \( x \). Prove that \( \bar{F}_r \) is a nondecreasing purely jump right-continuous function and that \( \bar{F}_l = F - \bar{F}_r \) is a nondecreasing left-continuous function. Mimic the above argument to construct a nondecreasing purely jump left-continuous function \( F_l \) such that \( F_l = F - F_r \) is a nondecreasing right-continuous function. Next, show that the measures \( m_F \) is actually equal to \( m_{F_l} + m_{F_r} \), and in view of part (5), deduce that actually \( m_F = m_{F_r} \). Calculate \( m_F(I) \), for every \( I \in \mathcal{I} \) and check that \( m_F^* \leq m_G^* \).
The reader interested in the Lebesgue measure on $\mathbb{R}^d$ may check the books either Gordon [54, Chapters 1 and 2, pp. 1–27] or Jones [65, Chapters 1 and 2, pp. 1-63] where a systematic approach of the Lebesgue measure and measurability is considered, in either a one dimensional or multi-dimensional settings. Also, see Shilov and Gurevich [106, Chapter 5, 88–110] and Taylor [114, Chapter 4, 177–225].

2.5. Lebesgue Measures

Invariant Under Translations

From the definition of the Lebesgue measure, we can check that $m$ is invariant under translations, i.e., for a given $h \in \mathbb{R}^d$ we have that $E$ measurable implies $E + h = \{ x \in \mathbb{R}^d : x - h \in E \}$ measurable and $m(E + h) = m(E)$. We will see later that the same is true for a rotation, i.e. if $r$ is an orthogonal $d$-dimensional matrix and $E$ is measurable then $r(E) = \{ x \in \mathbb{R}^d : r(x) \in E \}$ is measurable and $m(r(E)) = m(E)$. Moreover, we have

**Theorem 2.27** (invariance). Let $T$ be an affine transformation from $\mathbb{R}^d$ into itself with the linear part represented by a $d$-square matrix, also denoted by $T$. Then for every $A \subset \mathbb{R}^d$ we have $m^*(T(A)) = |\det(T)| m^*(A)$, where $\det(T)$ is the determinant of the matrix $T$ and $m^*$ is the Lebesgue outer measure on $\mathbb{R}^d$.

**Proof.** First the translation part of the affine transformation has already been considered, so only the linear part has to be discussed. Secondly, recall that an elementary matrix $E$ produces one of the following row operations (1) interchange rows, (2) multiply a row by a non zero scalar, (3) replace a row by that row minus a multiple of another. Next, any invertible matrix can be expressed as a finite product of elementary matrix of the type (1), (2) and (3). Thus, if $T$ is invertible, we need only to show the result for elementary matrix of type (2) and (3), since the expression of the Lebesgue measure is clearly invariant under a transformation of type (1).

Let $T$ be an elementary matrix and for the reference $d$-interval $J = [0,1] \times \cdots \times [0,1]$ define $\alpha = m(T(J))$. If $T$ is of type (2) and $c$ is the corresponding scalar then one (and only one) of the interval $[0,1]$ becomes either $[0,c]$ or $[c,0]$, i.e., $m(T(J)) = |c| = |\det(T)|$. On the other hand, if $T$ is of type (3) then we get also $\alpha = |\det(T)|$, e.g., $T$ replaces row 1 by the result of row 1 plus $c$ times row 2, and working with $d = 2$, the reference square for $J$ becomes a rhombus $T(J)$ with base and hight 1 (the $c$ only twist the square). Here, we need to verify that the measure of a right triangle is its area. This proves that $m(T(J)) = |\det(T)| m(J)$. By iteration, $T$ can be replaced by a product of elementary matrices. In particular, the case of a dilation $x \mapsto rx$ we have $m(rJ) = r^d m(J)$.

Let us now look at the general case $m^*(T(A))$ with $A \subset \mathbb{R}^d$ and $T$ elementary matrix. Again, to show this point we need to consider only the case of an open set $A$. Note that $T$ and it inverse $T^{-1}$ are continuous, so that $A$ is open (or compact) if and only if $T(A)$ is so. Thus, for a given open set $A$, first pave $\mathbb{R}^d$ with $d$-intervals $[a_1, a_1+1] \times \cdots \times [a_d, a_d+1]$, with $a_i$ integers, and select those $d$-intervals inside $A$. Then pave each unselected $d$-interval with $2^d$ $d$-intervals by
bisecting the edges of the original $d$-intervals, the resulting $d$-intervals have the form $[a_1/2, a_1/2 + 1/2] \times \cdots \times [a_d/2, a_d/2 + 1/2]$, with $a_i$ integers. Now, select those $d$-intervals inside $A$. By continuing this procedure, we have $A = \bigcup_{k=1}^{\infty} J_k$ where the $J_k$ are disjoint $d$-intervals and each of them is a translation of a dilation of the reference $d$-interval $J$, i.e., $J_k = t_k + r_k J$. As mentioned before, translation does not modify the measure and a $(r_k)$ dilation amplify the measure (by a factor of $|r_k|^d$), i.e., $m(J_k) = |r_k|^d m(J)$. Since $T(J_k) = T(t_k) + r_k (T(J))$, the previous argument shows that $m(T(J_k)) = |\det(T)| m(J_k)$. Hence, by the $\sigma$-additivity $m(T(A)) = m(A)$.

Finally, if $T$ is not invertible then $\det(T) = 0$ and the dimension of $T(\mathbb{R}^d)$ is strictly less than $d$. As mentioned early, any hyperplane perpendicular to any axis, e.g., $\pi = \{x \in \mathbb{R}^d : x_1 = 0\}$ has measure zero, and then for any invertible linear transformation (in particular orthogonal) $S$ we have $m(S(\pi)) = |\det(S)| m(\pi) = 0$, i.e., any hyperplane has measure zero. In particular, we have $m(T(\mathbb{R}^d)) = 0$.

**Remark 2.28.** As a consequence of Theorem 2.27, for any given affine transformation $T$ from $\mathbb{R}^d$ into itself, we deduce that $T(E)$ is $m^*$-measurable if and only if $E$ is $m^*$-measurable. Note that the situation is far more complicate for an affine transformation $T : \mathbb{R}^d \to \mathbb{R}^n$ and we use the Lebesgue (outer) measure $(m^*) m$ on $\mathbb{R}^d$ and $\mathbb{R}^n$, with $d \neq n$, see later sections on Hausdorff measure.

Another approach (to the Euclidean invariance property of the Lebesgue measure) is to establish first that a continuous functions from a closed set into $\mathbb{R}^d$ preserves $F_\sigma$-sets, and a function that preserves sets null sets is also measurable. Next, consider a Lipschitz transformation $f : \mathbb{R}^d \to \mathbb{R}^n$, $d \leq n$, i.e., $|f(x) - f(y)| \leq L|x - y|$ for every $x,y$ in $\mathbb{R}^d$, and verify the inequality $|m^*_n(f(Q))| \leq (2\sqrt{n})^n m^*_d(Q)$, for every cube $Q$ in $\mathbb{R}^d$, where $m^*_n$ and $m^*_d$ denote (Lebesgue) outer measure in $\mathbb{R}^n$ and $\mathbb{R}^d$. Finally, with arguments similar to those found in the following section, the inequality remains valid for any subset $Q = A$ of $\mathbb{R}^d$.

The reader may want to consult the book Stroock [112, Section 2.2, pp. 30–33].

**Exercise 2.31.** Let $D$ be a closed set in $\mathbb{R}^d$ and $f : D \to \mathbb{R}^n$ be a continuous function. Prove that if $A \subset \mathbb{R}^d$ is a $F_\sigma$-set (i.e., a countable union of closed sets) so is $f(D \cap A)$. Also show that if $f$ maps sets of (Lebesgue) measure zero into sets of measure zero, then $f$ also maps measurable sets into measurable sets.

**Exercise 2.32.** First give details on how to show that an hyperplane in $\mathbb{R}^d$ has zero Lebesgue measure. Second, verify that if $B$ and $\overline{B}$ denote the open and closed ball of radius $r$ and center $c$ in $\mathbb{R}^d$, then $m_d(B) = m_d(\overline{B}) = c_d r^d$, where $c_d$ is the Lebesgue measure of the unit ball.

**Vitali’s Covering**

Many constructions in analysis deal with a non-necessarily countable collection $\{B_i : i \in I\}$ of closed balls (where each ball satisfies a certain property) which cover a set $A$ in a special way, namely, $\inf \{r_i : x \in B_i, i \in I\} = 0$ for every $x$
in $A$, where $r_i$ denotes the radius of the ball $B_i$. The family \( \{B_i : i \in I\} \) of closed balls is called a fine cover (or a Vitali Cover or a cover in Vitali Sense) of a set $A$. In words this means that for every $x$ in $A$ and any $\varepsilon > 0$ there exists a ball in the family of radius at most $\varepsilon$ and containing $x$. It is clear that if what is given is a collection \( \{B_i : i \in I\} \) of open balls satisfying $\inf \{r_i : x \in B_i, i \in I\} = 0$ for every $x$ in $A$, then the collection \( \{B_i : i \in I\} \) of closed balls (i.e., $B_i$ is the closure of $B_i$) is indeed a fine cover of $A$, but the converse may not be true.

The interest is to be able to select a countable sub-family of disjoint closed balls that cover $A$ in some sense.

For instance, if $O$ is an open subset of $\mathbb{R}^d$ then $O$ can be expressed as a countable union on closed balls, e.g., take all balls included in $O$ with rational center and radii. Moreover,

\[
O = \bigcup \{a + rB^1 \subset O : a \in \mathbb{Q}^d, r \in \mathbb{Q} \cap (0, \varepsilon]\},
\]

for every $\varepsilon > 0$. However, two balls in the family could be overlapping, i.e., having an intersection with nonempty interior. Nevertheless, based on what follows, there exists a countable family of disjoint closed balls contained in $O$ such that $m(O \setminus \bigcup_i B_i) = 0$, where $m(\cdot)$ is the Lebesgue measure in $\mathbb{R}^d$.

In the sequel, it is convenient to denote by $B^1$ the closed unit ball in $\mathbb{R}^d$, and by $x_0 + rB^1$ the closed ball centered at $x_0$ with radius $r > 0$. First, a general selection argument

**Theorem 2.29.** Let \( \{B_i : i \in I\} \) be a nonempty collection of closed balls in $\mathbb{R}^d$ such that $B_i = x_i + r_i B^1$, $0 < r_i \leq C < \infty$, for every $i$ in $I$. Then there exists a countable sub-collection \( \{B_i : i \in J\} \), $J$ a countable subset of indexes in $I$, of disjoint balls such that $\bigcup_{i \in J} B_i \subset \bigcup_{i \in J} B^5_i$, where $B^5_i = x_i + 5r_i B^1$. Moreover, if $\{B_i : i \in I\}$ is also a fine cover of $A \subset \mathbb{R}^d$, i.e., $\inf \{r_i : x \in B_i, i \in I\} = 0$ for every $x$ in $A$, then the previous countable sub-collection $\{B_i : i \in J\}$ of disjoint balls is such that for every finite family $\{B_k : k \in K\}$, $K$ a finite subset of indexes in $I$, we have $A \setminus \bigcup_{k \in K} B_k \subset \bigcup_{j \in J \setminus K} B^5_j$.

**Proof.** For $n = 1, 2, \ldots$, set $I_n = \{i \in I : r 2^{-n} < r_i \leq r 2^{1-n}\}$, where $r = \sup \{r_i : i \in I\}$. Certainly, $I_1$ is nonempty. Now, define $J_n \subset I_n$ by induction as follows:

(a) Take any $i_0$ in $I_1$ and then keep choosing $i_1, i_2, \ldots$, in $I_1$ so that the sets $B_{i_0}, B_{i_1}, B_{i_2}, \ldots$ are disjoint closed balls. This procedure defines a countable set of indexes, finite or infinite, of closed disjoint balls. Thus, there exits a subset of indexes $J_1$ in $I_1$ such that $\{B_i : i \in J_1\}$ is a maximal disjoint collection of closed balls, i.e., for every $i$ in $I_1$ there exists $j$ in $J_1$ such that $B_i \cap B_j \neq \emptyset$.

(b) Given $J_1, J_2, \ldots, J_{n-1}$ we define $K_n$ as the subset of indexes in $I_n$ such that $B_i \cap B_k = \emptyset$, for every $i$ in $J_1 \cup \cdots \cup J_{n-1}$ and $k$ in $K_n$. Now, if $K_n$ is not empty then, as in (a), there exists a subset of indexes $J_n$ in $K_n$ such that $\{B_i : i \in J_n\}$ is a maximal disjoint collection of closed balls, i.e., for every $k$ in $K_n$ there exists $j$ in $J_n$ such that $B_k \cap B_j \neq \emptyset$. Moreover, for every $i$ in $I_n$ there exists $j$ in $J_1 \cup \cdots \cup J_n$ such that $B_i \cap B_j \neq \emptyset$.
Hence, for every $i$ in $I$ there exits some $n$ such that $i$ belongs to $I_n$. The maximality of $\{B_j : j \in J_n\}$ proves that there exits $j$ in $J_1 \cup \cdots \cup J_n$ such that $B_i \cap B_j \neq \emptyset$, which yields that the distance from $x_i$ to $x_j$ is less that $r_i + r_j$, with $B_i = x_i + r_i B^1$, and $B_j = x_j + r_j B^1$. Because $r_i \leq r 2^{1-n}$ and $r_j > r 2^{-n}$, we have $r_i \leq 2r_j$, which implies that the distance from $x_i$ to $x_j$ is less than $2r_j$. Now, for any point $x$ in $B_i$ has a distant to $x_j$ smaller than $r_i + 3r_j \leq 5r_j$, i.e., $B_i \subset B_j^5$. Thus, we obtain the desired subcover with $J = \bigcup_{n=1}^{\infty} J_n$.

Finally, let $A \subset \mathbb{R}^d$, $\{B_i : i \in I\}$ a fine cover of $A$, and $\{B_k : k \in K\}$ with $K$ a finite subset of indexes in $I$. Suppose $x$ in $A \setminus \bigcup_{k \in K} B_k$. Since the balls are closed and the cover is fine, there exits $i$ in $I$ such that $x$ belongs to $B_i$ for some $i$ in $I$ and $B_i \cap B_k = \emptyset$, for every $k$ in $K$. Since $i$ belongs to some $I_n$ there exits $j$ in $J_1 \cup \cdots \cup J_n$ such that $B_i \cap J_j \neq \emptyset$, i.e., $j$ must belong to $J \setminus K$ and $B_i \subset B_j^5$. Therefore, $A \setminus \bigcup_{k \in K} B_k \subset \bigcup_{j \in J \setminus K} B_j^5$.

- **Remark 2.30.** If the initial collections of closed balls is countable then to construct the indexes $J_n$, we do not need to invoke the (uncountable version) Axiom of Choice and Zorn’s Lemma to obtain a maximal element.

- **Remark 2.31.** The following is a useful variation on the statement of Theorem 2.29. Let $\{B_1, \ldots, B_n\}$ be a finite family of balls in a metric space $(X,d)$. Then there exists a subset collection $\{B_{i_1}, \ldots, B_{i_m}\}$ of disjoint balls such that

$$
\bigcup_{i=1}^{n} B_i \subset B_{i_1}^3 \cup \cdots \cup B_{i_m}^3.
$$

Indeed, first we choose a ball with largest radius, then a second ball of largest radius among those not intersecting the first ball, then a third ball of largest radius among those that not intersecting the union of the first and of the second ball, and so on. The process stops when either there is no ball left, or when the remaining balls intersect at least one of the balls already chosen. The family of chosen balls is disjoint by construction, and a ball was not chosen because it intersects some (previously) chosen ball with a larger (or equal) radius. Thus, if $x$ belongs to a ball $B_i$, centered at $x_i$ with radius $r_i$, and the ball $B_i$ has not been chosen, then there is a chosen ball $B_j$, centered at $x_j$ with radius $r_j \geq r_i$, intersecting $B_i$, which implies $d(x_i, x_j) \leq r_i + r_j \leq 2r_j$. Hence,

$$
d(x, x_j) \leq d(x, x_i) + d(x_i, x_j) \leq r_i + 2r_j \leq 3r_j,
$$

i.e., $x$ belongs to $B_j^3 = x_j + 3r_j B^1$, and the desired property is proved.

- **Remark 2.32.** If $\{B_i : i \in I\}$ is a family of open balls in a metric space $(X,d)$, and $B = \bigcup_{i \in I} B_i$ then for any compact set $K \subset B$ there exists a finite covering of $K$, to which the argument of the previous Remark 2.31 can be applied. Thus there exists a finite number of disjoint balls $\{B_{i_1}, \ldots, B_{i_n}\}$ such that $K \subset B_{i_1}^3 \cup \cdots \cup B_{i_n}^3$.

The following result (which has many applications) is usually called a simple Vitali Lemma, e.g., see Wheeden and Zygmund [119, Lemma 7.4, pp. 102–104].
Corollary 2.33. Let \( \{B_i : i \in I\} \) be a collection of closed balls covering a subset \( A \) of \( \mathbb{R}^d \) with finite outer Lebesgue measure \( m^*(A) < \infty \). Then for every \( \alpha > 5^d \) there exists a finite number of disjoint balls \( \{B_k : k \in K\} \), with \( K \) a finite subset of indexes in \( I \), such that \( m^*(A) \leq \alpha \sum_{k \in K} m(B_k) \).

Proof. Indeed, similar to the previous argument, by means of Theorem 2.29, there exists a countable collection of closed disjoint balls \( \{B_j : j \in J\} \), \( J \) a countable subset of indexes in \( I \), such that \( A \subset \bigcup_{j \in J} B_j^5 \). Thus

\[
m^*(A) \leq \sum_{j \in J} m(B_j^5) = 5^d \sum_{j \in J} m(B_j).
\]

Since \( m^*(A) < \infty \), if the series diverges then for any \( \alpha > 0 \) we can find a finite set \( K \subset J \) satisfying the required condition. However, if the series converges and \( \alpha > 5^d \) then there exists a finite set \( K \subset J \) such that

\[
\sum_{k \in K} m(B_k) > \frac{5^d}{\alpha} \sum_{j \in J} m(B_j),
\]

which yields \( m^*(A) \leq \alpha \sum_{k \in K} m(B_k) \) to conclude the argument. \( \square \)

- Remark 2.34. Note that given \( A \) with \( m^*(A) < \infty \) and \( \alpha > 3^d \) there exists a compact \( K \) such that \( \alpha m(K) > 3^d m^*(A) \). Hence, the argument of Remark 2.32 shows that if \( \{B_i : i \in I\} \) is a collection of open balls covering \( A \) then there exists a finite number of disjoint open balls \( \{B_{i_1}, \ldots, B_{i_n}\} \) such that \( m^*(A) \leq \alpha(m(B_{i_1}) + \cdots + m(B_{i_n})) \), e.g., see Taylor [115, Chapter 1, pp. 139–156]. \( \square \)

Now, another variation a little bit more delicate

Corollary 2.35. Let \( \{B_i : i \in I\} \) be a collection of closed balls which is a fine cover of subset \( A \) of \( \mathbb{R}^d \) with finite outer Lebesgue measure \( m^*(A) < \infty \). Then for every \( \varepsilon > 0 \) there exists a countable sub-collection \( \{B_j : j \in J\} \) of disjoint closed balls such that \( m^*(A \setminus \bigcup_{j \in J} B_j) = 0 \) and \( \sum_{j \in J} m(B_j) < (1 + \varepsilon)m^*(A) \).

Proof. Because \( m^*(A) < \infty \), applying Exercise 2.19 (or equivalently, the fact that the Lebesgue measure is a Borel measure), for every \( \varepsilon > 0 \) there exists an open set \( O \) such that \( O \supseteq A \) and \( m^*(O) < (1 + \varepsilon)m^*(A) \).

Now, let us show that for every \( \varepsilon > 0 \) and for any fixed constant \( 1 - 5^{-d} < a < 1 \), there exits a finite subcover of closed balls \( \{B_k : k \in K\} \), \( K \) finite, such that \( m(O \setminus \bigcup_{k \in K} B_k) \leq a m(O) \). Indeed, applying the first part of Theorem 2.29 to the collection \( \{B_i : i \in I, B_i \subset O\} \), we obtain a countable sub-family of disjoint closed balls \( \{B_i : i \in J\} \), \( B_i = x_i + r_i B^1 \), \( B_i \subset O \) such that \( O \subset \bigcup_{i \in I} B_i^5 = x_i + 5r_i B^1 \). Hence

\[
m(O) \leq \sum_{i \in I} m(B_i^5) = 5^d \sum_{i \in I} m(B_i) = 5^d m\left( \bigcup_{i \in I} B_i \right),
\]

which yields

\[
m(O \setminus \bigcup_{i \in I} B_i) \leq (1 - 5^{-d})m(O) < a m(O).
\]
Since $I$ is countable, we get a finite cover as required.

Now, we iterate the above argument. For the open set $O_n = O_{n-1} \setminus \bigcup_{k \in K} B_k$, with $O_0 = O$, we can find a finite family of closed balls $\{B_k : k \in K_n\}$, $K_n$ finite, with $B_k = x_k + r_kB^1$, $0 < r_k \leq \varepsilon$, such that

$$m(O_n \setminus \bigcup_{k \in K_n} B_k) \leq a m(O_n).$$

Therefore, for $J_n = K_1 \cup \cdots \cup K_n$ we have $O \setminus \bigcup_{k \in J_n} B_k = O_n \setminus \bigcup_{k \in K_n} B_k$ and we deduce

$$m(O \setminus \bigcup_{k \in J_n} B_k) \leq a m(O_n) \leq a^2 m(O_{n-1}) \leq \cdots \leq a^n m(O).$$

Since $a^n \to 0$ as $n \to \infty$ and $m(O) < \infty$, the countable collection of disjoint closed balls $\{B_j : j \in J\}$, $J = \bigcup_n J_n$ satisfies

$$B_j \subset O, \forall j \in J \text{ and } m(O \setminus \bigcup_{j \in J} B_j) = 0,$$

and because $A \subset O$ and $m(O) \leq (1 + \varepsilon)m^*(A)$, the argument is completed.

An alternative prove is to apply the second part of Theorem 2.29 to the collection $\{B_i : i \in I, B_i \subset O\}$ to obtain a countable sub-indexes $J$ of $\{i \in I : B_i \subset O\}$ such that $A \setminus \sum_{k \in K} B_k \subset \sum_{j \in J \setminus K} B_j^5$, for every finite set $K$ of subindexes of $J$. Since $\sum_{j \in J} m(B_j^5) = 5 \sum_{j \in J} m(B_j) \leq 5m(O) < \infty$, the remainder of the series approaches zero and therefore, $m^*(A \setminus \sum_{j \in J} B_j) = 0$.

Note that in the statement of this result, the condition $\sum_{i \in J} m(B_i) < (1 + \varepsilon)m^*(A)$ could be written as $\sum_{i \in J} m(B_i) < m^*(A) + \varepsilon$. \hfill \Box

**Remark 2.36.** If $A$ is a subset of $\mathbb{R}^d$ with $m^*(A) < \infty$ and $O$ is an open set as in Corollary 2.35, then we cannot say that $m^*(\bigcup_{i \in J} B_i \setminus A) = 0$. However, because $\bigcup_{i \in J} B_i$ is measurable

$$m^*(A) = m^*(A \setminus \bigcup_{i \in J} B_i) + m^*(A \cap \bigcup_{i \in J} B_i),$$

it follows that $m^*(A) \leq \sum_{i \in J} m(B_i) \leq (1 + \varepsilon)m^*(A)$. Hence, if $A$ is assumed to be measurable then $m(\bigcup_{i \in J} B_i \setminus A) < \varepsilon m(A)$. In any case, as a consequence of Corollary 2.35, for every $\varepsilon > 0$ there exists a finite sub-collection $\{B_k : k \in K\}$ of disjoint closed balls such that $m^*(A \setminus \bigcup_{k \in K} B_k) < \varepsilon$ and $m^*(A) - \varepsilon < \sum_{k \in K} m(B_k) < m^*(A) + \varepsilon$. Moreover, if $A$ is subset of an open set $\Omega \subset \mathbb{R}^d$ (not necessarily of finite Lebesgue measure) and $\{B_i : i \in I\}$ is a fine cover by balls (non necessarily closed) of $A$, then there exists a null set $N \subset A$ and a countable disjoint sub-family $\{B_i : i \in J\}$ of balls such that $A \setminus N \subset \bigcup_{i \in J} B_i \subset \Omega$. \hfill \Box

**Remark 2.37.** We may use any norm in $\mathbb{R}^d$, non necessarily the Euclidean norm, and the above argument holds. For instance, with the norm $|x| = \max\{|x_1|, \ldots, |x_d|\}$, the unit ball $B^1$ becomes the unit cube, i.e., the covering arguments (namely, Theorem 2.29 and Corollaries 2.35, 2.33) are valid also for cubes instead of balls. \hfill \Box
Remark 2.38. Recall that every open set in \( \mathbb{R}^d \) can be written as a countable union of non-overlapping closed cubes. Indeed, let \( \mathcal{K}_n \) be the collection of closed cubes in \( \mathbb{R}^d \) with edges parallel to the axis and size \( 2^{-n} \). Now, given an open set \( O \) let \( Q_0 \) the family of all cubes in \( \mathcal{K}_0 \) which lie entirely in \( O \). Next, for \( n \geq 1 \), let \( Q_n \) the family of cubes in \( \mathcal{K}_n \) which lie entirely in \( O \) but are not sub-cubes of any cube in \( Q_0, \ldots, Q_{n-1} \). Thus, the family of non-overlapping closed cubes \( Q = \bigcup_n Q_n \) is countable and \( O = \bigcup_{Q \in Q} Q \).

Remark 2.39. By means of the uniqueness extension property shown in Proposition 2.15, we deduce that the Lebesgue outer measure \( m^* \) in \( \mathbb{R}^d \) satisfies

\[
m^*(A) = \inf \left\{ \sum_{i=1}^{\infty} m(E_i) : A \subset \bigcup_{i=1}^{\infty} E_i, \ E_i \in \mathcal{E} \right\}, \quad \forall A \subset \mathbb{R}^d,
\]

where \( \mathcal{E} \) is for instance the class of all either (a) closed balls, or (b) all cubes with edges parallel to the axis.

Exercise 2.33. Actually, give more details relative to the statements in Remark 2.39, i.e., given a subset \( A \) of \( \mathbb{R}^d \) and a number \( r > m^*(A) \) there exist a sequence \( \{B_n\} \) of balls and a sequence \( \{Q_n\} \) of cubes (with edges parallel to the axis) such that \( A \subset \bigcup_n B_n \), \( A \subset \bigcup_n Q_n \), \( r > \sum_n m(B_n) \) and \( r > \sum_n m(Q_n) \). Moreover, relating to the above sequence of cubes, (1) can we make a choice of cubes intersecting only on boundary points (i.e., non-overlapping), and (2) can we take a particular type of cubes defining the class \( \mathcal{E} \) so that the cubes can be chosen disjoint? Finally, compare these assertions with the those in Exercises 2.27 and 2.21.

Remark 2.40. Recall that the diameter of a set \( A \) in Euclidean space \( \mathbb{R}^d \) is defined as \( d(A) = \sup \{|x - y| : x, y \in A\} \). If a set \( A \) is contained in a ball of diameter \( d(A) \) then the monotony of the Lebesgue outer measure \( m^* \) in \( \mathbb{R}^d \) implies

\[
m^*(A) \leq c_d \left(\frac{d(A)}{2}\right)^d, \quad \forall A \subset \mathbb{R}^d,
\]

where \( c_d \) is the volume of unit ball in \( \mathbb{R}^d \), calculated later as

\[
c_d = \pi^{-d/2} \Gamma(d/2 + 1), \quad \Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt,
\]

with \( \Gamma(\cdot) \) is the Gamma function. Certainly, any set \( A \) with diameter \( d(A) \) is contained in a ball of radius \( d(A) \), which yields the estimate \( m^*(A) \leq c_d (d(A)/2)^d \). However, an equilateral triangle \( T \) in \( \mathbb{R}^2 \) is not contained in a ball of radius \( d(A)/2 \). For instance, a carefully discussion on the isodiametric inequality (2.8) can be found in Evans and Gariepy [42, Theorem 2.2.1, pp. 69-70] or in Stroock [112, Section 4.2, pp. 74-79]. In the context of covering, this difficulty could be avoided by using the notation \( 5B = \bigcup \{B' : B' \text{ closed ball with } B' \cap B \neq \emptyset \text{ and } d(B') \leq 2d(B)\} \), which satisfies \( d(5B) \leq 5d(B) \), see Mattila [80, Chapter 2, pp. 23–43].
Remark 2.41. An interesting and useful variation of Vitali’s covering are the so-called Besicovitch Covering: If $A$ is a bounded subset of $\mathbb{R}^d$ and $\{Q_i : i \in I\}$ is a family of open cubes centered at points of $A$ and such that every point of $A$ is the center of some cube, then there exists a countable subset of indexes $J \subset I$ such that $A \subset \bigcup_{j \in J} Q_j$ and no point of $\mathbb{R}^d$ belongs to more than $4^d$ of the cubes in $\{Q_j : j \in J\}$. For instance, the reader is referred to Jones [65, Section 15.H, pp. 482–491] or Mattila [80, Chapter 2, pp. 23–43], for a detailed proof.

Some readers may benefit from taking a quick look at certain portions of Bridges [19, Chapter 2, pp. 70–122] for a concrete discussion on differentiation and Lebesgue integral.
Chapter 3

Measures and Topology

Depending on the background of the reader, this section could far more complicate than the precedent analysis, and perhaps, the first reading should be concentrated in understanding the main statements and then delay the proofs for later study. However, besides the Caratheodory (or outer) construction of measures, it may be convenient to mention the Geometric (Hausdorff) construction Theorem 2.25 (at the end of this chapter) and the inner measure construction of Section 2.3, which is now complemented with Theorem 3.19 below.

If the initial set Ω has a topological structure then it seems natural to consider measures defined on the Borel σ-algebra \( \mathcal{F} = \mathcal{B}(\Omega) \). Moreover, starting from an outer measure then we desire that all open sets (and then any Borel set) be measurable, see Definition 2.4. There are several good presentations of Borel measures, at various level, e.g., Bauer [10, Chapter IV, pp. 153–216], Bogachev [16, Vol 2, Chapter 7, pp. 67–174], Brown and Pearcy [20, Chapter 19, pp. 193–210], Halmos [57, Chapter X, pp. 216–249], Kubrusly [74, Chapter 11, pp. 201–222], Malliavin [79, Chapter II, pp. 55–100], Taylor [115, Chapter 13, pp. 179–191], among others. In any case, the reader may take a quick look at some pieces of Strichartz [111, Chapter 14, pp. 623–690] for a concise introduction to measures in metric spaces.

At this point, we could discuss the Lebesgue measure as the extension of the length, area or volume in \( \mathbb{R}^d \), the point to show is the \( \sigma \)-additivity. Thus, to fix some practical ideas, we may assume that \( \Omega = \mathbb{R}^d \), but most of the following arguments are more general.

3.1 Borel Measures

On a topological space \( \Omega \), an outer measure \( \mu^* \) (see Definition 2.4) is called a Borel outer measure if all Borel sets are \( \mu^* \)-measurable and a regular Borel outer measure if for every \( A \subset \Omega \) there exists \( B \in \mathcal{B}(\Omega) \) such that \( A \subset B \) and \( \mu^*(A) = \mu(B) \) (since \( \Omega \) is a Borel set with \( \mu^*(\Omega) \geq \mu^*(A) \), this condition regards only the case where \( \mu^*(A) < \infty \)). Remark that if \( \{A_n\} \) is a sequence
of \( \mu^* \)-measurable sets with finite measure \( \mu(A_n) < \infty \), and \( B_n \supset A_n \) are Borel sets satisfying \( \mu(B_n) = \mu(A_n) \), then for \( A = \bigcup_n A_n \) and \( B = \bigcup_n B_n \) we have \( B \setminus A \subset \bigcup_n (B_n - A_n) \), which implies \( \mu(B \setminus A) = 0 \), see Exercise 2.10. To make the name regular Borel outer measure more manageable, in many statement we omit the terms regular and/or outer, but unless explicitly stated, we really mean regular Borel outer measure. Moreover, it seems better in this context to simply call (regular) Borel measure what was just defined as (regular) Borel outer measure, without any risk of confusion.

Usually, an outer measure \( \mu^* \) is called \( \sigma \)-finite if there a sequence of sets \( \{\Omega_n\} \) satisfying \( \Omega = \bigcup_{n=1}^\infty \Omega_n \) with \( \mu^*(\Omega_n) < \infty \). In the case of a Borel outer measure we may require the sets \( \Omega_n \) to be Borel, however, more is necessary. A (regular) Borel outer measure \( \mu^* \) is called a \( \sigma \)-finite if there exists a sequence of open sets \( \{O_n\} \) such that \( \Omega = \bigcup_{n=1}^\infty O_n \) with \( \mu(O_n) < \infty \).

We denote by \( \mu \) the measure obtained by restricting \( \mu^* \) to measurable sets. Conversely, if we begin with a Borel measure \( \mu \), i.e., a (\( \sigma \)-finite) measure given only on the Borel \( \sigma \)-algebra \( \mathcal{B}(\Omega) \), then we may define an outer measure

\[
\mu^*(A) = \inf \{ \mu(B) : B \supset A, B \in \mathcal{B}(\Omega) \},
\]

which is a regular Borel outer measure by construction. Although, there may be another Borel outer measure \( \nu^* \) such that \( \nu^* = \mu \) on \( \mathcal{B}(\Omega) \), when \( \mu \) is not \( \sigma \)-finite. Thus, for \( \sigma \)-finite measures, Borel measure and Borel outer measure have the same meaning, via the above extension, i.e., any \( \sigma \)-finite regular Borel outer measure \( \mu^* \) can be regarded as an outer measure obtained via the Carathéodory construction of Theorem 2.9 with a Borel measure \( \mu \) and the class \( \mathcal{E} \) equals to the Borel sets with finite measure.

- **Remark 3.1.** If \( \mu^* \) and \( \nu^* \) are two regular Borel outer measures such that \( \mu(B) = \nu(B) \) for all Borel sets \( B \) then \( \mu^* = \nu^* \). Indeed, for any set \( A \subset \Omega \) there exists a Borel set \( B \supset A \) such that \( \mu^*(A) = \mu(B) \), which implies \( \mu^*(A) = \nu(B) \geq \nu^*(A) \), i.e., \( \mu^* \geq \nu^* \), and by symmetry the equality follows. However, if only \( \mu(B) = \nu(B) \) for all \( B \) in some \( \pi \)-class \( \mathcal{E} \) generating the Borel \( \sigma \)-algebra then to use Proposition 2.15, we need to know that \( \mu \) and \( \nu \) are \( \sigma \)-finite on \( \mathcal{E} \), i.e., \( \Omega = \bigcup_n E_n \) with \( \mu(E_n) = \nu(E_n) < \infty \) for every \( n \geq 1 \).

- **Remark 3.2.** If \( \mu^* \) is a regular Borel outer measure then for any \( \mu^* \)-measurable \( \sigma \)-finite subset \( A \) of \( \Omega \) (i.e., \( A = \bigcup_k A_k \), where \( A_k \) is \( \mu^* \)-measurable and \( \mu(A_k) < \infty \) there exist \( B_1, B_2 \in \mathcal{B}(\Omega) \) such that \( B_1 \supset A \supset B_2 \) with \( \mu(B_1 \setminus B_2) = 0 \). Indeed, first we assume that \( \mu(A) < \infty \), and we take \( B_1 \in \mathcal{B}(\Omega) \) such that \( A \subset B_1 \) and \( \mu(B_1 \setminus A) = \mu(B_1) - \mu(A) = 0 \). Second, we choose another Borel set \( B \supset B_1 \setminus A \) with \( \mu(B) = \mu(B_1 \setminus A) = 0 \) so that \( B_2 = B_1 \setminus B \) satisfies \( B_2 \subset A \), and \( A \setminus B_2 \subset B \setminus (B_1 \setminus A) \), i.e., \( \mu(A \setminus B_2) = 0 \), and the desired result follows. Next, assume that \( A \) is a \( \mu^* \)-measurable set and that there exists a sequence \( \{A_k\} \) of \( \mu^* \)-measurable sets with finite measure, \( \mu(A_k) < \infty \), such that \( A = \bigcup_k A_k \). For each \( A_k \) we apply the previous argument to find Borel sets \( B_{1,k} \) and \( B_{2,k} \) such that \( B_{1,k} \supset A_k \supset B_{2,k} \) with \( \mu(B_{1,k} \setminus B_{2,k}) = 0 \). Since, the Borel sets \( B_i = \bigcup_k B_{i,k}, i = 1,2 \) satisfy \( B_1 \supset A \supset B_2 \) and \( B_1 \setminus B_2 \subset \bigcup_k (B_{1,k} \setminus B_{2,k}) \), which implies \( \mu(B_1 \setminus B_2) = 0 \).
Theorem 3.3. Let $\mu^*$ be a regular Borel outer measure on a topological space $\Omega$, where the topology $\mathcal{T} \subset 2^\Omega$ is such that the Borel $\sigma$-algebra $\mathcal{B}(\Omega)$ results the smallest class which contains all closed sets and is stable under countable unions and intersections. Moreover, suppose that $\mu^*$ is $\sigma$-finite, i.e., there exists a sequence of open sets $\{O_n\}$ such that $\Omega = \bigcup_{n=1}^{\infty} O_n$ with $\mu(O_n) < \infty$. Then

$$\mu^*(A) = \inf \{\mu(O) : O \supset A, \; O \in \mathcal{T} \}, \; \forall A \in 2^\Omega,$$

(3.1)

and

$$\mu(B) = \sup \{\mu(C) : C \subset B, \; C^c \in \mathcal{T} \}, \; \forall B \in \mathcal{B}(\Omega),$$

(3.2)

where $\mu$ denotes the measure induces by $\mu^*$.

Proof. First we prove the following facts: suppose that $\mu$ is a measure on $\mathcal{B}(\Omega)$ and that $B$ is a given Borel set.

(a) If $\varepsilon > 0$ and $\mu(B) < \infty$ then there exists a closed set $C \subset B$ such that $\mu(B \setminus C) < \varepsilon$.

(b) If $\varepsilon > 0$ and $B \subset \bigcup_{i=1}^{\infty} O_i$ with $O_i$ open and $\mu(O_i) < \infty$ then there exists an open set $O \supset B$ such that $\mu(O \setminus B) < \varepsilon$.

To this purpose, define the measure $\nu(A) = \mu(A \cap B)$ for every $A \in \mathcal{B}(\Omega)$ and consider the class $\Phi$ of sets $A$ in $\mathcal{B}(\Omega)$ such that for every $\varepsilon > 0$ there exists a closed set $C \subset A$ satisfying $\nu(A \setminus C) < \varepsilon$. It is clear that $\Phi$ contains all closed sets. Now, if $\varepsilon > 0$ and $\{A_i\}$ is a sequence in $\Phi$ then there exist closed sets $C_i \subset A_i$ with $\nu(A_i \setminus C_i) < 2^{-i} \varepsilon$,

$$\nu \left( \bigcap_{i=1}^{\infty} A_i \setminus \bigcap_{i=1}^{\infty} C_i \right) \leq \sum_{i=1}^{\infty} \nu \left( A_i \setminus C_i \right) < \sum_{i=1}^{\infty} 2^{-i} \varepsilon = \varepsilon,$$

$$\lim_{n} \nu \left( \bigcup_{i=1}^{\infty} A_i \setminus \bigcup_{i=1}^{n} C_i \right) = \nu \left( \bigcup_{i=1}^{\infty} A_i \setminus \bigcup_{i=1}^{\infty} C_i \right) \leq \sum_{i=1}^{\infty} \nu \left( A_i \setminus C_i \right) < \varepsilon,$$

and $\bigcap_{i=1}^{\infty} C_i$ and $\bigcup_{i=1}^{n} C_i$ are closed. This proves that $\Phi$ is stable under countable unions and intersections. Hence, $\Phi = \mathcal{B}(\Omega)$, in particular for $A = B$, we deduce that (a) holds. To show (b), we use (a) to choose closed sets $C_i$ such that $C_i \subset O_i \setminus B$ with $\mu(A_i) < 2^{-i} \varepsilon$ for $A_i = (O_i \setminus B) \setminus C_i = (O_i \setminus C_i) \setminus B$. Since $B \cap O_i \subset O_i \setminus C_i$, we take $O = \bigcup_{i=1}^{\infty} (O_i \setminus C_i)$.

Therefore, given $A \in 2^\Omega$ choose $B \in \mathcal{B}(\Omega)$ such that $A \subset B$ with $\mu^*(A) = \mu(B)$. By means of (b), for every $\varepsilon > 0$ there exists an open set $O \supset B$ such that $\mu(O \setminus B) < \varepsilon$, i.e., $\mu(O) \leq \mu^*(A) + \varepsilon$ and the inf expression (3.1) follows. Similarly, given a set $B \in \mathcal{B}(\Omega)$ with $\mu^*(B) < \infty$, we apply (a), to see that for every $\varepsilon > 0$ there exists a closed set $C \subset B$ such that $\mu(B \setminus C) < \varepsilon$, i.e., we deduce $\mu(C) > \mu(B) - \varepsilon$ and the sup expression (3.2) holds true for any Borel set $B$ with finite measure. Finally when $\mu^*(B) = \infty$, we conclude by making use of an increasing sequence $\{B_k\}$ of Borel sets with $\mu^*(B_k) < \infty$. \qed
Note that if every open set is a countable union of closed sets then (see Proposition 1.8) the Borel $\sigma$-algebra satisfies the assumption in Theorem 3.3. Moreover, in this case, if the outer measure $\mu^*$ is $\sigma$-finite then there exists a sequence of closed sets $\{C_n\}$ such that $\Omega = \bigcup_{n=1}^{\infty} C_n$ and $\mu(C_n) < \infty$.

As discussed in the next section, a Borel measure is called a Radon measure if (a) the property (3.1) of Theorem 3.3, and the conditions (b) $\mu(O) = \sup\{\mu(K) : K \subset O, \text{ compact set}\}$, for every open set $O$, and (c) $\mu(K) < \infty$, for every compact set, are satisfied. The reader may want to take a look at Mattila [80, Chapter 1, pp. 7–22], for a quick summary.

In a topological space, a countable union of closed sets is called a $\mathfrak{F}_\sigma$-set and the class of all those sets is also denoted by $\mathfrak{F}_\sigma$. Similarly, a countable intersection of open sets is called a $\mathfrak{S}_\delta$-set and the class of all those sets is also denoted by $\mathfrak{S}_\delta$.

Under the assumptions of Theorem 3.3, equality (3.1) means that $\mu^*$ can be regarded as an outer measure obtained via the Caratheodory construction of Theorem 2.9 with the class $E$ equals to the open sets.

**Corollary 3.4.** Under the assumptions of Theorem 3.3, for every set $A$ there exists a there exists a $\mathfrak{S}_\delta$-set $G \supset A$ such that $\mu^*(A) = \mu(G)$. Moreover, if $A$ is $\mu^*$-measurable then (1) for every $\varepsilon > 0$ there exist an open set $O$ and a closed set $C$ such that $C \subset A \subset O$ and $\mu(O \setminus C) < \varepsilon$; and (2) there exist a $\mathfrak{S}_\delta$-set $G$ and a $\mathfrak{F}_\sigma$-set $F$ such that $F \subset A \subset G$ and $\mu(G \setminus F) = 0$.

*Proof.* For any set $A \subset \Omega$, in view of the equality (3.1), we can find a sequence $\{O_n\}$ of open sets such that $A \subset O_n$, $\mu^*(A) \leq \mu(O_n)$ and $\mu(O_n) \to \mu^*(A)$. If $G = \bigcap_n O_n$ and $O'_n = \bigcup_{i \leq n} O_n$ then $\mu(O'_n) \to \mu^*(A)$. The monotone continuity from below of the measure $\mu$ shows that $\mu(G) = \lim_n \mu(O'_n)$, provided $\mu(O'_n) < \infty$ for some $n$. However, if $\mu(O'_n) = \infty$ for every $n$ then $\mu(A) = \infty$ and the monotony of the outer measure $\mu^*$ shows that $\mu(G) = \infty$, i.e., in any case $\mu^*(A) = \mu(G)$.

Now, as in Remark 3.2, the claim of the $\mathfrak{S}_\delta$-set $G$ and $\mathfrak{F}_\sigma$-set $F$ need to be proved only for a Borel set $B$ with $\mu(B) < \infty$. Therefore, by means of the assertions (a) and (b) of Theorem 3.3 with $\varepsilon = 1/n$, we construct a sequence $\{C_n\}$ of closed sets and another sequence $\{O_n\}$ of open sets such that $C_n \subset B \subset O_n$ and $\mu(O_n \setminus C_n) < 2/n$. The set $F = \bigcup_n C_n$ belongs to $\mathfrak{F}_\sigma$ and the set $G = \bigcap_n O_n$ belongs to $\mathfrak{S}_\delta$, $C \subset B \subset G$ and $G \setminus F \subset O_n \setminus C_n$, which implies $\mu(G \setminus F) \leq 2/n$, i.e., $\mu(G \setminus F) = 0$.

Actually, the assertion (2) just proved also follows from assertion (1), which was not explicitly contained in the sup representation (3.2). However, the arguments in Theorem 3.3 can be modified to accommodate property (1). Indeed, consider the class $\Phi$ of sets $A$ in $\mathcal{B}(\Omega)$ such that for every $\varepsilon > 0$ there exists a closed set $C$ and an open set $O$ satisfying $C \subset A \subset O$ and $\mu(O \setminus C) < \varepsilon$. To show that $\Phi$ is the whole Borel $\sigma$-algebra we proceed as follows.

If $A$ is a closed set then take $C = A$ and apply assertion (b) in Theorem 3.3 (recall that $\mu$ is $\sigma$-finite) with $B = A$ to find an open set $O \supset A$ satisfying $\mu(O \setminus A) < \varepsilon$. Hence, the class $\Phi$ contains all closed sets.
By definition, it is also clear that \( \Phi \) is stable with respect to complements. Moreover, if \( \varepsilon > 0 \) and \( \{A_i : i \geq 1\} \) is a sequence of sets in \( \Phi \) then there exist a sequence \( \{C_i\} \) of closed sets and a sequence \( \{O_i\} \) of open sets such that \( C_i \subseteq A_i \subset O_i \) and \( \mu(O_i \setminus C_i) < 2^{-i-1}\varepsilon \). To show that \( A = \bigcap_i A_i \) belongs to \( \Phi \), first we see that \( C = \bigcap_i C_i \) is a closed set satisfying \( C \subset A \). Since the inclusion \( \bigcap_i O_i \setminus \bigcap_i C_i \subset \bigcup_i (O_i \setminus C_i) \) yields \( A \subset \bigcap_i O_i \subset B \), where \( B = \left( \bigcap_i C_i \right) \cup \left( \bigcup_i (O_i \setminus C_i) \right) \) is a Borel set satisfying

\[
\mu(B \setminus C) \leq \sum_i \mu(O_i \setminus C_i) \leq \sum_{i=1}^\infty 2^{-i-1}\varepsilon = \varepsilon/2.
\]

Hence, apply assertion (b) in Theorem 3.3 to the set \( B \) to find an open set \( O \supset B \) satisfying \( \mu(O \setminus B) < \varepsilon/2 \). Collecting all pieces, \( C \subset A \subset O \) and \( \mu(O \setminus C) < \varepsilon \), i.e., \( A \) belongs to \( \Phi \), proving that the class \( \Phi \) is stable under the formation of countable intersections. Therefore \( \Phi \) is a \( \sigma \)-algebra containing all closed sets, i.e., \( \Phi \) is the Borel \( \sigma \)-algebra.

\[\square\]

**Corollary 3.5.** Under the assumptions of Theorem 3.3, for any two subsets \( A, B \subset \Omega \) such that there exist open sets \( U \) and \( V \) satisfying \( A \subset U \) and \( B \subset V \) and \( U \cap V = \emptyset \), we have \( \mu^*(A \cup B) = \mu^*(A) + \mu^*(B) \).

**Proof.** Indeed, by means of Theorem 3.3, there is a sequence \( \{O_n\} \) of open sets \( O_n \supset A \cup B \) such that \( \lim_n \mu(O_n) = \mu^*(A \cup B) \). Thus, the open sets \( U_n = O_n \cap U \) and \( V_n = O_n \cap V \) satisfy \( U_n \cap V_n = \emptyset, A \subset U_n, B \subset V_n, \) and \( U_n \cup V_n \subset O_n \). Hence

\[
\mu(O_n) \geq \mu(U_n \cup V_n) = \mu(U_n) + \mu(V_n) \geq \mu^*(A) + \mu^*(B), \quad \forall n,
\]

which implies \( \mu^*(A \cup B) \geq \mu^*(A) + \mu^*(B) \), and the reverse inequality follows form the sub-additivity of \( \mu^* \).

\[\square\]

It is clear that from the above Remark 3.2, we deduce that the sup expression (3.2) remains valid for any \( \mu^* \)-measurable set \( B \), but in general, not necessarily true for any subset \( B \) of \( \Omega \).

• **Remark 3.6.** Note that if \( \mu \) is a Borel measure then the expression

\[
\mu^*(A) = \inf \{\mu(B) : B \supset A, B \in \mathcal{B}(\Omega)\}, \quad \forall A \in 2^\Omega,
\]

yields a regular Borel outer measure. Indeed, if \( \mu^* \) is obtained by (3.3) from a Borel measure \( \mu \) then for every set \( A \subset \Omega \) with \( \mu^*(A) < \infty \) we get a sequence \( \{B_n\} \) of Borel sets such that \( A \subset B_n \) and \( \mu^*(A) + 1/n \geq \mu(B_n) \), i.e., \( B = \bigcap_{n=1}^\infty B_n \) is a Borel set with \( B \supset A \) and \( \mu^*(A) = \mu(B) \).

\[\square\]

### 3.2 On Metric Spaces

The previous result can be applied to a metric space \( \Omega \), see Proposition 1.8. If an outer measure \( \mu^* \) is constructed as in Proposition 2.6 and all Borel sets are
µ*-measurable then we deduce expression (3.3) and µ* results a regular Borel outer measure, i.e., any Borel measure on a metric space becomes a regular Borel outer measure via the expression (3.3). In other words, if every Borel set of a given outer measure µ* is µ*-measurable then the expression (3.3) could be used to redefine µ* as a regular Borel outer measure. In this case, the σ-algebra \( \mathcal{B}^* \) of all µ*-measurable sets is the µ-completion of \( \mathcal{B}(\Omega) \) (see Remark 3.2), and the focus is on a criterium (applied to an outer measure) to know when Borel sets are µ*-measurable.

If \( (\Omega, d) \) is a metric space then \( d(A, B) \) denotes the distance between the sets \( A \) and \( B \), i.e., \( d(A, B) = \inf \{d(x, y) : x \in A, y \in B\} \). The following result is usually known as Caratheodory’s criterium of measurability, compare with Corollary 3.5.

**Proposition 3.7.** Let µ* be an outer measure on a metric space \( (\Omega, d) \). If \( \mu^*(A \cup B) = \mu^*(A) + \mu^*(B) \) for every \( A, B \subseteq \Omega \) such that \( d(A, B) > 0 \) then all Borel sets are µ*-measurable, i.e., µ* is a Borel outer measure.

**Proof.** For instance, we have to show that if \( C \) is a closed set then \( C \) satisfies µ*\((E) \geq \mu^*(E \cap C) + \mu^*(E \cap C^c)\), for every \( E \subseteq \Omega \) with \( \mu^*(E) < \infty \). To this purpose, define \( C_n = \{x \in \Omega : d(x, C) \leq 1/n\} \), for \( n = 1, 2, \ldots \) to have \( d(E \cap C_n, E \cap C) \geq 1/n \) and so by hypothesis

\[
\mu^*(E \cap C_n) + \mu^*(E \cap C^c) = \mu^*((E \cap C_n^c) \cup (E \cap C)) \leq \mu^*(E).
\]

Consider the sets \( R_k = \{x \in E : 1/(k + 1) < d(x, C) \leq 1/k\} \), which satisfy \( E \cap C^c = (E \cap C_n^c) \cup (\bigcup_{k \geq n} R_k) \). Since \( d(R_i, R_j) > 0 \) if \( j \geq i + 2 \), we have

\[
\sum_{k=1}^{m} \mu^*(R_{2k}) = \mu^*(\bigcup_{k=1}^{m} R_{2k}) \quad \text{and} \quad \sum_{k=1}^{m} \mu^*(R_{2k-1}) = \mu^*(\bigcup_{k=1}^{m} R_{2k-1}),
\]

which gives \( \sum_{k=1}^{\infty} \mu^*(R_k) \leq 2\mu^*(E) < \infty \). Based on the monotony and the sub σ-additivity of µ*, we obtain the inequality

\[
\mu^*(E \cap C_n^c) \leq \mu^*(E \cap C^c) \leq \mu^*(E \cap C_n^c) + \sum_{k=n}^{\infty} \mu^*(R_k),
\]

and as \( n \to \infty \), we deduce

\[
\mu^*(E \cap C^c) + \mu^*(E \cap C) = \lim_{n} \left[ \mu^*(E \cap C_n^c) + \mu^*(E \cap C) \right] \leq \mu^*(E),
\]

i.e., \( C \) is µ*-measurable. \( \square \)

**Remark 3.8.** It perhaps interesting to contrast the above Proposition 3.7 with the following assertion: if µ is a finite regular Borel outer measure then then a subset \( A \) of \( \Omega \) is µ*-measurable if and only if \( \mu^*(A) + \mu^*(A^c) = \mu^*(X) \), see Richardson [93, Theorem 2.4.2, pp. 28–29]. \( \square \)
The support of a Borel outer measure $\mu^*$ on a space $\Omega$ with a topology $\mathcal{T}$ is defined as the closed set

$$\text{supp}(\mu^*) = \Omega \setminus \bigcup \{O \in \mathcal{T} : \mu^*(O) = 0\}.$$ 

Note that we do not necessarily have $\mu^*(\Omega \setminus \text{supp}(\mu^*)) = 0$.

On the other hand, a regular Borel measure $\mu$ on a topological space $\Omega$ is called inner regular if

$$\mu(B) = \sup \{\mu(K) : K \subset B, \text{K compact}\}, \quad \forall B \in \mathcal{B}(\Omega). \quad (3.4)$$

Recall that a Polish space $\Omega$ is a complete separable metrizable space, i.e., for some metric $d$ on $\Omega$, the topology of the space $\Omega$ is also generated by a basis composed of open balls $B(x, r) = \{y \in \Omega : d(y, x) < r\}$ for $x$ in some countable dense set of $\Omega$, $r$ any positive rational, and any Cauchy sequence in the metric $d$ is convergence. In this context, a subset $K$ of $\Omega$ is compact if and only if $A$ is closed and totally bounded (i.e., it can be covered by a finite number of balls with arbitrary small radii).

**Proposition 3.9.** Any $\sigma$-finite regular Borel outer measure $\mu^*$ on a Polish space $\Omega$ induces an inner regular Borel measure $\mu$. Moreover we have $\mu^*(\Omega \setminus \text{supp}(\mu^*)) = 0$.

**Proof.** Since $\mu$ is $\sigma$-finite and each open set is a countable union of closed sets, there exists a sequence of closed set $C_k$ such that $\Omega = \bigcup_k C_k$ and $\mu(C_k) < \infty$. Thus, it suffices to check (3.4) only for $B \cap C_k$ instead of $B$. Hence, we are reduced to the case of finite measure, i.e., without loss of generality we assume $\mu(\Omega) < \infty$.

By means of the expression

$$\mu(B) = \sup \{\mu(C) : C \subset B, \text{C closed}\}, \quad \forall B \in \mathcal{B}(\Omega),$$

we need only to show

$$\forall \varepsilon > 0 \text{ there exists a compact } K_\varepsilon \text{ such that } \mu(\Omega \setminus K_\varepsilon) \leq \varepsilon. \quad (3.5)$$

Indeed, for every closed $C$ we have

$$\mu(C) = \mu(C \cap K_\varepsilon) + \mu(C \cap K_\varepsilon^c) \leq \mu(C \cap K_\varepsilon) + \varepsilon$$

and because $C \cap K_\varepsilon$ is compact, we deduce (3.4).

To construct the compact set satisfying (3.5) we recall the argument showing that a closed set is compact if and only if it is totally bounded, and we proceed as follows. For a sequence $\{\omega_i\}$ denote by $\bar{B}_i(r)$ the closed ball of center $\omega_i$ and radius $r > 0$. If $\{\omega_i\}$ is dense in $\Omega$ then we have $\bigcup_i \bar{B}_i(r) = \Omega$, for any $r > 0$. Now, the monotone continuity from above of $\mu$ implies that given any $\varepsilon > 0$ there exists a finite set of indexes $I = I(r, \varepsilon)$ such that $\mu(\Omega \setminus \bigcup_{i \in I} \bar{B}_i(r)) < \varepsilon$, in particular, for a fixed $\varepsilon > 0$ we can replace $r, \varepsilon$ with $1/n, 2^{-n}\varepsilon$, for an integer
n, to have $\mu(\Omega \setminus \bigcup_{i \in I_n} F_{i,n}) \leq 2^{-n}\varepsilon$, with $F_{i,n} = \overline{B}_i(1/n)$, for some finite set of indexes $I_n$. Thus, the closed set $K_\varepsilon = \bigcap_n \bigcup_{i \in I_n} F_{i,n}$ satisfies

$$\mu(\Omega \setminus K_\varepsilon) = \mu\left(\bigcup_n \left(\Omega \setminus \bigcup_{i \in I_n} F_{i,n}\right)\right) \leq \sum_n \mu(\Omega \setminus \bigcup_{i \in I_n} F_{i,n}) \leq \varepsilon.$$  

To check that $K_\varepsilon$ is indeed compact, take a sequence $\{x_j : j \in J\}$ in $K_\varepsilon$. Since $\{x_j : j \in J\} \subset \bigcup_{i \in I_n} F_{i,n}$ for each $n$, there exist a sequence of indexes $i_n \in I_n$ and a subsequence $\{x_j : j \in J_n\}$ such that $x_j \in \bigcap_{k=1}^n F_{i_k,k}$, for every $j \in J_n \subset J_{n-1}$. Thus, for the diagonal subsequence $\{x_j : j \in J_0\}$ we have $x_j \in \bigcap_{k=1}^n F_{i_k,k}$, for every $j \in J_0$ and integer $n$. Hence, $\{x_j : j \in J_0\}$ is a Cauchy sequence, which must be convergent since $\Omega$ is complete.

Finally, we observe that any compact subset $K$ of the open set

$$U = \Omega \setminus \text{supp}(\mu^*) = \bigcup \{O \in \mathcal{T} : \mu^*(O) = 0\}$$

is indeed contained in a finite union of open sets of measure zero, and thus, $\mu(K) = 0$ and the expression (3.4) yields $\mu^*(\Omega \setminus \text{supp}(\mu^*)) = 0$. 

- **Remark 3.10.** Note that the key property (3.5), referred to as tightness is actually related with the property of universally measurable of the metric space $\Omega$ (i.e., the property that any finite measure $\mu$ on a complete $\sigma$-algebra containing the Borel $\sigma$-algebra and any measurable set $E$ there exist two Borel sets $A$ and $B$ such that $A \subset E \subset B$ and $\mu(A) = \mu(B) = 0$). A separable metric space is universally measurable if and only if every finite measure is tight, i.e., the condition (3.5) holds, e.g., Dudley [37, Theorem 11.5.1, p. 402].

- **Remark 3.11.** It is clear that if $\mu^*$ is a regular Borel outer measure on a Polish space $\Omega$, not necessarily $\sigma$-finite, then the formula (3.4) holds for any $\sigma$-finite Borel sets, i.e., for any Borel set $B$ such that $B \subset \bigcup_k C_k$ for some sequence $\{C_k\}$ of closed sets with $\mu(C_k) < \infty$.

**Exercise 3.1.** Verify Remark 3.11 and give more details on the passage from a finite measure to a $\sigma$-finite measure in the proof of Proposition 3.9.

- **Remark 3.12.** The following assertion is commonly referred to as Ulam’s Theorem: any finite Borel measure on a complete separable metric space is inner regular, i.e., (3.5) is satisfied. In probability theory, $P(\Omega) = \mu(\Omega) = 1$ and (3.5) becomes

$$\forall \varepsilon > 0 \text{ there exists a compact } K_\varepsilon \text{ such that } P(K_\varepsilon) \geq 1 - \varepsilon,$$

which is referred to as tightness of the probability $P$ on a Polish space $\Omega$.

- **Remark 3.13.** Actually, Proposition 3.9 holds for non Polish spaces $\Omega$, the properties needed in the above proof are the following: $\Omega$ has a second-countable complete uniform topology, i.e., besides the convergence of any Cauchy sequence, there exists a countable basis, namely, a countable family of open sets $\mathcal{O} = \{O_{i,n} : i,n \geq 1\}$ such that the set $\bigcup_i \bigcap_n O_{i,n}$ is dense in $\Omega$ and for any open set
O and any x in O there exist O_{i,n} satisfying x \in O_{i,n} \subset O$. Thus, for any inner regular \(\sigma\)-finite Borel measure \(\mu\), for every \(\varepsilon > 0\) and any \(\mu^*\)-measurable set \(B\) with finite measure, \(\mu(B) < \infty\), there exist a compact set \(K \subset B\) and an open set \(O \supset B\) such that \(\mu(O \setminus K) < \varepsilon\). This implies that the countable family \(O_f\) of finite unions of open set in \(O\) is “\(\mu\)-dense”, i.e., there exists \(O_f\) in \(O_f\) such that \(\mu((B \setminus O_f) \cup (O_f \setminus B)) < \varepsilon\). \(\square\)

- **Remark 3.14.** Let \(A\) be the algebra generated by all open (or compact) sets. Given a \(\sigma\)-additive and \(\sigma\)-finite set function (pre-measure) \(\mu : A \to [0, \infty]\), we consider the set functions defined for every \(A\) in \(2^\Omega\)

\[
\mu^*(A) = \inf\{\mu(O) : O \supseteq A, O \text{ open}\},
\]

\[
\mu_*(A) = \sup\{\mu(K) : K \subset A, K \text{ compact}\}.
\]

The expression of \(\mu_*\) is called an inner measure, meaning, \(\sigma\) super-additive (i.e., \(\sum_{n} \mu_*(A_n) \leq \mu_*(A)\) whenever \(A = \sum_{n} A_n\)) and monotone, but not necessarily additive. Under the assumptions of Theorem 3.3, we deduce that \(\mu^*\) is an outer measure. Moreover, if \(\mu^*\) induces an inner regular Borel measure, see definition (3.4), then \(\mu^*(B) = \mu_*(B)\) for any Borel set \(B\). Then, measurable sets with finite measure are mainly identified by the condition \(\mu^*(A) = \mu_*(A)\). Inner measures can be defined in abstract measure space without any topology, see Halmos [57, Section III.14, pp. 58–62]. \(\square\)

**Exercise 3.2.** With the notation of the previous Remark 3.14, under the conditions of Theorem 3.3, and assuming the restriction \(\bar{\mu} = \mu^*|_B\) on the Borel \(\sigma\)-algebra \(B\) is inner regular:

1. Verify that \(\mu_*\) is \(\sigma\) super-additive, i.e., if \(A_i \in 2^\Omega\) and \(A = \sum_{i=1}^{\infty} A_i\) then \(\mu_*(A) \geq \sum_{i=1}^{\infty} \mu_*(A_i)\).

2. Show that if \(A \in 2^\Omega\) and \(\{C_i\}\) is a disjoint sequence of closed sets with \(C = \sum_i C_i\) then \(\mu_*(A \cap C) = \sum_i \mu_*(A \cap C_i)\).

3. Prove that if the topological space \(\Omega\) is countable at infinity (i.e., there exists a monotone increasing sequence \(\{K_n\}\), \(K_n \subset K_{n+1}\) of compact sets such that \(\Omega = \bigcup_n K_n\) then \(\bar{\mu}\) is inner regular.

4. Assuming that \(\bar{\mu}\) is inner regular, prove that if \(A\) is a \(\mu^*\)-measurable set then \(\mu^*(A) = \mu_*(A)\), and conversely, if \(A \in 2^\Omega\), \(A = \bigcup_k A_k\), \(\mu^*(A_k) = \mu_*(A_k) < \infty\) then \(A\) is \(\mu^*\)-measurable.

5. Prove that for any two disjoint subsets \(A\) and \(B\) of \(\Omega\) we have \(\mu_* (A \cup B) \leq \mu_*(A) + \mu^*(B) \leq \mu^*(A \cup B)\).

6. Assume that \(E\) is \(\mu^*\)-measurable. Show \(\mu_*(A) = \mu^*(E) - \mu^*(E \setminus A)\), that for every \(A \subset E\) with \(\mu^*(E \setminus A) < \infty\).

7. Discuss alternative definitions of \(\mu_*\), for instance, when \(\bar{\mu}\) is not necessarily inner regular or when \(\Omega\) is not a topological space. \(\square\)
3.3 On Locally Compact Spaces

On a locally compact Hausdorff topological space $\Omega$ we may consider the so-called Radon outer measure $\mu^*$, which are defined by means of the three following properties: (1) $\mu^*$ is a regular Borel outer measure, i.e., all Borel sets are $\mu^*$-measurable and for every $A \in 2^\Omega$ there exists a Borel set $B \supset A$ such that $\mu^*(A) = \mu(B)$; (2) any compact set $K$ has finite measure $\mu(K) < \infty$; (3) for any open set $O$ we have $\mu(O) = \sup\{\mu(K) : K \subset O, K \text{ compact}\}$. Similarly, the restriction of $\mu^*$ to a $\mathcal{B}(\Omega)$ (i.e., the measure $\mu$) is called a Radon measure. If Borel measure $\mu$ (instead of a Borel outer measure $\mu^*$) is initially given then a regular Borel outer measure $\mu^*$ is induced by $\mu$, and only conditions (2) and (3) regarding the measure $\mu$ are used as the definition of a Radon measure $\mu$.

Recall that any locally compact Hausdorff topological space satisfying the second axioms of separability (i.e., with a countable basis) is actually $\sigma$-compact, i.e., the whole space is a countable unions of compact sets. Thus, a Radon measure is $\sigma$-finite if the space has a countable basis. Also note that sometime, property (2) is also assumed for a regular Borel measure, and moreover, it is convenient (but not necessary) to begin with a locally compact spaces $\Omega$, what really matter is to assume that the above properties (1), (2) and (3) are satisfied. Moreover, as discussed later on, this is also related to the tightness of a finite Borel measure, e.g. Cohn [28, Chapter 7, 196–250].

Therefore, under the assumptions of Theorem 3.3, if $\mu^*$ is a Radon outer measure then for any $\varepsilon > 0$ and for any Borel set $B$ with $\mu(B) < \infty$ there exists a compact set $K \subset B$ such that $\mu(B \setminus K) < \varepsilon$. Indeed, first we choose open sets $U$ and $V$ such that $B \subset U$, $\mu(U \setminus B) < \varepsilon/2$ and $U \setminus B \subset V$, $\mu(V) < \varepsilon/2$. Next, we take a compact set $K \subset U$ with $\mu(U \setminus K) < \varepsilon/2$. Then $K \setminus V$ is a compact subset of $U \setminus V$, $B \setminus (K \setminus V) \subset (U \setminus K) \cup V$ and $\mu(B \setminus (K \setminus V)) < \varepsilon$.

Therefore, a Radon measure is inner regular, and in a locally compact Polish space, a regular Borel (outer) measure which is finite for every compact set is indeed a Radon measure (see Proposition 3.9).

The following result is mainly used for probability measures, e.g., Neveu [87, Proposition I.6.2, pp. 27–29].

**Proposition 3.15.** Let $\Omega$ be a topological space, $S \subset \mathcal{B}(\Omega)$ be a semi-ring (or semi-algebra) and $\mu : S \to [0, \infty]$ be an additive set function such that

$$\mu(S) = \sup\{\mu(K) : K \subset S, K \text{ is compact}, \mu(K) < \infty, K \in S\}, \quad (3.6)$$

for every $S$ in $S$. Then $\mu$ is indeed $\sigma$-additive and therefore $\mu$ can be uniquely extended to the ring (or algebra) $A$, generated by $S$, as a $\sigma$-additive measure on $A$ satisfying (3.6) with $(A,A)$ in lieu of $(S,S)$.

**Proof.** First, we show the validity of the representation (3.6) for the ring (or algebra) $A$ generated by $S$. Indeed, any $A \in A$ is a finite disjoint union of elements in $S$, i.e., $A = \sum_{i=1}^n S_i$ with $S_i \in S$, and the extension is given by $\mu(A) = \sum_{i=1}^n \mu(S_i)$. To prove the representation

$$\mu(A) = \sup\{\mu(K) : K \subset A, K \text{ is compact}, \mu(K) < \infty, K \in A\},$$
for every $A$ in $\mathcal{A}$, we denote by $\nu(A)$ the right-hand side, i.e., the expression with the sup. Thus, for any $\varepsilon > 0$ there exist compact sets $K_i \in \mathcal{S}$ with $K_i \subset S_i$ such that $\mu(S_i) = \nu(S_i) \leq \mu(K_i) + \varepsilon/n$. Since $K = \sum_{i=1}^{n} K_i$ is compact and belongs to $\mathcal{A}$, we deduce the $\mu(A) \leq \nu(A)$. For the opposite inequality, the monotony (or sub-additivity) of $\mu$ on $\mathcal{A}$ implies $\mu(K) \leq \mu(A)$, for any compact set $K$ in $\mathcal{A}$ with $K \subset A$, and by taking the supremum we obtain $\nu(A) \leq \mu(A)$.

To check the $\sigma$-additivity of $\mu$ on $\mathcal{A}$, we should prove that if $\{S_k\}$ is a sequence of disjoint sets in $\mathcal{S}$ such that $S = \sum_{k=1}^{\infty} S_k$ and $S \in \mathcal{S}$ then $\mu(S) = \sum_{k=1}^{\infty} \mu(S_k)$. We need only to discuss the case with $\mu(S) < \infty$, indeed, if $\mu(S) = \infty$ then for every $r > 0$ there exists a compact set $K$ in $\mathcal{S}$ such that $K \subset S$ and $r < \mu(K) < \infty$. If we assume that $\mu$ is $\sigma$-additive on sets of finite measure then $r < \mu(K) = \sum_{k} \mu(S_k \cap K)$, which implies $\mu(S) = \infty = \sum_{k} \mu(S_k)$.

Therefore, assuming $\mu(S) < \infty$, we have to prove that $\lim_{n} \mu(A_n) = 0$, where $A_n = S \setminus \sum_{k=1}^{n} S_k$, $A_n \in \mathcal{A}$, $A_n = \sum_{j=1}^{n} B_{n,j}$ with $B_{n,j} \in \mathcal{S}$, and $\cap_{k=1}^{\infty} A_n = \emptyset$. Thus, for any given $\varepsilon > 0$, the expression (3.6) applied to each $B_{n,j}$ implies that there exists a sequence $\{K_{n,j} : j = 1, \ldots, J_{n}\}$ of compact sets satisfying $K_{n,j} \subset B_{n,j}$ and $\mu(B_{n,j}) \leq \mu(K_{n,j}) + 2^{-n-j} \varepsilon$. Since $K_n = \cup_{j=1}^{J_n} K_{n,j} \subset A_n$ is compact and $\cap_{n=1}^{\infty} K_n = \emptyset$, there exists an index $N$ such that $\cap_{n=1}^{N} K_n = \emptyset$, which yields the inclusion

$$A_N = \bigcup_{n=1}^{N} (A_n \setminus K_n) \subset \bigcup_{n=1}^{N} (A_n \setminus K_n) \subset \bigcup_{n=1}^{N} \bigcup_{j=1}^{J_n} (B_{n,j} \setminus K_{n,j}).$$

Because $\mu$ is additive and $\mathcal{S}$ is a semi-ring, $\mu$ is also sub-additive. This implies

$$\mu(A_N) \leq \sum_{n=1}^{N} \sum_{j=1}^{J_n} (\mu(B_{n,j}) - \mu(K_{n,j})) \leq \sum_{n=1}^{N} \sum_{j=1}^{J_n} 2^{-n-j} \varepsilon < \varepsilon,$$

i.e., $\lim_{n} \mu(A_n) = 0$.  

\textbf{Remark 3.16.} A small variation on the argument used in the Proposition 3.15 allows us to substitute the sup representation (3.6) with

$$\mu(S) = \sup\{\mu(A) : A \subset K \subset S, K \text{ is compact}, \mu(A) < \infty, A \in \mathcal{S}\},$$

for every $S$ in $\mathcal{S}$. In this way, there is no implicit assumption that the semi-ring $\mathcal{S}$ contains compact sets. For instance, in the case of Lebesgue(-Stieltjes) measure in $\mathbb{R}^d$, we may begin by considering the hyper-volume $m$ as defined on the semi-ring of semi-open bounded $d$-intervals $[a, b]$, instead of using the semi-ring of all bounded $d$-intervals, and the verification of the sup representation takes the same effort.  

\textbf{Remark 3.17.} Under the assumptions of Proposition 3.15, we can prove that if $\mu$ is semi-finite (i.e., for every $S$ in $\mathcal{S}$ with $\mu(S) = \infty$ there exists an increasing sequence $\{S_k\} \subset \mathcal{S}$ such that $S_k \subset S_{k+1} \subset S$, $\mu(S_k) < \infty$ and $\mu(S_k) \to \infty$ (see Exercise 2.15), then the sup assumption (3.6) can be replaced by

$$\mu(S) = \sup\{\mu(K) : K \subset S, K \text{ is compact}, K \in \mathcal{S}\}, \forall S \in \mathcal{S}.$$
Indeed, for every $\varepsilon > 0$ and any $S_k$ in $S$ with $\mu(S_k) < \infty$ there exists a compact set $K_k \subset S_k$ in $S$ such $\mu(S_k) - \varepsilon < \mu(K_k) \leq \mu(S_k)$. Thus, both sup expressions coincide on any set with finite measure and if $\mu(S) = \infty$ then the sequence $\{K_k\}$ show that both sup expressions coincide also on any set with infinite measure. In particular, if $\mu$ is $\sigma$-finite (i.e., there exists a sequence $\{\Omega_k : k \geq 1\}$ of elements of the semi-ring $S$ such that $\Omega = \bigcup_k \Omega_k$ with $\mu(\Omega_k) < \infty$) then the condition $\mu(K) < \infty$ can be dropped in the sup expression (3.6).

**Remark 3.18.** It is clear that the sup representation (3.6) holds true for any $S = A$ belonging to the class $\mathcal{A}' = \{A \in 2^\Omega : A = \sum_{n=1}^{\infty} A_n, A_n \in S\}$ of countable disjoint unions of sets in $S$. As mentioned in Remark 1.7, this class $\mathcal{A}'$ is not necessarily stable under countable intersections. Nevertheless, assuming that the semi-ring $S$ generate the Borel $\sigma$-algebra $B$ in a metric space $\Omega$, we could apply Theorem 3.3 to deduce that $\mu$ satisfies (3.6) with $(A, B)$ in lieu of $(S, S)$, i.e., $\mu$ is inner regular.

**Exercise 3.3.** Elaborate the previous Remark 3.18, i.e., by means of Theorem 3.3 and Proposition 3.9 state and prove under which precise conditions a set function $\mu$ satisfying the assumptions of Proposition 3.15 can be extended to a (regular and) inner regular Borel measure.

Consider the definition: $\mathcal{K} \subset 2^\Omega$ is called a compact class if (a) $\mathcal{K}$ is stable under finite intersections and unions, and (b) for any sequence $\{K_i : i \geq 1\} \subset \mathcal{K}$ with $\bigcap_i K_i = \emptyset$ there exists an index $n$ such that $\bigcap_{i=1}^{n} K_i = \emptyset$.

**Exercise 3.4.** Regarding the sup-formula (3.6), prove that if $\bar{S}$ is the ring generated by the semi-ring $S$ and $\mu : S \to [0, \infty]$ is an additive set function (which is uniquely extended by additivity to the ring $\bar{S}$) and such that there exists a compact class $\mathcal{K} \subset \bar{S}$ satisfying

$$\mu(S) = \sup\{\mu(K) : K \subset S, \mu(K) < \infty, K \in \mathcal{K}\}, \quad \forall S \in \bar{S}. \quad (3.7)$$

then $\mu$ is necessarily $\sigma$-additive on $\bar{S}$.

A similar view regarding compact classes is discussed in the following

**Exercise 3.5.** In a topological space, (1) Verify that any family of compact sets is indeed a compact class of sets in the above sense. Now, let $\mu$ a finite countable additive (non-necessarily $\sigma$-additive, just additive) and $\sigma$-finite measure defined on an algebra $\mathcal{A} \subset 2^\Omega$. Suppose that there exists a compact class $\mathcal{K}$ such that for every $\varepsilon > 0$ and any set $A$ in $\mathcal{A}$ with $\mu(A) < \infty$ there exists $A_\varepsilon$ in $\mathcal{A}$ and $K_\varepsilon$ in $\mathcal{K}$ such that $A_\varepsilon \subset K_\varepsilon \subset A$ and $\mu(A \setminus A_\varepsilon) < \varepsilon$. (2) Using a technique similar to Proposition 3.15, show that $\mu$ is necessarily $\sigma$-additive. (3) Also, prove the representation

$$\mu(A) = \sup\{\mu(K) : K \subset A, K \in \mathcal{K}\}, \quad \forall A \in \mathcal{A},$$

provided $\mathcal{K} \subset \mathcal{A}$, see Bogachev [16, Section 1.4, pp. 13-16].
In Proposition 3.15, the additive set function $\mu$ is defined initially in a semi-ring. Since the difference of two compact sets is not necessarily a compact set, a semi-ring of just compact sets is not a reasonable assumption. Thus, a lattice (i.e., a class stable under the formation of finite intersections and finite unions, and containing the empty set) is a natural class of compact sets $\mathcal{K}$, and a method to construct a measure from an initial set function on a lattice seems useful.

In what follows, even if it is not assumed, the prototype of $\mathcal{K}$ is a family of compact sets so that $\mathcal{K}_F$ becomes a family of closed set, and eventually, to construct $\mu_*$ is an inner measure on the Borel $\sigma$-algebra. Recall that in a (Hausdorff) topological space $\Omega$, a compact set is closed (the converse is generally false), and any closed subset of a compact set is also compact.

Following the inner construction of Theorem 2.22, let $\mathcal{K} \subset 2^\Omega$ be a $\pi$-class (i.e., a class containing the empty set and stable under finite intersections) and $\mu : \mathcal{K} \to [0, \infty)$ be a finite-valued set function with $\mu(\emptyset) = 0$. The sup expression

$$\mu_*(A) = \sup \left\{ \sum_{i=1}^n \mu(K_i) : \sum_{i=1}^n K_i \subset A, K_i \in \mathcal{K} \right\}, \quad \forall A \subset \Omega. \quad (3.8)$$

defines an inner measure, namely, $\mu_*$ is monotone (i.e., $E \subset F$ implies $\mu_*(E) \leq \mu_*(F)$), and super-additive (i.e., $\mu_*(E \cup F) \geq \mu_*(E) + \mu_*(F)$ whenever $E \cap F = \emptyset$). Moreover, if $\mathcal{A}$ is the class of all sets $A$ satisfying $\mu_*(E) = \mu_*(E \cap A) + \mu_*(E \setminus A)$ for every $E \subset \Omega$ (which is referred to as the class of all Caratheodory $\mu_*$-measurable sets) then $\mathcal{A}$ is an algebra and that $\mu_*$ is additive on $\mathcal{A}$.

Next, assume that $\mu$ is $\mathcal{K}$-tight, i.e., for every $K$ and $K'$ in $\mathcal{K}$ with $K \supset K'$ we have $\mu(K) = \mu(K') + \mu_*(K \setminus K')$. Then

$$A \in \mathcal{A} \quad \text{iff} \quad \mu(K) \leq \mu_*(K \cap A) + \mu_*(K \setminus A), \quad \forall K \in \mathcal{K}, \quad (3.9)$$

and the algebra $\mathcal{A}$ contains the class

$$\mathcal{K}_F = \{ F \subset \Omega : F \cap K \text{ is a finite union of sets in } \mathcal{K}, \forall K \in \mathcal{K} \}. \quad (3.10)$$

Furthermore, the initial finite-valued set function $\mu$ is additive on $\mathcal{K}$ (i.e., $\mu(K) = \sum_{i=1}^n \mu(K_i)$ whenever $K = \sum_{i=1}^n K_i$ with all sets in $\mathcal{K}$) and $\mu(K) = \mu_*(K)$ for every $K$ in $\mathcal{K}$.

**Theorem 3.19.** Suppose that $\Omega$ is a (Hausdorff) topological space. Under the previous setting, for a $\mathcal{K}$-tight finite-valued set measure defined on a $\pi$-class $\mathcal{K}$ and, with $\mu_*$ and $\mathcal{K}_F$ given by (3.8) and (3.10), If $\mathcal{K}$ is a class of compact sets (i.e., each element in $\mathcal{K}$ is a compact set) then $\mu_*$ is a unique complete inner regular measure on the $\sigma$-algebra $\mathcal{A}$, and if the class $\mathcal{K}_F$ generates $\mathcal{B}(\Omega)$ the $\mathcal{A}$ contains the Borel $\sigma$-algebra $\mathcal{B}(\Omega)$, i.e., $\mu_*$ is an inner regular Borel measure.

**Proof.** The fact that $\mu = \mu_*$ is additive on $\mathcal{K}$ has been proved in Theorem 2.22. We retake the argument in the case under the assumption that $\mathcal{K}$ is a class of compact sets. Thus, Proposition 3.15 ensures that $\mu_*$ is $\sigma$-additive on the algebra $\mathcal{A}$ (which contains all closed sets), and therefore, it can be extended to
a unique measure on the \( \sigma \)-algebra generated by \( \mathcal{A} \), but a priori, the extension needs not to be an inner measure.

The point is to show that \( \mathcal{A} \) is a \( \mu_* \)-complete \( \sigma \)-algebra, independent of the fact that Carathéodory extension of \( (\mu_*, \mathcal{A}) \) yields a complete measure \( (\bar{\mu_*}, \mathcal{A}) \).

Hence, to check that \( \mathcal{A} \) is \( \mu_* \)-complete, take \( B \subset A \) with \( A \) in \( \mathcal{A} \) and \( \mu_*(A) = 0 \) to obtain, after using the monotony of \( \mu_* \), that

\[
\mu(K) \leq \mu_*(K \cap A) + \mu_*(K \setminus A) \leq 0 + \mu_*(K \setminus B) \leq \mu_*(K \cap B) + \mu_*(K \setminus B), \quad \forall K \in \mathcal{K},
\]
i.e., \( B \) belongs to \( \mathcal{A} \), as in Proposition 2.20.

To verify that the class of all \( \mu_* \)-measurable sets is a \( \sigma \)-algebra, we have to show only that \( \mathcal{A} \) is stable under the formation of countable intersections, i.e., if \( \{A_i\} \) is a sequence of sets in \( \mathcal{A} \) then we should show that \( A = \bigcap_i A_i \) also belongs to \( \mathcal{A} \). For this purpose, from the sup definition (3.8) of \( \mu_* \) and because \( \mathcal{A} \) contains any finite union of (compact) sets in \( \mathcal{K} \), for any \( \varepsilon > 0 \) and for any set \( K \) in \( \mathcal{K} \) there exist a (compact) set \( K' \subset K \cap A \) in \( \mathcal{A} \) such that

\[\mu_*(K \cap A) - \varepsilon < \mu_*(K').\]

Thus, define the sequence \( \{B_n\} \) with \( B_n = \bigcap_{i \leq n} A_i \) to have \( \bigcap_n (K \cap B_n \cap K') = K' \), and (after using the \( \sigma \)-additive of \( \mu_* \) on \( \mathcal{A} \)) to deduce

\[
\lim_n \mu_*(K \cap B_n) \geq \lim_n \mu_*(K \cap B_n \cap K') = \mu_*(K') > \mu_*(K \cap A) - \varepsilon.
\]

This proves that \( \lim_n \mu_*(K \cap B_n) = \mu_*(K \cap A) \). Recall that \( B_n \) is in \( \mathcal{A} \) to have

\[
\mu(K) \leq \mu_*(K \cap B_n) + \mu_*(K \setminus B_n) \leq \mu_*(K \cap B_n) + \mu_*(K \setminus A),
\]
and, after taking \( n \to \infty \) and invoking the condition (3.9), to deduce that \( A \) belongs to \( \mathcal{A} \), i.e., \( \mathcal{A} \) is a \( \sigma \)-algebra.

The uniqueness of \( \mu_* \) is not really an issue, i.e., any other measure \( \nu \) on a \( \sigma \)-algebra \( \mathcal{F} \subset \mathcal{A} \) containing the class \( \mathcal{K} \) and such that the sup representation (3.8) holds true for \( \nu \) replacing \( \mu_* \) is indeed equals to \( \mu_* \) on \( \mathcal{F} \). Indeed, they both agree on any set of finite measure, and for any set \( F \in \mathcal{F} \) with infinite measure there exists a sequence \( \{F_n\} \subset \mathcal{F} \) with \( \nu(F_n) < \infty \), \( F_n \subset F \) and \( \lim_n \nu(F_n) = \nu(F) \), i.e., \( \nu(F) = \mu_*(F) \) too.

Finally, since the class (of closed sets) \( \mathcal{K}_F \) is contained in \( \mathcal{A} \), it is clear that if the class \( \mathcal{K}_F \) generates \( \mathcal{B}(\Omega) \) then Borel \( \sigma \)-algebra \( \mathcal{B}(\Omega) \) is contained in \( \mathcal{A} \). \( \square \)

Note that abusing language, we are calling a measure on a \( \sigma \)-algebra \( \mathcal{A} \) inner regular if the sup representation (3.8) holds true for some \( \pi \)-class and \( \mu \) as above. It should be clear that additivity on a \( \pi \)-class (or lattice) \( \mathcal{K} \) (i.e., \( \mu(A \cap B) = \mu(A) + \mu(B) \), whenever \( A \) and \( B \) belongs to \( \mathcal{K} \) with \( A \cap B = \emptyset \)) is weak and not so useful (unless the class \( \mathcal{K} \) is a semi-ring). The good property on a lattice is the \( \mathcal{K} \)-tightness (i.e., for every \( \varepsilon > 0 \) and sets \( A \subset B \) in \( \mathcal{K} \) there exists a set \( K \subset B \setminus A \) in \( \mathcal{K} \) such that \( \mu(B) < \mu(A) + \mu(K) + \varepsilon \)), which implies additivity. In general, if \( \mathcal{K} \) is only a \( \pi \)-class, this \( \mathcal{K} \)-tightness property refers to the class of all disjoint finite unions of sets in \( \mathcal{K} \), not just the \( \pi \)-class \( \mathcal{K} \), see the
definition (3.8) of $\mu_*$. Moreover, in the case of a $\pi$-class $\mathcal{K}$ of compact sets, the lattice $\bar{\mathcal{K}}$ generated by $\mathcal{K}$ (i.e., the class of all finite unions of sets in $\mathcal{K}$) is also a class of compact sets. This compactness condition is used (via Proposition 3.15) to deduce that $\mu_*$ is monotone continuous from above on $\mathcal{A}$, meaning that again, assuming monotone continuity from above on the lattice $\bar{\mathcal{K}}$ is sufficient (but not only the $\pi$-class $\mathcal{K}$), see Exercise 2.17 and comments later on.

It could be interesting to realize that the topological context in Theorem 3.19 can be removed, actually, the assumption about a class $\mathcal{K}$ of compact sets can be replaced by the monotone continuity from below on a lattice $\mathcal{K}$, i.e., for any a monotone decreasing sequence $\{K_n\}$ of finite unions of sets in $\mathcal{K}$ with limit $K = \bigcap_n K_n$ in $\mathcal{K}$ we have $\mu(K) = \lim_n \mu(K_n)$, e.g., Pollard [90, Appendix A, pp. 289–300]. Also, Exercise 3.5 and Exercise 2.16 give a way of replacing the condition about a class $\mathcal{K}$ of compact sets with an assumption on a compact class, where topology is not involved.

- **Remark 3.20.** Proposition 3.15 and the previous Theorem 3.19 can be combined to show that a set function $\mu$ satisfying (3.6) on a semi-ring $\mathcal{S}$ (which generates the Borel $\sigma$-algebra) can be extended to an inner regular Borel measure. Indeed, take $\mathcal{K} = \{K \in \mathcal{S} : \mu(K) < \infty \text{ and compact}\}$ and note that for every $K$ in $\mathcal{K}$ and $A$ in $\mathcal{S}$ we have $K \setminus A = \sum_{i=1}^n B_i$ with $B_i$ in $\mathcal{S}$ and

$$
\mu(K) = \mu(K \cap A) + \sum_{i=1}^n \mu(B_i) = \mu_*(K \cap A) + \mu_*(K \setminus A),
$$

which shows that the semi-ring $\mathcal{S}$ is included in the algebra $\mathcal{A}$. \hfill \square

Perhaps, the prototype for Remark 3.20 is the Lebesgue(-Stieltjes) measure as discussed in Section 2.5. For instance, the additive finite-valued set function $\mu$ is initially defined on the semi-ring $\mathcal{I}$ of all semi-open bounded $d$-intervals of the form $]a, b]$ and $\mathcal{K}$ is the class of compact $d$-intervals of the form $[a, b]$, including the empty set. If $\mu([a, b]) = \prod_{i=1}^n (F_i(b) - F_i(a))$ then the right-continuity of $F_i$ is used to show that for any $I$ in $\mathcal{I}$ and ant $\varepsilon > 0$ there exists $K$ in $\mathcal{K}$ and $I'$ in $\mathcal{I}$ such that $I' \subset K \subset I$ and $\mu(I') + \varepsilon > \mu(I)$. As in Exercise 3.5, this implies the $\sigma$-additivity of $\mu$ on $\mathcal{I}$ and the sup representation (3.6). Alternatively, define $\mu([\{a\}) = \prod_{i=1}^n (F_i(a+) - F_i(a))$ to see that the initial additive finite-valued set function $\mu$ can be defined on the semi-ring of all bounded $d$-intervals $\mathcal{S}$ (which contains the compact class $\mathcal{K}$) and the sup representation (3.6) of Proposition 3.15 holds true. Finally, Theorem 3.19 shows that $\mu$ can be extended to a regular Borel measure on $\mathbb{R}^d$, which is inner regular.

- **Remark 3.21.** A dissecting system in a Hausdorff topological space $\Omega$ is a sequence $\mathcal{S} = \bigcup_n S_n$ of finite partitions $S_n = \{S_{ni} : i = 1, \ldots , k_n\}$ consisting of Borel sets satisfying:

1. (partition) $\Omega = S_{n1} \cup \cdots \cup S_{nk_n}$ and $S_{ni} \cap S_{nj} = \emptyset$ if $i \neq j$,
2. (nesting) for $m < n$ and any $i, j$ the intersection $A_{mj} \cap A_{ni}$ is either equal to $A_{ni}$ or empty,
3. (point-separating) for any $x \neq y$ in $\Omega$ there exists an integer $n = n(x, y)$
such that \( x \in S_{ni} \) and \( y \notin S_{ni} \).

This is useful to study atoms in general, e.g., for each \( x \) we can define a decreasing (or nested) sequence \( \{S_k(x)\} \subset S \) such that \( \bigcap_k S_k(x) = \{x\} \), so that for any finite measure \( \mu \) we have \( \mu(S_k(x)) \rightarrow \mu(\{x\}) \).

In a separable metric space \( \Omega \), a dissecting system can always be constructed, e.g., see the book by Daley and Vere-Jones [30, Appendixes A1 and A2, pp. 368–413] for further details.

Since \( \mathbb{R}^d \) is a complete separable metric space (Polish) then the Lebesgue measure \( m \) (actually any measure constructed as above, e.g., Lebesgue-Stieltjes measures) is a regular Borel outer measure and inner regular, i.e., for any \( \sigma \)-finite measure \( \mu \) on \( \mathcal{B}(\mathbb{R}^d) \) and for every Borel set \( B \) we have

\[
\mu(B) = \inf \{ \mu(O) : O \supset B, \text{ O open} \}
\]

and for any \( A \subset \mathbb{R}^d \) there exists a Borel set \( B \) such that \( A \subset B \) and \( \mu^*(A) = \mu(B) \). Indeed, this follows from Theorem 3.3 and Proposition 3.9. Moreover, the Lebesgue-(Stieltjes) measure \( m \) is a Radon measure, i.e., \( m(K) < \infty \) for every compact subset \( K \) of \( \mathbb{R}^d \). Clearly, the whole space \( \mathbb{R}^d \) is the support of Lebesgue measure \( m \), but in the case of a Lebesgue-Stieltjes measure, this is more complicated.

For instance, the reader may consult Konig [72, Chapter I-II, pp. 1-107] to understand better these two (exterior and interior) constructions of measures. Perhaps, reading part of the chapter on locally compact spaces, Baire and Borel sets of Halmos [57, Chapter X, pp. 216–249] may help, and certainly, checking Cohn [28, Chapter 7, pp. 196-250] could be a sequel for the reader.

### 3.4 Product Measures

We present a particular case to construct a finite product of inner regular Borel measures based on Proposition 3.15. Recall that an inner regular Borel measure is a regular Borel measure (i.e., a \( \sigma \)-finite measure on the Borel \( \sigma \)-algebra) satisfying the sup expression (3.4) with compacts.

**Proposition 3.22.** Let \( \mu_i \) be a \( \sigma \)-finite inner regular Borel measure on a topological space \( \Omega_i \), for \( i = 1, 2, \ldots, n \). Then there exists a unique \( \sigma \)-finite measure \( \mu \) on the product \( \sigma \)-algebra \( \prod_{i=1}^n \mathcal{B}(\Omega_i) \) such that

\[
\mu(B) = \prod_{i=1}^n \mu_i(B_i), \quad \forall B = \prod_{i=1}^n B_i, \quad B_i \in \mathcal{B}(\Omega_i),
\]

under the convention that \( 0 \times \infty = 0 \) in the above product. Moreover, if the Borel \( \sigma \)-algebra \( \mathcal{B}(\Omega) \) of the product topological space \( \Omega = \prod_{i=1}^n \Omega_i \) is equal to \( \prod_{i=1}^n \mathcal{B}(\Omega_i) \), e.g., each \( \Omega_i \) has a countable basis, then \( \mu \) is an inner regular Borel measure on \( \Omega \).
3.4. Product Measures

Proof. The key point is that the class

$$S = \left\{ S = \prod_{i=1}^{n} B_i : B_i \in B_i(\Omega_i) \right\}$$

is a semi-ring which generates the product Borel \(\sigma\)-algebra \(\prod_{i=1}^{n} B(\Omega_i)\). Thus, the above formula define \(\mu\) uniquely on \(S\) as an additive set function. To conclude, we need to show the sup expression (3.6) for the product (finitely additive, for now) measure \(\mu\).

The sub-additivity of \(\mu\) on \(S\) yields the \(\geq\) sign of the equality (3.6). For the reverse inequality (\(\leq\)), first, we assume \(0 < \mu(S) < \infty, S = \prod_{i=1}^{n} B_i\), and \(\varepsilon > 0\). Then, because \(\mu_i\) is inner regular, there exists a compact set \(K_i \subset B_i\) such that \(\mu_i(B_i) \leq \mu_i(K_i) + \varepsilon\). Hence

$$\mu(S) = \prod_{i=1}^{n} \mu_i(B_i) \leq \prod_{i=1}^{n} (\mu_i(K_i) + \varepsilon) \leq \mu(K) + [(c + \varepsilon)^n - c^n],$$

where \(K = \prod_{i=1}^{n} K_i\) and the constant \(c\) satisfies \(\mu_i(B_i) \leq c\), for every \(i\). Since the set \(K \in S\) is compact and \(\varepsilon\) is arbitrary, we deduce (3.6). Next, to examine the case \(\mu(S) = \infty\), we use a sequence \(\{S_k\} \subset S\) satisfying \(\mu(S_k) < \infty\) and \(\Omega = \bigcup_k S_k\). Because the sup equality holds for each \(S \cap S_k\), we conclude as \(k \to \infty\), i.e., the equality (3.6) holds true for the product measure \(\mu\) and the semi-ring \(S\).

Proposition 3.15 shows that \(\mu\) is \(\sigma\)-additive on the algebra \(A\) generated by \(S\), and by assumption \(\mu\) is \(\sigma\)-finite. Hence \(\mu\) is uniquely extended (see Theorem 2.9, Remark 3.6) to a regular Borel outer measure \(\mu^*\), which induces an inner regular Borel measure \(\bar{\mu}\) (abusing notation, also denoted by \(\mu\)).

- **Remark 3.23.** In Proposition 3.22, if the measures \(\mu_i\) are only semi-finite (i.e., for every \(S\) in \(S\) with \(\mu(S) = \infty\) there exists an increasing sequence \(\{S_k\} \subset S\) such that \(S_k \subset S_{k+1} \subset S\), \(\mu(S_k) < \infty\) and \(\mu(S_k) \to \infty\), see Remark 3.17 and Exercise 2.15) instead of \(\sigma\)-finite then the product set function \(\mu\) can be uniquely extended as a semi-finite measure on the product \(\sigma\)-algebra.

In general, the previous result (Proposition 3.22) degenerates for an infinite product of measures. Only the case of an infinite (possible uncountable) product of probability measures makes sense, the reader may check Halmos [57, Chapter VII, Section 38, pp. 154–160] and Ambrosio et al. [3, Section 6.3, pp. 90–94].

Recall that (1) a finite Borel measure \(\mu\) on a topological space \(\Omega\) is called tight if for every \(\varepsilon > 0\) there exists a compact set \(K_\varepsilon\) such that \(\mu(\Omega \setminus K_\varepsilon) \leq \varepsilon\); (2) a probability or probability measure \(\mu\) on \(\Omega\) is a (finite) measure space \((\mu, \Omega)\) satisfying \(\mu(\Omega) = 1\); (3) if \(\{(\Omega_i, B_i) : i \in I\}\) is any family of measurable spaces then a cylinder \(B\) in on the product space \(\Omega = \prod_{i \in I} \Omega_i\) is a set of the form \(B = \prod_{i \in I} B_i\), where \(B_i \in B_i, i \in I\) and \(B_i = \Omega_i, i \notin J\) with \(J \subset I\), finite. The number of indexes for which \(B_i \neq \Omega_i\) is call the dimension of the cylinder set \(B\) and the class \(B^I\) of all cylinders (or cylinder sets) is a semi-algebra, which generates the product \(\sigma\)-algebra \(\sigma\)-algebra \(\prod_{i \in I} B_i\). Similarly, if \(J \subset I\) is a
subset of indexes then the class $\mathcal{B}^J$ of all cylinders $B = \prod_{i \in I} B_i$ satisfying $B_i = \Omega_i$ for any $i \in I \setminus J$ is also a semi-ring, $\mathcal{B}^J \subset \mathcal{B}^I$. If $\pi_i : \Omega \to \Omega_i$, for every $i$ in $I$ are the projection mappings then a set $B \subset \Omega$ is cylinder if and only if $B = \bigcap_{i \in J} \pi_i(B_i)$ for some sets $B_i$ on $\mathcal{B}_i$ and some finite set of subindexes $J$.

Our interest is when $\mathcal{B}_i$ is the Borel $\sigma$-algebra. As briefly discussed in Section 1.2, if the set of indexes $I$ is countable and each space $\Omega_i$ has a countable basis then the Borel $\sigma$-algebra $\mathcal{B}$ on the product space $\Omega$ (with the product topology) coincides with the product $\sigma$-algebra $\prod_{i \in I} \mathcal{B}_i$.

If the topological space $\Omega$ is such that the Borel $\sigma$-algebra $\mathcal{B}(\Omega)$ results the smallest class which contains all closed sets and is stable under countable unions and intersections then Theorem 3.3 and Corollary 3.4 ensure that for any regular Borel measure $\mu$ has the following property: for any Borel set $B$ and for any $\varepsilon > 0$ there exist an open set $O$ and a closed set $C$ such that $C \subset B \subset O$ and $\mu(O \setminus C) < \varepsilon$. Recall that a topological space where each open set is a countable union of closed set (see Proposition 1.8) enjoys the satisfies the above condition, in particular, any metrizable space does. Also, recall that a (Hausdorff) topological space with a countable basis (i.e., satisfying the second axiom of countability) is metrizable and separable, e.g., see Kelley [67, Urysohn Theorem 10, Chapter 4, 115].

**Proposition 3.24.** Let $\{\Omega_i : i \geq 1\}$ be a sequence of topological spaces satisfying the second axiom countability. If $\mu_i$ is a tight regular Borel probability on $\Omega_i$ with $\mu_i(\Omega_i) = 1$, for any $i \geq 1$, then there exists a unique tight regular Borel probability $\mu$ on product topological space $\Omega = \prod_{i=1}^{\infty} \Omega_i$ such that

$$
\mu(B) = \prod_{i=1}^{n} \mu_i(B_i), \quad \forall B = \prod_{i=1}^{\infty} B_i, \quad B_i \in \mathcal{B}(\Omega_i), \quad B_i = \Omega_i, \quad \forall i \geq n + 1,
$$

and for any $n \geq 1$.

**Proof.** First note that under these assumptions Theorem 3.3 can be applied, i.e., for every $\varepsilon > 0$ and for every set $B$ in the Borel $\sigma$-algebra $\mathcal{B}(\Omega_i)$ there exist an open set $O$ and a closed set $C$ satisfying $C \subset B \subset O$ and $\mu_i(O \setminus C) < \varepsilon/2$; and in view of the tightness property, there exists a compact set $K \subset C$ satisfying $\mu_i(C \setminus K) < \varepsilon/2$, i.e., $\mu_i(O \setminus K) < \varepsilon$.

Now, denote by $\mathcal{B}^\infty(\mathcal{B}^r)$ the semi-algebra of all cylinder sets (with dimension at most $r$) in product space $\Omega$. It is clear that the product expression defines an additive set function $\mu$ on the semi-algebra $\mathcal{B}^\infty$, which can be extended, preserving additivity, to the algebra $A^\infty$ generated by all cylinder sets, i.e., the class of finite unions of disjoint cylinder sets. Note that a set $B$ (or $A$) belongs to $\mathcal{B}^\infty$ (or $A^\infty$) if and only if $B$ (or $A$) belongs to $\mathcal{B}^r$ (or $A^r$) for some finite index $r$.

Proposition 3.22 ensures that the restriction of the additive set function $\mu$ to the semi-algebra $\mathcal{B}^n$ is $\sigma$-additive. Our intension is to show that $\mu$ is actually $\sigma$-additive on $\mathcal{B}^\infty$, i.e., for any sequence $\{B_n\}$ in $\mathcal{B}^\infty$ such that $B = \sum_n B_n$ with $B$ in $\mathcal{B}^\infty$ we should establish that $0 < \mu(B) \leq \sum_n \mu(B_n)$.
To this purpose, for every $\varepsilon > 0$ and $B_n = \prod_i B_{n,i}$ with $B_{n,i}$ in $\mathcal{B}(\Omega_i)$ there exist open sets $O_{n,i} \supset B_{n,i}$ in $\Omega_i$ such that $\ln (\mu(O_{n,i})) < \ln (\mu(B_{n,i})) + \varepsilon 2^{-i}$, if $\mu(B_{n,i}) > 0$ and $\mu_i(O_{n,i}) < \varepsilon 2^{-i-n}$, if $\mu(B_{n,i}) = 0$, i.e., $B_n \subset O_n = \prod_i O_{n,i}$ is an open set in $\Omega$ belonging to $\mathcal{B}^\infty$ and either $\mu(O_n) < \varepsilon 2^{-n}$ or

$$\ln (\mu(O_n)) = \sum_i \ln (\mu_i(O_{n,i})) < \sum_i \ln (\mu_i(B_{n,i})) + \varepsilon = \ln (\mu(B_n)) + \varepsilon,$$

i.e., $\mu(O_n) < e^\varepsilon \mu(B_n) + \varepsilon 2^{-n}$. Similarly, if $B = \prod_i B_i$ then there exist compact sets $K_i$ in $\Omega_i$ such that $\ln (\mu_i(K_i)) > \ln (\mu_i(B_i)) - \varepsilon 2^{-i}$, i.e.,

$$\sum_i \ln (\mu_i(K_i)) > \sum_i \ln (\mu_i(B_i)) - \varepsilon = \ln (\mu(B)) - \varepsilon.$$

Tychonoff’s Theorem implies that the product set $K = \prod_i K_i$ is compact in $\Omega$, but $K$ is not necessarily in $\mathcal{B}^\infty$. Since $B = \sum_n B_n$, the sequence $\{O_n\}$ of open sets is a cover of the compact set $K$ in $\Omega$, and therefore, there exists a finite subcover, i.e., $K \subset \bigcup_{n \leq N} O_n$, for a finite index $N$. The finite union $\bigcup_{n \leq N} O_n = O$ is a set in the algebra $\mathcal{A}^\infty$, and thus, $O$ must belong to $\mathcal{A}^r$ for some finite index $r$, i.e., $O$ contains the cylinder set $K' = \prod_i K'_i$ with $K'_i = K_i$ if $i \leq r$ and $K'_i = \Omega_i$ if $i > r$. Thus, the set $K'$ belongs to $\mathcal{B}^r$ and $\ln (\mu(K')) > \ln (\mu(B)) - \varepsilon$. Since $\mu$ is additive on $\mathcal{A}^r$, we obtain $\mu(K') \leq \mu(O) \leq \sum_{n \leq N} \mu(O_n)$. Collecting all pieces, we deduce

$$\varepsilon + e^\varepsilon \sum_n \mu(B_n) \geq \sum_{n \leq N} \mu(O_n) \geq \mu(K') > e^{-\varepsilon} \mu(B),$$

which implies the $\sigma$-additivity as $\varepsilon \to 0$.

At this point, by means of Caratheodory’s extension Proposition 2.11, the $\sigma$-additive set measure $\mu$ can be extended to a probability measure defined on the product $\sigma$-algebra $\prod_i \mathcal{B}(\Omega_i)$. Since there is a countable basis for each topological space $\Omega_i$, the product $\sigma$-algebra is indeed the whole Borel $\sigma$-algebra $\mathcal{B}(\Omega)$.

To verify that $\mu$ is tight, for each $i$ we find a compact set $K_i \subset \Omega_i$ such that $\ln (\mu(K_i)) > -\varepsilon 2^{-i}$. Define $K = \prod_{i=1}^\infty K_i$ and $B_n = \prod_{i=1}^\infty B_{n,i}$ with $B_{n,i} = K_i$ for $i = 1, \ldots, n$ and $B_{n,i} = \Omega_i$ for $i > n$. Since $\bigcap_n B_n = K$ we have $\lim_n \mu(B_n) = \mu(K)$, and

$$\lim_n \ln (\mu(B_n)) = \lim_n \sum_{i=1}^n \ln (\mu(K_i)) \geq - \sum_i \varepsilon 2^{-i} = \ln (e^{-\varepsilon}),$$

we obtain $\mu(K) \geq e^{-\varepsilon}$, which means that $\mu$ is tight. Alternatively, Theorem 3.19 could be used to obtain, directly, the extension to an inner measure, which is necessarily tight. \qed

In the previous Proposition 3.24, the existence of a countable basis for each topological space $\Omega_i$ was used only to ensure that the product probability measure is defined on the whole Borel $\sigma$-algebra $\mathcal{B}(\Omega)$. If the topological space $\Omega$
is such that the Borel $\sigma$-algebra $\mathcal{B}(\Omega)$ results the smallest class which contains all closed sets and is stable under countable unions and intersections then product probability measure is defined only on the product $\sigma$-algebra $\prod_i \mathcal{B}(\Omega_i)$. If the topological spaces $\Omega_i$ are Polish spaces (i.e., complete separable metrizable spaces) then the tightness assumption on each probability $\mu_i$ is not necessary, indeed, Proposition 3.9 implies that any Borel probability measure on a Polish space is tight.

**Remark 3.25.** The setting of Proposition 3.24 can be twisted as follows: Suppose that $\{S^n, n \geq 1\}$ is an increasing sequence of semi-rings of the Borel $\sigma$-algebra of some topological space $\Omega$ such that the union semi-ring $S^\infty$ (i.e., $S$ belongs to $S^\infty$ if and only if $S$ belongs to $S^n$ for some finite index $n$) generates $\mathcal{B}(\Omega)$. Denote by $\mathcal{A}^n$ and $\mathcal{A}^\infty$ the algebras generated by $\{S^n\}$ and $S^\infty$ (i.e., the classes of finite disjoint unions of elements in the semi-ring). If $\nu_n$ is an additive finite-valued set function defined on the semi-ring $S^n$ then, as mentioned early, $\nu$ can be extended to an additive finite-valued set function defined the algebra $\mathcal{A}^n$. Moreover, if the sequence $\{\nu_n\}$ satisfies a compatibility condition, like $\nu_{n+1}(A) = \nu(A)$ for every $A$ in $\mathcal{A}^n$, then there exists a unique additive finite-valued set function $\nu$ defined on the algebra $\mathcal{A}^\infty$ such that $\nu(A) = \nu_n(A)$ for every $A$ in $\mathcal{A}^n$. The tightness assumption can be translated into

\[
\forall \varepsilon > 0 \text{ and } \forall A \in \mathcal{A}^\infty \text{ there exist } B \in \mathcal{A}^\infty \text{ and a compact set } K \text{ such that } B \subset K \subset A \text{ and } \nu(A \setminus B) < \varepsilon. \tag{3.12}
\]

The point is that if each $\nu_n$ satisfies this tightness (3.12) with $(\nu, \mathcal{A}^\infty)$ replaced with $(\nu_n, \mathcal{A}^n)$ then (1) $\nu_n$ is $\sigma$-additive and can be extended to a measure on the $\sigma$-algebra $\sigma(\mathcal{A}^n)$, and (2) the same argument applies to $\nu$ on $\mathcal{A}^\infty$, i.e., $\nu$ is $\sigma$-additive and can be extended to a measure on the $\sigma$-algebra $\sigma(\mathcal{A}^\infty) = \mathcal{B}(\Omega)$. The assertion (1) or (2) follows by proving that the additive finite-valued set function is monotone continuous from below at $\emptyset$, as in the proof of Proposition 3.15, see also Exercise 3.5. It is clear that $\nu_n = \prod_{i=1}^n \mu_i$, as in the previous Proposition 3.24, is the typical example.

**Exercise 3.6.** Recall that a compact metrizable space is necessarily separable and so it satisfies the second axiom of countability (i.e., there exists a countable basis). A topological space $\Omega$ is called a *Lusin space* if it is homeomorphic (i.e., there exists a bi-continuous bijection function between them) to a Borel subset of a compact metrizable space. Certainly, any Borel set in $\mathbb{R}^d$ is a Lusin space and actually, any Polish space (complete separable metrizable space) is also a Lusin space. Similarly to Proposition 3.24, if $\{\Omega_i : i = 1, 2, \ldots, n, \ldots\}$ is a sequence of Lusin spaces then verify (1) that the product space $\Omega = \prod_i \Omega_i$ is also a Lusin space. Let $\{\nu_n\}$ be a sequence of $\sigma$-additive set functions defined on the algebra $\mathcal{A}^n$, generated by all cylinder sets of dimension at most $n$. Verify (2) that $\nu^n$ can be (uniquely) extended to a measure on the $\sigma$-algebra $\sigma(\mathcal{A}^n)$ generated by $\mathcal{A}^n$, and that the particular case of a finite product measures $\prod_{i \leq n} \mu_i$ can be taken as $\nu^n$. Assume that $\{\nu_n\}$ is a compatible sequence of probabilities, i.e., $\nu_{n+1}(A) = \nu_n(A)$ for every $A$ in $\mathcal{A}^n$ and $\nu_n(\Omega) = 1$, for every $n \geq 1$. Now, prove
(3) that if each $\Omega_i$ is a compact metrizable space then there exists a unique probability measure $\nu$ defined on the product $\sigma$-algebra $\prod_i B(\Omega_i)$ such that $\nu(A) = \nu_n(A)$, for every $A$ in the algebra $\mathcal{A}^n$. Finally, show (4) the same result, except for the uniqueness, when each $\Omega_i$ is only a Lusin space. In probability theory, this construction is referred to as Daniell-Kolmogorov Theorem, e.g., see Rogers and Williams [97, Vol 1, Sections II.3.30-31, pp. 124–127].

- **Remark 3.26.** If should be clear that the choice of the natural index $i = 1, 2, \ldots$ is unimportant in Proposition 3.24, what really count is to have a denumerable index set $I$. Indeed, for any cylinder sets of the form $B = \prod_{i \in I} B_i$, where $B_i \in B_i$, $i \in I$ and $B_i = \Omega_i$, $i \notin J$ with $J \subset I$, finite, we can use $\mu(B) = \prod_{i \in J} \mu_i(B_i)$ to define the infinite product probability measure $\mu = \prod_{i \in I} \mu_i$. Actually, the same argument remains valid for uncountable index sets (see next Exercise), and the construction of the infinite product of probability on the product $\sigma$-algebra can be accomplished without any topological assumptions, e.g., see Halmos [57, Section VII.38, pp. 157–160]. However, product $\sigma$-algebra becomes too small for an uncountable product of topological spaces.

**Exercise 3.7.** In a product space $\Omega = \prod_{i \in I} \Omega_i$ the projection mappings $\pi_J : \Omega \rightarrow \Omega_J = \prod_{i \in J} \Omega_j$ are defined as $\pi_J((\omega_i : i \in J)) = (\omega_i : i \in J)$, for any subindex $J$ of $I$. Assume that the index $I$ is uncountable, first (1) prove that a set $A$ belongs to the product $\sigma$-algebra of the uncountable product space $\Omega$ if and only if for some countable subset $J$ of indexes the projection $\pi_J(A)$ belongs to the product $\sigma$-algebra of the countable product space $\Omega_J$ and $\pi_{I \setminus J}(A) = \Omega_{I \setminus J}$. Next (2) show that Proposition 3.24 can be extended to the case of an uncountable product of probability spaces. Finally, (3) discuss how to extend Remark 3.25 and Exercise 3.6 to the uncountable case.

Given a measure space $(\Omega, \mathcal{F}, \mu)$, every atom $A$ of $\mathcal{F}$ is not necessarily relevant for $\mu$, only those with positive measure count. Thus, an atom on $(\Omega, \mathcal{F}, \mu)$ is a measurable set $A \in \mathcal{F}$ satisfying (1) $\mu(A) > 0$ and (2) any $B \in \mathcal{F}$ with $B \subset A$ is such that $\mu(B) = 0$ or $\mu(A \setminus B) = 0$. Moreover, atoms $A$ and $B$ such $\mu((A \setminus B) \cup (B \setminus A)) = 0$ are considered equals, i.e., atom becomes a family $[A] \subset \mathcal{F}$ of sets (equal almost everywhere) which are considered as single sets. Thus, for any measurable set $F$ such that $0 < c < \mu(F) < \infty$ there is at most a finite family of atoms $A \subset F$ satisfying $c \leq \mu(A) \leq \mu(F)$. Moreover, if $\mu$ is $\sigma$-finite then there are no atoms of infinite measure and there is only countable many atoms with finite measure. If $\{A_n\}$ is the sequence (possible finite or empty) of all atoms with finite measure then $B \mapsto \mu_a(B) = \sum_n \mu(A_n \cap B)$ is the atomic or discrete part of $\mu$ and $B \mapsto \mu_c(B) = \mu(B \setminus \bigcup_n A_n)$ is a measure without any atoms. It is possible to show that for any $\varepsilon > 0$ and any $F \in \mathcal{F}$ with $\mu(F) < \infty$ there exists a finite partition $\{F_i : i = 1, \ldots, k\}$ of $F$ such that either $\mu(F_i) \leq \varepsilon$ or $F_i$ is an atom with $\mu(F_i) > \varepsilon$, e.g., see Neveu [87, Exercise I.4.3, p. 18] or Dunford and Schwartz [40, Vol I, Lemma IV.9.7, pp. 308-309].

**Exercise 3.8.** Formalize the previous comments, e.g., (1) prove that for any measurable set $F$ such that $0 < c < \mu(F) < \infty$ there is at most a finite family
of atoms $A \subset F$ satisfying $c \leq \mu(A) \leq \mu(F)$; and, assuming that $\mu$ is $\sigma$-finite, (2) deduce that there are no atoms of infinite measure and there is a countable family (possible finite or empty) containing all atoms, moreover (3) discuss in some details the existence of the partition $\{F_i : i = 1, \ldots, k\}$.  

Sometimes, we prefer to skip the whole previous Chapter 3 on measures and topology, and as a consequence, we are forced to establish formula (3.11) directly for the Lebesgue measure $\mu = m$. Thus, at this point, the reader may benefice of reconstructing the key elements necessary to obtain the inf and sup expressions of (3.11) for the Lebesgue measure (e.g., in $\mathbb{R}^d$ with $d = 1$ to simplify) without a direct reference to Chapter 3. However, the treatment of the product measure based on Theorem 3.22 is rarely presented in this way, it is customary to construct the product measure later, together with the integration on product spaces, where the $\sigma$-additivity is easily shown.
Chapter 4

Integration Theory

Recall that a simple function \( \varphi : \Omega \to \mathbb{R} \) is a measurable functions assuming a finite number of values, i.e., a linear (finite with real coefficients) combination of characteristic functions. Any simple function has a standard represented as \( \varphi(x) = \sum_{i=1}^{n} a_i \mathbb{1}_{E_i}(x) \), with \( a_i \neq a_j \) for \( i \neq j \) and \( \{E_i\} \) a finite sequence of disjoint measurable sets. Denote by \( \mathcal{S} = \mathcal{S}(\Omega, \mathcal{F}) \) the set of all simple functions on a measurable space \( (\Omega, \mathcal{F}) \). Clearly \( \mathcal{S} \) is stable under the addition, multiplication, \( \max (\vee) \) and \( \min (\wedge) \), i.e., if \( \varphi, \psi \in \mathcal{S} \) and \( a, b \in \mathbb{R} \) then \( a\varphi + b\psi, \varphi \vee \psi, \varphi \wedge \psi \in \mathcal{S} \). Also, we have seen in Theorem 1.9, that simple functions can be used to approximate pointwise any measurable function.

Once a measure space \( (\Omega, \mathcal{F}, \mu) \) has been given, it is clear that for any measurable set \( F \) we should assign the value \( \mu(F) \) as the integral of the characteristic function \( \mathbb{1}_F \). Then, by imposing linearity, for a simple function \( \varphi(x) \) we should have

\[
\int_{\Omega} \varphi(x) \mu(dx) = \sum_{i=1}^{n} a_i \mu(\varphi^{-1}(\{a_i\}))
\]

under the convention that the sum is possible, i.e., we set \( a \times (+\infty) = 0 \) if \( a = 0 \), \( a \times (+\infty) = +\infty \) if \( a > 0 \), \( a \times (-\infty) = +\infty \) if \( a < 0 \), and the case \( \pm\infty \pm\infty \) is forbidden. Hence, we can approximate any nonnegative measurable function \( f \) by an increasing sequence of simple functions to have

\[
\lim_n \int_{\Omega} f \mathbb{1}_{f<2^n} \, d\mu = \lim_n \sum_{k=0}^{2^{2n}-1} k2^{-n} \mu(\varphi^{-1}([k2^{-n}, (k+1)2^{-n}]))
\]

which is always meaningful, and then writing \( f = f^+ - f^- \) we treat the general case. Details of these arguments follow.

4.1 Definition and Properties

Let \((\Omega, \mathcal{F}, \mu)\) be a measurable space. If \( \varphi : \Omega \to [0, \infty) \) is a simple function with standard represented as \( \varphi(x) = \sum_{i=1}^{n} a_i \mathbb{1}_{E_i}(x) \), with \( a_i \neq a_j \) for \( i \neq j \) and \( \{E_i\} \)
a finite sequence of disjoint measurable sets, then we define the integral of \( \varphi \) over \( \Omega \) with respect to the measure \( \mu \) as

\[
\int_{\Omega} \varphi \, d\mu = \int_{\Omega} \varphi(\omega) \mu(d\omega) = \int_{\Omega} \varphi(\omega) \, d\mu(\omega) = \sum_{i=1}^{n} a_i \mu(F_i),
\]

under the only convention \( 0 \times (+\infty) = 0 \), since \( \varphi \geq 0 \).

**Proposition 4.1.** If \( \varphi \) and \( \psi \) are nonnegative simple functions then

(a) \( \int_{\Omega} c \varphi \, d\mu = c \int_{\Omega} \varphi \, d\mu, \quad \forall c \geq 0 \),

(b) \( \int_{\Omega} (\varphi + \psi) \, d\mu = \int_{\Omega} \varphi \, d\mu + \int_{\Omega} \psi \, d\mu, \)

(c) if \( \varphi \leq \psi \), then \( \int_{\Omega} \varphi \, d\mu \leq \int_{\Omega} \psi \, d\mu \) (monotony),

(d) the function \( A \to \int_{A} \varphi \, d\mu \) is a measure on \( \mathcal{F} \).

**Proof.** The property (a) follows directly from the definition of the integral.

To check the identity (b) take standard representations \( \varphi = \sum_{i=1}^{n} a_i 1_{F_i} \) and \( \psi = \sum_{j=1}^{m} b_j 1_{G_j} \). Since \( F_i = \bigcup_{j=1}^{m} F_i \cap G_j \) and \( G_j = \bigcup_{i=1}^{n} F_i \cap G_j \), both disjoint unions, the finite additivity of \( \mu \) implies

\[
\int_{\Omega} (\varphi + \psi) \, d\mu = \int_{\Omega} \sum_{j=1}^{m} \sum_{i=1}^{n} (a_i + b_j) 1_{F_i \cap G_j} \, d\mu =
\]

\[
= \sum_{j=1}^{m} \sum_{i=1}^{n} (a_i + b_j) \mu(F_i \cap G_j) =
\]

\[
= \sum_{j=1}^{m} \sum_{i=1}^{n} a_i \mu(F_i \cap G_j) + \sum_{j=1}^{m} \sum_{i=1}^{n} b_j \mu(F_i \cap G_j) =
\]

\[
= \sum_{i=1}^{n} a_i \mu(F_i) + \sum_{j=1}^{m} b_j \mu(G_j) = \int_{\Omega} \varphi \, d\mu + \int_{\Omega} \psi \, d\mu.
\]

as desired.

To show (c), if \( \varphi \leq \psi \) then \( a_i \leq b_j \) each time that \( F_i \cap F_j \neq \emptyset \), hence

\[
\int_{\Omega} \varphi \, d\mu = \sum_{j=1}^{m} \sum_{i=1}^{n} a_i \mu(F_i \cap G_j) \leq \sum_{j=1}^{m} \sum_{i=1}^{n} b_j \mu(F_i \cap G_j) = \int_{\Omega} \psi \, d\mu.
\]
For (d), we have to prove only the countable additivity. If \( \{A_j\} \) are disjoint and \( A = \bigcup_{j=1}^{\infty} A_j \) then

\[
\int_A \varphi \, d\mu = \sum_{i=1}^{n} a_i \mu(A \cap F_i) = \sum_{i=1}^{n} \sum_{j=1}^{\infty} a_i \mu(A_j \cap F_i) = \sum_{j=1}^{\infty} \sum_{i=1}^{n} a_i \mu(A_j \cap F_i) = \sum_{j=1}^{\infty} \int_{A_j} \varphi \, d\mu,
\]

and we conclude.

**Definition 4.2.** If \( f \) is a nonnegative measurable function then we define the integral of \( f \) of \( \Omega \) with respect to \( \mu \) as

\[
\int_{\Omega} f \, d\mu = \sup \left\{ \int_{\Omega} \varphi \, d\mu : \varphi \text{ simple, } 0 \leq \varphi \leq f \right\},
\]

which is nonnegative and perhaps \(+\infty\). If \( f \) is a measurable function with valued in \([-\infty, +\infty]\), writing \( f = f^+ - f^- \), then

\[
\int_{\Omega} f \, d\mu = \int_{\Omega} f^+ \, d\mu - \int_{\Omega} f^- \, d\mu < \infty,
\]

whenever the above expression is defined (i.e., \( \pm \infty \mp \infty \) is not allowed), and in this case \( f \) is called quasi-integrable. If both integrals are finite then we say that \( f \) is integrable.

By means of the previous proposition, part (c), implies that both definitions agree on simple functions, and parts (a) and (c) remain valid if \( \varphi = f \) and \( \psi = g \) for any integrable functions. To check the linearity, we use the following result. Since \( f^+, f^- \leq |f| = f^+ + f^- \), given a measurable functions \( f \), we deduce that \( f \) is integrable if and only if \( |f| \) is integrable.

Sometimes, an integrable function (as above, with finite integral) is called summable, while a quasi-integrable function (as above, with possible infinite integral) is called integrable.

We keep the notation

\[
\int_A f \, d\mu = \int_{\Omega} 1_A \, d\mu, \quad \forall A \in \mathcal{F}
\]

and the inequality

\[
c \mu(\{|f| \geq c\}) \leq \int_{\Omega} |f| 1_{\{|f| \geq c\}} \, d\mu \leq \int_{\Omega} |f| \, d\mu, \quad \forall c \geq 0,
\]

shows that if \( f \) is integrable then the set \( \{|f| \geq c\} \) has finite \( \mu \)-measure, for every \( c > 0 \), and so the set \( \{f \neq 0\} \) is \( \sigma \)-finite. On the other hand, a measurable function \( f \) is allowed to assume the values \(+\infty\) and \(-\infty\), but an integrable function is finite almost everywhere, i.e., \( \mu(\{|f| = \infty\}) = 0 \).
• Remark 4.3. Instead of initially defining the integral for nonnegative simple functions with the convention \(0 \infty = 0\), we may consider only (nonnegative) integrable simple functions in Proposition 4.1. In this case, only (nonnegative) measurable functions which vanish outside of a \(\sigma\)-finite set can be expressed as a (monotone) limit of integrable (nonnegative) integrable simple functions, see Proposition 1.9.

A key point is the monotone convergence

**Theorem 4.4** (Beppo Levi). If \(\{f_n\}\) is a monotone increasing sequence of nonnegative measurable functions then

\[
\int f_n \, d\mu = \lim_n \int f_n \, d\mu \quad \text{or} \quad \int (\sup_n f_n) \, d\mu = \sup_n \{\int f_n \, d\mu\}.
\]

**Proof.** Since \(f_n \leq f_{n+1}\) for every \(n\), the limiting function \(f\) is defined as taking values in \([0, +\infty]\) and the monotone limit of the integral exists (finite or infinite). Moreover

\[
\int f_n \, d\mu \leq \int f \, d\mu \quad \text{and} \quad \lim_{n \to \infty} \int f_n \, d\mu \leq \int f \, d\mu.
\]

To check inverse inequality, for every \(\alpha \in (0, 1)\) and every simple function \(\varphi\) such that \(0 \leq \varphi \leq f\) define \(F_n = \{x : f_n(x) \geq \alpha \varphi(x)\}\). Thus \(\{F_n\}\) is an increasing sequence of measurable sets with \(\bigcup_n F_n = \Omega\) and

\[
\int f_n \, d\mu \geq \int f_n \, d\mu \geq \alpha \int F_n \varphi \, d\mu.
\]

By means of Proposition 4.1, part (d), and the continuity from below of a measure, we have

\[
\lim_n \int F_n \varphi \, d\mu = \int \varphi \, d\mu, \quad \text{and} \quad \lim_{n \to \infty} \int f_n \, d\mu \geq \alpha \int \varphi \, d\mu.
\]

Since this holds for any \(\alpha < 1\), we can take \(\alpha = 1\). Taking the sup in \(\varphi\) we deduce

\[
\lim_n \int f_n \, d\mu \geq \int f \, d\mu,
\]

as desired inequality. \(\square\)

The additivity follows from Beppo Levi Theorem, i.e., if \(\{f_n\}\) is a finite or infinite sequence of nonnegative measurable functions and \(f = \sum_n f_n\) then

\[
\int f \, d\mu = \sum_n \int f_n \, d\mu.
\]

Indeed, first for any two functions \(g\) and \(h\), we can find two monotone increasing sequences \(\{g_n\}\) and \(\{h_n\}\) of nonnegative simple functions pointwise convergent
to \( g \) and \( h \). Thus \( \{g_n + h_n\} \) is a monotone increasing sequence pointwise convergent to \( g + h \), and by means of Theorem 4.4

\[
\int_{\Omega} (g + h) \, d\mu = \lim_{n} \int_{\Omega} (g_n + h_n) \, d\mu = \lim_{n} \int_{\Omega} g_n \, d\mu + \lim_{n} \int_{\Omega} h_n \, d\mu = \int_{\Omega} g \, d\mu + \int_{\Omega} h \, d\mu.
\]

Hence, by induction we deduce

\[
\int_{\Omega} \left( \sum_{n=1}^{m} f_n \right) \, d\mu = \sum_{n=1}^{m} \int_{\Omega} f_n \, d\mu,
\]

and applying again Theorem 4.4 as \( m \to \infty \) follows the desired equality.

**Remark 4.5.** Because the integral is unchanged when the integrand is modified in a negligible set, the results of Beppo Levi Theorem 4.4 remain valid for an almost monotone sequence \( \{f_n\} \), i.e., when \( f_{n+1} \geq f_n \) a.e., of measurable functions non necessarily nonnegative, but such that \( f_1^- \) is integrable. \( \square \)

Based on the monotone convergence, we deduce two results on the passage to the limit inside the integral. First, Fatou lemma or lim inf convergence

**Theorem 4.6 (Fatou).** If \( \{f_n\} \) is a sequence of nonnegative measurable functions then

\[
\int_{\Omega} \lim \inf_{n} f_n \, d\mu \leq \lim \inf_{n} \int_{\Omega} f_n \, d\mu.
\]

**Proof.** For each \( k \) we have that \( \inf_{n \geq k} f_n \leq f_j \) for every \( j \geq k \), which implies that

\[
\int_{\Omega} \inf_{n \geq k} f_n \, d\mu \leq \int_{\Omega} f_j \, d\mu \quad \text{and} \quad \int_{\Omega} \inf_{n \geq k} f_n \, d\mu \leq \inf_{j \geq k} \int_{\Omega} f_j \, d\mu.
\]

Hence, applying Theorem 4.4 as \( k \to \infty \) we have

\[
\int_{\Omega} \lim \inf_{n} f_n \, d\mu = \int_{\Omega} \lim \inf_{k} \sum_{n \geq k} f_n \, d\mu = \lim_{k} \int_{\Omega} \inf_{n \geq k} f_n \, d\mu \leq \lim \inf_{n} \int_{\Omega} f_n \, d\mu,
\]

i.e., the desired result. \( \square \)

Secondly, Lebesgue or dominate convergence

**Theorem 4.7 (Lebesgue).** Let \( \{f_n\} \) be a sequence of measurable functions such that there exists an integrable function \( g \) satisfying \( |f_n(x)| \leq g(x) \), for every \( x \) in \( \Omega \) and any \( n \). Then the functions \( \overline{f} = \lim \sup_n f_n \) and \( \underline{f} = \lim \inf_n f_n \) are integrable and

\[
\int_{\Omega} f \, d\mu \leq \lim \inf_{n} \int_{\Omega} f_n \, d\mu \leq \lim \sup_{n} \int_{\Omega} f_n \, d\mu \leq \int_{\Omega} \overline{f} \, d\mu.
\]
In particular,
\[ \lim_{n} \int_{\Omega} f_n \, d\mu = \int_{\Omega} f \, d\mu. \]

provided \( \overline{f} = \underline{f} = f \), i.e., \( f_n \) converges pointwise to \( f \).

Proof. First, note that the condition \(|f_n(x)| \leq g(x)\) (valid also for the limit \( \overline{f} \) or \( \underline{f} \)) implies that \( f_n \) (and the limit \( \overline{f} \) or \( \underline{f} \)) is integrable. Next, apply Fatou lemma to \( g + f_n \) and \( g - f_n \) to obtain
\[
\int_{\Omega} (g + f) \, d\mu \leq \lim \inf_{n} \int_{\Omega} (g + f_n) \, d\mu
\]
and
\[
\lim \sup_{n} \int_{\Omega} (g + f_n) \, d\mu \leq \int_{\Omega} (g + \overline{f}) \, d\mu,
\]
Finally, using the fact that \( g \) is integrable, we deduce (4.1), which implies the desired equalities. \qed

• Remark 4.8. We could re-phase the previous Theorem 4.7 as follows: If \( \{f_n\} \) and \( \{g_n\} \) are sequences of measurable functions satisfying \(|f_n| \leq g_n\), a.e. for any \( n \), and
\[
g_n \rightarrow g \text{ a.e. and } \int_{\Omega} g_n \, d\mu \rightarrow \int_{\Omega} g, d\mu < \infty,
\]
then the inequality (4.1) holds true. Indeed, applying Fatou lemma to \( g_n + f_n \) we obtain
\[
\int_{\Omega} (g + f) \, d\mu = \int_{\Omega} \lim \inf_{n} (g_n + f_n) \, d\mu \leq \lim \inf_{n} \int_{\Omega} (g_n + f_n) \, d\mu = \int_{\Omega} g \, d\mu + \lim \inf_{n} \int_{\Omega} f_n \, d\mu,
\]
which yields the first part of the inequality (4.1), after simplifying the (finite) integral of \( g \). Similarly, by using \( g_n - f_n \), we conclude. \qed

In the above presentation, we deduced Fatou and Lebesgue Theorems 4.6 and 4.7 from Beppo Levi Theorem 4.4, actually, from any one of them, we can obtain the other two.

In the previous arguments, we have followed Lebesgue’s recipe to extend the definition of the integral, i.e., assuming the integral defined for any simple function we extend its definition to measurable functions with finite integral. Alternatively, we may use the Daniell-Riesz approach, namely, assuming the integral defined for step functions we extend its definition to a larger class of functions, see Exercises 4.4, 4.5, 4.6 and 4.22 (later on). In this context, a real-valued step function \( \varphi \) has the form \( \varphi = \sum_{i=1}^{n} r_i \mathbb{1}_{E_i} \), with \( E_i \) in \( \mathcal{E} \), where \( \mathcal{E} \)
is a semi-ring covering the whole space \( \Omega \), i.e., such that \( \Omega = \bigcup_i \Omega_i \) for some sequence \( \{\Omega_i\} \) in \( \mathcal{E} \). The set of all step functions \( \mathcal{E} \) forms a vector lattice space, where the integral is naturally defined. The intrinsic difference is that to define the integral for any simple function we need a measure \( \mu \) defined on a \( \sigma \)-algebra, while to define the integral for any step function we only need a (finite) measure \( \mu \) defined on a semi-ring, covering the whole space. In this case, instead of Definition 4.2, we define the superior integral

\[
\int_\Omega f \, d\mu = \inf \left\{ \lim_n \int_\Omega f_n \, d\mu : \lim_n f_n \geq f, \ f_n \leq f_{n+1} \in \mathcal{E} \right\}
\]

and the inferior integral

\[
\int_\Omega f \, d\mu = -\int_\Omega (-f) \, d\mu,
\]

where \( f \) is any extended real-valued function, \( \{f_n\} \) is any monotone increasing sequence in \( \mathcal{E} \), and the inequality is considered pointwise everywhere. If

\[
\int_\Omega f \, d\mu = \int_\Omega f \, d\mu < \infty
\]

then the function \( f \) is called integrable.

Some serious work is needed to show the linearity of the integral and even more to obtain the three convergence theorems, in particular, a delicate point is to study the concept of sets of zero-measure. In Daniell-Riesz approach, we bypass the Caratheodory’s construction of the measure \( \mu \), and we construct first the integral, and a posteriori, we obtain a measure define on the \( \sigma \)-algebra generated by the integrable sets, i.e., subsets \( A \) of \( \Omega \) such that \( 1_A \) is an integrable function. For instance, the interested reader may take a look at the book by Royden [98, Chapter 4].

**Exercise 4.1.** Let \((\Omega, \mathcal{F}, \mu)\) be a measure space.

1. Show that if \( f \) is an integrable function and \( N \) is a set of measure zero then

\[
\int_N f \, d\mu = 0.
\]

2. Prove that if \( f \) is a strictly positive integrable function and \( E \) is a measurable set such that

\[
\int_E f \, d\mu = 0
\]

then \( E \) is a set of measure zero.

3. Suppose that an integrable function \( f \) satisfies

\[
\int_E f \, d\mu = 0,
\]

for every measurable set \( E \), deduce that \( f = 0 \) a.e.
(4) If $f$ is a measurable function and $g$ is an integrable function such that $|f| \leq g$ a.e., then $f$ is also an integrable function.

**Remark 4.9.** As mentioned early, the use of the concept “almost everywhere” for a pointwise property in a measure space $(\Omega, \mathcal{F}, \mu)$ is very import, essentially, insisting in a pointwise property could be unwise. For instance, the statement $f = 0$ a.e. means strictly speaking that the set $\{x : f(x) \neq 0\}$ belongs to $\mathcal{F}$ and $\mu(\{x : f(x) \neq 0\}) = 0$, but it also could be understood in a large sense as requiring that there exists a set $N$ in $\mathcal{F}$ such that $\mu(N)$ and $f(x) = 0$ for every $x$ in $\Omega \setminus N$. Thus, the large sense refers to the strict sense when $(\mathcal{F}, \mu)$ is complete and certainly, both concepts are the same if the measure space $(\Omega, \mathcal{F}, \mu)$ is complete. Sometimes, we may build-in this concept inside the definition of the integral by adding the condition almost everywhere, i.e., using

$$
\int_{\Omega} f \, d\mu = \sup \left\{ \int_{\Omega} \varphi \, d\mu : \varphi \text{ simple}, 0 \leq \varphi \leq f \text{ a.e.} \right\}
$$

as the definition of integral (where the a.e. inequality is understood in the large sense) for any nonnegative “almost” measurable function, i.e., any function $f$ such that there exist a negligible set $N$ and a nonnegative measurable function $g$ such that $f(x) = g(x)$, for every $x$ in the complement $N^c$. Therefore, parts (1) and (2) of the Exercise 4.1 are necessary to prove that the above definition of integral (with the a.e. inequality) is indeed meaningful and non-ambiguous.

**Exercise 4.2.** Give examples of sequences $\{f_n\}$ of real-valued measurable functions satisfying one following conditions (a) $\{f_n\}$ is pointwise decreasing to 0, but the integral of $f_n$ does not converges to 0; (b) $f_n \geq 0$ is integrable for every $n$ and the inequality in Fatou’s Lemma holds strictly; (c) the integral of $f_n \leq 0$ is a bounded numerical sequence, $f_n$ converges pointwise to an integrable function $f$, but the integral of $f_n$ does not converges to the integral of $f$.

**Exercise 4.3.** Given a nonnegative measurable function $h$ on a measure space $(\Omega, \mathcal{F}, \mu)$, define the set function

$$
\lambda(A) = \int_{A} h \, d\mu, \quad \forall A \in \mathcal{F}.
$$

Prove that $\lambda$ is a measure and that

$$
\int_{\Omega} f \, d\lambda = \int_{\Omega} fh \, d\mu,
$$

for every nonnegative measurable function $f$.

**Exercise 4.4.** Let $\mathcal{E}$ be a semi-ring of measurable space $(\Omega, \mathcal{F})$ and $\mathcal{E}$ be the space of $\mathcal{E}$-step functions, i.e., functions of the form $\varphi = \sum_{i=1}^{n} r_i \mathbb{1}_{E_i}$, with $E_i$ in $\mathcal{E}$ and $r_i$ in $\mathbb{R}$. (1) Verify that $\mathcal{E}$ is a vector lattice and any $\mathcal{R}$-step function belongs to $\mathcal{E}$, where $\mathcal{R}$ is the ring generated by $\mathcal{E}$. Given an additive and finite
measure $\mu$ on $E$ we define the integral for functions in $E$ by the formula

$$I(\varphi) = \sum_{i=1}^{n} r_i \mu(E_i), \quad \text{if } \varphi = \sum_{i=1}^{n} r_i \mathbb{1}_{E_i}.$$  

(2) Prove that $\mu$ is $\sigma$-additive on $E$ if and only if for any decreasing sequence $\{\varphi_k\}$ in $E$ such that $\varphi_k(x) \downarrow 0$ for every $x$, we have $I(\varphi_k) \downarrow 0$. \hfill $\Box$

**Exercise 4.5.** Let $E$ be a semi-ring of measurable sets in a measure space $(X, \mathcal{F}, \mu)$, and $E$ be the class of $E$-step functions, see Exercise 4.4. Suppose that any function $\varphi$ in $E$ is integrable and define

$$I(\varphi) = \int_{\Omega} \varphi d\mu,$$

A subset $N$ of $X$ is called a $I$-null or $I$-negligible set if there exists a decreasing sequence $\{\varphi_k\} \subseteq E$ and a constant $C$ such that (a) $\varphi_k(x) \uparrow +\infty$ for every $x$ in $N$ and (b) $I(\varphi_k) \leq C$, for every $k \geq 1$.

(1) Prove that (a) if $\varphi \in E$ with $\varphi \geq 0$ outside of a $I$-null set then $I(\varphi) \geq 0$, and (b) if $\varphi \in E$ with $\varphi \geq 0$ and $I(\varphi) = 0$ then $\varphi = 0$ outside of a $I$-null set.

(2) Show that a set $N$ is $I$-null if and only if for every $\varepsilon > 0$ there exists a sequence $\{E_k\}$ in $E$ such that (a) $N \subseteq \bigcup_{k} E_k$ and (b) $\sum_{k} I(\mathbb{1}_{E_k}) < \varepsilon$. \hfill $\Box$

**Exercise 4.6.** Let $E$ be a vector lattice of real-valued function defined on $X$ and $I : E \to \mathbb{R}$ be a linear and monotone functional satisfying the condition: if $\{\varphi_n\}$ is a decreasing sequence in $E$ such that $\varphi_n(x) \downarrow 0$ for every $x$ in $X$, then $I(\varphi) \downarrow 0$. A functional $I$ as above is called a pre-integral or a Daniell integral (functional) on $E$. Now, let $\bar{E}$ be the semi-space of extended real-valued functions which can be expressed as the pointwise limit of a monotone increasing sequence of functions in $E$. Verify that repeating this procedure does not add more functions, i.e., any pointwise limit of a monotone increasing sequence of functions in $\bar{E}$ is a function in $\bar{E}$. Moreover, also verify that if $\varphi$ and $\psi$ belong to $\bar{E}$ and $c$ is a positive real number then $\varphi + c\psi$, $\varphi \land \psi$ and $\varphi \lor \psi$ belong to $\bar{E}$.

(1) Define $I$-null sets and prove assertion (1) as in Exercise 4.5, with $E$ replaced with $\bar{E}$, i.e., (a) if $\{\varphi_n\}$ is an increasing sequence in $E$ such that $\lim_n \varphi_n(x) < 0$ on a $I$-null set then $\lim_n I(\varphi_n) \geq 0$, and (b) if $\{\varphi_n\}$ is an increasing sequence in $E$ such that $\lim_n \varphi_n \geq 0$ and $\lim_n I(\varphi_n) = 0$ then $\lim_n \varphi_n = 0$ except in a $I$-null set. Similarly, show that a countable union of $I$-null sets is a $I$-null set.

(2) Prove that the limit $I(\varphi) = \lim_n I(\varphi_n)$ if $\varphi = \lim_n \varphi_n$ provides a unique extension of $I$ as a semi-linear mapping from $\bar{E}$ into $(-\infty, \infty]$, i.e., (a) for any two increasing sequences $\{\varphi_n\}$ and $\{\psi_n\}$ in $E$ pointwise convergent to $\varphi$ and $\psi$ with $\varphi \leq \psi$ we have $\lim_n I(\varphi_n) \leq \lim_n I(\psi_n)$, and (b) for every $\varphi, \psi$ in $\bar{E}$ and $c$ in $[0, \infty)$ we have $I(\varphi + c\psi) = I(\varphi) + cI(\psi)$, under the convention that $0\infty = 0$. Also show the monotone convergence, i.e., if $\{\varphi_n\}$ is an increasing sequence of functions in $\bar{E}$ then $I(\lim_n \varphi_n) = \lim_n I(\varphi_n)$. How about if the sequence $\{\varphi_n\}$ is increasing only outside of a $I$-null set?
(3) Check that if \( \varphi \) belongs to \( \mathbb{E} \) and \( I(\varphi) < \infty \) then \( \varphi \) assumes (finite) real values except in a \( I \)-null set. A function \( f \) belongs to the class \( L^+ \) if \( f \) is equal, except in a \( I \)-null set, to some nonnegative function \( \varphi \) in \( \mathbb{E} \). Next, verify that \( I \) can be uniquely extended to \( L^+ \), by setting \( I(f) = I(\varphi) \).

(4) The class \( L \) of \( I \)-integrable functions is defined as all functions \( f \) that can be written in the form \( f = g - h \) (except in a \( I \)-null set) with \( g \) and \( h \) in \( \mathbb{E} \) satisfying \( I(g) + I(h) < \infty \); and by linearity, we set \( I(f) = I(g) - I(h) \). Verify that (a) \( I \) is well-defined on \( L \), (b) \( L \) is a vector lattice space, and (c) \( I \) is a linear and monotone. Moreover, show that any function in \( L \) is (except in a \( I \)-null set) a pointwise decreasing limit of a sequence \( \{ f_n \} \) of functions in \( \mathbb{E} \) such that \( |I(f_n)| \leq C < \infty \), for every \( n \). Furthermore, prove that \( f \) belongs to \( L \) if and only if \( f^+ \) and \( f^- \) belong to \( L^+ \) with \( I(f^+) + I(f^-) < \infty \). A posteriori, we write \( f = f^+ - f^- \) with \( f^\pm \) in \( L^+ \) and if either \( I(f^+) \) or \( I(f^-) \) is finite then \( I(f) = I(f^+) - I(f^-) \), which may be infinite.

(5) Show that if \( f = f^+ - f^- \) with \( f^+ \) and \( f^- \) belonging to \( L^+ \) and \( I(f^-) < \infty \) then for every \( \varepsilon > 0 \) there exist two functions \( g \) and \( h \) in \( \mathbb{E} \) satisfying \( f = g - h \), except in a \( I \)-null set, \( h \geq 0 \) and \( I(h) < \varepsilon \). Moreover, deduce a Beppo Levi monotone convergence result, i.e., if \( \{ f_k \} \) is a sequence of nonnegative (except in a \( I \)-null set) functions in \( L \) then \( \sum_k I(f_k) = I(\sum_k f_k) \), in particular, if \( \sum_k I(f_k) < \infty \) then \( \sum_k f_k \) belongs to \( L \).

Hint: the identities \( g \wedge h = (g + h - |h - g|)/2 \), \( g \vee h = (g + h + |h - g|)/2 \) and \( |h - g| = g \vee h - g \wedge h \) may be of some help.

Exercise 4.7. On a measurable space \((\Omega, \mathcal{A})\), let \( f \) be a nonnegative measurable, \( \mu \) be a measure and \( \{ \mu_n \} \) be a sequence of measures. Prove that (1) if \( \lim \inf_n \mu_n(A) \geq \mu(A) \), for every \( A \in \mathcal{A} \) then
\[
\int_{\Omega} f \, d\mu \leq \lim \inf_n \int_{\Omega} f \, d\mu_n,
\]

(2) if \( \lim_n \mu_n(A) = \mu(A) \), for every \( A \) in \( \mathcal{A} \) and \( \mu_n(A) \leq \mu_{n+1}(A) \), for every \( A \) in \( \mathcal{A} \) and \( n \), then
\[
\int_{\Omega} f \, d\mu = \lim_n \int_{\Omega} f \, d\mu_n.
\]

Finally, along the lines of the dominate convergence, what conditions we need to impose on the sequence of measure to ensure the validity of the above limit? Hint: use the monotone approximation by simple functions given in Proposition 1.9, e.g., see Dshalalow [36, Section 6.2, pp. 312–326] for more comments and details.

Exercise 4.8. Let \((X, \mathcal{X}, \mu)\) be a measure space, \((Y, \mathcal{Y})\) be a measurable space and \( \psi : X \to Y \) be a measurable function. Verify that the set function \( \mu_\psi : B \mapsto \mu(\psi^{-1}(B)) \) is a measure on \( \mathcal{Y} \), called image measure. Prove that
\[
\int_X g(\psi(x)) \mu(dx) = \int_Y g(y) \mu_\psi(dy),
\]
for every nonnegative measurable function \( g \) on \((Y, \mathcal{Y})\). In particular, if \( \psi \) is also bijective with measurable inverse then

\[
\int_X f(x)\mu(dx) = \int_Y f(\psi^{-1}(y))\mu_\psi(dy),
\]

for every nonnegative measurable function \( f \) on \((X, \mathcal{X})\).

The reader may take a look at Taylor [114, Chapter 5, 226–280].

4.2 Cartesian Products

First, it is clear that we can change the values of an integrable function in a set of measure zero without any changes in its integral, however, we need to know that the resulting function is measurable, e.g., we should avoid the situation \( g = f\mathbb{1}_{N^c} \), where \( N \) is a nonmeasurable subset of a set of measure zero. In other words, it is convenient to assume that the measure space is complete (or complete it if necessary), see also Remark 4.9.

Let \((X, \mathcal{X}, \mu)\) be a \(\sigma\)-finite measure space and \((Y, \mathcal{Y})\) be a measurable space. A function \( \nu : X \times Y \to [0, +\infty] \) is called a \(\sigma\)-finite regular transition measure if

(a) the mapping \( x \to \nu(x, B) \) is \(\mathcal{X}\)-measurable for every \( B \in \mathcal{Y} \),

(b) the mapping \( B \to \nu(x, B) \) is a measure on \( \mathcal{Y} \) for every \( x \in X \),

(c) there exists increasing sequences \( \{X_n\} \subset \mathcal{X} \) and \( \{Y_n\} \subset \mathcal{Y} \) such that \( \bigcup_{n=1}^\infty X_n = X \), \( \bigcup_{n=1}^\infty Y_n = Y \),

\[
\nu(x, Y_n) < \infty, \quad \forall x \in X, \quad \int_{X_n} \mu(dx) \nu(x, Y_n) < \infty, \quad \forall n.
\] (4.2)

If \( \mu(X) = 1 \) and \( \nu(x, Y) = 1 \) for every \( x \in X \) then \( \nu \) is called a transition probability measure. The qualification regular is attached to the condition (b), a non regular transition measure would satisfy almost everywhere the \(\sigma\)-additivity property, i.e., besides the condition \( \nu(x, \emptyset) = 0 \), for every sequence of disjoint set \( \{B_k\} \subset \mathcal{Y} \) there exists a set \( A \) in \( \mathcal{X} \) with \( \mu(A) = 0 \) such that \( \nu(x, \bigcup_k B_k) = \sum_k \nu(x, B_k) \), for every \( x \) in \( X \setminus A \).

Note the following two particular cases: (1) \( \nu(x, B) = \nu(B) \) independent of \( x \), for a given \(\sigma\)-finite measure \( \nu \) on \( \mathcal{Y} \), and (2) \( \nu(x, B) = \sum_{k=1}^\infty a_k(x)\mathbb{1}_{f_k(x) \in B} \), for sequences \( \{a_k\} \) and \( \{f_k\} \) of measurable functions \( a_k : X \to [0, \infty) \) and \( f_k : X \to Y \), i.e., a sum of Dirac measures \( \nu = \sum_k a_k(x)\delta_{f_k(x)} \). Remark that, for the case (2), the assumptions on transition measure \( \nu \) are equivalent to the measurability of the functions \( a_k \) and \( f_k \), for every \( k \).

For any \( E \subset X \times Y \) and any \( x \in X \), we define the sections as the sets \( E_x = \{y \in Y : (x, y) \in E\} \subset Y \) (similarly \( E^y \), by exchanging \( X \) with \( Y \)). Note that \((E \cup F)_x = E_x \cup F_x \) and \((E \setminus F)_x = E_x \setminus F_x \), but we may have \( E \cap F = \emptyset \) with \( E_x \cap F_x \neq \emptyset \). Recall that the product \(\sigma\)-algebra \( \mathcal{X} \times \mathcal{Y} \) is generated by the semi-algebra of rectangle \( A \times B \) with \( A \in \mathcal{X} \) and \( B \in \mathcal{Y} \).
Proposition 4.10. Let $\nu(\cdot, \cdot)$ be a $\sigma$-finite regular transition measure from $\sigma$-finite measure $(X, \mathcal{X}, \mu)$ into $(Y, \mathcal{Y})$ as above, and let $E$ be a set in the product $\sigma$-algebra $\mathcal{X} \times \mathcal{Y}$. Then (a) all sections are measurable, i.e., $E_x \in \mathcal{Y}$, for every $x \in X$; (b) the mapping $x \mapsto \nu(x, E_x)$ is $\mathcal{X}$-measurable and (c) the mapping

$$E \rightarrow (\mu \times \nu)(E) = \int_X \mu(dx) \nu(x, E_x), \quad \forall E \in \mathcal{X} \times \mathcal{Y}$$

is a $\sigma$-finite measure, in particular the expression

$$(\mu \times \nu)(A \times B) = \int_A \nu(x, B) \mu(dx), \quad \forall A \in \mathcal{X}, B \in \mathcal{Y},$$

uniquely determines the values of the product measure $\mu \times \nu$.

Proof. First remark that for any $E = A \times B$ the sections satisfy $E_x = B$ if $x \in A$ and $E_x = \emptyset$ if $x \notin A$. Hence $\nu(x, E_x) = 1_A \nu(x, B)$, for any rectangle $E$ and with the convention that $0 \cdot \infty = 0$.

Take increasing sequences $\{X_n\} \subset \mathcal{X}$ and $\{Y_n\} \subset \mathcal{Y}$ as in (4.2). It is clear that if the conditions (a) and (b) hold for $E \cap (X_n \times Y_n)$ instead of $E$, for every $n$, then they should be valid for $E$. Thus we may assume

$$\nu(x, Y) < \infty, \quad \forall x \in X, \quad \int_X \nu(dx) \nu(x, Y) < \infty,$$

without any loss of generality.

Let $\mathcal{D}$ be the class of sets $E$ in $\mathcal{X} \times \mathcal{Y}$ for which the conditions (a) and (b) are satisfied. Because $(F \cup E)_x = F_x \cup E_x$ and $(F \setminus E)_x = F_x \setminus E_x$, the family $\mathcal{D}$ is a $\lambda$-class, which contains the $\pi$-class of all rectangle. Hence, a monotone argument (see Proposition 1.6) shows that $\mathcal{D} = \mathcal{X} \times \mathcal{Y}$.

To check (c), we need to verify that the product $\mu \times \nu$ is $\sigma$-additive on the semi-algebra of measurable rectangle. To this purpose, note that if $E = \sum_{k=1}^{\infty} E_k$, $E = A \times B$ and $E_k = A_k \times B_k$ then

$$1_A(x)1_B(y) = \sum_{k=1}^{\infty} 1_{A_k}(x)1_{B_k}(y), \quad \forall x, y$$

Thus, the $\sigma$-additivity of the measure $\nu(x, \cdot)$ implies

$$1_A(x) \nu(x, B) = \sum_{k=1}^{\infty} 1_{A_k}(x) \nu(x, B_k), \quad \forall x \in X,$$

and the monotone convergence (Theorem 4.4) yields

$$\int_A \mu(dx) \nu(x, B) = \sum_{k=1}^{\infty} \int_{A_k} \mu(dx) \nu(x, B_k), \quad \forall A \in \mathcal{X}, B \in \mathcal{Y}.$$
At this point, either by Proposition 2.11 or repeating the above argument with any \( E \in \mathcal{X} \times \mathcal{Y} \) and remarking that \( 1_E(x, y) = 1_{E_n}(y) \), we deduce
\[
E \to (\mu \times \nu)(E) = \int_X \mu(dx) \nu(x, E_x) = \int_X \mu(dx) \int_Y 1_E(x, y) \nu(x, dy),
\]
for every \( E \in \mathcal{X} \times \mathcal{Y} \), is a \( \sigma \)-finite measure. \( \square \)

- **Remark 4.11.** It is clear that in Proposition 4.10 we also proved that the function \( y \mapsto \mu(E_y) \) is \( \mathcal{Y} \)-measurable, and if the transition measure \( \nu \) is actually a measure on \((Y, \mathcal{Y})\) then we deduce the equality
\[
\int_X \nu(E_x) \mu(dx) = \int_Y \nu(E_y) \nu(dy), \quad \forall E \in \mathcal{X} \times \mathcal{Y}
\]
as expected. \( \square \)

By means of Proposition 1.9, we can approximate a measurable functions by a pointwise convergence sequence of simple functions to deduce from Proposition 4.10 that if \( f : X \times X \to \bar{\mathbb{R}} \) is a \( \mathcal{X} \times \mathcal{Y} \)-measurable function then for every \( y \) in \( Y \), the section function \( x \mapsto f(x, y) \) is \( \mathcal{X} \)-measurable. Certainly, we may replace the extended real \( \bar{\mathbb{R}} \) by any separable metric space and use the approximation given by Corollary 1.10 to deduce that the sections of a product-measurable functions are indeed measurable. Note that the converse is not valid in general, i.e., although if a contra-example is not easy to get, we may have a non measurable subset \( E \) of \( X \times Y \) such that the sections \( E_x \) and \( E_y \) are measurable, for every fixed \( x \) and \( y \).

Moreover, for any \( N \in \mathcal{X} \times \mathcal{Y} \) we have \( (\mu \times \nu)(N) = 0 \) if and only if its sections \( N_x \) have \( \nu(x, \cdot) \)-measure zero, for \( \mu \)-almost every \( x \), i.e., there exists a set \( A_N \in \mathcal{X} \) such that \( \mu(A_N) = 0 \) and \( \nu(x, N_x) = 0 \), for every \( x \in X \setminus A_N \). Hence, if \( (\lambda, \mathcal{F}) \) is the completion of the product measure \( \mu \times \nu \), and if \( f : X \times Y \to \bar{\mathbb{R}} \) is \( \mathcal{F} \)-measurable then there exists a \( \mathcal{X} \times \mathcal{Y} \)-measurable function \( \tilde{f} \) such that \( \lambda(\{(x, y) : f(x, y) \neq \tilde{f}(x, y)\}) = 0 \). Moreover, there exists a set \( N \in \mathcal{X} \times \mathcal{Y} \) with \( \lambda(N) = 0 \) such that \( f(x, y) = \tilde{f}(x, y) \) for every \( (x, y) \notin N \). Thus we have

**Corollary 4.12.** Let \( (\lambda, \mathcal{F}) \) be the completion of the product measure \( \mu \times \nu \), as given by Proposition 4.10. If \( f : X \times Y \to \bar{\mathbb{R}} \) is \( \mathcal{F} \)-measurable then there exists a set \( A_f \) in \( \mathcal{X} \) with \( \mu(A_f) = 0 \) such that the function \( y \mapsto f(x, y) \) is \( \mathcal{Y} \)-measurable, for every \( x \in X \setminus A_f \).

**Proof.** In view of the approximation by simple functions (see Proposition 1.9), we need to show the result only for \( f = 1_E \) with \( E \in \mathcal{F} \).

Now, for a \( \lambda \)-measurable set \( E \) there exists sets \( E', N \in \mathcal{X} \times \mathcal{Y} \) such that \( (E \setminus E') \cup (E' \setminus E) \subset N \), i.e., \( \|1_E - 1_{E'}\| \leq 1_N \). Because \( \nu(x, \cdot) \) is \( \sigma \)-finite regular transition measure, there is an increasing sequence \( \{Y_n\} \subset \mathcal{Y} \) such that \( \nu(x, Y_n) < \infty \) for every \( x \in X \), for every \( n \). Thus, \( \nu(x, Y_n \cap E_x) < \infty \) and \( |\nu(x, Y_n \cap E_x) - \nu(x, Y_n \cap E'_x)| \leq \nu(x, N_x) \), for every \( n \) and every \( x \in X \). Since
\[
0 = \lambda(N) = \int_X \mu(dx) \nu(x, N_x) = \int_X \mu(dx) \int_Y 1_N(x, y) \nu(x, dy),
\]
there exists a set \( A_E \in \mathcal{X} \) with \( \mu(A_E) = 0 \) such that \( \nu(x, N_x) = 0 \) for every \( x \in X \setminus A_E \). Hence \( \nu(x, E_x) = \nu(x, E_x') \), for every \( x \notin A_E \).

- **Remark 4.13.** Recall that the approximation of measurable functions by integrable simple functions (as in Proposition 1.9) can only be used a in \( \sigma \)-finite space, i.e., if the space is not \( \sigma \)-finite then there are nonnegative measurable functions which are nonzero on a non \( \sigma \)-finite set, and therefore, they can not be a pointwise limit of integrable simple functions. On the other hand, there are ways of dealing with product of non \( \sigma \)-finite measures, essentially, only \( \sigma \)-finite measurable sets (i.e., covered by a sequence \( \{A_k \times B_k : k \geq 1\} \) of rectangles where each \( A_k \) and \( B_k \) is measurable and the product measure of \( \bigcup_k A_k \times B_k \) is finite) are considered and the product measure is defined on a \( \sigma \)-ring, instead of a \( \sigma \)-algebra. The difficulty is the measurability of the mapping \( x \mapsto \nu(x, E_x) \), for an arbitrary measurable set \( E \) on the product space, for instance see Pollard [90, Section 4.5, pp. 93-95].

**Theorem 4.14** (Fubini-Tonelli). Let \( \lambda \) be the completion of the product measure \( \mu \times \nu \) defined in Proposition 4.10 and let \( f : X \times Y \to [0, \infty] \) be a \( \mathcal{X} \times \mathcal{Y} \)-measurable (respect., \( \lambda \)-measurable) function. Then (a) the function \( f(x, \cdot) \) is \( \mathcal{Y} \)-measurable for every \( x \in X \) (respect., for \( \mu \)-almost everywhere \( x \) in \( X \)); (b) the function

\[
\begin{align*}
\text{for every } x \in X, \quad \int_Y f(x, y) \nu(x, dy) & \text{ is } \mathcal{X} \text{-measurable} \\
\int_{X \times Y} f(x, y) \lambda(dx, dy) & = \int_X \mu(dx) \int_Y f(x, y) \nu(x, dy).
\end{align*}
\]

(4.3)

**Proof.** Let \( E \subset X \times Y \) be a \( \lambda \)-measurable set, i.e., there exists \( E', N \in \mathcal{X} \times \mathcal{Y} \) such that \( (E \setminus E') \cup (E' \setminus E) \subset N \) and \( (\mu \times \nu)(N) = 0 \). If \( f = 1_E \) then Proposition 4.10 and Corollary 4.12 proves the validity of the assertions for this particular case, and so for any simple function. Next, we conclude by approximating \( f \) by a monotone sequence of nonnegative simple functions.

If \( f : X \times Y \to \overline{\mathbb{R}} \) is \( \lambda \)-integrable then \( f \) takes finite valued outside of a set \( N \in \mathcal{X} \times \mathcal{Y} \) with \( (\mu \times \nu)(N) = 0 \). Applying (a), (b) and (c) for \( f^+ \) and \( f^- \) we deduce that (1) \( f(x, \cdot) \) is \( \nu(x, \cdot) \)-integrable for \( \mu \)-almost everywhere \( x \) in \( X \); (2) the integral of \( f(x, y) \) with respect to \( \nu(x, dy) \) is \( \mu \)-integrable; (3) the iterate integral reproduces the double integral, i.e., (4.3) holds.

In the particular case of a constant transition measure \( \nu(x, \cdot) = \nu(\cdot) \), we may consider also \( \nu \times \mu \) and we deduce from (4.3) the exchange of the integration order, i.e.,

\[
\begin{align*}
\int_{X \times Y} f(x, y) \lambda(dx, dy) & = \int_X \mu(dx) \int_Y f(x, y) \nu(dy) = \\
& = \int_Y \nu(dy) \int_X f(x, y) \mu(dx),
\end{align*}
\]
4.2. Cartesian Products

for every $f$ either nonnegative and measurable or integrable in the product space. This is the traditional Fubini-Tonelli Theorem.

- **Remark 4.15.** Among the assumptions in Fubini-Tonelli Theorem 4.14, the $\sigma$-finiteness of the measures is essential (as in Proposition 4.10), for the product measure and for the equality of the iterated integrals. A typical contra-example is the case of the diffuse measure $\mu$ (like the Lebesgue measure, which satisfies $\mu(\{x\}) = 0$ for any point $x$) and the counting measure $\nu$ in $[0,1]$ (which counts the points), where

$$
\int_{X} 1_{D}(x,y) \mu(dx) = 0, \quad \text{and} \quad \int_{Y} 1_{D}(x,y) \mu(dy) = 1,
$$

for the diagonal $D = \{(x,y) : x = y \in [0,1]\}$, $X = Y = [0,1]$. \hfill $\square$

It is clear that these arguments extend to a finite product, with suitable transition measures. The reader may take a look at Ambrosio et al. [3, Chapter 6, pp. 83–118] and Taylor [114, Chapter 7, 324–347].

**Exercise 4.9.** Let $(\Omega, \mathcal{F}, \lambda)$ be a probability measure space and $\mathcal{G} \subset \mathcal{F}$ be a sub $\sigma$-algebra. Suppose that $\Omega = X \times Y$, where $(X, \mathcal{X}, \mu)$ is another probability measure space and $\mathcal{G}$ is the $\sigma$-algebra generated by the projection $p$ from $\Omega$ into $X$, i.e., $\mathcal{G} = p^{-1}(\mathcal{X})$, and that $\lambda$ restricted to $\mathcal{G}$ coincides with $p^{-1}(\mu)$, i.e., $\mu(A) = \lambda(p^{-1}(A))$, for every $A$ in $\mathcal{X}$. First, show that a real-valued $\mathcal{F}$-measurable function $f = f(\omega)$ is $\mathcal{G}$-measurable if and only if $f$ is independent of the variable $y$, i.e., $f(\omega) = g(x)$, for any $\omega = (x,y)$. Next, suppose that $\nu : X \times Y \rightarrow [0,1]$ is a probability transition measure (i.e., $\nu(x,Y) = 1$ for every $x$) such that $\lambda = \mu \times \nu$ as in Proposition 4.10, and for any nonnegative $\mathcal{F}$-measurable function $f$ define

$$
\nu(f)(\omega) = \int_{Y} f(x,y) \nu(x,dy), \quad \text{with} \quad \omega = (x,y) \in X \times Y = \Omega.
$$

Prove that

$$
\int_{\Omega} fg d\lambda = \int_{\Omega} \nu(f) g d\lambda,
$$

for every nonnegative $\mathcal{G}$-measurable function $g$. In probability terms, $\nu(f)$ is called the conditional expectation of $f$ given $\mathcal{G}$, and the transition measure $\nu$ is a regular conditional probability measure given $\mathcal{G}$. \hfill $\square$

The converse of the product measure construction (disintegration problem) is as follows: given a $\sigma$-finite measure $(\lambda, \mathcal{F})$ on a product space $\Omega = X \times Y$, $\mathcal{F} = \mathcal{X} \times \mathcal{Y}$, and a $\sigma$-finite measure $(\mu, \mathcal{X})$ on $X$, we want to find a $\sigma$-finite transition measure $\nu(x,dy)$ on $X \times \mathcal{Y}$ such that $\lambda = \mu \times \nu$. The construction of the transition measure $\nu(x,dy)$ is more delicate than one may expect, the relation

$$
\lambda(A \times B) = \int_{A} \nu(x,B) \mu(dx), \quad \forall A \in \mathcal{X}, B \in \mathcal{Y}
$$
identify \( \nu(x, B) \) for \( \mu \)-almost every \( x \), but it is convenient to have \( \nu(x, B) \) defined \( \mu \)-almost everywhere in \( x \) as a (\( \sigma \)-additive) measure for \( B \) in \( \mathcal{Y} \), instead of just being \( \mu \)-almost everywhere defined for each \( B \), i.e., we need to make a selection of all the possible values of \( \nu(x, B) \) so that \( \nu(x, \cdot) \) results a (\( \sigma \)-additive) measure for \( \mu \)-almost every \( x \). For instance, the reader could check the book by Pollard [90, Section 5.3, pp. 116-118, and Appendix F, pp. 339-346] for a quick treatment and appropriated references.

**Exercise 4.10.** Let \((X, \mathcal{X}, \mu)\) be a complete measure space and \((Y, \mathcal{Y})\) be a measurable space. Consider a function \( \nu : X \times \mathcal{Y} \to [0, \infty] \) satisfying the conditions of a \( \sigma \)-finite transition measures, except in a set of \( \mu \)-measure zero, e.g., (b) means that the mapping \( B \to \nu(x, B) \) is a measure on \( \mathcal{Y} \) for almost every \( x \in X \). Give details to show that for any set \( E \in \mathcal{X} \times \mathcal{Y} \) the function \( x \mapsto \nu(x, E_x) \) is \( \mathcal{X} \)-measurable. Next, verify that the product measure \( \mu \times \nu \) is constructed even if (b) is weaken as follows: besides the condition \( \nu(x, \emptyset) = 0 \), for every sequence of disjoint sets \( \{B_k\} \subset \mathcal{Y} \) there is a \( \mu \)-null set \( A \) such that \( \nu(x, \sum_k B_k) = \sum_k \nu(x, B_k) \) for every \( x \in X \setminus A \). If necessary, make adequate modifications to Fubini-Tonelli Theorem 4.14 to include this new situation. \( \square \)

### 4.3 Convergence in Measure

For functions from a measure space into a topological space we may think of various modes of convergence. For instance, (1) \( f_n \to f \) pointwise a.e. (almost everywhere) if there exists a set \( N \in \mathcal{F} \) with \( \mu(N) = 0 \) such that \( f(x) \to f(x) \) for every \( x \in \Omega \setminus N \); or (2) \( f_n \to f \) pointwise quasi-uniform (quasi-uniformly) if for every \( \varepsilon > 0 \) there exists a set \( \Omega_\varepsilon \in \mathcal{F} \) with \( \mu(\Omega \setminus \Omega_\varepsilon) \leq \varepsilon \) such that \( f_n(x) \to f(x) \) uniformly in \( \Omega_\varepsilon \). It is clear that (2) implies (1) and the converse is not necessarily true. Also we have

**Definition 4.16.** Let \((\Omega, \mathcal{F}, \mu)\) be a measure space and \((E, d)\) be a metric space. A sequence \( \{f_n\} \), \( f_n : \Omega \to E \), of measurable functions is a Cauchy sequence in measure (or in probability if \( \mu(\Omega) = 1 \)) if for every \( \varepsilon > 0 \) there exists \( n(\varepsilon) \) such that \( \mu(\{x \in \Omega : d(f_n(x), f_m(x)) \geq \varepsilon\}) < \varepsilon \) for every \( n, m \geq n(\varepsilon) \). Similarly, \( f_n \to f \) in measure, if for every \( \varepsilon > 0 \) there exists \( n(\varepsilon) \) such that \( \mu(\{x \in \Omega : d(f_n(x), f(x)) \geq \varepsilon\}) < \varepsilon \) for every \( n \geq n(\varepsilon) \). \( \square \)

Note that we may use the distance \( d(x, y) = |\arctan(x) - \arctan(y)| \), for any \( x, y \in E = [-\infty, +\infty] \), when working with extended-valued measurable functions, i.e., the mapping \( z \mapsto \arctan z \) transforms the problem into real-valued functions. It is clear that for any sequence \( \{x_n\} \) of real numbers we have \( x_n \to x \) if and only if \( \arctan(x_n) \to \arctan(x) \), but the usual distance \( (x, y) \to |x - y| \) and \( d(x, y) \) are not equivalent in \( \mathbb{R} \). Actually, consider the sequence \( \{f_n(x) = (x + 1/n)^2\} \) on Lebesgue measure space \((\mathbb{R}, \mathcal{L}, \ell)\) and the limiting function \( f(x) = x^2 \) to check that

\[
\ell(\{x \in \mathbb{R} : |f_n(x) - f(x)| \geq \varepsilon\}) = \ell(\{x \in \mathbb{R} : |x + (1/n)| \geq n\varepsilon\}) = \infty,
\]

[108] Chapter 4. Integration Theory
for every $\varepsilon > 0$ and $n \geq 1$, i.e., $f_n$ does not converge in measure to $f$. However, if $g_n(x) = \arctan(f_n(x))$ and $g(x) = \arctan(x^2)$ then $|g_n(x) - g(x)| \leq 1/n$, i.e., $g_n$ converges to $g$ uniformly in $\mathbb{R}$. Thus, on the Lebesgue measure space $(\mathbb{R}, \mathcal{L}, \ell)$, we have $g_n \to g$ in measure, i.e., the convergence in measure depends not only on the topology given to $\mathbb{R}$, but actually, on the metric used on it. Moreover, as typical examples of these three modes of convergences in $(\mathbb{R}, \mathcal{L}, \ell)$ let us mention:

(a) $\mathbb{1}_{[0,1/n]} \to 0$, (c) $\mathbb{1}_{[n,n+1/n]} \to 0$, and (c) $\mathbb{1}_{[n,n+1]} \to 0$, where the convergence in (a), (b), (c) is pointwise almost everywhere (but not pointwise everywhere), the convergence in (a), (b) is also in measure (but (c) does not converge in measure), the convergence in (a) is also pointwise quasi-uniform (but (b), (c) does not converge pointwise quasi-uniform).

**Exercise 4.11.** (a) Verify that if the sequence $\{f_n\}$, $f_n : \Omega \to E$, of measurable functions is convergent (or Cauchy) in measure, $(Z, d_Z)$ is a metric space and $\psi : E \to Z$ is a uniformly continuous function then the sequence $\{g_n\}$, $g_n(x) = \psi(f_n(x))$ is also convergent (or Cauchy) in measure. (b) In particular, if $(E, |\cdot|_E)$ is a normed space then for any sequences $\{f_n\}$ and $\{g_n\}$ of $E$-valued measurable functions and any constants $a$ and $b$ we have $af_n + bg_n \to af + bg$ in measure, whenever $f_n \to f$ and $g_n \to g$ in measure. Moreover, assuming that the sequence $\{g_n\}$ takes real (or complex) values, (c) if the sequences are also quasi-uniformly bounded, i.e., for any $\varepsilon > 0$ there exists a measurable set $F$ with $\mu(F) < \varepsilon$ such that the numerical series $\{|f_n(x)|_E\}$ and $\{|g_n(x)|\}$ are uniformly bounded for $x$ in $F^c$, then deduce that $f_n g_n \to fg$ in measure. Furthermore, (d) if $g_n(x)g(x) \neq 0$ a.e. $x$ and the sequences $\{f_n\}$ and $\{1/g_n\}$ are also quasi-uniformly bounded then show that $f_n/g_n \to f/g$ in measure. Finally, (e) verify that if the measure space $\Omega$ has finite measure then the conditions on quasi-uniformly bounded are automatically satisfied.

$$
\lim_n \mu(\{x \in \Omega : |f_n(x) - f(x)| \geq \varepsilon\}) = 0, \quad \forall \varepsilon > 0,
$$

and if $f_n(x) = \mathbb{1}_{\{|x| > n\}}$ then $f_n(x) \to 0$ for every $x$ in $\mathbb{R}^d$, but $\ell(\{x \in \mathbb{R}^d : |f_n(x)| \geq \varepsilon\}) = \infty$, with the Lebesgue measure $\ell$, i.e., the pointwise almost everywhere convergence does not necessarily yields the convergence in measure. However, we have

**Theorem 4.17.** Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, $E$ be a complete metric space and $\{f_n\}$ be a Cauchy sequence in measure of measurable functions $f_n : \Omega \to E$. Then there exist (1) a subsequence $\{f_{n_k}\}$ such that $f_{n_k} \to f$ pointwise a.e. and (2) a measurable function $f$ such that $f_n \to f$ in measure. Moreover, if $f_n \to g$ in measure then $g = f$ a.e.

**Proof.** Given $\varepsilon > 0$ define $X(\varepsilon, n, m) = \{x \in \Omega : d(f_n(x), f_m(x)) \geq \varepsilon\}$ to see that for $\varepsilon = 2^{-1} > 0$ we can find $n_1$ such that $\mu(X(\varepsilon, n_1, m)) < \varepsilon$ for every $m \geq n_1$. Next, for $\varepsilon = 2^{-2} > 0$ again, we can find $n_2 > n_1$ such that $\mu(X(\varepsilon, n_2, m)) < \varepsilon$ for every $m \geq n_2$. By induction, we get $n_k < n_{k+1}$ and $A_k = X(2^{-k}, n_k, n_{k+1})$ with $\mu(A_k) < 2^{-k}$, for every $k \geq 1$. 

---

Now, if $F_k = \bigcup_{i=k}^{\infty} A_i$ then $\mu(F_k) \leq \sum_{i=k}^{\infty} 2^{-i} = 2^{1-k}$. On the other hand, if $x \notin F_k$ then for any $i \geq j \geq k$ we have

$$d(f_{n_j}(x), f_{n_i}(x)) \leq \sum_{r=j}^{i-1} d(f_{n_{r+1}}(x), f_{n_r}(x)) \leq \sum_{r=j}^{i-1} 2^{-r} \leq 2^{1-k}, \quad (4.4)$$

i.e., $\{f_{n_i}(x)\}$ is a Cauchy sequence in $E$, for every $x \notin F_k$.

Define $F = \bigcap_k F_k$ to have $\mu(F) \leq \mu(F_k)$, for every $k$, i.e., $\mu(F) = 0$. If $x \notin F$ then $x$ belongs to a finite number of $F_k$ and therefore, because $E$ is complete, there exists the limit of $\{f_{n_k}(x)\}$, which is called $f(x)$. If $x \in F$ we set $f(x) = 0$. Hence $f_{n_k} \to f$ almost everywhere.

Let $i \to \infty$ in (4.4) to have $d(f_{n_k}(x), f(x)) \leq 2^{1-k}$ for every $x \notin F_k$. Since $\mu(F_k) \leq 2^{1-k} \to 0$, we deduce that $f_{n_k} \to f$ in measure, and in view of the inclusion

$$\{x : d(f(x), f(x)) \geq \varepsilon\} \subset \{x : d(f(x), f_{n_k}(x)) \geq \varepsilon/2\} \cup \{x : d(f_{n_k}(x), f(x)) \geq \varepsilon/2\}, \quad \forall \varepsilon > 0,$$

the whole sequence $f_n \to f$ in measure. Moreover, in view of

$$\{x : d(f(x), g(x)) \geq \varepsilon\} \subset \{x : d(f(x), g(x)) \geq \varepsilon/2\} \cup \{x : d(f(x), f(x)) \geq \varepsilon/2\}, \quad \forall \varepsilon > 0,$$

if $f_n \to g$ in measure then $f = g$ a.e. \hfill \square

\textbf{Remark 4.18.} In a measure space $(\Omega, \mathcal{F}, \mu)$, take a measurable set $A \in \mathcal{F}$ with $0 < \mu(A) \leq 1$ and find a finite partition $A = \bigcup_{i=1}^{k} A_{k,i}$ with $0 < \mu(A_{k,i}) \leq 1/k$, for every $i$. If $\{a_k\}$ and $\{b_k\}$ are two sequences of real numbers then we construct a sequence of functions $\{f_n\}$ as follows: the sequence of integers $\{1, 2, 3, \ldots, 10, 11, \ldots\}$ is grouped as $\{(1) ; (2, 3) ; (4, 5, 6) ; (7, 8, 9, 10) ; \ldots\}$ where the $k$ group has exactly $k$ elements, i.e., for any $n = 1, 2, \ldots$, we select first $k = 1, 2, \ldots$, such that $(k-1)k/2 < n \leq k(k+1)/2$ and we write (uniquely) $n = (k-1)k/2 + i$ with $i = 1, 2, \ldots, k$ to define

$$f_n(x) = \begin{cases} a_k & \text{if } x \in A \setminus A_{k,i}, \\ b_k & \text{if } x \in A_{k,i}. \end{cases}$$

Assuming that $a_k \to a$ as $k \to \infty$ and $|b_k - a| \geq c > 0$ for any $k$, we have $\mu(\{|f_n - a| \geq \varepsilon\}) = \mu(A_{k,i}) \leq 1/k \leq 2/\sqrt{n}$ for every $0 < \varepsilon < c$, i.e., $f_n \to f$ in measure with $f(x) = a$ for every $x$. However, for every $x \in A$ there exist $i, k$ such that $x \in A_{k,i}$ and $f_n(x) = b_k$, i.e., $f_n(x)$ does not converge to $f(x)$. Moreover, for any given $b \leq a \leq \overline{b}$, we can choose $b_k$ so that $\liminf_n f_n(x) = b$ and $\limsup_n f_n(x) = \overline{b}$, for every $x \in A$. \hfill \square

Sometimes we begin with a known notion of convergence to define closed sets in a space $X$. For instance, if we know that the “convergence $x_n \to x$” satisfies the following (Kuratowski) three axioms (1) uniqueness of the limit;
(2) For every $x$ in $X$, the constant sequence $\{x, x, \ldots\}$ converges to $x$; (3) Given a sequence $\{x_n\}$ convergent to $x$, every subsequence $\{x_{n'}\} \subset \{x_n\}$ converges to the same limit $x$; then we can define the open sets in the topology $\mathcal{T}$ as the complement of closed sets, where a set $C$ is closed if for any sequence $\{x_n\}$ of point in $C$ such that $x_n \to x$ results $x \in C$. Next, knowing the topology $\mathcal{T}$ we have the “convergence $x_n \xrightarrow{\mathcal{T}} x$,” i.e., for any open set $O$ (element in $\mathcal{T}$) with $x \in O$ there exists an index $N$ such that $x_n \in O$ for any $n \geq N$. Actually, this means that $x_n \xrightarrow{\mathcal{T}} x$ if and only if for any subsequence $\{x_{n'}\}$ of $\{x_n\}$ there exists another subsequence $\{x_{n''}\} \subset \{x_{n'}\}$ such that $x_{n''} \to x$. Clearly, if $x_n \to x$ then $x_n \xrightarrow{\mathcal{T}} x$. If the initial convergence $x_n \to x$ comes from a metric, then we can verify that $x_n \to x$ is equivalent to $x_n \xrightarrow{\mathcal{T}} x$, but, in general, this could be false. For instance, let $(\Omega, \mathcal{F}, \mu)$ be a measure space with $\mu(\Omega) < \infty$, and consider the space $X$ of real-valued measurable functions (actually, equivalent classes of functions because we have identified functions almost everywhere equal), with the almost everywhere convergence $x_n(\omega) \to x(\omega)$ a.e. $\omega$. By means of Theorem 4.17, we see that $x_n \xrightarrow{\mathcal{T}} x$ if and only if $x_n \to x$ in measure.

**Exercise 4.12.** Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, $(E, d)$ a metric space and $\{f_n\}$ a sequence of measurable functions $f_n : \Omega \to E$. Show that if $\{f_n\}$ converges to some function $f$ pointwise quasi-uniform then $f_n \to f$ in measure. Prove or disprove the converse. \[\square\]

Let us compare the pointwise almost everywhere convergence with the pointwise uniform convergence and the convergence in measure. Recall that a Borel measure means a measure on a topological space $\Omega$ such that all Borel sets are measurable. We have

**Theorem 4.19 (Egorov).** If $\mu(\Omega) < \infty$ then pointwise almost everywhere convergence implies pointwise quasi-uniform convergence, i.e., if a sequence $\{f_n\}$ of measurable functions taking values in a metric space $(E, d)$ satisfies $f_n(x) \to f(x)$ a.e. in $x$, then for every $\varepsilon > 0$ there exists an index $n_\varepsilon$ and a set $F \in \mathcal{F}$ with $\mu(F) < \varepsilon$ such that $d(f_n(x), f(x)) < \varepsilon$ for every $n \geq n_\varepsilon$ and $x \in F^c = \Omega \setminus F$. Moreover, if $\mu$ is a Borel measure then $F = O$ is an open set of $\Omega$.

**Proof.** Even if this is not necessary, we first prove that assuming a finite measure, pointwise almost everywhere convergence implies convergence in measure. Indeed, given a sequence $\{f_n\}$ and a function $f$, define $X(\varepsilon, f_n, f) = \{x \in \Omega : d(f_n(x), f(x)) \geq \varepsilon\}$ to check that $f_n(x) \to f(x)$ if and only if $x \not\in F_\varepsilon = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} X(\varepsilon, f_k, f)$ for every $\varepsilon > 0$. Since $X(\varepsilon, f_n, f) \subset F_{\varepsilon,n} = \bigcap_{k=1}^{\infty} \bigcup_{i=k}^{\infty} X(\varepsilon, f_i, f)$, we have $\mu(X(\varepsilon, f_n, f)) \leq \mu(F_{\varepsilon,n})$, and therefore

$$\lim_{n} \sup \mu(X(\varepsilon, f_n, f)) \leq \lim_{n} \mu(F_{\varepsilon,n}), \quad \forall \varepsilon > 0.$$ 

If $f_n \to f$ pointwise almost everywhere then $\mu(F_\varepsilon) = 0$ for every $\varepsilon > 0$, and if also $\mu$ is a finite measure then $\mu(F_{\varepsilon,n}) \to \mu(F_\varepsilon) = 0$. 

[Preliminary] 

To show the quasi-uniform convergence, let \( k, n \) be positive integers and set
\[
A_k(n) = \bigcup_{m=n}^{\infty} \{ x : d(f_m(x), f(x)) \geq 1/k \} = \bigcup_{m=n}^{\infty} X(1/k, f_m, f).
\]

It is clear that \( A_k(n) \supset A_k(n + 1) \) for any \( k, n \), and the almost everywhere convergence implies that \( \mu(B_k) = 0 \) with \( \bigcap_{n=1}^{\infty} A_k(n) = B_k \). Since \( \mu(\Omega) < \infty \) we deduce \( \mu(A_k(n)) \to 0 \) as \( n \to \infty \). Hence, given \( \varepsilon > 0 \) and \( k \), choose \( n_k \) such that \( \mu(A_k(n_k)) < \varepsilon 2^{-k} \) and define \( F = \bigcup_{k=1}^{\infty} A_k(n_k) \). Thus \( \mu(F) < \varepsilon \), and \( d(f_n(x), f(x)) < 1/k \) for any \( n > n_k \) and \( x \notin F \). This yields \( f_n \to f \) uniformly on \( \mathbb{F}^c \).

Finally, if \( \mu \) is a Borel measure then we conclude by choosing (see Theorem 3.3) an open set \( O \supset F \) with \( \mu(O) < 2\varepsilon \). \( \Box \)

As mentioned early, if the measure is not finite then pointwise almost everywhere convergence does not necessarily implies convergence in measure. The converse is also false. It should be clear (see Exercise 4.12) that quasi-uniform convergence implies the convergence in measure, so that Theorem 4.19 also affirms that if the space has finite measure then pointwise almost everywhere convergence implies convergence in measure.

**Exercise 4.13.** Assume that \( \mu(\Omega) < \infty \) and by means of arguments similar to those of Egorov’s Theorem 4.19, prove (a) if \((E, | \cdot |_E)\) is a normed space and the numerical sequence \( \{|f_n(x)|_E\} \) is bounded for almost every \( x \) in \( \Omega \) then for every \( \varepsilon > 0 \) there exists a measurable subset \( F \) of \( \Omega \) such that \( \mu(F) < \varepsilon \) and the sequence \( \{|f_n(x)|_E\} \) is uniformly bounded for any \( x \in \mathbb{F}^c \). Moreover, (b) show that if \( E = [-\infty, +\infty] \), \( \bar{f}(x) = \limsup_n f_n(x) \) (or \( \bar{f}(x) = \liminf_n f_n(x) \)) and \( \bar{f} \) (or \( \bar{f} \)) is a real valued (finite) a.e. then for every \( \varepsilon > 0 \) there exists a measurable subset \( F \) of \( \Omega \) such that \( \mu(F) < \varepsilon \) and the \( \limsup \) (or \( \liminf \)) is uniformly for \( x \) in \( \mathbb{F}^c \). Moreover, if \( \mu \) is a Borel measure then \( F = O \) is an open set of \( \Omega \). \( \Box \)

**Definition 4.20.** A sequence \( \{f_n\} \) of (extended) real-valued integrable functions on measure space \((\Omega, \mathcal{F}, \mu)\) is a Cauchy sequence in mean if for every \( \varepsilon > 0 \) there exists \( n(\varepsilon) \) such that
\[
\int_{\mathbb{F}^c} |f_n(x) - f_m(x)| \mu(dx) < \varepsilon, \quad \forall n, m \geq n(\varepsilon).
\]

Similarly, we define \( f_n \to f \) in mean. \( \Box \)

**Exercise 4.14.** Prove that if a sequence \( \{f_n\} \) of (extended) real-valued integrable functions on measure space \((\Omega, \mathcal{F}, \mu)\) converges in mean to \( f \) then \( f_n \to f \) in measure. Moreover, give an example of a sequence of functions defined on the Lebesgue measure space \(([0,1], \ell)\) which converges in mean but it does not converge in mean. \( \Box \)

**Exercise 4.15.** The assumption that \( \mu(\Omega) < \infty \) is essential in Egorov’s Theorem 4.19. Indeed, let \( \{A_k\} \) be a sequence of disjoint measurable sets such that
$A = \sum_k A_k$, $\mu(A) = \infty$ and $\mu(A_k) \to 0$. Consider the sequence of measurable functions $f_k = 1_{A_k}$. Prove that (1) $f_k \to 0$ in mean (and in measure, in view of Exercise 4.14), (2) $f_k(x) \to 0$ for every $x$, and (3) $f_k(x) \to 0$ uniformly for $x$ in $B$ if and only if there exists an index $n_B$ such that $B \cap A_k = \emptyset$, for every $k \geq n_B$. Finally, (4) deduce that $\{f_k\}$ does not converge pointwise quasi-uniform and (5) make an explicit construction of the sequence $\{A_k\}$. □

- **Remark 4.21.** Another consequence of Egorov Theorem 4.19 is the approximation of any measurable function by a sequence of continuous functions. Indeed, if $\mu$ is a finite Borel measure on $\Omega$ and $f$ is $\mu^*$-measurable function with values in $\mathbb{R}^d$ then there exists a sequence $\{f_n\}$ of continuous functions such that $f_n \to f$ almost everywhere, see Doob [35, Section V.16, pp. 70-71]. □

Recall that if $\mu$ is a Borel measure on the metric space $\Omega$ then $\mu$ is defined on the Borel $\sigma$-algebra $B = B(\Omega)$, the outer measure $\mu^*$ (induced by $\mu$) is a Borel regular outer measure, and a subset $A$ of $\Omega$ is called $\mu$-measurable to shorter the expression $\mu^*$-measurable in the Caratheodory’s sense, see Definition 2.4. As stated in Corollary 3.4, for a Borel measure (recall, where all Borel sets are measurable, e.g., the Lebesgue measure in $\mathbb{R}^d$), for any $\mu^*$-measurable $A$ and for every $\varepsilon > 0$, there exist an open set $O$ and a closed set $C$ such that $C \subset A \subset O$ and $\mu(O \setminus C) < \varepsilon$. This last property is relative simple to show when $A$ has a finite $\mu^*$-measure, but a little harder in the general case.

**Theorem 4.22 (Lusin).** Let $\mu$ be a $\sigma$-finite regular Borel measure $\mu$ on a metric space $\Omega$. If $\varepsilon > 0$ and $f : \Omega \to E$ is a $\mu$-measurable function with values in a separable metric space $E$ then there exists a closed set $C$ such that $f$ is continuous on $C$ and $\mu(\Omega \setminus C) < \varepsilon$.

**Proof.** Since $E$ is a separable metric space, for every integer $i$ there exists a sequence $\{E_{i,j}\}$ of disjoint Borel sets of diameters not larger that $1/i$ such that $E = \sum_j E_{i,j}$. By means of Theorem 3.3 and Corollary 3.4 (when $\mu$ is a $\sigma$-finite regular Borel measure), there exist a sequence $\{C_{i,j}\}$ of disjoint closed sets such that $C_{i,j} \subset f^{-1}(E_{i,j})$ and $\mu(f^{-1}(E_{i,j}) \setminus C_{i,j}) < 2^{-i-j}\varepsilon$. Hence

$$\mu(\Omega \setminus \bigcup_{j=1}^{\infty} C_{i,j}) = \sum_{j=1}^{\infty} \mu(f^{-1}(E_{i,j}) \setminus C_{i,j}) < 2^{-i}\varepsilon,$$

and there exists an integer $n = n(i)$ such that $C_i = \bigcup_{j=1}^{n(i)} C_{i,j}$ and $\mu(\Omega \setminus C_i) < 2^{-i}\varepsilon$. Now, by choosing $e_{i,j}$ in $E_{i,j}$, we can define a function $g_i : C_i \to E$ as $g_i(x) = e_{i,j}$ whenever $x$ belongs to some $C_{i,j}$ with $j = 1, \ldots, n$. This function $g_i$ is continuous because $\{C_{i,j} : j \geq 1\}$ are closed and disjoint, and the distance from $g_i(x)$ to $f(x)$ is not larger then $1/i$. Thus, we have

$$C = \bigcap_{i=1}^{\infty} C_i = \bigcap_{i=1}^{n(i)} \bigcup_{j=1}^{\infty} C_{i,j}, \quad \mu(\Omega \setminus C) < \varepsilon,$$

and $g_i(x) \to f(x)$ uniformly for every $x$ in $C$, i.e., the restriction of $f$ to the closed set $C$ is continuous. □

If $E = \mathbb{R}^d$ then the restriction of $f$ to $C$, i.e., $f|_C : C \to \mathbb{R}^d$, can be extended to a continuous function $h : \Omega \to \mathbb{R}^d$, see Tietze’s extension Proposition 0.2. Thus, Lusin Theorem affirms that there exists a continuous function $h$ such that $\mu(f \neq h) < \varepsilon$. Moreover, the spaces $\Omega$ and $E$ may be more general topological spaces, non necessarily metric spaces.

If $\Omega$ is a Polish space or $\mu$ is a Radon measure (and $\Omega$ is locally compact) then instead of a closed set $C$, we can find a compact set $K$ such that $f$ is continuous on $K$ and $\mu(\Omega \setminus K) < \varepsilon$.

For instance, the interested reader may consult Bauer [10, Sections 30 and 31, pp 188–216] for more details about convergence of Radon measures.

• **Remark 4.23.** It should be clear that the requirement of having a $\sigma$-finite measure $\mu$ and a separable space $E$ are not really necessary in Lusin Theorem 4.22. Actually, it suffices to know that the set $\{x \in \Omega : f(x) \neq 0\}$ is $\sigma$-finite and the range $\{f(x) : x \in \Omega\}$ is separable.

**Exercise 4.16.** Let $\Omega$ and $E$ be two metric spaces. Suppose that $\mu$ is a regular Borel measure on $\Omega$ and $f : \Omega \to E$ is a $\mu$-measurable function such that for some $e_0$ in $E$ the set $\{x \in \Omega : f(x) \neq e_0\}$ is a $\sigma$-finite and the range $\{f(x) \in E : x \in \Omega\}$ is contained in a separable subspace of $E$. With arguments similar to those of Lusin Theorem 4.22, (a) show that for any $\varepsilon > 0$ there exists a close set $C$ such that $f$ is continuous on $C$ and $\mu(\Omega \setminus C) < \varepsilon$; and (b) establish that there exists a $\mathfrak{F}_\sigma$ set (i.e., a countable union of closed sets) $F$ such that $f$ is continuous and $\mu(F^c) = 0$. Next (c) deduce that any $\mu$-measurable function as above is almost everywhere equal to a Borel function. Is there any other way (without using Lusin Theorem) of proving part (c)?

**Exercise 4.17.** Given a closed subset $F$ of $\mathbb{R}^d$ and a real-valued function $f$ defined on $\mathbb{R}^d$, what does the assertion “the restriction of $f$ to $F$ is continuous” actually means, in term of convergent sequences? Let $(\mathbb{R}^d, \mathcal{L}, \ell)$ be the Lebesgue measure space and $\Omega$ a measurable subset of $\mathbb{R}^d$. Prove that a function $f : \Omega \to \mathbb{R}$ is measurable if and only if for every $\varepsilon > 0$ there exists a closed set $F \subset \Omega$ such that $f$ restricted to $F$ is continuous and $\ell(F^c) < \varepsilon$.

**Exercise 4.18.** Let $\mu$ be a Radon measure on a locally compact space $X$ and $f : X \to \mathbb{R}$ be a measurable function that vanishes outside a set of finite measure. Prove that for any $\varepsilon > 0$ there exist a closed set $F$ with $\mu(F^c) < \varepsilon$ and a continuous function with compact support $g$ such that $f = g$ on $F$ and if $|f(x)| \leq C$ a.e. then $|g(x)| \leq C$ for every $x$, see Folland [45, Section 7.2, pp. 216–221].

### 4.4 Almost Measurable Functions

For a given measure space $(\Omega, \mathcal{F}, \mu)$, we denote by $\mathcal{L}^0 = \mathcal{L}^0(\Omega, \mathcal{F}; E)$ the space of measurable functions $f : \Omega \to E$, where $E$ is a measurable space. However, once a measure $\mu$ is defined on $\mathcal{F}$ and a measure space $(\Omega, \mathcal{F}, \mu)$ is constructed, we may complete the $\sigma$-algebra $\mathcal{F}$ to get a complete measure space $(\Omega, \mathcal{F}^\mu, \mu)$ and to
make use of $\mathcal{L}^0(\Omega, \mathcal{F}; E)$, also denoted by $\mathcal{L}^0(\Omega, \mu; E)$, instead of $\mathcal{L}^0(\Omega, \mathcal{F}; E)$. If $E$ is a vector space, to check that $\mathcal{L}^0(\Omega, \mathcal{F}; E)$ is indeed a vector space we need to know that the sum and the scalar multiplication on $E$ are Borel (or continuous) operations, e.g., when $E$ is topological vector space or when $E$ is separable metric space of real (or complex) functions. Actually, to facilitate the understanding of this section, it may be convenient for the student to assume that $E = \mathbb{R}$ or $\mathbb{R}^n$, with the Borel $\sigma$-algebra, and perhaps in the second reading, reconsider more general situations.

Recall that the abbreviation a.e. means almost everywhere, i.e., there exists a set $N$ (which can be assumed to be $\mathcal{F}$-measurable even if $\mathcal{F}$ is not $\mu$-complete) such that the equality (or in general, the property stated) holds for any point $\omega$ in $\Omega \setminus N$. Thus, assuming that $\mathcal{F}$ is complete with respect to $\mu$ we have: (a) if $f$ is measurable and $f = g$ a.e. then $g$ is also measurable; (b) if $\{f_n\}$ are measurable and $f_n \to f$ a.e. then $f$ is also measurable. If $\mathcal{F}$ is not necessarily $\mu$-complete then a function $f$ measurable with respect to $\mathcal{F}^\mu$, the $\mu$-completion of $\mathcal{F}$, is called $\mu$-measurable. Now, if $\varphi$ is a $\mu$-measurable simple function then by the definition of the completion $\mathcal{F}^\mu$ there exists another $\mathcal{F}$-measurable simple function $\psi$ such that $\varphi = \psi$ a.e., and since any measurable function is a pointwise limit of a sequence of simple functions, we conclude that for every $\mu$-measurable function $f$ there exists a $\mathcal{F}$-measurable function $g$ such that $f = g$ a.e.

Therefore, our interest is to study measurable functions defined (almost everywhere) outside of an unknown set of measure zero, i.e., $f : \Omega \setminus N \to E$ measurable with $\mu(N) = 0$. To go further in this analysis, we use $E = \mathbb{R}^n$, $n \geq 1$ or $\mathbb{R} = [-\infty, +\infty]$, or in general a (complete) metric (or Banach) space $E$ with its Borel $\sigma$-algebra $\mathcal{E}$. Clearly, the case $E = \mathbb{R}^n$, $n \geq 1$ is of main interest, as well as when $E$ is an infinite dimensional Banach space.

We endow $\mathcal{L}^0(\Omega, \mathcal{F}; E)$ with the topology induced by convergence in measure. This topology does not separate points, so to have a Hausdorff space we are forced to consider equivalence class of functions under the relation $f \sim g$ if and only if there exists a set $N \in \mathcal{F}$ with $\mu(N) = 0$ and $f(\omega) = g(\omega)$ for every $\omega \in \Omega \setminus N$. Thus, the quotient space $\mathcal{L}^0 = \mathcal{L}^0/\sim$ or $\mathcal{L}^0(\Omega, \mathcal{F}; \mu; E)$ becomes a Hausdorff topological space with the convergence in measure. Actually, we regard the elements of $\mathcal{L}^0$ as measurable functions defined almost everywhere, so that even if $\mathcal{L}^0(\Omega, \mathcal{F}; E)$ may not be equal to $\mathcal{L}^0(\Omega, \mathcal{F}^\mu; E)$, we are really looking at $\mathcal{L}^0 = \mathcal{L}^0(\Omega, \mathcal{F}^\mu; \mu; E) = \mathcal{L}^0(\Omega, \mu; E)$. Note that for the quotient space $\mathcal{L}^0$ (where the elements are equivalence classes) we may omit the $\sigma$-algebra $\mathcal{F}$ from the notation, while for the initial space $\mathcal{L}^0$ we may use the whole measure space $(\Omega, \mathcal{F}, \mu)$. Also if $\Omega_0$ is a measurable subset in a measure space $(\Omega, \mathcal{F}, \mu)$ then we may define the restriction to $\Omega_0$, of $\mathcal{F}$ and $\mu$ to form the measure space $(\Omega_0, \mathcal{F}_0, \mu_0)$, and for instance, we may talk about functions measurable on $\Omega_0$.

**Definition 4.24.** When the space $E$ is not separable, we need to modify the concept of measurability as follows: on a measure space $(\Omega, \mathcal{F}, \mu)$ a function with values in a Borel space $(E, \mathcal{E})$ is called measurable if (a) $f^{-1}(B)$ belongs to $\mathcal{F}$ for every $B$ in $\mathcal{E}$ and (b) $f(\Omega)$ is contained in a separable subspace of $E$. Also, functions measurable with respect to the completion $\mathcal{F}^\mu$ are called $\mu$-
measurable. An equivalence class of $\mu$-measurable functions is called an almost measurable function, which is considered defined only almost everywhere, i.e., a function whose restriction to the complement of a null set is a measurable function. This space $L^0(\Omega, \mathcal{F}, \mu; E) = L^0(\Omega, \mathcal{F}^\mu, \mu; E)$ of $E$-valued measurable functions defined almost everywhere is denoted by $L^0(\Omega, \mu; E)$ and by $L^0$, when the meaning is clear from the context. Certainly, “equality” in $L^0$ means $\mu$-almost everywhere pointwise equality.

In most of the cases, $E$ is a metric space and $\mathcal{E}$ is its Borel $\sigma$-algebra. The imposition of a separable range $f(\Omega)$ is rather technical, but very convenient. Most of the time, we have in mind the typical case of $E$ being a Polish space (mainly, the extended $\mathbb{R}^d$), so that this condition is always satisfied.

**Proposition 4.25.** Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $(E, d_E)$ be a metric space. If $f$ and $g$ are two almost measurable functions from $\Omega$ into $E$, we define

$$d_\mu(f, g) = \inf \{ r > 0 : \mu(\{ \omega \in \Omega : d_E(f(\omega), g(\omega)) > r \}) \leq r \}.$$ 

Then (1) the map $(f, g) \rightarrow d_\mu(f, g)$ is a metric on $L^0 = L^0(\Omega, \mathcal{F}, \mu; E)$; (2) one has $d_\mu(f_n, f) \to 0$ if and only if $f_n \to f$ in measure; (3) the metric $d_\mu$ is complete in $L^0$ whenever $d_E$ is complete in $E$.

**Proof.** Note that to have $d_\mu(f, g)$ fully define, we should contemplate the possibility of having $\mu(\{ \omega \in \Omega : d_E(f(\omega), g(\omega)) > r \}) = \infty$ for every $r > 0$, in which case, we define $d_\mu(f, g) = \infty$. Thus, to make a proper distance we could replace $d_\mu(f, g)$ with $d_\mu(f, g) \vee 1$, or equivalently re-define

$$d_\mu(f, g) = \inf \{ r \in [0, 1] : \mu(\{ \omega \in \Omega : d_E(f(\omega), g(\omega)) > r \}) \leq r \},$$

with the understanding that $\inf \{0\} = 1$.

First, we can check that $d_\mu$ satisfies the triangular inequality and becomes a metric (or distance) in $L^0$. Now, by definition, there exists a decreasing sequence $r_n = r_n(f, g)$ such that $r_n \to d_\mu(f, g)$ and $\mu(\{ \omega \in \Omega : d_E(f(\omega), g(\omega)) > r_n \}) \leq r_n$, the monotone continuity from below of the measure $\mu$ shows that

$$\mu(\{ \omega \in \Omega : d_E(f(\omega), g(\omega)) > d_\mu(f, g) \}) \leq d_\mu(f, g),$$

i.e., convergence in measure is given as the convergence in the metric $d_\mu$. Finally, we conclude by applying Theorem 4.17.

Consider $S^0 = S^0(\Omega, \mathcal{F}; E) \subset L^0$ and $S^0 = S^0(\Omega, \mu; E) \subset L^0$, the subspaces of all simple functions, (i.e., measurable functions assuming only a finite number of values). We may also consider $S^0(\Omega, \mathcal{F}^\mu; E)$ if needed. Clearly, $S^0$ is not closed (nor complete) in $L^0$. For instance, if $E$ is a separable metric space then Corollary 1.10 shows that for any element $f$ in $L^0(\Omega, \mu; E)$ there exists a sequence $\{f_n\} \subset L^0(\Omega, \mu; E)$ and a null set $N$ such that $f_n$ is a measurable function assuming only a finite number of values (i.e., $f_n$ is an almost everywhere simple function), and $d_E(f_n(\omega), f(\omega))$ decreases to 0 as $n \to \infty$ for every $x \in$
\( \Omega \setminus N. \) Hence, if \( \mu(\Omega) < \infty \) then \( f_n \to f \) in measure, i.e., \( d_\mu(f_n, f) \to 0 \) as \( n \to \infty. \)

Because it is desirable to approximate any function in \( L^0 \) by a sequence of function in \( S^0 \), we have modified a little the definition of measurable functions when \( E \) is not separable, by adding almost separability of the range. Moreover, the topology in \( L^0 \) should be slightly modified, i.e., convergence in measure on every set of finite measure.

Even when the (complete) metric space \( E \) and the \( \sigma \)-algebra \( \mathcal{F} \) are separable, the separability of the (complete) metric space \( L^0 \) is an issue, because some property of the measure \( \mu \) are also involved.

- **Remark 4.26.** If \( \Omega \) is a Polish space, \( \mu \) is a regular Borel finite measure and \( E \) is separable, then \( L^0(\Omega, \mu; E) \) is separable. Indeed, if \( \Omega \) is a separable complete metric space then the arguments in Remark 3.13 can be used to show that there exists a countable basis of open sets \( \mathcal{O} \) such that for every \( \varepsilon > 0 \) and any Borel set \( B \) with finite measure, \( \mu(B) < \infty \), there exists \( \mathcal{O} \) in \( \mathcal{O} \) satisfying \( \mu((B \setminus \mathcal{O}) \cup (\mathcal{O} \setminus B)) < \varepsilon. \) Hence, if \( \mu(\Omega) < \infty \) and \( E \) is separable metric space then the space \( S^0 = S^0(\Omega, \mu; E) \) of simple functions is separable, e.g., the countable family of simple functions \( f \) such that \( f^{-1}(c) \) belongs to \( \mathcal{O} \) for every \( c \) is a dense set. Since \( S^0 \) is dense in \( L^0 \), we deduce the separability of \( L^0. \)

If \( E \) is a Banach space (i.e., complete normed space) with norm \( | \cdot |_E \) then the function

\[
 d_\mu(f, 0) = \inf \{ r > 0 : \mu(\{ \omega \in \Omega : |f(\omega)|_E > r \}) \leq r \}
\]

is not necessarily homogeneous, for instance if \( f = 1_F \) with \( F \in \mathcal{F} \) then \( d(cf, 0)_\mu = c \wedge \mu(F) \), for every \( c \geq 0. \) Nevertheless, \( d_\mu(cf, 0) \leq (1 \vee |c|)d_\mu(f, 0) \) and therefore \( cf \to 0 \) if \( f \to 0. \) Moreover, if \( \mu(\{ \omega \in \Omega : f(\omega) \neq 0 \}) < \infty \) then for every \( \varepsilon > 0 \) there exist \( \delta > 0 \) such that \( \mu(\{ \omega \in \Omega : d_\mu(f(\omega), 0) > 1/\delta \}) < \varepsilon \) and therefore \( d_\mu(cf, 0) \leq \varepsilon \) whenever \( |c| < \varepsilon\delta. \) Thus, besides \( L^0(\Omega, \mu; E) \) being a complete metric space, it is not quite a topological vector space, i.e., the vector addition is continuous but the scalar multiplication is continuous only on functions vanishing outside of a set of finite measure.

If \( E = \mathbb{R} \) then \( L^1(\Omega, \mathcal{F}, \mu) = L^1(\Omega, \mathcal{F}, \mu; \mathbb{R}) \) is the vector space of real-valued integrable functions, where the expression

\[
 \|f\|_1 = \int_\Omega |f|d\mu
\]

defines a semi-norm, i.e., we need to consider equivalence class of functions and consider the quotient space \( L^1(\Omega, \mu) \) as a subspace of \( L^0(\Omega, \mu) \), and \( | \cdot |_1 \) becomes a norm on \( L^1(\Omega, \mu). \) It is simple to verify that \( L^1(\Omega, \mu) \) is a closed subspace of the complete space \( L^0(\Omega, \mu) \), therefore \( L^1(\Omega, \mu) \) is complete, i.e., \( L^1(\Omega, \mu) \) results a Banach space. Note that if \( \mathbb{R} = [-\infty, +\infty] \) then \( L^0(\Omega, \mu; \mathbb{R}) \) is not necessarily equal to \( L^0(\Omega, \mu; \mathbb{R}) \), but, since any integrable function is finite almost everywhere, we do have \( L^1(\Omega, \mu; \mathbb{R}) = L^1(\Omega, \mu; \mathbb{R}). \)

Denote by \( S^1 \subseteq S^1(\Omega, \mu; E) \subseteq L^0 \) the subspace of all (almost) simple functions with finite-measure support, i.e., almost measurable functions \( f \) assuming a finite number of values and satisfying \( \mu(x \in \Omega : f(x) \neq 0) < \infty. \)
Proposition 4.27. Let \((\Omega, \mathcal{F}, \mu)\) be a measure space and \((E, |\cdot|_E)\) be a Banach space. Then for every \(f\) in \(L^0(\Omega, \mu; E)\) such there exists a sequence \(\{f_n\}\) of almost everywhere simple functions almost everywhere pointwise convergent to \(f\) satisfying
\[
|f_{n+1}(\omega) - f(\omega)|_E \leq |f_n(\omega) - f(\omega)|_E \leq |f(\omega)|_E, \quad \forall n \geq 1, \text{ a.e.} \omega.
\]
Moreover, if \(\mu(\{\omega \in \Omega : |f(\omega)|_E \geq c\}) < \infty\) for every \(c > 0\), then \(f_n \to f\) in measure. Furthermore, if \(\mu(\Omega) < \infty\) then the subspace \(S^1(\Omega, \mu; E) = S^0(\Omega, \mu; E)\) is dense in the complete topological vector metric space \((L^0, d_\mu)\).

Proof. Suppose \(f\) in \(L^0\) defined outside of a negligible set \(N\) with \(f(\Omega \setminus N)\) contained into the closure of a countable subset \(\{e_i : i \geq 1\} \subset E\). Include \(e_0 = 0\) and use the construction in Corollary 1.10 to define \(f_n(\omega) = e_{i(n,\omega)}\), where
\[
i(\omega, n) = \min\{i \leq n : q(\omega, n) = |f(\omega) - e_i|_E\},
q(\omega, n) = \min\{|f(\omega) - e_i|_E : i = 0, 1, \ldots, n\}.
\]
Since
\[
|f_{n+1}(\omega) - f(\omega)|_E = q(\omega, n+1) \leq q(\omega, n) = |f_n(\omega) - f(\omega)|_E \leq |e_0 - f(\omega)|_E = |f(\omega)|_E, \quad \forall \omega \in \Omega \setminus N,
\]
we obtain the estimate. The almost pointwise convergence follows form the density of subset \(\{e_i : i \geq 0\}\).

Now, for every \(\varepsilon > 0\) we have
\[
A_{n,\varepsilon} = \{\omega \in \Omega : |f_n(\omega) - f(\omega)|_E \geq \varepsilon\} \subset \{\omega \in \Omega : |f(\omega)|_E \geq \varepsilon\} = B_\varepsilon
\]
and \(\mu(B_\varepsilon) < \infty\). Since \(\{A_{n,\varepsilon} : n \geq 1\}\) is a decreasing sequence with \(\bigcap_n A_{n,\varepsilon} \subset N\) we deduce \(\mu(A_{n,\varepsilon}) \to 0\), i.e., \(f_n \to f\) in measure. \(\square\)

Remark 4.28. Let \((E, |\cdot|_E)\) be a Banach space (not necessarily separable) and \(f\) be an element in \(L^0(\Omega, \mu; E)\) (including separable range as mentioned early) such that \(\mu(\{\omega \in \Omega : |f(\omega)|_E \geq c\}) < \infty\) for every \(c > 0\). Then, for a sequence \(\{f_n\}\) of almost everywhere simple functions convergent (pointwise) almost everywhere to \(f\), we may define \(\tilde{f}_n(\omega) = f_n(\omega)\mathbb{1}_{\Omega \setminus F_n}(\omega)\) with \(F_n = \{\omega \in \Omega : n|f(\omega)|_E \geq 1\}\) to have
\[
\mu(\{\omega \in \Omega : |\tilde{f}_n(\omega) - f(\omega)|_E \geq \varepsilon\}) \leq \mu(\{\omega \in F_m : |\tilde{f}_n(\omega) - f(\omega)|_E \geq 1/m\}),
\]
for any \(n \geq m\). Since \(\mu(F_m) < \infty\), we deduce that \(\tilde{f}_n \to f\) in measure. \(\square\)

Exercise 4.19. A real-valued function \(f\) on a measure space \((\Omega, \mathcal{F}, \mu)\) belongs to weak \(L^1\) if
\[
\mu(\{\omega \in \Omega : r|f(\omega)| > 1\}) \leq Cr, \quad \forall r > 0,
\]
for some finite constant $C = C_f$. Prove (1) any integrable function belongs to weak $L^1$ and verify (2) that the function $f(x) = |x|^{-d}$ is not integrable in $\Omega = \mathbb{R}^d$ with the Lebesgue measure $\mu = \ell$, but it does belong to weak $L^1$. Now, consider the map

$$
\|f\|_1 = \sup_{r>0} \{ r \mu(\{ \omega \in \Omega : |f(\omega)| > r \}) \},
$$

for every $f$ in $L^1(\Omega, \mu; \mathbb{R})$. Prove that (3) $\|cf\|_1 = |c| \|f\|_1$, for any constant $c$; (4) $\|f+g\|_1 \leq 2\|f\|_1 + 2\|g\|_1$; and (5) if $\|f_n - f\|_1 \to 0$ then $f_n \to f$ in measure. Finally, if $L^1_w = L^1_w(\Omega, \mu; \mathbb{R})$ is the subspace of all almost measurable functions $f$ satisfying $\|f\|_1 < \infty$ (i.e., the weak $L^1$ space) then prove that (6) $(L^1_w, \| \cdot \|_1)$ is a topological vector space, i.e., besides being a vector space with the topology given by the balls (even if $\|f\|_1$ is not a norm and therefore does not induce a metric) makes the scalar multiplication and the addition continuous operations. See Folland [45, Section 6.4, pp. 197–199] and Grafakos [55, Section 1.1].

Another subspaces of interest is $L^\infty = L^\infty(\Omega, \mu; E) \subset L^0$ of all almost measurable and almost bounded functions, i.e., $f$ defined outside of a negligible set $N$ with $f(\Omega \setminus N)$ contained into the closure of a countable bounded subset of the Banach space $(E, | \cdot |_E)$. Thus, $(L^\infty, \| \cdot \|_\infty)$, where

$$
\|f\|_\infty = \inf \{ C \geq 0 : |f(\omega)|_E \leq C, \text{ a.e. } \omega \}.
$$

is a Banach space. The elements in $L^\infty$ are called essentially bounded measurable functions and $\|f\|_\infty$ is the essential sup-norm of $f$. In general, it is clear that $L^\infty$ is non separable.

In particular, if $f$ belongs to $L^\infty(\Omega, \mu; \mathbb{R})$ then the approximation arguments in Proposition 1.9 applied to $\omega \mapsto f^\pm(\omega)$ yield sequences $\{f^\pm_n : n \geq 1\}$ of almost everywhere simple functions such that $0 \leq f^\pm_n \leq f^\pm_{n+1} \leq f^\pm$ and $0 \leq f^\pm(\omega) - f^\pm_n(\omega) \leq 2^{-n}$ if $f^\pm(\omega) \leq 2^n$, and $f^\pm_n(\omega) = 0$ if $f^\pm(\omega) < 2^{-n}$. Therefore, the function $f_n = f^+_n - f^-_n$ belongs to $S^0(\Omega, F, \mu; \mathbb{R})$ and it satisfies $|f_n(\omega)| \leq |f(\omega)|$, $|f(\omega) - f_n(\omega)| \leq |f(\omega)|$ and also $|f(\omega) - f_n(\omega)| \leq 2^{-n}$ if $|f(\omega)| \leq 2^n$. Hence, $\|f - f_n\|_\infty \to 0$, i.e., $S^0(\Omega, \mu; \mathbb{R}^d)$ is dense in $(L^\infty(\Omega, \mu; \mathbb{R}^d), \| \cdot \|_\infty)$, for every $d \geq 1$. However, $(S^0, \| \cdot \|_\infty)$ is not separable in general.

Finally, we say that an almost measurable function $f$ has $\sigma$-finite support if the set $\{ \omega \in \Omega : f(\omega) \neq 0 \}$ is $\sigma$-finite, i.e., there exists a sequence $\{A_n : n \geq 1\}$ of measurable sets satisfying $\{ \omega \in \Omega : f(\omega) \neq 0 \} = \bigcup_n A_n$ and $\mu(A_n) < \infty$, for every $n$. The subspace of all almost measurable functions with $\sigma$-finite support is denoted by $L^0_\sigma = L^0_\sigma(\Omega, F, \mu; E)$ and endowed with the convergence in measure on every set of finite measure (also called stochastic convergence), i.e., $f_n \to f$ in $L^0_\sigma$ if for every $\varepsilon > 0$ and any set $F$ in $\mathcal{F}$ with $\mu(F) < \infty$ there exists an index $N = N(\varepsilon, F)$ such that $\mu(\{ \omega \in F : |f_n(\omega) - f(\omega)|_E \geq \varepsilon \}) < \varepsilon$, for every $n > N$. If $(\Omega, \mathcal{F}, \mu)$ is a $\sigma$-finite measure space then $L^0_\sigma$ is a metrizable space, which is complete if $E$ is so. Note that the sequence constructed in Proposition 4.27 satisfies $\{ \omega \in \Omega \setminus N : f_n(\omega) \neq 0 \} \subset \{ \omega \in \Omega \setminus N : f(\omega) \neq 0 \}$, with $\mu(N) = 0$, thus $f_n$ belongs to $L^0_\sigma$ if $f$ do so. Hence, $S^0$ is dense in $L^0_\sigma \subset L^1_w$.

For instance, the interested reader may consult the books by Bauer [10, Section 20, pp. 112–121] and Federer [44, Section 2.3, pp. 72–80], among others.
Exercise 4.20. Paying special attention to the almost everywhere concept, show that if a sequence of functions \( \{f_n\} \) in \( L^1(\Omega; \mu; \mathbb{R}) \) satisfies \( \sum_n \|f_n\|_1 < \infty \) then \( \sum_n f_n \) converges to a function \( f \) in \( L^1(\Omega; \mu; \mathbb{R}) \) and
\[
\int_{\Omega} f \, d\mu = \sum_n \int_{\Omega} f_n \, d\mu.
\]
Again, deduce that \( L^1(\Omega; \mu; \mathbb{R}) \) is a complete space, i.e., a Banach space. \( \square \)

Exercise 4.21. Let \( \mu \) be a Borel measure on a Polish space \( \Omega \). Give details on most of the statements related to the spaces \( L^1(\Omega; \mu; \mathbb{R}), L^0(\Omega; \mu; \mathbb{R}), L^0_0(\Omega; \mu; \mathbb{R}), L^\infty(\Omega; \mu; \mathbb{R}), L^\infty_0(\Omega; \mu; \mathbb{R}), S^0(\Omega; \mu; \mathbb{R}) \) and \( S^1(\Omega; \mu; \mathbb{R}) \), recall that the \( \sigma \) refers to the \( \sigma \)-finite support. In particular, define a metric (or norm), specify when the space is separable and/or complete, and state any topological inclusion. Moreover, for \( \Omega = \mathbb{R}^d \) and \( \mu = \ell \) the Lebesgue measure, if possible, give examples of functions in each of the above spaces not belonging to any of the others. \( \square \)

Exercise 4.22. Let \( E \) be a vector lattice of real-valued function defined on \( X \) and \( I : E \to \mathbb{R} \) be a linear and monotone functional satisfying the condition: if \( \{\varphi_n\} \) is a decreasing sequence in \( E \) such that \( \varphi_n(x) \downarrow 0 \) for every \( x \) in \( X \), then \( I(\varphi) \downarrow 0 \). Consider the vector lattice \( L \) as defined in Exercise 4.6, which are \( I \)-integrable function defined outside of an \( I \)-null set, see Exercise 4.5. Let \( M^+ \) be the semi-space of functions \( f : X \to [0, +\infty] \) such that \( f \land \varphi \) belongs to \( L \), for every nonnegative \( \varphi \) in \( L \). Functions \( f \) such that \( f^+ \) and \( f^- \) belong to \( M^+ \) are called \( I \)-measurable.

(1) Prove that (a) \( M^+ \) is stable under the pointwise convergence of sequences, and that (b) \( M^+ \) is a semi-vector lattice, i.e., for every positive constant \( c \) and any functions \( f_1 \) and \( f_2 \) in \( M^+ \), the functions \( f_1 + cf_2, f_1 \land f_2 \) and \( f_1 \lor f_2 \) are also in \( M^+ \). Moreover, show that for any \( f \) and \( g \) in \( M^+ \) with \( f - g \geq 0 \) we have \( f - g \) in \( M^+ \). For a \( f \) in \( M^+ \) we define
\[
I(f) = \sup \left\{ I(f \land \varphi) : \varphi \in L, \varphi \geq 0 \right\}.
\]
Verify that Beppo Levi’s Theorem holds true within the semi-space \( M^+ \) and \( I \) is semi-linear and monotone on \( M^+ \).

(2) Consider the class \( \mathcal{A} \) of subsets of \( \Omega \) such that \( 1_A \) belongs to \( M^+ \). Prove that \( \mathcal{A} \) is a \( \sigma \)-ring and the set function \( A \mapsto \mu(A) = I(1_A) \) is a measure on \( \mathcal{A} \).

(3) Assume that the function \( 1 = 1_X \) is \( I \)-measurable and verify that \( \mathcal{A} \) is a \( \sigma \)-algebra. Prove that a function \( f \) is almost everywhere \( (\mathcal{A}, \mu) \)-measurable if and only if \( f \) is \( I \)-measurable.

(4) Suppose that the initial vector lattice \( E \) is the vector space generated by all functions of the form \( 1_A \) with \( A \) in \( \mathcal{E} \), where \( \mathcal{E} \) is a semi-ring in a measure space \( (X, \mathcal{F}, \mu) \), and \( I \) is defined as in Exercise 4.5. Assume that \( \mathcal{F} = \sigma(\mathcal{E}) \) and that there is a sequence \( \{E_n\} \) of sets in \( \mathcal{E} \) such that \( X = \bigcup_n E_n \) with \( \mu(E_n) < \infty \). Can we affirm that any almost everywhere measurable function \( f \) is an almost everywhere pointwise limit of a sequence of functions in \( E \) or perhaps in \( E \)?
Hint: the identities \((x - y) \land z = x \land (z + y \land z) - y \land z\), for real numbers \(x, y, z\) with \(x \geq 0\), \((a - 1)^+ \land b = (a \land (b + a \land 1) - 1)^+\), for any real numbers \(a, b\) with \(b \geq 0\), and the monotone limit \(\mathbb{1}_A = \lim_n (n[f(\cdot)/a - 1]^+) \land 1\), with \(A = \{x \in \Omega : f(x) > a > 0\}\) may be of some help.

\[\square\]

**Exercise 4.23.** Let \(f\) be a function defined on \(\mathbb{R}^d\) and let \(B(x, r)\) denote the open ball \(\{y \in \mathbb{R}^d : |y - x| < r\}\). (1) Prove that the functions \(\underline{f}(x, r) = \inf \{f(y) : y \in B(x, r)\}\) and \(\overline{f}(x, r) = \sup \{f(y) : y \in B(x, r)\}\) are lower semi-continuous (lsc) and upper semi-continuous (usc) with respect to \(x\). (2) Can we replace the open ball with the closed ball \(\bar{B}(x, r) = \{y \in \mathbb{R}^d : |y - x| \leq r\}\)? (3) Why the functions \(\overline{f}(x) = \inf_{r>0} \overline{f}(x, r)\) and \(\underline{f}(x) = \sup_{r>0} \underline{f}(x, r)\) are also usc and lsc, respectively? (4) Discuss what could be the meaning of \(\overline{f}(x, r)\) and \(\underline{f}(x, r)\) if the function \(f\) is only almost everywhere defined, see Exercises 1.22 and 5.1. In this case, can we deduce that \(\overline{f}(x, r)\) and \(\underline{f}(x, r)\) are usc and lsc, almost everywhere? \[\square\]

## 4.5 Typical Function Spaces

Let \((\Omega, \mathcal{F}, \mu)\) be a measure space, \((E, \mathcal{E})\) be a measurable space, and \(v : \Omega \rightarrow E\) be a measurable function. We can define the image measure \(\mu_v(B) = \mu(v^{-1}(B))\), for every \(B\) in \(\mathcal{E}\). Thus, \((E, \mathcal{E}, \mu_v)\) becomes a measure space, which carried the combined “information” about \(\mu\) and \(v\).

For instance, if \(v\) is a real-valued measurable function then the distribution of \(v\) with respect to \(\mu\) is defined by

\[\mu_v(r) = \mu(\{x : v(x) > r\}), \quad \forall r \in \mathbb{R}.\]

It is clear that \(\mu_v : \mathbb{R} \rightarrow [0, +\infty]\) is an increasing function, and then \(\mu_v\) defines a unique measure on \(\mathbb{R}\), the image of \(\mu\) through \(v\). For any measurable function \(h : \mathbb{R} \rightarrow \mathbb{R}\) we have

\[\int_{\Omega} h(v(x)) \mu(dx) = \int_{\mathbb{R}} h(r) \mu_v(dr).\]

Indeed, if \(h\) is a characteristic function \(\mathbb{1}_A\) then the above equality is the definition of the image measure. Next, approximating \(h^+\) and \(h^-\) be an increasing sequence of simple functions we conclude. In particular, the function \(r \mapsto \mu_{|v|}(r)\) can be considered only as defined for \(r \geq 0\) and it satisfies

\[\int_{\Omega} |v|^p \, d\mu = -\int_{0}^{\infty} r^p \, d\mu_{|v|}(r) = p \int_{0}^{\infty} r^{p-1} \mu_{|v|}(r) \, dr,\]

for every measurable function \(v\). Note that

\[r \mu_{|v|}(r) \leq \int_{\Omega} |v(x)| \mu(dx),\]

and the fact that \(\mu_v(r)\) or \(\mu_{|v|}(r)\) are considered functions while, the \(\mu_v(dr)\) or \(\mu_{|v|}(dr)\) means the corresponding measures, see Exercise 5.6. In any case, the reader may check other books, e.g., Yeh [120, Chapter 4, pp. 323–480].
Some Inequalities

Now, let $L^0 = L^0(\Omega, \mathcal{F}, \mu; E)$ be the space of all almost measurable $E$-valued functions, where $(E, |\cdot|_E)$ is a Banach space. For $1 \leq p \leq \infty$ and any $f \in L^0$ we consider

$$\|f\|_p = \left( \int_{\Omega} |f|^p_{E} \, d\mu \right)^{1/p} < \infty, \quad \forall 1 \leq p < \infty,$$

(4.5)

$$\|f\|_\infty = \inf \{ C \geq 0 : |f|_E \leq C, \text{ a.e.} \},$$

where $\|f\|_\infty = \infty$ if $\mu(\{x : |f(x)|_E \geq C\}) > 0$, for every $C > 0$. We define $L^p = L^p(\Omega, \mathcal{F}, \mu; E)$ as the subspace of $L^0(\Omega, \mathcal{F}, \mu; E)$ such that $\|f\|_p < \infty$ and for $p = \infty$ we add the condition (which is already included if $p < \infty$) that $\{f \neq 0\}$ is a $\sigma$-finite (i.e., a countable union of sets with finite measure). Recall that elements $f$ in $L^0$ are equivalence classes (i.e., functions defined almost everywhere), and that $f$ takes valued in some separable subspace of $E$, when $E$ is not separable.

Most of what follows is valid for a (separable) Banach space $E$, but to simplify, we consider only the case $E = \mathbb{R}$ or $E = \mathbb{R}^d$, with the Euclidean norm denoted by $|\cdot|$.

We have already shown that $(L^1, \|\cdot\|_1)$ and $(L^\infty, \|\cdot\|_\infty)$ are Banach spaces. The general case $1 < p < \infty$ requires some estimates to prove that $\|\cdot\|_p$ is indeed a norm.

First, recalling that the $-\ln$ function is a strictly convex function,

$$\ln(ax + by) \geq a \ln x + b \ln y, \quad \forall a, b, x, y > 0, \quad a + b = 1,$$

we check that the arithmetic mean is larger that the geometric mean, i.e.,

$$x^a y^b \leq ax + by, \quad \forall a, b, x, y > 0, \quad a + b = 1,$$

(4.6)

where the equality holds only if $x = y$.

(a) Hölder inequality: for any $p, q \geq 1$ with $1/p + 1/q = 1$ (where the limit case $1/\infty = 0$ is used) we have

$$\|fg\|_1 \leq \|f\|_p \|g\|_q, \quad \forall f \in L^p, \quad g \in L^q,$$

(4.7)

where the equality holds only if for some constant $c$ we have $|f|^p = c |g|^q$, almost everywhere. Indeed, if $\|fg\|_1 > 0$ then $\|f\|_p > 0$ and $\|g\|_q > 0$. Taking $a = 1/p$, $b = 1/q$, $x = |f|^p/\|f\|_p^p$ and $y = |g|^q/\|g\|_q^q$ in (4.6) and integrating in $\mu$, on deduce (4.7).

If $1 \leq p < r < q \leq \infty$ and $f$ belongs to $L^p \cap L^q$ then $f$ belongs to $L^r$ and

$$(1/p - 1/q) \ln \|f\|_r \leq (1/r - 1/q) \ln \|f\|_p + (1/p - 1/r) \ln \|f\|_q.$$ 

Indeed, for some $\theta$ in $(0, 1)$ we have $1/r = \theta/p + (1 - \theta)/q$ and Hölder inequality yields

$$\|f\|_r = \|f^{\theta} f^{1-\theta}\|_r = \|f^{r\theta} f^{r(1-\theta)}\|_1^{r} \leq \left\{ \|f^{r\theta}\|_{p/r\theta} \|f^{r(1-\theta)}\|_{q/r(1-\theta)} \right\}^{1/r} = \left\{ \|f\|_p^{\theta} \|f\|_q^{1-\theta} \right\}^{1/r} = \|f\|_p^{\theta} \|f\|_q^{1-\theta},$$

and the desired estimate follows.

(b) **Minkowski inequality:** if $1 \leq p \leq \infty$ then

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p, \quad \forall f, g \in L^p. \quad (4.8)$$

Indeed, only the case $1 < p < \infty$ need to be considered. Thus the inequality $|f + g|^p \leq (|f| + |g|)^p \leq 2^p (|f|^p + |g|^p)$ shows that $f + g$ belongs to $L^p$. With $q = p/(p - 1)$ we have $\|f + g\|^{p-1}_q = (\|f + g\|^p)^{p-1}$. Next, applying (4.7) to $|f + g|^p = |f + g| |f + g|^{p-1} \leq |f| |f + g|^{p-1} + |g| |f + g|^{p-1}$ we obtain (4.8).

Therefore $(L^p, \| \cdot \|_p)$ is a normed space, and the inequality

$$\varepsilon^p \mu(\{|f| \geq \varepsilon\}) \leq \|f\|_p^p,$$

shows that if $\{f_n\}$ is a Cauchy sequence in $L^p$ then it is also a Cauchy sequence in $L^0$. Hence $L^p$ is complete, i.e., it is a Banach space.

• **Remark 4.29.** A discrete version of the above Hölder inequality (4.7) can be written as

$$ab \leq \frac{a^p}{p} + \frac{a^q}{q}, \quad \forall a, b \geq 0, \quad 1 \leq p, q \leq \infty, \quad \frac{1}{p} + \frac{1}{q} = 1,$$

or more general,

$$a_1 \ldots b_n \leq \frac{a_1^{p_1}}{p_1} + \cdots + \frac{a_n^{p_n}}{p_n}, \quad \forall a_i \geq 0, \quad \frac{1}{p_1} + \cdots + \frac{1}{q_n} = 1,$$

with $1 \leq p_i, q_i \leq \infty$. Similarly, we have

$$\|f_1 \ldots f_n\|_1 \leq \|f_1\|_{p_1} \cdots \|f_n\|_{p_n}, \quad \forall f_i \in L^{p_i},$$

with $1/p_1 + \cdots + 1/p_n = 1$ and finite $n$.

• **Remark 4.30.** If $0 < p < 1$ and $f, g$ belongs to $L^p$ then $f + g$ belongs to $L^p$ and

$$\|f + g\|_p^p \leq \|f\|_p^p + \|g\|_p^p.$$

This follows from the elementary inequality $(a + b)^p \leq a^p + b^p$, for every $a, b$ in $[0, \infty)$ and $0 < p < 1$, which is deduced from $[a/(a + b)]^p + [b/(a + b)]^p \geq a/(a + b) + b/(a + b) = 1$. Hence $L^p$ with the distance $d_p(f, g) = \|f - g\|_p$, $0 < p < 1$, is a (complete metric) topological vector space. Also we have

$$\|f + g\|_p \geq \|f\|_p + \|g\|_p, \quad \forall f, g \in L^p, \quad 0 < p < 1,$$

$$\|fg\|_1 \leq \|f\|_p \|g\|_q, \quad \forall f \in L^p, \quad g \in L^q,$$
again $1/p + 1/q = 1$, but in this case $q < 0$. It is possible to show that

$$\lim_{p \to 0} \|f\|_p^p = \mu(\{\omega \in \Omega : f(\omega) \neq 0\}),$$

$$\lim_{p \to 0} \|f\|_p = \exp \left( \int_{\Omega} \ln |f| \, d\mu \right), \quad \text{if } \mu(\Omega) = 1 \text{ and } f \neq 0 \text{ a.e.,}$$

provided $f$ belongs to some $L^p(\Omega, \mathcal{F}, \mu)$ with $p > 0$. Indeed, the first inequality follows after splitting the integral over the regions $0 < |f(x)| \leq 1$ and $|f(x)| > 1$. To check the second inequality, we assume $|f| > 0$ a.e. to show (with the help of the mean value theorem) that

$$\ln \|f\|_p = \|f\|_q^{-q} \int_{\Omega} |f|^q \ln |f| \, d\mu,$$

for some $q$ in $(0, p)$. Hence, as in the argument to prove first inequality, we have

$$\|f\|_q^{-q} \int_{\Omega} |f|^q \ln |f| \, d\mu \to \int_{\Omega} \ln |f| \, d\mu.$$

Notice that if $|f| > 0$ on a set $\Omega_0$ with $0 < \mu(\Omega_0) < 1$ then we use the previous argument on the space $\Omega_0$ with the measure $A \mapsto \mu(A)/\mu(\Omega_0)$.

**Exercise 4.24.** Let $(X, \mathcal{X}, \mu)$ and $(Y, \mathcal{Y}, \nu)$ be two $\sigma$-finite measure spaces. Prove Minkowski’s integral inequality, i.e.,

$$\left[ \int_X \left( \int_Y |f(x, y)| \nu(dy) \right)^p \mu(dx) \right]^{1/p} \leq \int_Y \left( \int_X |f(x, y)|^p \mu(dx) \right)^{1/p} \nu(dy),$$

for any real-valued $(\mu \times \nu)$-measurable function $f$.

**Remark 4.31.** As mentioned in the previous Exercise 4.24, Minkowski inequality can be generalized in the following way. If $f(x, y)$ is a nonnegative measurable function on the product space $X \times Y$ and $1 \leq p \leq \infty$ then

$$\left\| \int_Y f(\cdot, y) \nu(dy) \right\|_{L^p(X)} \leq \int_Y \|f(\cdot, y)\|_{L^p(X)} \nu(dy),$$

where the integral in $Y$ is regarded as a limit of sums, i.e., approximating $f$ by an increasing sequence of simple measurable functions and taking limit. This is usually referred to as Minkowski inequality for integrals.

**Exercise 4.25.** Actually, Hölder inequality (4.7) can be generalized as follows

$$\|fg\|_r \leq \|f\|_p \|g\|_q, \quad \forall f \in L^p, \ g \in L^q,$$

for any $p, q, r > 0$ and $1/p + 1/q = 1/r$. Indeed, (1) make an argument to reduce the inequality to the case $r = 1$ and $\|f\|_p = \|g\|_q = 1$; and (2) use calculus to get first $|fg| \leq |f|^{p/p} + |g|^{q/q}$ and next the conclusion.
Based on Hölder inequality, we can define the duality paring

$$\langle f, g \rangle = \int_{\Omega} f g \, d\mu, \quad \forall f \in L^p, \ g \in L^q, \ \frac{1}{p} + \frac{1}{q} = 1,$$  \hspace{1cm} (4.9)

which has the property $$|\langle f, g \rangle| \leq \|f\|_p \|g\|_q.$$

**Proposition 4.32 (dual norm).** For any function $$f$$ in $$L^0(\Omega, \mathcal{F}, \mu)$$ with $$\sigma$$-finite support $$\{f \neq 0\}$$ we have

$$\|f\|_p = \sup \{ \langle f, g \rangle : g \in L^q, \ with \ \|g\|_q = 1 \}, \quad 1 \leq p \leq \infty,$$ \hspace{1cm} (4.10)

where $$\langle \cdot, \cdot \rangle$$ is the duality paring (4.9), and the supremum is attained with $$g = \text{sign}(f)|f|^{p-1}|f|^{1-p}$$, if $$p < \infty$$ and $$0 < \|f\|_p < \infty$$.

**Proof.** Temporarily denote by $$\|f\|_p$$ the right-hand term of (4.10). Thus Hölder inequality yields $$\|f\|_p \leq \|f\|_p$$.

For $$p < \infty$$ and $$0 < \|f\|_p < \infty$$ define $$g = \text{sign}(f)|f|^{p-1}|f|^{1-p}$$ to get $$\|g\|_q = 1, 1/p + 1/q = 1$$, and $$\langle f, g \rangle = \|f\|_p$$. On the other hand, if $$0 < a < \|f\|_\infty$$ then define the function $$g = \text{sign}(f)1_A/\mu(A)$$ with $$A = \{x : |f(x)| > a\}$$ to get $$\|g\|_1 = 1$$ and $$\langle f, g \rangle \geq a$$. Hence we have the reverse inequality $$\|f\|_p \leq \|f\|_p$$, provided $$p = \infty$$ or $$\|f\|_p < \infty$$.

If $$\|f\|_p = \infty$$ then $$f$$ is a pointwise limit of a bounded $$\mu$$-measurable bounded functions $$f_n$$ such that $$\mu(\{f_n \neq 0\}) < \infty$$ and $$|f_n| \leq |f_{n+1}| \leq |f|$$. Then $$\|f_n\|_p$$ increases to $$\|f\|_p = \infty$$ and $$\|f_n\|_p = \|f_n\|_p \leq \|f\|_p$$, i.e., $$\|f\|_p = \infty$$. \hfill \Box

**Remark 4.33.** The above proof shows that we may replace the condition $$\|g\|_q = 1$$ by $$\|g\|_q \leq 1$$ and the equality (4.10) remain true. Moreover, we may take the supremum only over simple functions $$g$$ in $$L^q$$ satisfying $$\|g\|_q = 1$$, i.e.,

$$\|f\|_p = \sup \{ \langle f, \varphi \rangle : \varphi \in S^1, \ with \ \|\varphi\|_q = 1 \},$$

where $$S^1 = S^1(\Omega, \mathcal{F}, \mu)$$ is the space of simple functions, $$\varphi = \sum_{i=1}^n a_i 1_{A_i}$$, with $$\{A_i\}$$ measurable and $$\mu(A_i) < \infty$$, for every $$i$$. \hfill \Box

**Exercise 4.26.** Let $$(\Omega, \mathcal{F}, \mu)$$ be a measure space with $$\mu(\Omega) < \infty$$. Prove (a) that $$\|f\|_p \to \|f\|_\infty$$ as $$p \to \infty$$, for every $$f$$ in $$L^\infty$$. On the other hand, on a $$\mathbb{R}^d$$ with the Lebesgue measure and a function $$f$$ in $$L^p$$, (b) verify that the function $$x \mapsto \|f(\cdot)1_{|x|<r}\|_p$$ is continuous, for any $$r > 0$$. Next, show that (c) the function $$x \mapsto f^\sharp(x, r) = \text{ess-sup}|y \setminus |y-x|<r|f(y)|$$ is lower semi-continuous (lsc) for every $$r > 0$$, and therefore (d) the function $$x \mapsto f^\sharp(x) = \lim_{r \to 0} f^\sharp(x, r)$$ is Borel measurable, see related Exercises 7.9, 4.23 and 1.22. \hfill \Box

**Exercise 4.27.** Let $$(X, \mathcal{X}, \mu)$$ and $$(Y, \mathcal{Y}, \nu)$$ be two $$\sigma$$-finite measure spaces. Suppose $$k(x, y)$$ is a real-valued ($$\mu \times \nu$$)-measurable function such that

$$\int_X |k(x, y)|\mu(dx) \leq C, \ a.e. \ y \quad and \quad \int_Y |k(x, y)|\nu(dy) \leq C, \ a.e. \ x,$$
for some constant $C > 0$. Prove (1) that the expression

$$(Kf)(x) = \int_Y k(x,y)f(y)\nu(dy), \quad \forall f \in L^p(\nu),$$

defines a linear continuous operator from $L^p(\nu)$ into $L^p(\mu)$, for any $1 \leq p \leq \infty$, i.e., the functions (a) $y \mapsto k(x,y)f(y)$ is $\nu$-integrable for almost every $x$, (b) $x \mapsto (Kf)(x)$ belongs to $L^p(\mu)$, and (c) the mapping $f \mapsto Kf$ is linear and there exists a constant $C > 0$ such that $\|Kf\|_p \leq C\|f\|_p$, for every $f$ in $L^p(\nu)$. Now, same setting, but under the assumption that the kernel $k(x,y)$ belongs to the product space $L^p(\mu \times \nu)$, prove (2) that $K$ is again a linear continuous operator from $L^q(\nu)$ into $L^p(\mu)$ and $\|Kf\|_p \leq \|k\|_p \|f\|_q$, with $1/p + 1/q = 1$. □

**Exercise 4.28.** Let $\{\psi_n\}$ be a sequence of functions in $L^1(\mathbb{R}^d)$ such that $\sup_n \|\psi_n\|_1 < \infty$ and

$$\lim_n \int_{\mathbb{R}^d} \psi_n(x)dx = 1 \quad \text{and} \quad \lim_n \int_{|x| > \delta} |\psi_n(x)|dx = 0, \quad \forall \delta > 0.$$ 

Prove that for every $f$ in $L^p(\mathbb{R}^d)$, $1 \leq p < \infty$, we have $\|f * \psi_n - f\|_p \to 0$. □

For instance, the reader may consult Folland [45, Section 193–197], Jones [65, Chapter 12, pp. 277–291] for more details.
Chapter 5

Integrals on Euclidean Spaces

Now our interest turns on measures defined on the Euclidean space $\mathbb{R}^d$. First, comparisons with the Riemann and Riemann-Stieltjes integrals are discussed and then the diadic approximation is presented. Next, a more geometric construction of measures is used, area (on manifold) and Hausdorff measures is briefly discussed.

5.1 Multidimensional Riemann Integrals

Similar to the semi-open $d$-dimensional intervals used in Section 2.5, we can consider open ($d$-dimensional) intervals $]a, b[ = ]a_1, b_1[ \times \cdots \times ]a_d, b_d[$ or closed $[a, b] = [a_1, b_1] \times \cdots \times [a_d, b_d]$, or even other possibilities like closed on the right or on the left on certain coordinates $i$, not all the same choices. A partition of a given interval $I$ is a finite collection $\{I_1, \ldots, I_n\}$ of non-overlapping intervals whose union in $I$, i.e., $\hat{I}_i \cap I_j = \emptyset$ when $i \neq j$, and $I = \bigcup_{i=1}^n \hat{I}_i$. This means that the boundary of the intervals are ignored, or alternatively, we may use the semi-ring of semi-open intervals. Now, recall that a step function on a compact interval $I$ is a function $\alpha : I \to \mathbb{R}$ such that $\alpha$ is constant on the each open interval $\hat{I}_k$ of some partition $\{I_1, \ldots, I_n\}$ of $I$, i.e., $\alpha(x) = \alpha_k$ for every $x \in \hat{I}_k$, $k = 1, \ldots, n$, and the values on $I_k \setminus \hat{I}_k$ are ignored (a negligible set) and any pointwise property such as “equality” is considered everywhere “except on the boundaries”. Denote by $S(I)$ the class of all steps functions on a given interval $I$. It is clear that if $\alpha$ and $\beta$ belong to $S(I)$ then $\alpha + \beta$, $\alpha \beta$, $\max\{\alpha, \beta\}$ and $\min\{\alpha, \beta\}$ also belong to $S(I)$. Even without the knowledge of the Lebesgue measure, we may define the integral

$$\int_I \alpha(x) dx = \sum_{k=1}^n \alpha_k m(I_k), \quad \forall \alpha \in S(I),$$
where \( m(I_k) \) is the \( d \)-volume of the interval \( I_k \). The notation \( dx \) is temporarily used for the Riemann (or Darboux) integral.

**Semi-continuous Functions**

First, it may be convenient to give some comments on semi-continuous functions. Recall that a function \( f : \mathbb{R}^d \rightarrow [-\infty, +\infty] \) is lower semi-continuous (in short lsc) at a point \( x \) if and only if for any \( r < f(x) \) there exists \( \delta > 0 \) such that for every \( y \in \mathbb{R}^d \) with \( |y - x| < \delta \) we have \( r < f(y) \). This is equivalent to the condition \( f(x) \leq \liminf_{y \rightarrow x} f(y) \), i.e., for every \( x \) and any \( \varepsilon > 0 \) there exists \( \delta > 0 \) such that \( |y - x| < \delta \) implies \( f(y) \geq f(x) - \varepsilon \). Similarly, \( f \) is upper semi-continuous (in short usc) if \( -f \) is lsc, i.e., if and only if \( f(x) \geq \limsup_{y \rightarrow x} f(y) \), i.e., for every \( x \) and any \( \varepsilon > 0 \) there exists \( \delta > 0 \) such that \( |y - x| < \delta \) implies \( f(y) \leq f(x) + \varepsilon \).

Given a function \( f : \mathbb{R}^d \rightarrow [-\infty, +\infty] \) we define its lsc envelop \( \underline{f} \) and its usc envelop \( \overline{f} \) as follow

\[
\underline{f}(x) = f(x) \land \liminf_{y \rightarrow x} f(y) \quad \text{and} \quad \overline{f}(x) = f(x) \lor \limsup_{y \rightarrow x} f(y).
\]  

(5.1)

We have \( f \leq f \leq \overline{f} \) and if \( f(x, r) = \inf\{f(y) : |y - x| \leq r\} \) and \( \overline{f}(x, r) = \sup\{f(y) : |y - x| \leq r\} \) then

\[
\underline{f}(x) = \lim_{r \rightarrow 0^+} f(x, r) = \sup_{r > 0} f(x, r), \quad \overline{f}(x) = \lim_{r \rightarrow 0^+} \overline{f}(x, r) = \inf_{r > 0} \overline{f}(x, r).
\]

Moreover, we do not necessarily need to use the Euclidian norm on \( \mathbb{R}^d \), the ball (either open or closed), i.e., \( \{y : |y - x| < r\} \) or \( \{y : |y - x| \leq r\} \), may be a cube, i.e., \( \{y = (y_i) : |y_i - x_i| < r\} \), or \( \{y = (y_i) : |y_i - x_i| \leq r\} \), and the values of either \( \underline{f} \) and \( \overline{f} \) are unchanged. Actually, \( \mathbb{R}^d \) can be replaced by any metric space \((X, d)\) and all properties are retained.

Note that a subset \( A \subset \mathbb{R}^d \) is open if and only if \( 1_A \) is lsc. Clearly, \( 1_A = 1_{\bar{A}} \) and \( \overline{1_A} = 1_{\bar{A}} \), where \( A \) and \( \bar{A} \) are the interior and the closure of a set \( A \).

The following list of properties hold:

1. If \( f(x) = -\infty \) (or \( f(x) = +\infty \)) then \( f \) is lsc (or usc) at \( x \);
2. If \( f(x) = +\infty \) then \( f \) is lsc at \( x \) if and only if \( \lim_{y \rightarrow x} f(y) = +\infty \);
3. \( f \) (or \( \overline{f} \)) is the largest lsc (or smallest usc) function above (below) \( f \);
4. \( f \) is continuous at \( x \) if and only if \( f \) is lsc and usc at \( x \);
5. \( f \) is lsc (or usc) at \( x \) if and only if \( f(x) \leq \underline{f}(x) \) (or \( f(x) \geq \overline{f}(x) \));
6. If \( f_i \) is lsc (or usc) for all \( i \) then \( \sup_i f_i \) (or \( \inf_i f_i \)) is also lsc (or usc), moreover, if the family \( I \) is finite then \( \inf_i f_i \) (or \( \sup_i f_i \)) results lsc (or usc);
7. \( f \) is lsc (or usc) if and only if \( f^{-1}([a, +\infty]) \) (or \( f^{-1}([\infty, a]) \)) is an open set for every real number \( a \), as a consequence, any lsc (or usc) function \( f \) is Borel measurable;
(8) If \( f = g \) except in a set \( D \) of isolated points then \( \underline{f}(x) = \underline{g}(x) \) and \( \overline{f}(x) = \overline{g}(x) \), for every \( x \) not in \( D \);

(9) If \( K \) is a compact set and \( f \) is lsc (usc) then the infimum (supremum) of \( f \) in \( K \) is attained at a point in \( K \).

It is clear that if \( f \) and \( g \) are lsc (or usc) and \( c > 0 \) is a constant, then \( f + cg \) is lsc (or usc), provided it is well defined, i.e., no indetermination \( \infty - \infty \). Also, \( f \) is continuous at \( x \) if and only if \( f(x) = \underline{f}(x) = \overline{f}(x) \).

**Exercise 5.1.** Prove the above claims (1), . . . , (9).

The reader may check Gordon [54, Chapter 3, pp. 29–48], or Jones [65, Chapters 3-8, pp. 65–198], or Taylor [114, Chapters 4-6, pp. 177–323], or Yeh [120, Chapter 6, pp. 597–675] for a comprehensive treatment of the Lebesgue integral in either \( \mathbb{R} \) or \( \mathbb{R}^d \). Also, a more concise discussion can be found in Stroock [112, Chapter V, pp. 80–113] or Giaquinta and G. Modica [50, Chapter 2, pp. 67–136]. In a way, the lsc property is very useful in convex analysis, e.g., see Borwein and Vanderwerff [18], Ekeland and Temam [41], or Rockafellar [95]. On the other hand, reading pieces of the textbook by Carothers [24, Chapters 10–20] may prove very useful in relation to function spaces and integration.

**Definition of Integral**

A bounded function \( f \) defined on a compact interval \( I \) is said to be **Riemann integrable** if for every \( \varepsilon > 0 \) there exist step functions \( \alpha \) and \( \beta \) defined on \( I \) such that

\[
\alpha \leq f \leq \beta \quad \text{and} \quad \int_I (\beta - \alpha) \, dx \leq \varepsilon.
\]

The upper and lower integrals are defined by

\[
\int_I f \, dx = \inf \left\{ \int_I \beta \, dx : \beta \geq f \right\} \quad \text{and} \quad \int_I f \, dx = \sup \left\{ \int_I \alpha \, dx : \alpha \leq f \right\},
\]

which is a common number (the integral) when \( f \) is Riemann integrable. Alternatively, we may define the upper and the lower Riemann sums for \( d \)-dimensional intervals, i.e., we use

\[
\alpha_{f,\{I_k\}}(x) = \inf_{y \in \hat{I}_k} f(y), \quad \forall x \in \hat{I}_k \quad \text{and} \quad \Sigma(f,\{I_k\}) = \int_I \alpha_{f,\{I_k\}}(x) \, dx,
\]

\[
\beta_{f,\{I_k\}}(x) = \sup_{y \in \hat{I}_k} f(y), \quad \forall x \in \hat{I}_k \quad \text{and} \quad \Sigma(f,\{I_k\}) = \int_I \beta_{f,\{I_k\}}(x) \, dx.
\]

Note that we have \( \alpha_{f,\{I_k\}} \leq f \leq \beta_{f,\{I_k\}} \) if the boundary \( \bigcup_k \partial I_k \) is ignored or if we define \( \alpha_{f,\{I_k\}}(x) = \inf_I f \) and \( \beta_{f,\{I_k\}}(x) = \sup_I f \) for \( x \) belonging to \( \bigcup_k \partial I_k \). Moreover, if \( \{J_i\} \) is a refinement of \( \{I_k\} \) (i.e., for each \( I_k \) there is a sub-collection of \( \{J_i\} \) which is a partition of \( I_k \) then \( \alpha_{f,\{I_k\}} \leq \alpha_{f,\{J_i\}} \) and \( \beta_{f,\{I_k\}} \leq \beta_{f,\{J_i\}} \).
Thus a bounded function $f$ on $I$ is Riemann integrable if for every $\varepsilon > 0$ there exists a partition $\{I_k\}$ such that $\Sigma(f, \{I_k\}) - \Sigma(f, \{I_k\}) < \varepsilon$.

Since the boundary $\partial I_k$ of a $d$-dimensional interval is contained in a finite number of “vertical” hyperplane, for any sequence of partitions $\{I^n_k\}$ the set of boundary points $\bigcup_{k,n} \partial I^n_k$ has Lebesgue measure zero. The mesh or norm of a partition is given by $\max_k \{d(I_k)\}$, where $d(I_k)$ is the diameter of $I_k$. Remarking that given a partition $\{I_k\}$ on $I$, any point in $I \setminus \bigcup_k I_k$ is an interior point of some interval $I_k$, we deduce that if $\{I^n_k\}$ is a sequence of partitions of $I$ with $\max_k \{d(I^n_k)\} \to 0$ as $n \to \infty$, then the usc and lsc envelops functions, as in (5.1), satisfy

$$\bar{f}(x) = \lim_n \beta_{f, \{I^n_k\}}(x) \quad \text{and} \quad \underline{f}(x) = \lim_n \alpha_{f, \{I^n_k\}}(x),$$

for every $x$ in $I \setminus \bigcup_{k,n} \partial I^n_k$.

However, if ignoring the boundary is not desired then replace the interior $\dot{I}_k$ with the closure $\bar{I}_k$ when taking the infimum and supremum to define the functions $\alpha$ and $\beta$. Actually, the only point to observe is that any bounded functions $f$ which vanishes except in a finite number of hyperplane is Riemann integrable and its integral is zero. Indeed, for such a function there is a partition $J_k$ of intervals satisfying $f(x) = 0$ for any $x$ in the finite union of interiors $\bigcup_k J_k$. Hence, for any $\varepsilon > 0$, take a finite cover by intervals of the boundary $\bigcup_k \partial J_k$ with volume less that $\varepsilon > 0$ and complete it to a partition $\{I_1, \ldots, I_n\}$ to check that there are step functions $\alpha$ and $\beta$ (this time defined everywhere) such that $\alpha \leq f \leq \beta$ everywhere, $\alpha = \beta$ except on the intervals containing some boundary $\partial J_k$, and $\Sigma(f, \{I_k\}) - \Sigma(f, \{I_k\}) \leq \varepsilon \max|f|$, proving that $f$ is Riemann integrable with integral equal to zero. Moreover, we have

**Theorem 5.1.** Every Riemann integrable function is Lebesgue measurable and both integrals coincide. Moreover, a bounded function is Riemann integrable if and only if it is continuous almost everywhere.

**Proof.** Assume $f$ is Riemann integrable. By definition, there exists sequences of partitions of $I$ with $\max_k \{I^n_k\} \to 0$ as $n \to \infty$ such that

$$\alpha_{f, \{I^n_k\}} \leq f \leq \beta_{f, \{I^n_k\}} \quad \text{and} \quad \int_I \left(\beta_{f, \{I^n_k\}} - \alpha_{f, \{I^n_k\}}\right) dx < \frac{1}{n}, \ \forall n.$$ 

Because the usc and the lsc envelopes of a step function may modify the function only on the boundary points, we deduce

$$\alpha_{f, \{I^n_k\}} \leq f \leq \beta_{f, \{I^n_k\}},$$

except on the boundary points, which is a of Lebesgue measure zero.

Since $f$ and $I$ are bounded, if $\ell$ denotes the Lebesgue measure then

$$\int_I \ell \, d\ell = \lim_n \int_I \beta_{f, \{I^n_k\}} dx = \lim_n \int_I \alpha_{f, \{I^n_k\}} dx = \int_I f \, d\ell,$$
which implies \( f = \overline{f} \) a.e., i.e. \( f \) is continuous almost everywhere and both integrals coincide.

To show the converse, let \( f \) be a bounded function which is continuous almost everywhere. Partition the interval \( I \) into \( 2^{dn} \) congruent intervals (or rectangles of equal size) \( \{I_k^n\} \) and consider the step functions \( \alpha_{f,I_k^n} \) and \( \beta_{f,I_k^n} \) used to define the lower and the upper Riemann sums \( \Sigma(f, \{I_k^n\}) \) and \( \Sigma(f, \{I_k^n\}) \), for each \( n = 1, 2, \ldots \). Since \( \{I_k^{n+1}\} \) is a refinement of \( \{I_k^n\} \), the monotone limits below exist and we have

\[
\underline{f} = \lim_n \alpha_{f,I_k^n} \leq \lim_n \beta_{f,I_k^n} = \overline{f},
\]

outside of the boundaries, i.e., for points not in \( \bigcup_{n,k} \partial I_k^n \), which is a set of Lebesgue measure zero. Moreover, because \( f \) is continuous almost everywhere, we have \( \underline{f} = \overline{f} = f \) almost everywhere. Hence, taking the integral and limit we deduce

\[
\int_I f \, d\ell \leq \lim_n \int_I \alpha_{f,I_k^n} \, dx \leq \lim_n \int_I \beta_{f,I_k^n} \, dx \leq \int_I f \, d\ell,
\]

i.e., \( f \) is Riemann integrable. \( \square \)

Note that \( f \) is continuous almost everywhere if and only if the sets of points where \( f \) is not continuous has Lebesgue measure zero, which is not the same as there exists a continuous function \( g \) such that \( f = g \) almost everywhere.

**Exercise 5.2.** Let \( \{r_n\} \) be the sequence of all rational numbers and define the function \( f(x) = \sum_{n \geq 1} 2^{-n}(x-r_n)^{-1/2} \mathbb{1}_{\{x-r_n \in (0,1)\}} \) for every \( x \) in \( \mathbb{R} \). Prove (a) \( f \) is Lebesgue integrable in \( \mathbb{R} \), (b) \( f \) is unbounded in any interval of \( \mathbb{R} \), and (c) \( f^2 \) is not Lebesgue integrable in \( \mathbb{R} \). \( \square \)

**Exercise 5.3.** Calculate the limit

\[
A = \lim_{n \to \infty} \int_a^\infty f_n(x) \, dx,
\]

for \( f_n(x) = n/(1 + n^2 x^2) \), \( a > 0 \), \( a = 0 \) and \( a < 0 \). \( \square \)

**Spherical Coordinates**

Spherical coordinates can be used in \( \mathbb{R}^d \), i.e., every \( x \) in \( \mathbb{R}^d \) \( \setminus \{0\} \) can be written uniquely as \( x = r \, x' \), where \( 0 < r < \infty \) and \( x' \) belongs to \( S^{d-1} = \{ x \in \mathbb{R}^d : |x| = 1 \} \).

**Theorem 5.2.** The Lebesgue measure \( dx \) in \( \mathbb{R}^d \) can be expressed as a product measure \( dr \times dx' \), where \( dr \) is the Lebesgue measure on \( (0, \infty) \) and \( dx' \) is a (surface) measure on \( S^{d-1} \). Moreover, for every nonnegative measurable function in \( \mathbb{R}^d \) we have

\[
\int_{\mathbb{R}^d} f(x) \, dx = \int_{(0, \infty) \times S^{d-1}} f(r x') \, r^{d-1} \, dr \times dx'.
\]
In particular, if \( f \) is homogeneous, i.e., \( f(x) = g(|x|) \), then
\[
\int_{\mathbb{R}^d} f(x) \, dx = \omega_{d-1} \int_0^{\infty} g(r) r^{d-1} \, dr,
\]
where the value \( \omega_{d-1} = 2\pi^{d/2}/\Gamma(d/2) \) is the surface area of the unit ball, i.e., 
\( dx'(S^{d-1}) \).

Proof. It is clear that for \( d = 2 \) (or \( d = 3 \)) this is called polar (or spherical) coordinates. Moreover, the crucial point is to define the surface measure \( dx' \) on \( S^{d-1} \), which will agree with the \((d-1)\)-dimensional Hausdorff measure in \( \mathbb{R}^d \) (except for a multiplicative constant) as discussed later on.

It is clear that
\[
\psi: \mathbb{R}^d \setminus \{0\} \to (0, \infty) \times S^{d-1}, \quad \psi(x) = (r, x'), \quad r = |x|, x' = x/|x|
\]
is a continuous bijection mapping with \( \psi^{-1}(r, x') = rx' \). Then, given a Borel set \( B \) in \( S^{d-1} \) we define \( B_a = \{rx': x' \in B, r \in (0, a]\} \), i.e., \( B_a = \psi^{-1}(\{0, a]\times E) \). Thus, for the desired surface measure \( dx' \) we must satisfy (5.2) for \( f = \mathbb{1}_{E_1} \), i.e.,
\[
\ell(E_1) = \int_{\mathbb{R}^d} \mathbb{1}_{E_1}(x) \, dx = dx'(E) \int_0^1 r^{d-1} \, dr,
\]
and therefore we can define \( dx'(E) = dx(E_1) \), which results in a measure on \( S^{d-1} \). On the other hand, Theorem 2.27 shows that \( dx(E_a) = a^d dx(E_1) \) and thus
\[
dx([a, b] \times E) = dx(E_b \setminus E_a) = \frac{b^d - a^d}{d} dx'(E) = dx'(E) \int_a^b r^{d-1} \, dr,
\]
i.e., with \( dx' \) defined as above, we have the validity of equality (5.2) for any function \( f = \mathbb{1}_{[a, b] \times E} \). We conclude approximating any nonnegative measurable function by a sequence of simple functions. \( \square \)

For instance, the interested reader may consult the book by Folland [45, Section 2.7, pp. 77–81] for more details. Note that
\[
\int \mathbb{1}_{\{|x| \leq r\}} \, dx = \frac{\pi^{d/2}}{\Gamma(d/2 + 1)} r^d \quad \text{and} \quad \int \mathbb{1}_{\{|x| = r\}} \, dx' = \frac{2\pi^{d/2}}{\Gamma(d/2)} r^{d-1}
\]
are the volume and the surface area of a ball radius \( r \).

Exercise 5.4. Prove that the function \( x \mapsto |x|^{-\alpha} \) is Lebesgue integrable (a) on the unit ball \( B = \{x \in \mathbb{R}^d : |x| < 1\} \) if and only if \( \alpha < d \) and (b) outside the unit ball \( \mathbb{R}^d \setminus B \) if and only if \( \alpha > d \). \( \square \)

1Recall the Gamma function \( \Gamma(x) \) satisfying \( \Gamma(n+1) = n(n-1) \ldots 1 \), for any integer \( n \), and \( \Gamma(1/2) = \sqrt{\pi} \)
Change of Variables

First recall that ∂₁{T}(x) denotes the matrix whose entries are the partial derivative ∂₁T_j of the given mapping x ↦ T(x), between subset of a Euclidean space; and det(·) is the determinant of a matrix. Thus, we have

**Theorem 5.3 (Change of variable).** Let X and Y be open subsets of \( \mathbb{R}^d \) and \( T : X \to Y \) be a homeomorphism of class \( C^1 \). A function \( y \mapsto f(y) \) is Lebesgue measurable on \( (Y, \mathcal{L}_y, dy) \) if and only if \( x \mapsto f(T(x)) \) is Lebesgue measurable on \( (X, \mathcal{L}_x, dx) \). In this case, we have

\[
\int_Y f(y) \, dy = \int_X f(T(x)) J_T(x) \, dx,
\]

where \( J_T(x) = |\det(\partial_x T(x))| \) denotes the Jacobian of \( T \).

There are several ways of expressing the assumptions necessary for a smooth change of variables, e.g., see Strichartz [111, Section 15.2, pp. 705–726] and many other. Now, based on Theorem 2.27, we can easily prove the change of variable formula for an affine transformation \( T \). Indeed, it suffices to approximate \( f \) by a sequence of simple functions. Some more preparation is required for a nonlinear homeomorphism of class \( C^1 \), e.g., see Ambrosio et al. [3, Chapter 8, pp. 129–136] or Apostol [4, Sections 15.9–15.13, pp. 416–430] or Jones [65, Chapter 15, pp. 494–510] or Knapp [71, Section VI.5, pp. 320–326] or Schilling [104, Chapter 15, pp. 142–162]. Actually, essentially with the same arguments, we can prove the following estimate: For any function \( T : X \to \mathbb{R}^d \) with \( X \) an open subset of \( \mathbb{R}^d \), and for any set \( E \subset X \) where \( T \) is differentiable at every point of \( E \), we have

\[
\ell^*(T(E)) \leq (\sup_E J_T) \ell^*(E),
\]

where \( \ell^* \) denotes the Lebesgue outer measure on \( \mathbb{R}^d \). This implies Sard’s Theorem, i.e., the set of point \( x \), where the function \( T(x) \) is differentiable and the Jacobian \( J_T(x) = 0 \), is indeed negligible. Moreover, if \( T \) is a measurable function from an open set \( X \subset \mathbb{R}^d \) into \( \mathbb{R}^d \), i.e., \( T : X \to \mathbb{R}^d \), which is differentiable at every point of a measurable set \( E \subset \mathbb{R}^d \) then

\[
\ell^*(T(E)) \leq \int_E J_T(x) \, dx,
\]

which implies a one-side inequality \( \leq \) in the Theorem 5.3, under the sole assumption that \( T \) is only differentiable and \( f \) is a nonnegative Borel function. The reader may take a look at Cohn [28, Chapter 6, pp. 167–195] for a carefully discussion, and to Duistermaat and Kolk [38, Chapter 6, pp. 423–486] for a number of details in the change of variable formula for the Riemann integral.

**Exercise 5.5.** Prove the first part of the above Theorem 5.3, namely, on the Lebesgue measure space \( (\mathbb{R}^d, \mathcal{L}, \ell) \), verify that if \( T : \mathbb{R}^d \to \mathbb{R}^d \) is a homeomorphism of class \( C^1 \) then a set \( N \) is negligible if and only if \( T(N) \) is negligible. Hence, deduce that a set \( E \) (or a function \( f \)) is a (Lebesgue) measurable if and only if \( T(E) \) (or \( f \circ T \)) is (Lebesgue) measurable.
By means of the change of variables formula, we can define the surface measure of a $n$-dimensional $C^1$-manifold $M$ with local coordinates chart $T: \mathcal{O} \to \mathbb{R}^n$ and metric tensor given locally by a positive definite matrix $a = (a_{ij})$, $a_{ij} = (\partial_k T_i)(\partial_k T_j)$. Indeed, the expression

$$
\mu(\mathcal{O}) = \int_{T(\mathcal{O})} \sqrt{\det(a(x))} \, dx, \quad \forall \mathcal{O} \text{ open subset of } M
$$

is well defined and invariant within the manifold. For instance, if $M$ is the graph of a real-valued continuously differentiable function $y = u(x)$ with $x \in \Omega \subset \mathbb{R}^n$ then $M$ is an $n$-dimensional manifold in $\mathbb{R}^{n+1}$ and the map $T(x) = (u(x), x)$ provides a natural (local) coordinates with metric tensor given locally by the matrix $a_{ij} = \delta_{ij} + \partial_i u \partial_j u$. Thus $\sqrt{\det(a(x))} = \sqrt{1 + |\nabla u(x)|^2}$, and

$$
\mu(M) = \int_{\Omega} \sqrt{1 + |\nabla u(x)|^2} \, dx
$$

is the surface measure of $M$, in particular this is valid for the unit sphere $S^{n-1} = \{x \in \mathbb{R}^n : |x| = 1\}$.

For instance, the interested reader may consult Taylor [115, Chapter 7, pp. 83–106] for more details. In general, the reader may take a look at the textbooks by Apostol [4, Chapters 14 and 15, pp. 388–433] and Duistermaat and Kolk [39], for a detail account of the multidimensional Riemann integral.

**Exercise 5.6.** Let $\ell$ be the Lebesgue measure in $\mathbb{R}^d$, $E$ a measurable set with $\ell(E) < \infty$, and $f : \mathbb{R}^d \to [0, \infty]$ be a measurable function. Define $F_{f,E}(r) = \ell(\{x \in E : f(x) \leq r\})$ and prove (a) $r \mapsto F_{f,E}(r)$ is a right-continuous increasing function and (b) that the Lebesgue-Stieltjes measure $\ell_F$ induced by $F_{f,E}$ is equal to the $f$-image measure of the restriction of $\ell$ to $E$, i.e., $\ell_F = m_{f,E}$, with $m_{f,E}(B) = \ell(f^{-1}(B) \cap E)$, for every Borel set $B$ in $\mathbb{R}$. Now, (c) prove that for any Borel measurable function $g : [0, \infty] \to [0, \infty]$ we have the equality

$$
\int_E g(f(x)) \ell(dx) = \int_0^\infty g(r)m_{f,E}(dr) = \int_0^\infty g(r)\ell_F(dr).
$$

In particular, if $\lambda_f(r) = \ell(\{x \in \mathbb{R}^d : |f(x)| > r\})$, $m_f(B) = \ell(f^{-1}(B))$, for every Borel set $B$ in $\mathbb{R}$, and $f$ is any measurable then deduce that

$$
\int_{\mathbb{R}^d} |f(x)|^p \ell(dx) = \int_0^\infty r^p m_f(dr) = \int_0^\infty p r^{p-1} \lambda_f(r)dr, \quad \forall p > 0.
$$

Actually, verify this assertion holds for any measure space $(\Omega, \mathcal{F}, \mu)$, any measurable function $f : \Omega \to \mathbb{R}$, and any $\sigma$-finite measurable set $E$, e.g., see Wheeden and Zygmund [119, Section 5.4, pp. 77–83] for more details. Hint: consider first the case $g = 1_B$, for a Borel measurable set $B$, and regard both sides of the equal sign as measures. \qed
Exercise 5.7. Consider a measurable function \( f : \mathbb{R}^d \times \mathbb{R}^k \to \mathbb{R} \) such that the expression
\[
F(x) = \int_{\mathbb{R}^d} f(x,y) \, dy, \quad \forall x \in \mathbb{R}^d.
\]
produces a continuous function. Give precise assumptions on the function \( f \) to be able to use the dominate convergence and to show that \( F \) is differentiable at a point \( x_0 \).

Exercise 5.8. If \( E \) is a non empty subset of \( \mathbb{R}^d \) then (1) prove that the distance function \( x \mapsto d(x,E) = \inf \{|y - x| : y \in E\} \) is a Lipschitz continuous function, moreover, \( |d(x,E) - d(y,E)| \leq |x - y| \), for every \( x,y \in \mathbb{R}^d \). Now, consider Marcinkiewicz integral
\[
M_{E,\alpha}(x) = \int_{Q} d^\alpha(y,E) |x - y|^{-d-\alpha} \, dy,
\]
where \( E \) is a compact subset of \( \mathbb{R}^d \), \( \alpha \) is a positive constant and \( Q \) is a cube containing \( E \). Using Fubini-Tonelli Theorem 4.14 and spherical coordinates prove that (2) \( M_{E,\alpha}(x) \) is integrable on \( E \), and that (3)
\[
\alpha \int_{E} M_{E,\alpha}(x) \, dx \leq \omega_d \ell(Q \setminus E),
\]
where \( \ell \) is the Lebesgue measure in \( \mathbb{R}^d \) and \( \omega_d \) is the measure of the unit sphere in \( \mathbb{R}^d \), see DiBenedetto [32, Section III.15, pp. 148–151].

5.2 Riemann-Stieltjes Integrals

Usually, the Riemann-Stieltjes integral of a bounded function \( f \) with respect to another bounded function \( g \) is defined as the limit of the partial sums, i.e., for a partition \( 0 = t_0 < t_1 < \cdots < t_n = T \) choose points \( t'_i \) in \([t_{i-1}, t_i]\) to form the partial sum
\[
\Sigma(f, dg, \{t'_i, t_i\}) = \sum_{i=1}^{n} f(t'_i)[g(t_i) - g(t_{i-1})],
\]
and when the limit exists (being independent of the choice of the points \( \{t'_i\} \), as the mesh of the partition vanishes, the function \( f \) (integrand) is said to be Riemann-Stieltjes integrable (RS-integrable) with respect to (wrt) the function \( g \) (integrator). Almost the same arguments used for the Riemann integral can be used to show that a continuous integrand \( f \) is RS-integrable wrt to any monotone integrator \( g \) (this implies also wrt to a difference of monotone functions, i.e., wrt a bounded variation integrator). Next, the integration by parts shows that a monotone integrand \( f \) (non necessarily continuous) is RS-integrable wrt to any continuous integrator \( g \). Recalling that there is a one-to-one correspondence between right-continuous non-decreasing functions and Lebesgue-Stieltjes
measures, our focus is on integrators \( g = a \) which are right-continuous nondecreasing functions, and integrand with possible discontinuities. Note that contrary to the previous section on Riemann integrals, the values on the boundary of the partitions cannot be ignored.

For instance, the reader is referred to Apostol [4, Chapters 7, pp. 140–182] or Riesz and Nagy [94, Section 58, pp. 122–128] or Rudin [99, Chapter 6, pp. 120–142] or Shilov and Gurevich [106, Chapter 4 and 5, 61–110] or Stroock [112, Section 1.2, pp. 7–16], among others, to check classic properties such that

(a) any continuous integrand is RS-integrable wrt to any monotone integrator (which implies wrt to any function of bounded variation); (b) the integration by parts formula, i.e., if \( u \) is RS-integrable wrt to \( v \) then \( v \) is RS-integrable wrt to \( u \) and

\[
\int_0^T u \, dv = \left. uv \right|_0^T - \int_0^T v \, du, \quad \forall T > 0;
\]

(c) the reduction to the Riemann integral, i.e., if the integrator \( v \) is continuously differentiable (actually, absolutely continuous suffices) then any Riemann integrable function \( u \) is RS-integrable wrt to the integrator \( v \) and

\[
\int_0^T u \, dv = \int_0^T u \, \hat{v} \, dt, \quad \forall T > 0,
\]

where \( \hat{v} \) denotes the derivative of \( v \). Note that the integrator \( g \) may have jump discontinuities, or may have derivative zero almost everywhere while still being continuous and increasing (e.g., \( g \) could be the Cantor function), in these cases, the reduction of the Riemann-Stieltjes integral to the Riemann integral is not valid. Another classical result states if \( f \) is \( \alpha \)-Hölder continuous and \( g \) is \( \beta \)-Hölder continuous with \( \alpha + \beta > 1 \) then \( f \) is RS-integrable wrt \( g \). There are other ways of setting up integrals, e.g., using the Moore-Smith limit on the directed set of partitions, e.g., see in general the book by Gordon [54], Natanson [86], among others.

**Problems with Jumps**

The simple case \( f = 1_{]0,\tau]} \) (left-continuous) and \( a = 1_{[\tau,\infty[} \) with \( \tau > 0 \) presents some difficulties, i.e., give a partition as above with \((\tau = t_\eta \text{ for some } 1 \leq \eta \leq n)\) to deduce that \( \Sigma(f, dg, \{t'_i, t_i\}) = 1 \) if \( t'_\eta < \tau \) and \( \Sigma(f, dg, \{t'_i, t_i\}) = 0 \) if \( t'_\eta \geq \tau \). This means that as the mesh vanishes, the choice \( t'_i = t_i \) and \( t'_i = t_{i-1} \) yield different values, i.e., \( 1_{]0,\tau]} \) is not RS-integrable wrt to \( 1_{[\tau,\infty[} \). On the contrary, if \( f = 1_{]0,\tau]} \) then \( \Sigma(f, dg, \{t'_i, t_i\}) = 1 \) for any choice of \( t'_i \), i.e., an integrand \( f \) which is not left-continuous at some point \( \tau \) cannot be RS-integrable wrt any monotone right-continuous integrator (similarly, when the integrand \( f \) is right-continuous and the integrator \( a \) is left-continuous). Therefore, an equivalent definition (for monotone right-continuous integrators) of the Riemann-Stieltjes
integrated by saying that the RS-integral of a left-continuous piecewise constant function \( \varphi = \sum_{i=1}^{n} \varphi_i \mathbb{1}_{[\tau_{i-1}, \tau_i]} \), \( 0 = \tau_0 < \tau_1 < \cdots < \tau_n = T \) wrt a monotone integrator \( a \) is given by the partial sum

\[
\int \varphi \, da = \int_{[0,T]} \varphi \, da = \sum_{i=1}^{n} \varphi_i [a(t_i) - a(t_{i-1})].
\]

Next, for any bounded function \( f \), the upper and lower Riemann-Stieltjes integrals are defined by

\[
\overline{\int f \, \! da} = \inf \left\{ \int \beta \, \! da : \beta \geq f \right\} \quad \text{and} \quad \underline{\int f \, \! da} = \sup \left\{ \int \alpha \, \! da : \alpha \leq f \right\},
\]

and, a bounded function \( f \) is RS-integrable wrt a monotone right-continuous integrator \( a \) if for every \( \varepsilon > 0 \) there exists two left-continuous piecewise constant functions satisfying

\[
\alpha \leq f \leq \beta \quad \text{and} \quad \int_{[0,T]} (\beta - \alpha) \, da \leq \varepsilon,
\]

i.e., if the upper and lower Riemann-Stieltjes integrals agree in a common number, which is called the RS-integral of integrand \( f \) wrt the monotone integrator \( a \). The notation as integral over the semi-open interval \([0,T]\) instead of the integral from \( 0 \) to \( T \), for the RS-integral is intensional, due to the used of left-continuous integrand and right-continuous integrators.

As in the case of the Riemann integrals, any left-continuous piecewise continuous integrand (i.e., for any \( T > 0 \) there exists a partition \( 0 = t_0 < t_1 < \cdots < t_n = T \) such that the restriction of the function to the semi-open interval \([t_{i-1}, t_i]\) may be rendered continuous on the closed interval \([t_{i-1}, t_i]\)) is Riemann-Stieltjes integrable on \([0,T]\) wrt any monotone integrator \( a \). Indeed, if \( f \) is a left-continuous piecewise continuous function, \( 0 = t_0 < t_1 < \cdots < t_n = T \) and \( \varepsilon > 0 \) then there exists a refinement \( 0 = t'_0 < t'_1 < \cdots < t'_n = T \) such that the oscillation of \( f \) on the semi-open little intervals \([t'_{i-1}, t'_i]\) is smaller than \( \varepsilon \), i.e.,

\[
\sup \{|f(s) - f(s')| : s, s' \in [t'_{i-1}, t'_i]\} < \varepsilon.
\]

Hence, define \( \alpha_f(t) = \inf \{ f(s) : s \in [t'_{i-1}, t'_i]\} \) and \( \beta_f(t) = \sup \{ f(s) : s \in [t'_{i-1}, t'_i]\} \) for \( t \) in \([t'_{i-1}, t'_i]\), to deduce (5.3) as desired. The class of left-continuous piecewise continuous integrands is not suitable for making limits, a larger class is necessary.

**Relation with LS-Integral**

A left-continuous function having right-hand limits \( f : [0, \infty) \to \mathbb{R} \) is called *cag-lag*, i.e., if \( f(t) = f(t^-) \), for every \( t > 0 \), and the right-hand limit \( f(t^+) \) exists as a finite value, for every \( t \geq 0 \). Similarly, a right-continuous function having left-hand limits is called *cad-lag*, i.e., if \( a(t) = a(t+) \), for every \( t \geq 0 \), and the left-hand limit \( a(t^-) \) exists as a finite value, for every \( t \geq 0 \).

For any nonnegative Borel function \( f \) defined on \( \mathbb{R} \), denote by

\[
\int_{[0,+\infty[} f(s) \, da(s) \quad \text{and} \quad \int_{\{0\}} f(s) \, da(s) = f(0)a(0),
\]

(5.4)
the Lebesgue-Stieltjes integral, i.e., the integral of \( f \) relative to the one dimen-
sional Lebesgue-Stieltjes measure generated by the right-continuous nondecreasing
function \( a \). Note that if \( f \) is a left-continuous piecewise constant function
then the above Lebesgue-Stieltjes integral coincides with the Riemann-Stieltjes
integral, i.e., both integral agree on the class of piecewise continuous functions
if we choose a left-continuous representation.

Regarding the Riemann-Stieltjes and the Lebesgue-Stieltjes integrals,

**Proposition 5.4.** If \( \varphi \) is a cag-lad function and \( T = 1/\varepsilon > 0 \) then \( \varphi \) is bounded
on \( [0,T] \) and there exists a left-continuous piecewise constant function \( \phi_\varepsilon \) and
a \( \varepsilon \)-decomposition of the form \( 0 = \tau_0 < \tau_1 < \cdots < \tau_{n-1} < \tau_n = T \) with
\( \tau_i - \tau_{i-1} < \varepsilon \) such that the oscillation of \( \varphi - \phi_\varepsilon \) within any closed interval
\( [\tau_{i-1}, \tau_i] \) is smaller than \( \varepsilon \). Moreover, any cag-lad integrand is RS-integrable
with respect to any cad-lag monotone nondecreasing integrator and both, the
Lebesgue-Stieltjes and the Riemann-Stieltjes integrals agree for such a pair of
integrand-integrator.

**Proof.** Denote by \( \text{osc}(f,I) = \sup\{|f(s) - f(s')| : s, s' \in I\} \) the oscillation of a
function \( f \) on an interval \( I \). For any cag-lad function \( \varphi \) and \( T = 1/\varepsilon \), consider
a \( \varepsilon \)-decomposition of the form \( 0 \leq S < s_1 < \cdots < s_{n-1} < s_n = T \) such that
\( \text{osc}(\varphi,[s_{i-1},s_i]) \leq \varepsilon \), for every \( i = 1, \ldots, n \). Since \( \varphi(T) = \varphi(T-) \) there
exists \( S > 0 \) sufficiently small so that such a decomposition (with \( n=1 \)) is
possible. Now, define \( S_* \) the infimum of all those \( 0 \leq S \leq T \), where a finite \( \varepsilon \)-
decomposition \( (S,T] = \bigcup_i (s_{i-1}, s_i] \) is possible. If \( S_* > 0 \) then, because \( \varphi(S_*-) \)
exists, we can decomposes \( (S_*,T] \) and since \( \varphi(S_*) = \varphi(S_*-) \), we would be able
to decompose some larger interval \( (S,T] \supset (S_*,T] \), which is a contradiction.
Hence \( S_* = 0 \), i.e, there exists a \( \varepsilon \)-decomposition \( 0 = \tau_0 < \tau_1 < \cdots < \tau_{n-1} < \tau_n = T \)
with \( \tau_i -\tau_{i-1} < \varepsilon \) such that the oscillation \( \text{osc}(\varphi,[\tau_{i-1}, \tau_i]) < \varepsilon \), for every
\( i = 1, \ldots, n \). Consider the left-continuous piecewise constant function given by

\[
\phi_\varepsilon(t) = \sum_{i=1}^n \left( \varphi(\tau_i+) - \varphi(\tau_i) \right) 1_{\tau_i < t},
\]

with \( \varphi(0-) = 0 \), to deduce that \( \varphi - \phi_\varepsilon \) is continuous at each points \( \tau_i \). This
implies that \( \text{osc}(\varphi - \phi_\varepsilon,[\tau_{i-1}, \tau_i]) < \varepsilon \), for any \( i = 1, \ldots, n \).

To show that \( \varphi \) is RS-integrable with respect to \( a \), given a \( \varepsilon > 0 \) consider an
\( \varepsilon \)-decomposition \( 0 = \tau_0 < \tau_1 < \cdots < \tau_{n-1} < \tau_n = T \) as above and define the
following left-continuous piecewise constant functions \( \alpha(s) = \sum_{i=1}^n (\varphi(\tau_i+) - \varphi(\tau_i)) 1_{\tau_i < t} \) and \( \beta(s) = \sum_{i=1}^n (\varphi(\tau_i+) - \varphi(\tau_i) + \beta_i) 1_{\tau_i < t} \), where \( \alpha_i = \inf\{\varphi(s) : s \in [\tau_{i-1}, \tau_i]\} \) and \( \beta_i = \sup\{\varphi(s) : s \in [\tau_{i-1}, \tau_i]\} \). Since \( \alpha \leq \varphi \leq \beta \) and
\( |\beta - \alpha| \leq \varepsilon \) we deduce

\[
\int_{[0,T]} (\beta - \alpha) \, da \leq \varepsilon |a(T) - a(0)|,
\]

which proves that \( \varphi \) is RS-integrable wrt \( a \).

To calculate the value of the RS-integral, note that the previous argument
applied to \( \varepsilon = 1/k \) yields sequences \( \{\varphi_k\} \) and \( \{\varphi_k\} \) of left-continuous piecewise
constant functions satisfying \( \varphi_k(s) \leq \varphi(s) \leq \varphi_k(s) \) and \( |\varphi_k(s) - \varphi(s)| < 1/k \) for every \( s \) in any fixed bounded interval \([0, T] \). By definition of the Riemann-Stieltjes integral we have

\[
\int_{[0,T]} \varphi(s)da(s) = \lim_k \int_{[0,T]} \varphi_k(s)da(s) = \lim_k \int_{[0,T]} \varphi_k(s)da(s),
\]

and because both, Riemann-Stieltjes and Lebesgue-Stieltjes integrals agree on left-continuous piecewise constant functions, we conclude. \( \Box \)

Multidimensional Riemann-Stieltjes can be considered. Given an additive expression defined on any semi-open \( d \)-dimensional intervals \([a_1, b_1] \times \cdots \times [a_d, b_d] \) via a “monotone” function, e.g., \( (v_1(b_1) - v_1(a_1)) \cdots (v_d(b_d) - v_d(a_d)) \), the construction can be developed. Essentially, this is the construction of the Lebesgue-Stieltjes measure from a additive (actually, \( \sigma \)-additivity is necessary) defined on the semi-ring of semi-open \( d \)-dimensional intervals. Moreover, most of the previous definitions can be used when either the integrand \( f \) or the integrator \( g \) takes values in a Banach space, however, this issue is not discussed further.

**Exercise 5.9** (cag-lad modulo of continuity). Suggested by the arguments in Proposition 5.4, a modulo of continuity for a cag-lad function \( f: [a, b] \to \mathbb{R} \) (i.e., \( f(t+) = \lim_{s \to t, s > t} f(s) \) exists and is finite for every \( t \) in \([a, b[ \), \( f(t-) = \lim_{s \to t, s < t} f(s) \) exists and is finite, and \( f(t) = f(t-) \) for every \( t \) in \([a, b]) \) should be defined as

\[
\rho(r; f, [a, b]) = \inf \left\{ \max_{i=1, \ldots, n} \operatorname{osc}(f, [s_{i-1}, s_i]) : a = s_0 < s_1 < \cdots < s_n = b, \; s_i - s_{i-1} > r \right\}.
\]

Verify that a function \( f \) is cag-lad if and only if \( \rho(r) \to 0 \) as \( r \to 0 \). Moreover, a more convenient modulo of continuity could be the following

\[
\rho'(r; f, [a, b]) = \sup \left\{ |f(t') - f(t)| \wedge |f(t'') - f(t)| : t' \leq t \leq t'', t'' - t' \leq r, t', t'' \in [a, b] \right\},
\]

with \( \wedge \) denoting the min. Prove that \( \rho'(r; f, [a, b]) \leq \rho(r; f, [a, b]) \), but the converse inequality does not hold. Finally, state an analogous for result for cad-lag functions. See Billingsley [14, Chapter 3, p. 109-153] and Amann and Escher [2, Chapter VI, Theorem 1.2, pp. 6-7]. \( \Box \)

Any monotone function \( v \) has finite-valued left-hand and right-hand limits at any point, which implies that only jump-discontinuities (i.e., so-called of the first class) may exist. A jump occurs at \( s \) in \([a, b[ \) when \( 0 < |v(s+) - v(s-)| \leq |v(b) - v(a)| \), and therefore, only a finite number of jumps larger that a positive constant may appear within a bounded interval. Hence, there are only a countable number of jumps and the series \( v_j(I) = \sum_{s \in I} [v(s+) - v(s-)] \), collecting all jumps within the interval \( I \), is convergent, and \( |v_j([a, b])| \leq |v(b) - v(a)| \) when \( v \) is nondecreasing. Thus, the function \( v_c(t) = v(t) - v_j([a, t]) \) with \( t \)
in \([a,b]\) can be rendered continuous on \([a,b]\), and if \(v\) is continuous either from left or from right at each point \(t\) then \(v_c\) is continuous in \([a,b]\).

If \(\{t_i\}\) is a sequence of positive numbers and \(\sum_i a_i\) is a convergence series of positive terms then the expression \(a(t) = \sum_i a_i \mathbb{1}_{t_i \leq t}\) yields an nondecreasing function called a nondecreasing purely jump function. If \(\varphi\) is a cag-lad integrand then

\[
\int_{[0,T]} \varphi(s)da(s) = \sum_i (\varphi(t_i^+) - \varphi(t_i)) a_i \mathbb{1}_{t_i \leq T}, \quad \forall T > 0.
\]

Note that the integrand and the integrator are defined on the whole semi-line \([0,\infty)\), with the convention that \(f(0^-) = f(0)\) for a left-continuous function \(f\) and also \(g(0^-) = 0\) a right-continuous function \(g\). This convention is not used if the functions are assumed to be defined on the whole real line \(\mathbb{R}\).

**A Particular Change of Variable**

Let \(a\) be a right-continuous nondecreasing function, \(a: [0, \infty] \to [0, \infty]\), i.e., \(a(s) = \lim_{r \to s, r > s} a(r)\) for every \(s\) in \([0, \infty]\) and \(a(s) \leq a(s')\) for every \(s \leq s'\) in \([0, \infty]\). Define

\[
a^{-1}(t) = \inf \{ s \in [0, \infty] : a(s) > t \}, \quad \forall t \in [0, \infty], \quad (5.5)
\]

with the convention that \(\inf \{\emptyset\} = \infty\). The function \(a^{-1}: [0, \infty] \to [0, \infty]\) is called the right-inverse of \(a\). Note that \(a^{-1}(t) = 0\) for any \(t\) in \([0, a(0)]\), and it may be convenient to define

\[
r^*(a) = \sup \{ a(s) : s \in [0, \infty[, a(s) < \infty \} \quad (5.6)
\]

with the convention \(\sup \{\emptyset\} = 0\), so that the function \(a\) maps \([0, \infty]\) into \([a(0), r^*(a)]\). Therefore, the expression \(a^{-1}(t) = \inf \{ s \in [0, \infty] : a(s) > t \}\) is properly defined for every \(t\) in \([0, r^*(a)]\), while \(a^{-1}(t) = \infty\) for every \(t\) in \([r^*(a), \infty]\). It is also clear that the mapping \(a \mapsto a^{-1}\) is monotone decreasing, i.e., if \(a_1(s) \leq a_2(s)\) for every \(s \geq 0\) then \(a_1^{-1}(t) \geq a_2^{-1}(t)\) for every \(t \geq 0\), in particular, \(a(s) \leq s\) (or \(\geq\)) for every \(s \geq 0\) implies \(a^{-1}(t) \geq t\) (or \(\leq\)) for every \(t \geq 0\).

For instance, if \(a\) has only one jump, i.e., \(a(s) = c_0\) for any \(s < s_0\) and \(a(s) = c_1\) for any \(s \geq s_0\), with some constants \(c_0 \leq c_1\) and \(s_0\) in \([0, \infty[\), then \(a^{-1}(t) = 0\) for every \(t < c_0\), \(a^{-1}(t) = s_0\) for every \(t\) in \([c_0, c_1[\) and \(a^{-1}(t) = \infty\) for every \(t \geq c_1\).

Consider \(a_-\) and \(a_-^{-1}\) the left-continuous functions obtained from \(a\) and \(a^{-1}\), i.e., \(a_-(0) = 0\), \(a_-^{-1}(0) = 0\), \(a_- (s) = \lim_{r \to s, r < s} a(r)\) and \(a_-^{-1}(t) = \lim_{t \to t^{-}, t < t} a^{-1}(r)\) for every \(s\) and \(t\) in \([0, \infty]\). The functions \(a, a_-, a^{-1}\) and \(a_-^{-1}\) can have only a countable number of discontinuities.

**Proposition 5.5.** Firstly, (1) if \(a\) is a right-continuous non-decreasing function then the right-inverse \(a^{-1}\) is right-continuous non-decreasing function. Also, for
every $t$ in $[0, \infty]$, we have $a^{-1}(t) < +\infty$ if and only if $t < r^*(a)$. Secondly, (2) alternative definition of the right-inverse $a^{-1}$ is the following expression

$$a^{-1}(t) = \sup \{ s \in [0, \infty[ : a(s) \leq t \}, \quad \forall t \in [0, +\infty), \quad (5.7)$$

with the convention $\sup \{ \emptyset \} = 0$, i.e., $a^{-1}(t) = 0$, for any $t$ in $[0, a(0)[$. Thirdly, (3) the relation between $a$ and $a^{-1}$ is symmetric, namely, $a$ is the right-inverse of $a^{-1}$, i.e., $a(s) = \inf \{ t \in [0, +\infty] : a^{-1}(t) > s \}$, for every $s$ in $[0, \infty]$, and

$$a_-(a^{-1}(t)) \leq a_-(a^{-1}(t)) \leq t \leq a(a^{-1}(t)) \leq a(a^{-1}(t)), \quad \text{for every } t \in [0, r^*(a)[,$n

e.g., if $a$ is continuous at the point $s = a^{-1}(t)$ with $t$ in $[0, r^*(a)]$ then $a(s) = a(a^{-1}(t)) = t$. Fourthly, (4) for any nonnegative Borel measurable function $f$ the Lebesgue-Stieltjes integral can be calculated as

$$\int_{[0, +\infty[} f(s) \, da(s) = \int_{[0, +\infty[} f(c(t)) \, 1_{c(t)<\infty} \, dt, \quad (5.8)$$

where $c = a^{-1}$ except in a set of Lebesgue measure zero.

Proof. Since $t \leq t'$ implies $\{ s : a(s) > t \} \supset \{ s : a(s) > t' \}$ the function $a^{-1}$ is a non-decreasing function. To show that $a^{-1}$ is right-continuous at any $t \geq 0$ take $\varepsilon > 0$ and, by definition, there exists $s \geq 0$ with $a(s) > t$ and $s < a^{-1}(t) + \varepsilon$. If $t_n > t$ and $t_n \to t$ then there exists an index $N$ such that $a(s) > t_n \geq t$ for every $n \geq N$, which implies that $a^{-1}(t_n) \leq s < a^{-1}(t) + \varepsilon$, i.e., $\lim_{n} a^{-1}(t_n) \leq a^{-1}(t)$.

Moreover, from the definition, $a^{-1}(t) < +\infty$ if and only if $t < a(+)\infty$, for every $t$ in $[0, \infty]$. To check the alternative definition for the right-inverse of $a$, fix $t$ and, if $\varepsilon > 0$ and temporarily $b(t)$ stand for the alternative above sup expression, then $a(a^{-1}(t) - \varepsilon) \leq t$ implies $b(t) \geq a^{-1}(t) - \varepsilon$, i.e., $b(t) \geq a^{-1}(t)$. The argument is completed by mentioning that because $a$ is monotone non-decreasing, any $s$ satisfying $a(s) \leq t$ has to be smaller than or equal to any $s'$ satisfying $a(s') > t$, i.e., $b(t) \leq a^{-1}(t)$.

These alternative definitions show that

(a) $s < a^{-1}(t)$ implies $a(s) \leq t$, and that $s > a^{-1}(t)$ implies $a(s) > t$,

as well as their symmetric counterparts, $a(s) > t$ implies $s \geq a^{-1}(t)$, and that $a(s) \leq t$ implies $s \leq a^{-1}(t)$, for any $t, s$ in $[0, +\infty]$. Now, use the right-continuity of $a$ to obtain that

(b) $s \geq a^{-1}(t)$ implies $a(s) \geq t$.

Thus, if temporarily $b = (a^{-1})^{-1}$ denotes the iteration of infimum (5.5) on the function $a^{-1}$, then (a) yields $a(s) \leq b(s)$, and (b) yields $a(s) \geq b(s)$. Hence the relation between $a$ and $a^{-1}$ is symmetric, i.e., $a$ is the right-inverse of $a^{-1}$. Moreover, take a sequence $\{ s_n \}$ such that $s_n < a^{-1}(t)$ and $s_n \to a^{-1}(t)$ to deduce from (a) that $a(s_n) \leq t$, i.e., $a_-(a^{-1}(t)) \leq t$. Similarly, take another a sequence $\{ t_n \}$ with $t_n < t$, $t_n \to t$, $s_n = \varepsilon + a^{-1}(t_n)$ and $\varepsilon > 0$ to obtain from
(b) \(a(s_n) \geq t_n\), i.e., \(a-(a^{-1}(t) + \epsilon) \geq t\), which implies \(a(a^{-1}(t)) \geq t\). Since \(a^{-1} \leq a^{-1}\), the parts (1), (2) and (3) are proved.

To show (4) assume that \(c = a^{-1}\). Take first \(f = 1_{\{0\}}\), i.e., \(f(c(t)) = 1\) if and only if \(c(t) = 0\). From the infimum definition of \(a^{-1}(t)\) follows that \(c(t) = 0\) if \(0 \leq t < a(0)\) and the continuity of \(a\) at \(s = 0\) implies that \(c(t) > 0\) if \(t > a(0)\). This proves the validity of the equality (5.8) when \(f = 1_{\{0\}}\) with integral equals to \((0+) = 0\) if \(0 < c\). Also, by definition of the infimum \(a^{-1}(t)\), if \(t < a(\tau)\) then \(\tau \geq a^{-1}(t)\), i.e., (b) \(f(c(t)) = 1\) if \(t < a(\tau)\). Thus, from (a) and (b), the equality (5.8) is valid for \(f = 1_{[0,\tau]}\), where the integral is equal to \(a(\tau) - a(0)\). Hence, by linearity, the equality remain true for any left-continuous piecewise constant functions \(f\).

Because both sides of the equality are measures, again the equality holds true for any functions \(f = 1_B\), with \(B\) a Borel set, and therefore for any nonnegative Borel function \(f\).

\[\square\]

- **Remark 5.6.** Given a right-continuous nondecreasing function \(a\) form \([0, +\infty[\) into itself define \(a_* = \sup\{s \geq 0 : a(s) = a(0)\}\) and \(a^* = \sup\{a(s) : s > 0\}\). The function \(a\) maps \([0, \infty[\) into \([a(0), a^*]\) and its right-inverse \(a^{-1}\), given by either the infimum (5.5) or the supremum (5.7), maps \([a(0), a^*]\) into \([a_*, +\infty[\), with \(a_* = a^{-1}(a(0))\) and \(a^* = r^*(a)\), and also maps \([0, a(0)]\) into \(\{0\}\). Therefore, equality (5.8) can be rewritten as

\[\int_{[0,T]} f(s) \, da(s) = \int_{a(0)}^{a(T)} f(a^{-1}(t)) \, dt,\]

or in the more familiar form

\[\int_{[0,T]} f(a(s)) \, da(s) = \int_{a(0)}^{a(T)} f(s) \, ds,\]

for every \(0 < T < \infty\).

Let us remark that the expressions

\[a^{-1}(t) = \inf\{s \in [0, +\infty] : a(s) \geq t\}, \quad \forall t \in [0, \infty],\]

again with the convention \(\inf\{\emptyset\} = \infty\), i.e., \(a^{-1}(t) = +\infty\), for any \(t\) in \([r^*(a), \infty]\), as well as

\[a^{-1}(t) = \sup\{s \in [0, +\infty[ : a(s) < t\}, \quad \forall t \in ]0, \infty[,\]

(5.9)

again with the convention \(\sup\{\emptyset\} = 0\), i.e., \(a^{-1}(t) = 0\), for any \(t\) in \([0, a(0)]\), are alternative definitions for \(a^{-1}\). Indeed, an argument similar to the above shows that the inf and sup expressions agree, and essentially the same technique used to prove that \(a^{-1}\) is right-continuous can be applied to deduce that \(a^{-1}\) is left-continuous. Note that the left-continuity yields no change in the value
of \( a_-(t) \) if \( a \) is replaced by \( a_- \) in (5.9), i.e., if the initial monotone function \( a \) is assumed to be left-continuous (instead of right-continuous) then (5.9) could be used to define the right-inverse of a left-continuous monotone function, even other approaches are possible, e.g., see Leoni [75, Chapter 1].

**Exercise 5.10** (change-of-time). Let \( \nu \) be a \( \sigma \)-finite measure on the measurable space \((X, \mathcal{F})\), and \( c \) be a nonnegative real-valued Borel measurable function on \( X \times [0, \infty) \) and, for every fixed \( x \) in \( X \), define

\[
\tau^{-1}(x, s) = \int_0^s c(x, r) dr \quad \text{and} \quad \tau(x, t) = \inf\{s > 0 : \tau^{-1}(x, s) > t\},
\]

and \( \tau(x, t) = \infty \) if \( \tau^{-1}(x, s) \leq t \) for every \( s \geq 0 \). Assume that \( \tau^{-1}(x, s) \) is finite for every \((x, s) \) in \( X \times [0, \infty) \), and that \( \tau^{-1}(x, s) \to \infty \) as \( s \to \infty \). Consider \( \tau^{-1} \) and \( \tau \) as functions from \( X \times [0, \infty) \) into \( [0, \infty) \) and verify (1) that both functions are Borel measurable, and also cad-lag increasing functions in the second variable. Now, consider the product measure \( \mu = \nu(dx)dt \) on the product space \( X \times [0, \infty) \) and apply the transformation \((x, t) \mapsto (x, \tau(t))\) to obtain the following set function

\[
\mu_\tau(A \times [0, s]) = \mu\{(x, t) : x \in A, 0 < \tau(x, t) \leq s\}.
\]

Verify (2) that \( \mu_\tau \) extends to a unique measure defined on \( \mathcal{F} \times \mathcal{B} \), where \( \mathcal{B} \) is the Borel \( \sigma \)-algebra on \([0, \infty)\). Prove (3) that \( \mu_\tau = c(x, t)\nu(dx)dt \), i.e.,

\[
\mu_\tau(A \times [0, t]) = \int_{A \times [0, t]} c(x, r) \nu(dx)dr,
\]

for every measurable set \( A \) and any \( t \geq 0 \).

## Multidimensional RS-Integral

Instead of insisting in left- or right-continuous functions as one-dimensional problem, consider a finitely additive set function \( \mu = \mu_g \) defined on the semi-ring \( \mathcal{I}_d \) of semi-open (or semi-closed) \( d \)-dimensional intervals \([a, b]\) in \( \mathbb{R}^d \), which is given through a \( d \)-monotone function \( g : \mathbb{R}^d \to \mathbb{R} \), i.e., \( \mu_g([a, b]) = \Delta_{a_1}^{b_1} \cdots \Delta_{a_d}^{b_d} g \), with

\[
\Delta_{a_i}^{b_i} g(x_1, \ldots, x_i, \ldots, x_d) = g(x_1, \ldots, b_i, \ldots, x_d) - g(x_1, \ldots, a_i, \ldots, x_d).
\]

Note that \( d \)-monotone means \( \Delta_{a_i}^{b_i} g(x_1, \ldots, x_i, \ldots, x_d) \geq 0 \), for every \( b_i > a_i \), \( x = (x_1, \ldots, x_i, \ldots, x_d) \), and that \( \mu_g \) generates a (Lebesgue-Stieltjes) measure \( \mu_g^* \), but the function \( g \) needs to be right-continuous by coordinate to deduce that \( \mu_g^* = \mu_g \) on \( \mathcal{I}_d \). Moreover, the simple example in mind is the product form \( f(x_1, \ldots, x_d) = f_1(x_1) \cdots f_d(x_d) \).

To define the multidimensional Riemann-Stieltjes integrable, repeat the arguments of previous Section 5.1 on Riemann integrable, the only difference is
that partitions are with \(d\)-intervals in \(I_d\), and so are the definition of the infima \(\alpha_f(I_k)\) and suprema \(\beta_f(I_k)\). The mesh (or norm) of a partition \(\{I_k\}\) is again \(\max_k\{d(I_k)\}\) (\(d(\cdot)\) is the diameter of \(I_k\)), but another quantity intervene, namely, \(\max_k\{\mu_g(I_k)\}\). Therefore, the upper and the lower multidimensional Riemann-Stieltjes integrable are defined (as the smallest bound of the upper Darboux sums and the largest bound of the lower Darboux sums), which is a common (finite) value (the integral) when \(f\) is RS-integrable with respect to \(g\).

Also, the multidimensional Riemann-Stieltjes sums are defined, i.e., if \(I = ]a,b[\) is partitioned with a finite sequence \(\{I_k\}\) of disjoint \(d\)-intervals in \(I_k \in I_d\) then

\[
\Sigma(f, g, \{I_k\}, \{y_k\}) = \sum_k f(y_k)\mu_g(I_k), \quad y_k \in I_k,
\]

\[
\underline{\Sigma}(f, g, \{I_k\}) = \sum_k \alpha_f(I_k)\mu_g(I_k), \quad \alpha_f(I_k) = \inf_{y \in I_k} f(y),
\]

\[
\overline{\Sigma}(f, g, \{I_k\}) = \sum_k \beta_f(I_k)\mu_g(I_k), \quad \beta_f(I_k) = \sup_{y \in I_k} f(y).
\]

Remark that the \(d\)-monotonicity of \(f\) is necessary for the definition of the lower \(\underline{\Sigma}(f, g, \{I_k\})\) and upper \(\overline{\Sigma}(f, g, \{I_k\})\) Darboux sums, but the definition of the Riemann-Stieltjes sums \(\Sigma(f, g, \{I_k\}, \{y_k\})\) does not actually need \(f\) to be \(d\)-monotone. In both cases, \(f\) should be bounded and \(g\) should satisfy \(\sup_k\{\mu_g(I_k)\} \leq C < \infty\), for any partition \(\{I_k\}\) of \(I = ]a,b[\).

Always under the assumptions of \(f, g\) are bounded and \(g\) is \(d\)-monotone, if \(\{J_n\}\) is a refinement of the partition \(\{I_k\}\) of the \(d\)-interval \(]a,b[\) (i.e., each \(I_k\) is equal to a disjoint union of some \(J_n\)) then

\[
\Sigma(f, g, \{I_k\}) \leq \Sigma(f, g, \{J_k\}) \leq \overline{\Sigma}(f, g, \{I_k\}) \leq \underline{\Sigma}(f, g, \{I_k\}),
\]

and

\[
\max \{\overline{\Sigma}(f, g, \{I_k\}) - \underline{\Sigma}(f, g, \{J_k\}), \underline{\Sigma}(f, g, \{J_k\}) - \overline{\Sigma}(f, g, \{I_k\})\} \leq \max_{x \in ]a,b[}\{|f(x)|\} \max_k \{\mu_g(I_k)\}.
\]

The first estimate is sufficient to define the Riemann-Stieltjes integrability of the integrand \(f\) with respect to the integrator \(g\) (in short \(f\) is RS-integrable wrt \(g\) on \(]a,b[\)) when the upper and the lower Darboux sums coincide, and the common value \(\overline{\Sigma}(f, g, \{I_k\}) = \underline{\Sigma}(f, g, \{I_k\})\) be taken as the value of the integral. Next, the second estimate and the inequality

\[
\Sigma(f, g, \{I_k\}) \leq \Sigma(f, g, \{I_k\}, \{y_k\}) \leq \overline{\Sigma}(f, g, \{I_k\}),
\]

are used to show that both Darboux sums and the Riemann-Stieltjes sums converges to the Riemann-Stieltjes integral as the the mesh (or norm) of the partition \(\max_k\{d(I_k)\}\) and the quantity \(\max_k\{\mu_g(I_k)\}\), both vanish.

A more detailed analysis is necessary and estimate

\[
\Sigma(f, g, \{I_k\}) - \overline{\Sigma}(f, g, \{I_k\}) \leq \max_k \{|\beta_f(I_k) - \alpha_f(I_k)|\} \mu_g([a,b]),
\]
is used to obtain a characterization of Riemann-Stieltjes integrable functions similar to Theorem 5.1, i.e., denoting by \( D \) the set of discontinuities on \( [a, b] \), (a) if \( \mu_g^*(D) = 0 \) then \( f \) is Riemann-Stieltjes integrable and \( \mu_g^* \)-integrable, and both integrals coincide, (b) if \( f \) is Riemann-Stieltjes integrable and \( g \) is continuous (i.e., the measure \( \mu_g \) is diffuse, namely, each single point in \( [a, b] \) has zero \( \mu_g^* \)-measure) then \( \mu_g^*(D) = 0 \). On the other hand, if \( \max_k \{ \mu_g(I_k) \} \to 0 \) as \( \max_k \{ d(I_k) \} \to 0 \) (i.e., as seen in a later chapter, this is the case of an absolutely continuous function \( g \)) then everything work as in the case of the multidimensional Riemann integral. In fact, the RS-integral of \( f \) wrt \( g \) is the Riemann integral of \( fg' \), where \( g' \) is the Radon-Nikodym derivative (see Theorem 6.3 later) of the measure \( \mu_g^* \) with respect to the Lebesgue measure.

### 5.3 Diadic Riemann Integrals

In a simple way, integration in \( \mathbb{R} \) (or \( \mathbb{R}^d \)) can be regarded as a series. Let \( \mathbb{R}_n = \{ i2^{-n} : i + 1, \ldots, 4^n \} \) be the sets of (positive) \( n \)-dyadic numbers (or of order \( n \)) and let \( \mathbb{R} = \bigcup_n \mathbb{R}_n \) be all (positive) dyadic numbers (or points). Using a sequence dyadic intervals \( [(i - 1)2^{-n}, i2^{-n}] \) with \( i = 1, 2, \ldots, 4^n \) and \( n \geq 1 \), we have \( [0, 2^n] = \sum_{i=1}^{4^n} [(i - 1)2^{-n}, i2^{-n}] \). Moreover, if

\[
\begin{align*}
    d_n(t) &= \max\{i2^{-n} \leq t : i = 1, \ldots, 4^n\}, \quad \text{(lower dyadic function)} \\
    \overline{d}_n(t) &= \min\{i2^{-n} \geq t : i = 1, \ldots, 4^n\}, \quad \text{(upper dyadic function)}
\end{align*}
\]

then \( d_n(t) \leq t \leq \overline{d}_n(t) \), \( \overline{d}_n(t) - d_n(t) < 2^{-n} \), for every \( t \) in \( [0, 4^n] \), and \( \overline{d}_n(t) = d_n(t) \) when \( t \) belongs to \( \mathbb{R}_n \).

**Proposition 5.7.** First, with the above notation we have

\[
t = \sum_n 4^{-n} \sum_{i=1}^{4^n} 1_{i2^{-n} \leq t}, \quad \forall t \in [0, \infty[.
\]

Moreover

\[
\int_0^T \varphi(t) dt = \sum_n 4^{-n} \sum_{i=1}^{4^n} \varphi(i2^{-n}) 1_{i2^{-n} \leq T}, \quad \forall T > 0,
\]

for any Riemann integrable function \( \varphi \) on \( [0, T] \). Furthermore, if \( \ell \) denotes the Lebesgue measure on \( (0, \infty) \) then

\[
\int_0^T \psi(t) \ell(dt) = \sum_n 4^{-n} \sum_{i=1}^{4^n} \ell(\{ s \in (0, T) : \psi(s) > i2^{-n} \}), \quad \forall T > 0,
\]

for any nonnegative measurable function \( \psi \) on \( (0, \infty) \).
Proof. To show (5.11), if \( t = k2^{-m} = (k2^{-m})2^{-n}, 1 \leq k \leq 4^m \) then \( k2^{-m} \leq 4^n \), \( \mathbb{1}_{i2^{-n} \leq t} = 1 \) if and only if \( i = 1, \ldots, k2^{-m} \) with \( k2^{-m} \geq 1 \). This implies \( \sum_{i=1}^{4^n} \mathbb{1}_{i2^{-n} \leq t} = k2^{-m} = t2^n \) when \( k2^{-m} = t2^n \geq 1 \) and \( \sum_{i=1}^{4^n} \mathbb{1}_{i2^{-n} \leq t} = 0 \) when \( k2^{-m} = t2^n < 1 \), i.e., for every \( t \in \mathbb{R}_m \) we have \( \sum_{i=1}^{4^n} \mathbb{1}_{i2^{-n} \leq t} = t2^n \) if \( t2^n \geq 1 \) and equal to zero otherwise. Hence, for any \( t > 0 \), because \( d_m(t) \leq t \leq m(t) \) with the previous notation (5.10), we deduce

\[
d_m(t) = \sum_n 2^{-n} d_m(t) \leq \sum_n 4^{-n} \sum_{i=1}^{4^n} \mathbb{1}_{i2^{-n} \leq t} \leq \sum_n 2^{-n} \bar{d}_m(t) = \bar{d}_m(t),
\]

which yields (5.11), after letting \( m \to \infty \). Note that replacing \( \mathbb{1}_{i2^{-n} \leq t} \) with the strict \( < \) in \( \mathbb{1}_{i2^{-n} < t} \) changes nothing only when \( t \) is not a dyadic point (number).

Moreover, if \( t \geq s > 0 \) then

\[
2^n (d_m(t) - \bar{d}_m(s)) \leq \sum_{i=1}^{4^n} \mathbb{1}_{s < i2^{-n} \leq t} \leq 2^n (\bar{d}_m(t) - d_m(s)),
\]

which implies that the equality (5.12) holds true for any piecewise constant function \( \varphi(t) = a_i \) for any \( t \) in \([t_{i-1}, t_i[\) with \( t_0 < t_1 < \cdots < t_k \) which is left-continuous at any dyadic point.

In general, if \( \varphi \) is a bounded integrable function on any compact set \([0, T] \) then define \( \underline{\varphi}_m(t) = c_i \) and \( \overline{\varphi}_m(t) = C_i \) for \( t \) in \( I_{i,m} = ](i-1)2^{-m}, i2^{-m} [ \) with \( c_i = \sup\{ \varphi(s) : s \in I_{i,m} \} \) and \( C_i = \sup\{ \varphi(s) : s \in I_{i,m} \} \) to check that

\[
\int_0^T \varphi_m(t)dt \leq \sum_n 4^{-n} \sum_{i=1}^{4^n} \varphi(i2^{-n}) \mathbb{1}_{i2^{-n} \leq T} \leq \int_0^T \overline{\varphi}_m(t)dt.
\]

Hence, if \( \varphi \) is Riemann integrable on \([0, T] \) then the validity of (5.12) is verified.

Finally, if \( \psi = \mathbb{1}_B \) for a Borel set in \((0, \infty) \) then

\[
\int_0^T \psi(t)dt = \int_0^\infty rm_{\psi}(dr) = \int_0^\infty \lambda_{\psi}(r)dr,
\]

where the distribution functions of \( \psi \) are \( m_{\psi}(A) = \ell(\{t \in (0, T) : \psi(t) \in A \}) \) and \( \lambda_{\psi}(r) = \ell(\{t \in (0, T) : \psi(t) > r \}) \). This equality remains true for any \( \psi \) nonnegative measurable function on \((0, \infty) \), see also Exercise 5.6. Hence, take \( \varphi = \lambda_{\psi}(\cdot) \) in (5.12), initially with \( \psi \) bounded, to deduce (5.13).\( \square \)

The contrast between both ideas of integration are really seen in the previous Proposition 5.7. The Riemann integral uses a partition on the domain, while the Lebesgue integral uses a partition on the range, which requires first, the notion of measure.

The expression (5.12) extends easily to vector-valued functions \( \varphi \) (or functions with values into a Banach space) and the series is convergent in the corresponding norm. Moreover, with a heavier notation, namely, for \( d \)-dimensional

[Initial Page]
nonnegative integers \( j = (j_1, \ldots, j_d) \),
\[
\int_{(a,b]} f(x) dx = \sum_{n} 4^{-dn} \sum_{1 \leq j_i \leq 4^n} f(j2^{-n}) 1_{a < j2^{-n} \leq b},
\]
for every \( d \)-dimensional interval \( (a,b] \subset \mathbb{R}^d \), \( a = (a_1, \ldots, a_d) \), \( b = (b_1, \ldots, b_d) \), with \( a \leq b \) and \( a_1, \ldots, a_d \geq 0 \). While, the expression in (5.13) can be easily adapted to \( \mathbb{R}^d \), i.e.,
\[
\int_{\mathbb{R}^d} f^\pm(x) \ell_d(dx) = \sum_{n} 4^{-n} \sum_{i=1}^{4^n} \ell_d(\{x \in \mathbb{R}^d : f^\pm(x) > i2^{-n}\}),
\]
and actually, the Lebesgue measure \( \ell_d \) on \( \mathbb{R}^d \) can be replaced by a measure \( \mu \) on an abstract space \( \Omega \).

As a direct consequence of (5.8) in the previous Proposition 5.5 and the last Proposition 5.7, if \( t \mapsto \varphi(c(t)) \) is Riemann integrable then
\[
\int_{[0,T]} \varphi(t) da(t) = \sum_{n} 4^{-n} \sum_{i=1}^{4^n} \varphi(c(i2^{-n})) 1_{i2^{-n} \leq a(T)}, \tag{5.14}
\]
for every \( T > 0 \).

The dyadic grid can be used to define the so-called Jordan content. Indeed, the countable collection \( Q_n \) of cubes in \( \mathbb{R}^d \) whose side length is \( 2^{-n} \) and whose vertices are in \( 2^{-n}\{\ldots, -1, 0, 1, \ldots\} \) is called the \( n \)-dyadic cubes, i.e., a typical cube \( Q \) in \( Q_n \) have the form
\[
Q = [i_12^{-n}, (i_1 + 1)2^{-n}] \times \cdots \times [i_d2^{-n}, (i_d + 1)2^{-n}],
\]
for any integer numbers \( i_1, \ldots, i_d \). These cubes may be taken closed, open or semi-closed (i.e., closed on the left and open in the right).

If \( A \) is a bounded set in \( \mathbb{R}^d \) then the inner and the outer Jordan content of \( A \) is defined as follows:
\[
\kappa(A) = \lim_{n \to \infty} \kappa_n(A) \quad \text{and} \quad \tilde{\kappa}(A) = \lim_{n \to \infty} \tilde{\kappa}_n(A), \tag{5.15}
\]
where \( \kappa_n(A) \) or \( \tilde{\kappa}_n(A) \) are the series of the volume of all \( n \)-dyadic cubes inside or touching \( A \), i.e,
\[
\kappa_n(A) = \sum_{Q \in Q_n, Q \subset A} 2^{-dn} \quad \text{and} \quad \tilde{\kappa}_n(A) = \sum_{Q \in Q_n, Q \cap A \neq \emptyset} 2^{-dn}.
\]

If the limits \( \kappa(A) = \tilde{\kappa}(A) \) then the common number \( \kappa(A) \) is called the Jordan contain of \( A \). Note that the number \( \kappa_n(A) \) increases and the \( \tilde{\kappa}_n(A) \) decreases with \( n \), i.e., the limits (5.15) are monotone and finite (because \( A \) is bounded).

Certainly, the Jordan content is usually defined using general \( d \)-intervals or rectangles, but the result is the same. Even if the definition make sense for any
set $A$, the Jordan content is meaningful when $A$ is bounded since $\tilde{\kappa}(A) = \infty$ for every unbounded set $A$. It is clear that the Caratheodory’s arguments and the inner approach give suitable generalization.

It is clear that if

$$A_n = \bigcup_{Q \in \mathcal{Q}_n, Q \subseteq A} Q \quad \text{and} \quad \tilde{A}_n = \bigcup_{Q \in \mathcal{Q}_n, Q \cap A \neq \emptyset} Q,$$

then $\kappa_n(A)$ and $\tilde{\kappa}_n(A)$ are the Euclidean volumes (or the Lebesgue measure) of the sets $A_n$ and $\tilde{A}_n$, respectively. Actually,

**Proposition 5.8.** If $A$ is a bounded set then

$$\bigcup_n A_n = A \subset A \subset \tilde{A} = \bigcap_n \tilde{A}_n,$$

(5.16)

$A$ and $\tilde{A}$ are Borel sets, and $\ell(A) = \kappa(A)$ and $\ell(\tilde{A}) = \tilde{\kappa}(A)$, where $\ell$ denotes the Lebesgue measure on $\mathbb{R}^d$. Moreover, if $A = O$ is an open set then $O = O$ and $O$ can be expressed as a countable union of closed dyadic cubes with disjoint interiors (or semi-closed dyadic disjoint cubes). Furthermore, $\kappa(O) = \ell(O)$ and $\tilde{\kappa}(K) = \ell(K)$ for any open set $O$ and compact set $K$ in $\mathbb{R}^d$.

**Proof.** The relation (5.16) is clear and the first part follow from the monotone convergence. Moreover, this imply that the Jordan content of $A$ exists if and only if $\ell(\tilde{A} \setminus A) = 0$, which implies that $A$ is measurable and $\kappa(A) = \ell(A)$.

It is also clear that $\kappa(O) = \ell(O)$, for any bounded open set $O$. Now, given a compact $K$, there is a large compact cube $C$ with interior $\hat{C} \supset K$. Certainly, the content of the boundary of $C$ is zero, and since, any $n$-dyadic cube inside $C$ is either inside $C \setminus K$ or is intersecting $K$, we deduce that $\tilde{\kappa}_n(K) + \kappa_n(C \setminus K) = \ell(C)$, and as $n \to \infty$, the equality

$$\tilde{\kappa}(K) + \kappa(C \setminus K) = \ell(C),$$

follows. Because the boundary of the cube $C$ has content zero, we get

$$\kappa(C \setminus K) = \kappa(\hat{C} \setminus K) = \ell(C \setminus K),$$

i.e., $\tilde{\kappa}(K) = \ell(K)$. \qed

Therefore, we find again the two-step approximation process given in either Caratheodory’s arguments (outer-inner) and the inner approach (inner-outer).

### 5.4 Lebesgue Measure on Manifolds

First we recall the concept of manifold. If $U$ and $V$ are two open sets in $\mathbb{R}^d$ then a bijective mapping $\Phi : U \to V$ which is continuously differentiable up to the order $k$ together with its inverse $\Phi^{-1} : V \to U$ is called a homeomorphism of class $C^k$ (or a $C^k$ diffeomorphism). If $k = 0$ then $\Phi$ and its inverse are just
continuous, and a (locally) Lipschitz homeomorphism (or a (local) bi-Lipschitz mapping) is when $\Phi$ and $\Phi^{-1}$ are both (local) Lipschitz continuous functions, i.e., for some constant $C \geq c > 0$,

$$c|x - y| \leq |\Phi(x) - \Phi(y)| \leq C|x - y|, \quad \forall x, y \in U$$

or if the ‘locally’ prefix is used, for any $x$ and $y$ in $K$, for any compact set $K \subset U$, where the constants $C$ and $c$ may depend on $K$. Also the case $k = \infty$ (i.e., continuously differentiable of any order) is included. In this context, a homeomorphism is also called a (local) change-of-variables or coordinates. The interested reader may take a look at Taylor [115, Appendix B, pp. 267–275] for quick review on diffeomorphisms and manifolds.

In $\mathbb{R}^3$, a curve (or surface) is regarded as the graph of 1-variable (or 2-variable) mapping with values in $\mathbb{R}^3$. Certainly, several mappings (or parameterizations) can produce the same curve (or surface), and the geometric concept of curves (or surfaces) is independent of the given parameterization. In general, the graph of a mapping $\psi$ is the set $\{(x, \psi(x))\}$, and the name manifold is used to signify a local graph. Submanifolds preserve the structure of manifolds (they are themselves manifolds) and our interest is only on (sub)manifolds embedded in the Euclidean space $\mathbb{R}^d$. Indeed,

**Definition 5.9.** In the Euclidean space $\mathbb{R}^d$, a set $S \subset \mathbb{R}^d$ is called a $C^k$ submanifold at $x \in S$ of dimension $1 \leq m < d$ if there exists an open neighborhood $U$ of $x$ such that $S \cap U$ is the graph of a mapping $\psi$ of class $C^k$ from an open set $V \subset \mathbb{R}^m$ into $\mathbb{R}^{d-m}$, i.e., for some orthogonal change-of-variables $y = (y_1, \ldots, y_d)$, $y' = (y_1, \ldots, y_m)$,

$$S \cap U = \{(y', \psi(y')) \in \mathbb{R}^d : y' \in V \subset \mathbb{R}^m\},$$

and $\psi$ is continuously differentiable up to the order $k$. If this property holds for every $x$ in $S$, with the same constants $k$ and $m$, but possibly a different choice of the orthogonal coordinates (and mapping $\psi$), then $S$ is called a $C^k$ submanifold of dimension $m$. The $m$-dimensional linear space of all tangent vectors, i.e., the graph of the $(d - m) \times m$ matrix gradient $\nabla\psi$, namely,

$$\text{graph}(\nabla\psi(x')) = \{(y', \nabla\psi(x')y') : y' \in \mathbb{R}^m\}, \quad \text{with } x = (x', \psi(x')),$$

is called the tangent space at the point $x$. The strictly positive function

$$y = (y', \psi(y')) \mapsto J_\psi(y') = \sqrt{\det(\nabla\phi(y')^*\nabla\phi(y'))}, \quad \phi(y') = (y', \psi(y'))$$

defined on $S \cap U$ is called the Euclidean $m$-dimensional density function, where $(\cdot)^*$ means the transposed matrix, and $\det(\cdot)$ is the determinant of a $m \times m$ matrix. With obvious changes, continuous submanifolds ($k = 0$), $C^\infty$ submanifolds, and (locally) Lipschitz submanifolds ($\psi$ is locally Lipschitz) are also defined. For (locally) Lipschitz submanifolds, the tangent space and the Euclidean density may not be defined at every points.

Similarly, any open subset of $\mathbb{R}^d$ and any point in $\mathbb{R}^d$ can be regarded as submanifolds of dimension $m = d$ and $m = 0$, respectively. As mentioned above, submanifolds (or manifolds) are considered as embedded in Euclidean space $\mathbb{R}^d$. Therefore, instead of calling $S$ a $C^k$ submanifold (of $\mathbb{R}^d$) of dimension $m$, we may call $S$ a manifold of dimension $m$ (in $\mathbb{R}^d$).

A typical example of a $C^\infty$ manifold is the sphere $S^{d-1} = \{x \in \mathbb{R}^d : |x| = 1\}$. Indeed, for any $x^0$ in $S^{d-1}$ there is at least one coordinate nonzero, e.g., $x_1^0 > 0$, and so,

$$x_1 = \psi(x_2, \ldots, x_d) = \sqrt{1 - x_2^2 - \cdots - x_d^2}$$

yields a local description. It is clear from the definition that a homeomorphism $\Phi$ of class $C^k$ preserves submanifolds, i.e., if $S$ is a manifold then $\Phi(S)$ is also a manifold.

**Remark 5.10.** The mapping $\phi(y') = (y', \psi(y'))$ from $V$ into $\mathbb{R}^d$ is injective and of class $C^k$, and its inverse $(y', \psi(y')) \rightarrow y'$ is necessarily continuous (actually, of class $C^k$) and the $d \times m$ matrix gradient $\nabla \phi = (I_m, \nabla \psi)^*$ is injective. Similarly, the mapping $g(y) = g(y', r) = r - \psi(y')$ from $U$ into $\mathbb{R}^{d-m}$ satisfies $S \cap U = \{y \in U : g(y) = 0\}$ and the $(d - m) \times d$ matrix gradient $\nabla g = (-\nabla \psi | I_{d-m})$ is surjective, indeed, the number $d - m$ of equation requires for a local description of $S$ is called the co-dimension. Moreover, these $d - m$ coordinates can be flattened, i.e., $\Phi(S \cap U) = O \times 0_{d-m}$, for a suitable homeomorphism $\Phi$ from $U$ into $\mathbb{R}^d$ and an open subset $O$ of $\mathbb{R}^m$. Actually, as long as we work within the class $C^k$ with $k \geq 1$, these three functions $\phi$, $g$ and $\Phi$ are of class $C^k$ and they provide an equivalent definition of submanifold, via the implicit and the inverse function theorems (which are not valid for Lipschitz functions). For instance, a set $S$ is a $C^k$ submanifold of $\mathbb{R}^d$ at $x \in S$ of dimension $m$ if there exists an open set $V$ of $\mathbb{R}^m$ and an injective function $\phi : V \rightarrow \mathbb{R}^d$ such that (a) $x$ belongs to $\phi(V)$, (b) $\phi$ and its inverse $\phi^{-1}$ are of class $C^k$, $k \geq 1$, and (c) the matrix $\nabla \phi(x)$ has rank $m$. Indeed, from such a function $\phi$ the implicit function theorem applies, and after re-ordering the variables, the equation $\phi(y') = z$ can be solved, locally, as $z = (y', \psi(y'))$ to fit Definition 5.9. A function $\phi$ satisfying (a), (b), (c) is called a local chart of $S$, and a family such functions is called an atlas of $S$. Essentially, any property of an object acting on a manifold is defined in term of an atlas and should be independent of the particular atlas used. It is clear that atlas are preserved by homeomorphism of the same regularity.

Also note that the tangent space and the Euclidean density are independent of the particular local coordinates (i.e., the choice of the $m$ independent coordinates and the function $\psi$) chosen. Setting $\phi(y') = (y', \psi(y'))$, this means that if $\tilde{\phi}(\tilde{y}')$ is another local coordinates (or charts) on an open subset $V$ of $\mathbb{R}^m$ then the tangent space at the point $x = \phi(x') = \tilde{\phi}(\tilde{x}')$ is given by

$$\{\nabla \phi(x')y' \in \mathbb{R}^d : y' \in \mathbb{R}^m\} = \{\nabla \tilde{\phi}(\tilde{x}')\tilde{y}' \in \mathbb{R}^d : \tilde{y}' \in \mathbb{R}^m\},$$

while, for the Euclidean density $J_\psi(y')$ the invariance is expressed by the relation

$$J_{\tilde{\phi}}(\tilde{y}') = J_{\phi}(\phi^{-1} \circ \tilde{\phi}(\tilde{y}')) | \det (\nabla (\phi^{-1} \circ \tilde{\phi}(\tilde{y}'))) |,$$
for any $\tilde{y}'$ in $\bar{\phi}^{-1}(\phi(V) \cap \bar{\phi}(V))$. Actually, any nonnegative function $\rho$ defined on $S$ by local coordinates $\rho_\phi(y') = \rho(\phi(y'))$ that follows the above invariance is called a density on $S$.

In particular, if $m = d - 1$ (i.e., the hyper-area) then $\psi$ is real-valued, $\phi(y') = (y', \psi(y'))$, $y'$ in $\mathbb{R}^{d-1}$, and

$$J_\psi(y') = \sqrt{\det (\nabla \phi(y')^* \nabla \phi(y'))} = \sqrt{1 + |\nabla \psi(y')|^2},$$

where $\nabla \psi$ is the gradient of $\psi$, i.e., the $(d - 1)$-dimensional vector of all partial derivatives. This means that if $y' = (y_1, \ldots, y_{d-1})$ and $y_d = \psi(y_1, \ldots, y_{d-1})$ then the vector

$$n(y', \psi(y')) = \pm \frac{(- \partial_1 \psi(y'), \ldots, - \partial_{d-1} \psi(y'), 1)}{\left[1 + (\partial_1 \psi(y'))^2 + \cdots + (\partial_{d-1} \psi(y'))^2\right]^{1/2}},$$

represents the unit normal vector (field) to the surface $S$. This yields $d - 1$ independent tangential unit vectors $t_i$, for $i = 1, \ldots, d - 1$, i.e.,

$$t_1 = \frac{(1,0,\ldots,0, \partial_1 \phi(y'))}{\left[1 + (\partial_1 \phi(y'))^2\right]^{1/2}}, \ldots, t_{d-1} = \frac{(0,0,\ldots,1, \partial_1 \phi(y'))}{\left[1 + (\partial_{d-1} \phi(y'))^2\right]^{1/2}},$$

which are orthogonal to $n$ as expected.

- **Remark 5.11.** The concept of manifold applied to an open set $\Omega \subset \mathbb{R}^d$ with boundary $\partial \Omega$ could reads as follows: either (a) the boundary $\partial \Omega = S \subset \mathbb{R}^d$ is a $(d - 1)$-dimensional manifold satisfying Definition 5.9 and

$$\Omega \cap U = \{(y', y_d) \in \mathbb{R}^d : y_d < \psi(y'), \ y' \in V \subset \mathbb{R}^{d-1}\},$$

or (b) the closure $\bar{\Omega}$ is a $d$-dimensional manifold with boundary $\partial \Omega = S$, i.e., as in Definition 5.9 with $\phi : U \rightarrow \mathbb{R}^d$,

$$\Omega \cap U = \{y = (y', y_d) \in \mathbb{R}^d : \phi_d(y) < 0\} \quad \text{and} \quad S \cap U = \{y = (y', y_d) \in \mathbb{R}^d : \phi_d(y) = 0\}.$$

In this case, the normal direction $n$ is one-sided, i.e., the “graph” cannot traverses the tangent plane. As mentioned early, both viewpoints (a) and (b) are equivalent within the class $C^k$, $k \geq 1$, but for only continuous or Lipschitz manifolds, (a) implies (b), but (b) does not necessarily implies (a). For instance, the reader is referred to Grisvard [56, Section 1.2, pp. 4–14].

Similarly, if $m = 1$ (i.e., the arc-length) then $\psi$ takes values in $\mathbb{R}^{d-1}$, $\phi(y') = (y', \psi(y'))$, $y'$ in $\mathbb{R}^1$, and

$$J_\psi(y') = \sqrt{\det (\nabla \phi(y')^* \nabla \phi(y'))} = \sqrt{1 + |d \psi(y')|^2},$$
where \( d\psi(y') \) is the \((d - 1)\)-dimensional vector of the first derivative of \( \psi \). This means that if \( y' = y_1, \psi = (\psi_2, \ldots, \psi_d) \) and \( \psi' \) denotes the derivative, then the vector
\[
\mathbf{t}(y_1, \psi(y_1)) = \pm \frac{(1, \psi_2'(y_1), \ldots, \psi_d'(y_1))}{\left[1 + (\psi_2'(y_1))^2 + \cdots + (\psi_d'(y_1))^2\right]^{1/2}},
\]
represents the unit tangent vector (field) to the curve \( S \). This means that for \( d = 3 \), we have the arc-length with \( m = 1 \) and the area with \( m = 2 \), as expected.

To patch all the pieces of a submanifold we need a partition of the unity:

**Theorem 5.12** (continuous PoU). Let \( \{O_\alpha : \alpha\} \) be an open cover of \( S \subset \mathbb{R}^d \), i.e., \( O_\alpha \) are open sets and \( \bigcup_\alpha O_\alpha \supset S \). Then there exists a continuous partition of the unity subordinate to \( \{O_\alpha : \alpha\} \), i.e., there exists a sequence of continuous functions \( \chi_i : \mathbb{R}^d \to [0, 1], \ i = 1, 2, \ldots \), such that the support of each function \( \chi_i \) is a compact set contained in some element \( O_\alpha \) of the cover, and for any compact set \( K \) of \( \bigcup_\alpha O_\alpha \) there exists a finite number \( k \) such that \( \sum_{i=1}^k \chi_i = 1 \) on \( K \).

**Proof.** First, if \( I \) and \( J \) are \( d \)-intervals (or \( d \)-rectangles) in \( \mathbb{R}^d \) such that \( I \) is compact, \( J \) is open with compact closure and \( I \subset J \) then there exists a continuous function \( \varpi : \mathbb{R}^d \to [0, 1] \) satisfying \( \varpi(x) = 1 \) for every \( x \) in \( I \) and \( \varpi(x) = 0 \) for every \( x \) outside of \( J \), actually, an explicit construction of the \( \varpi \) is clearly available.

Second, consider a sequence of compact set \( K_n \) such that \( \bigcup_n K_n = \bigcup_\alpha O_\alpha \) to deduce that for every \( x \) in \( K_n \) must belong to some open set \( O_\alpha \), and so, there are \( d \)-intervals \( I \) compact and \( J \) open with compact closure such that \( x \) belongs to the interior of \( I \) and \( I \subset J \subset J \subset O_\alpha \). By compactness, there exists a finite number of \( I_i \subset J_i \) with the above property which form a finite cover of \( K_n \), i.e., \( \bigcup_{i=1}^k I_i \supset K_n \). Hence, there is a sequence of \( d \)-intervals \( I_i \subset J_i, I_i \) is compact, \( J_i \) is open with a compact closure contained in some \( \Omega_\alpha \), and such that \( \bigcup_{i=1}^k I_i = \bigcup_\alpha O_\alpha \).

Next, denote by \( \varpi_i \) a continuous function such that \( \varpi = 1 \) on \( I_i \) and \( \varpi = 0 \) outside \( J_i \) to define \( \chi_1 = \varpi_1 \) and
\[
\chi_i = (1 - \varpi_1)(1 - \varpi_2) \cdots (1 - \varpi_{i-1}) \varpi_i,
\]
for \( i \geq 2 \). The support of \( \chi_i \) is certainly contained in some \( \Omega_\alpha \) and since
\[
\sum_{i=1}^k \chi_i = 1 - \prod_{i=1}^k (1 - \varpi_i), \quad \forall k \geq 1,
\]
we deduce that \( \sum_{i=1}^\infty \chi_i = 1 \) on any compact \( K \) of \( \bigcup_\alpha O_\alpha \), where the series is locally finite, i.e., only a finite number of \( \chi_i \) have support in \( K \). \( \square \)

It is not hard to modify the argument so that the functions \( \chi_i \) are of class \( C^k \), but to actually see that \( \chi_i \) may be chosen of class \( C^{\infty} \), we make use of the
fact that
\[
g(x) = \begin{cases} 
0 & \text{if } x \leq 0 \\
e^{-1/x} & \text{if } x > 0
\end{cases}
\]
is a function of class $C^\infty$. Actually, the reader may check Theorem 7.8 later on in the text.

Therefore, apply Theorem 5.12 to the open cover $\{U\}$ of the submanifold $S$ as in Definition 5.9 to find a continuous partition of the unity $\{\chi_i\}$ with a compact support contained in the open set $U_i \subset \mathbb{R}^d$, and charts $\psi_i$ defined on an open set $V_i \subset \mathbb{R}^m$ such that
\[
S \cap U_i = \{\phi_i(y') = (y',\psi_i(y')) \in \mathbb{R}^d : y' \in \mathbb{R}^m\}.
\]

Now, the Lebesgue measure defined on $\mathbb{R}^m$ can be transported to $S$. Indeed, a nonnegative function $f$ defined on a submanifold $S$ in $\mathbb{R}^d$ of dimension $m$ is called integrable with respect to the surface Lebesgue measure $\sigma(dx)$ if the function
\[
y' \mapsto \sum_i \chi_i(\phi_i(y')) f(\phi_i(y')) \sqrt{\det (\nabla \phi_i(y')^* \nabla \phi_i(y'))}
\]
is Lebesgue integrable on $\mathbb{R}^m$ and
\[
\int_S f(x) \sigma(dx) = \sum_i \int_{\mathbb{R}^m} \chi_i(\phi_i(y')) f(\phi_i(y')) \sqrt{\det (\nabla \phi_i(y')^* \nabla \phi_i(y'))} \, dy'
\]
is the definition of the integral. These definition are independent of the particular partition of the unity and the charts chosen. Indeed, if $\phi$ and $\bar{\phi}$ have a common image $S_0$ then the formula for the change-of-variables $y' \mapsto \bar{\phi}^{-1}(\phi(y'))$ in the $m$-dimensional integral yields
\[
\int_{\phi^{-1}(S_0)} f(\phi_i(y')) \sqrt{\det (\nabla \phi_i(y')^* \nabla \phi_i(y'))} \, dy' = \\
= \int_{\bar{\phi}^{-1}(S_0)} f(\bar{\phi}_i(y')) \sqrt{\det (\nabla \bar{\phi}_i(y')^* \nabla \bar{\phi}_i(y'))} \, dy',
\]
as expected. In particular, any linear (affine) submanifold $S$ in $\mathbb{R}^d$ of dimension $m$ can be represented as
\[
S = \{(y',\psi(y')) \in \mathbb{R}^d : y' \in \mathbb{R}^m\}, \text{ with } \psi(y') = a + y'A,
\]
where $A$ is a $m \times (d-m)$ matrix $A$ of maximal rank and $a$ is a row vector in $\mathbb{R}^{d-m}$. Hence, $\phi(y') = (y',a + y'A)$, $\nabla \phi(y') = (I_m,A)^*$, and $\det((\nabla \phi(y'))^* \nabla \phi(y')) = \det(I_m + AA^*)$, independent of $y'$, and it represents the $m$-volume of the image $\{y'A \in \mathbb{R}^d : y' \in Q' \subset \mathbb{R}^m\}$, where $Q'$ is the unit cube in $\mathbb{R}^m$. In general
\[
\sigma(\phi(Q)) = \int_Q \sqrt{\det (\nabla \phi(y')^* \nabla \phi(y'))} \, dy',
\]
for any cube $Q \subset \mathbb{R}^m$ inside the open set $D_\phi$ where the local chart $\phi$ is defined. Actually, the above equality holds true for any Lebesgue measurable set $A = Q \subset D_\phi$ in $\mathbb{R}^m$ and $\sigma$ becomes a Borel measure on $S \subset \mathbb{R}^d$, and except for a multiplicative constant, this surface Lebesgue measure agrees with the $m$-dimensional Hausdorff measure as discussed in next section.

For the case of the hyper-area ($m = d - 1$), if for instance, the local charts are taken

$$\phi(y_1, \ldots, y_{d-1}) = (y_1, \ldots, y_{d-1}, \psi(y_1, \ldots, y_{d-1}))$$

then the surface Lebesgue measure has locally the form

$$\int_{\mathbb{R}^{d-1}} f(y', \psi(y')) \sqrt{1 + |\nabla \psi(y')|^2} \, dy', \quad y' = (y_1, \ldots, y_{d-1}).$$

If the submanifold $S$ is only (locally) Lipschitz then the Euclidean density is defined almost everywhere in $\mathbb{R}^m$, and the surface Lebesgue measure $\sigma$ still makes sense as a Borel measure on $S$.

In particular, if $\Omega$ is an open subset of $\mathbb{R}^d$ with a Lipschitz boundary $\partial \Omega$ (see Remark 5.11) then the surface Lebesgue measure $d\sigma$ is can be used to define the space $L^1(\partial \Omega)$, i.e., the space of functions $f : \Omega \to \mathbb{R}$ such that the composition function

$$y' \mapsto \sum_i \chi_i(\phi_i(y')) f(\phi_i(y')) \sqrt{\det (\nabla \phi_i(y')^* \nabla \phi_i(y'))}$$

is integrable in $\mathbb{R}^{d-1}$, for some local coordinates $\psi_i : V_i \to \mathbb{R}^d$, $\phi_i(y') = (y' \psi(y'))$, and a subordinate partition of the unity $\{\chi_i\}$. As mentioned early, all properties of function defined on the boundary $\partial \Omega$ are studied by local coordinates. Moreover, if $\Omega$ a bounded domain as above and $F$ is a continuously differentiable functions defined on the closure $\overline{\Omega}$ with values in $\mathbb{R}^d$ then the divergent theorem, i.e.,

$$\int_{\Omega} \nabla F \, d\ell = \int_{\partial \Omega} F \cdot \mathbf{n} \, d\sigma$$

holds true, where $\mathbf{n}$ is the outward unit normal vector defined almost everywhere with respect to the surface Lebesgue measure $\sigma$. Similarly, with the integration-by-parts or Green formula.

Recall that by definition complex-valued measures are finite measure, i.e., a complex measure $\mu$ has a real-part $\Re(\mu)$ and an imaginary-part $\Im(\mu)$ both of which are finite real-valued measures on a measurable space $(\Omega, \mathcal{F})$. Thus, following the complex numbers arithmetic, a real- or complex-valued measurable function $f$ is integrable with respect to a complex valued measure $\mu$ if and only if the real-valued function $|f|$ is integrable with respect to the real-valued measures $\Re(\mu)$ and $\Im(\mu)$.

In general, every integral with complex values is reduced to its real and imaginary parts, and then each one is studied separately and put back together.
when the result make sense, i.e., both parts are finite and 5h3 complex plane is identified with \( \mathbb{R}^2 \) for all practical use. Hence, of particular interest is the integral over a complex Lipschitz curve, which treated as a generalization of the Riemann-Stieltjes contour integral over a complex \( C^1 \)-curve, i.e., the complex line integral

\[
\int_C f(z)\,dz = \int_a^b f(x(t) + iy(t))(a'(t) + iy'(t))\,dt,
\]

where the curve \( C \) is parameterized as \( z = x(t) + iy(t) \), with \( t \) from \( a \) to \( b \) and Lipschitz functions \( t \mapsto x(t) \) and \( t \mapsto y(t) \).


### 5.5 Hausdorff Measure

Recalling that \( d(E) \) means the diameter of \( E \), i.e., \( d(E) = \sup \{d(x, y) : x, y \in E \} \), the second construction of outer measure, given in Section 2.4, is used with the choice \( E = 2^X \), \( b(E) = (d(E))^s = d^s(E) \), with \( 0 \leq s < \infty \) and the conventions \( d(\emptyset)^s = 0 \) and \( d^0(E) = 1 \) if \( E \neq \emptyset \), yields the so-called \( s \)-dimensional Hausdorff measure, which is denoted by \( h_s \) (most of the times \( H^s \)) and \( h_s(A) = \lim_{\delta \to 0} h_{s,\delta}(A) \), with

\[
h_{s,\delta}(A) = \inf \left\{ \sum_n d^s(E_n) : A \subset \bigcup_n E_n, \ E_n \subset X, \ d(E_n) \leq \delta \right\}.
\]

In view of Theorem 2.25, \( h_s \) is a regular Borel (outer) measure and as \( \delta \) decreases, the infimum extends over smaller classes, so that \( h_{s,\delta}(A) \) does not decrease, i.e., \( h_{s,\delta}(A) \uparrow h_s(A) \), and therefore, if \( h_s(A) = 0 \) then \( h_{s,\delta}(A) = 0 \) for any \( \delta > 0 \).

Note that, we have no distinction in notation between the outer measure \( h_s^* \) and the corresponding measure \( h_s \), i.e., we drop the star * to simplify. Actually, from the context, we use the outer measure on non measurable sets. The cases of an integer \( s \) are of great importance. For instance, \( h^0(A) \) is equal to the number of points in \( A \) and when \( X = \mathbb{R} \) we easily show that \( h_1 = \ell_1^* \), the Lebesgue outer measure on the line.

Note that for any \( A \subset \mathbb{R}^d \), \( a \in \mathbb{R}^d \) and \( r > 0 \) we have \( h_s(A + a) = h_s(A) \) (invariance under translations) and \( h_s(rA) = r^s h_s(A) \) (\( s \)-homogeneous under dilations). Moreover, if \( \vartheta : \mathbb{R}^d \to \mathbb{R}^d \) is an affine isometry then \( h_s(\vartheta(A)) = h_s(A) \). In a separable metric space \( X \), the measure \( h_s \) is unchanged if we use the class (for the coverings) of closed sets (or open sets) instead of \( 2^X \). Moreover,
for a normed space $X$, we may also use the class of all convex sets. Indeed, it suffices to remark that $d(\overline{A}) = d(A)$, $d(A_\varepsilon) = d(A) + 2\varepsilon$ and $d(\text{co}(A)) = d(A)$, where $\overline{A}$ is the closure of $A$, $A_\varepsilon = \{x \in X : d(x,A) < \varepsilon\}$ and $\text{co}(A) = \{y = tx + (1-t)y : x,y \in A, t \in [0,1]\}$.

**Exercise 5.11.** Consider the 1-dimensional Hausdorff measure $h_1$ on $\mathbb{R}^1$ and $\mathbb{R}^2$. Verify that $h_{1,\delta}(A)$ is independent of $\delta$ and is equal to the Lebesgue measure $\ell_1(A)$, for any $A \subset \mathbb{R}^1$. Now, show that if $S_a = \{(x,a) : x \in [0,b]\} \subset \mathbb{R}^2$ is an isometric copy of the interval $[0,b]$ then $h_1(S_a) = b$, for every $b > 0$. Since the $S_a$ are disjoint, we should deduce that $h_1$ is not a $\sigma$-finite measure in $\mathbb{R}^2$. Next, consider $S_0$ and $S_a$ with $a < \delta < 1$ and discuss the role of the parameter $\delta$ in the above definition. Check that $h_{1,\delta}(S_0 \cup S_a) = 2b$ if $\delta < a$, but the diameter of $S_0 \cup S_a$ is $\sqrt{b^2 + a^2}$.

**Proposition 5.13.** Consider the $s$-dimensional Hausdorff (outer) measure $h_s$ on $X$ and let $A$ be a subset of $X$. We have (a) if $h_s(A) < \infty$ then $h_t(A) = 0$ for $t > s$ or equivalently (b) if $h_s(A) > \infty$ then $h_t(A) = \infty$ for $t < s$.

**Proof.** Indeed, let $A = \bigcup_n E_n$ with $d(E_n) < \delta$. If $t > s$ then

$$h_{t,\delta}(A) \leq \sum_n d^t(E_n) \leq \delta^{t-s} \sum_n d^s(E_n),$$

which yields $h_{t,\delta}(A) \leq \delta^{t-s} h_{s,\delta}(A)$. Hence, as $\delta$ vanishes, we deduce $h_{t,\delta}(A) = 0$ if $h_{s,\delta}(A) < \infty$.

Based on this result, we define

$$\dim(A) = \sup\{s \geq 0 : h_s(A) = \infty\} = \inf\{s \geq 0 : h_s(A) = 0\}$$

as Hausdorff dimension of a subset $A$ of $X$. Certainly this unique value $0 \leq \dim(A) \leq \infty$ is such that $s < \dim(A)$ implies $h_s(A) = \infty$ and $t > \dim(A)$ implies $h_t(A) = 0$. As expected, we may prove (see below) that $\dim(\mathbb{R}^d) = d$, and we give example of sets with non-integer Hausdorff dimension.

**Exercise 5.12.** Regarding the Hausdorff dimension prove for any Borel sets (1) if $A \subset B$ then $\dim(A) \leq \dim(B)$ and (2) $\dim(A \cup B) = \max\{\dim(A), \dim(B)\}$

Moreover, prove that (3) for a sequence $\{A_k\}$ of Borel sets we have $\dim(\bigcup_k A_k) = \sup_k \{\dim(A_k)\}$.

Let us give a couple of sequential comments before comparing the Hausdorff measure $h_d$ and the Lebesgue measure $\ell_d$ in $\mathbb{R}^d$.

(1) For $s > d$ we have $h_s(\mathbb{R}^d) = 0$. Indeed, the unit cube $Q^1$ in $\mathbb{R}^d$ can be decomposed into $k^d$ cubes of side $1/k$ and diameter $\delta = \sqrt{d}/k$. Hence

$$h_{s,\delta}(Q^1) \leq \sum_{i=1}^{k^d} \delta^s = d^s/2 k^{d-s},$$

which tends to zero as $k \to \infty$ if $d < s$, i.e., $h_s(Q^1) = 0$, and then $h_s(\mathbb{R}^d) = 0$. [Preliminary]
(2) Let \( Q \) be the collection of all cubes in \( \mathbb{R}^d \) with edges parallel to the axes, precisely, of the form \( a_i \leq x_i < a_i + \alpha \), with \( \alpha > 0 \) and \( \alpha = (a_1, \ldots, a_d) \) in \( \mathbb{R}^d \). Recall that \( h_s(A) = h_s(A, 2^d) \) and consider \( \tilde{h}_s(A) = h_s(A, Q) \). We have

\[
\begin{align*}
    h_s(A) & \leq \tilde{h}_s(A) \leq (2\sqrt{d})^s h_s(A), \quad \forall A \subset \mathbb{R}^d. \quad (5.17)
\end{align*}
\]

Indeed, because any set of diameter strictly less than \( r \) is contained in a cube with edges of size \( 2r \), i.e., with diameter \( 2r\sqrt{d} \); for each cover \( A \subset \bigcup_n E_n \) with \( d(E_k) < \delta \) we can find cubes \( Q_n \supset E_n \), with \( d(Q_n) = 2\sqrt{d}d(E_n) \). Hence

\[
\sum_n d^s(E_n) = (2\sqrt{d})^{-s} \sum_n d^s(Q_n) \geq (2\sqrt{d})^{-s} \tilde{h}_{s,\tilde{\delta}}(A),
\]

with \( \tilde{\delta} = 2\sqrt{d} \delta \). Hence \( h_{s,\delta}(A) \geq (2\sqrt{d})^{-s} \tilde{h}_{s,\tilde{\delta}}(A) \), which yields the inequality \( h_s(A) \geq (2\sqrt{d})^{-s} h_s(A) \).

(3) Similarly, if \( \tilde{h}_s(A) = h_s(A, B) \), where \( B \) is the collection of all balls in \( \mathbb{R}^d \), then with the same argument (remarking that any set \( A \) is contained in a ball \( B \) of radius equal to the diameter of \( A \)) we obtain

\[
\begin{align*}
    h_s(A) & \leq \tilde{h}_s(A) \leq 2^s h_s(A), \quad \forall A \subset \mathbb{R}^d. \quad (5.18)
\end{align*}
\]

Thus, the requiring \( \delta \to 0 \) in Theorem 2.25 forces the covering to incorporate into the definition of the Hausdorff measure \( h_s(A) \) some information about the local geometry of the set \( A \).

(4) If \( \ell_d \) denotes the Lebesgue measure in \( \mathbb{R}^d \) then a subset of \( \mathbb{R}^d \) has Lebesgue (outer) measure zero if and only if it has \( d \)-dimensional Hausdorff measure zero, moreover we have

\[
\begin{align*}
    d^{-d/2}h_d(A) & \leq \ell_d(A) \leq 2^d h_d(A), \quad \forall A \subset \mathbb{R}^d. \quad (5.19)
\end{align*}
\]

Indeed, for any cube \( Q \), the quantity \( \delta(Q)^d \) is proportional to the volume (measure) of \( Q \), i.e., \( \delta^d(Q) = d^{d/2}\ell_d(Q) \). Thus, due to the additivity of the measure we see that \( \tilde{h}_{s,\delta} \) is independent of \( \delta \), i.e.,

\[
\tilde{h}_s(A) = \inf \left\{ \sum_n d^s(Q_n) : \bigcup_n Q_n \supset A \right\}, \quad \text{with} \quad s = d, \quad \forall A \subset \mathbb{R}^d,
\]

where \( \{Q_n\} \) is any collection of cube covering \( A \), without restriction on the size of the diameters. Hence, essentially by definition of the outer Lebesgue measure \( \ell_d \) (see Remark 2.39), we deduce \( \tilde{h}_d(A) = d^{d/2}\ell_d(A) \), for every \( A \subset \mathbb{R}^d \). Combining this with the inequality (5.17), we conclude. Note that in particular, estimate (5.19) and Proposition 5.13 imply that for any subset \( A \) of \( \mathbb{R}^d \) with positive outer Lebesgue measure \( \ell_d^*(A) > 0 \) we have \( h_s(A) = \infty \) for any \( s < d \), and certainly, \( h_t(A) \leq h_t(R^d) = 0 \) for any \( t > d \). This shows that Hausdorff dimension of \( A \) is \( d \), i.e., \( \dim(A) = d \).
(5) One step further is to realize that for every cube $Q$ we have $\mathcal{H}_d(a + rQ) = r^d \mathcal{H}_d(Q)$, $\mathcal{H}_d(a + rQ) = r^d \mathcal{H}_d(Q)$ and $\ell_d(a + rQ) = r^d \ell_d(Q)$. Hence, for every cube $Q$ we deduce $\mathcal{H}_d(Q) = q_d \ell_d(Q)$, where $q_d = \mathcal{H}_d(Q^1)$ with $Q^1 = [0, 1]^d$ being the unit cube. Next, the equality $\mathcal{H}_d = q_d \ell_d$ remains valid for any $d$-interval $[a_1, b_1] \times \cdots \times [a_d, b_d]$ written as a countable disjoint union of semi-open cubes, and finally, Proposition 2.15 shows the equality of both measures, i.e., $\mathcal{H}_d = q_d \ell_d$ on the Borel $\sigma$-algebra. To calculate the constant $q_d$ we need either the isodiametric inequality (2.8) or to know that $\mathcal{H}_d(B^1) = 2^d$, with $B^1$ the unit ball in $\mathbb{R}^d$.

(6) An alternative argument is to know that if $m$ is a translation-invariant Borel measure on $\mathbb{R}^d$ then $m = c \ell_d$, where $c = m(Q^1)$, with $Q^1$ the unit cube, e.g., see Dshalalow [36, Theorem 3.10, pp. 266-269].

(7) It is now clear that if the class $\mathcal{E}$ is other than $2^X$ then the Hausdorff measure obtained may be not the same (often, a multiplicative constant is involved). However, it is interesting to remark that if the class $\mathcal{E}$ is chosen to be (a) all closed sets or (b) all open sets or (c) all the convex sets of $X$ then the same Hausdorff measure is obtained. Indeed, (a) and (c) follow from the fact that the closure and the convex hull operations preserves the diameter of any set, while (b) follows by arguing that for any $\varepsilon > 0$ the set $\{x : d(x, E) < \varepsilon\}$ is open and has diameter at most $d(E) + 2\varepsilon$, see Mattila [80, Chapter 4, pp. 54-74].

Now, under the assumption of the isodiametric inequality (2.8), we have

Proposition 5.14. If $X = \mathbb{R}^d$ then $\ell_d = c_d \mathcal{H}_d$, where $\ell_d$ is the (outer) Lebesgue measure in $\mathbb{R}^d$ and $c_d = 2^{-d} \pi^{d/2} / \Gamma(d/2 + 1)$.

Proof. If $\{E_n\}$ is a cover of $A$ then the sub $\sigma$-additivity of the Lebesgue outer measure ensures that $\ell_d(A) \leq \sum_n \ell_d(E_n)$. Thus, assuming the validity of the isodiametric inequality (2.8), we have $\ell_d(E_n) \leq c_d (d(E_n))^d$, which implies that $\ell_d(A) \leq c_d \mathcal{H}_d(A)$, for every $A \subset \mathbb{R}^d$.

For the converse inequality, we recall that $\ell_d(A) = 0$ implies $\mathcal{H}_d(A) = 0$ and that

$$\ell_d(A) = \inf \{ \sum_n \ell_d(Q_n) : \bigcup_n Q_n \supset A, \ d(Q_n) \leq \delta \}, \ \forall A \subset \mathbb{R}^d,$$

for every $\delta > 0$ and cubes with edges parallel to the axis. Hence, we know that given any $\delta, \varepsilon > 0$ and $A \subset \mathbb{R}^d$ with $\ell_d(A) < \infty$, there exists a sequence of cubes $\{Q_n\}$ such that

$$A = \bigcup_n Q_n, \quad d(Q_n) < \delta \quad \text{and} \quad \sum_n \ell_d(Q_n) \leq \ell_d(A) + \varepsilon.$$

In view of Corollary 2.35, for each $n$ there exists a disjoint sequence of balls $\{B_{n,k}\}$ contained in the interior of the cube $Q_n$ such that

$$d(B_{n,k}) < \delta, \quad \ell_d\left(Q_n \setminus \bigcup_k B_{n,k}\right) = 0, \ \forall n.$$
Therefore \( h_d(Q_n \setminus \bigcup_k B_{n,k}) = 0 \). Hence

\[
h_{d,\delta}(A) \leq \sum_n h_{d,\delta}(Q_n) = \sum_n h_{d,\delta}\left(\bigcup_k B_{n,k}\right) \leq \sum_{n,k} h_{d,\delta}(B_{n,k})
\]

which yields

\[
c_d h_{d,\delta}(A) \leq c_d \sum_{n,k} (d(B_{n,k}))^d = \sum_{n,k} \ell_d(B_{n,k}) = \sum_n \ell_d(Q_n),
\]

i.e., \( c_d h_{d,\delta}(A) \leq \ell_d(A) + \varepsilon \), and as \( \delta \) and \( \varepsilon \) vanish, we obtain \( c_d h_d(A) \leq \ell_d(A) \), for every \( A \subset \mathbb{R}^d \).

Note that \( c_d = 2^{-d} \ell_d(B) \), where \( \ell_d(B) \) is the hyper-volume of the unit ball in \( \mathbb{R}^d \), i.e., \( \ell_d(B) = \pi^{d/2}/\Gamma(d/2 + 1) \), with \( \Gamma \) being the Gamma function, e.g., \( c_1 = 1, c_2 = 4/\pi, c_3 = 6/\pi \) and \( c_4 = 32/\pi^2 \).

**Remark 5.15.** To simplify the above proof, we could use in the definition of the Hausdorff measure a class \( \mathcal{E} \) other than \( 2^{\mathbb{R}^d} \) (e.g., either the class of \( d \)-intervals of the form \( [a,b] \), or the class of cubes with edges parallel to the axis) where the isodiametric inequality (2.8) is easily verified. Moreover, if \( \mathcal{E} \) is the class of balls then the above proof is almost immediate, after recalling that the Lebesgue outer measure \( \ell_d^* \) in \( \mathbb{R}^d \) satisfies

\[
\ell_d^*(A) = \inf \left\{ \sum_{i=1}^{\infty} \ell_d(E_i) : A \subset \bigcup_{i=1}^{\infty} E_i, E_i \in \mathcal{E} \right\}, \quad \forall A \subset \mathbb{R}^d,
\]

see Remark 2.39. However, we should prove later that the change of the classes \( \mathcal{E} \) does not change the values of the outer measure \( h_{d,\delta} \), which fold back to the isodiametric inequality (2.8). A back door alternative is to define the diameter of a set as \( d(A) = \inf\{d(B) : A \subset B\} \), where \( B \) is any closed ball with diameter \( d(A) \), instead of the usual \( d(A) = \sup\{d(x,y) : x,y \in A\} \). In this context, the isodiametric inequality (2.8) is not necessary to prove Proposition 5.14, but the diameter is not longer what is expected, e.g., an equilateral triangle \( T \) in \( \mathbb{R}^2 \) cannot be covered with a ball of radius \( \sup\{|x-y|/2 : x,y \in T\} \).

**Exercise 5.13.** It is simple to establish that if \( T : \mathbb{R}^d \rightarrow \mathbb{R}^n \), with \( d \leq n \), is a linear map and \( T^* : \mathbb{R}^n \rightarrow \mathbb{R}^d \) is its adjoint then \( T^* T \) is a positive semi-definite linear operator on \( \mathbb{R}^n \), i.e., \( \langle T^* Tx, x \rangle \geq 0 \), for every \( x \in \mathbb{R}^d \) and with \( \langle \cdot, \cdot \rangle \) denoting the scalar product on the Euclidean space \( \mathbb{R}^d \). Now, use Theorem 2.27 and Proposition 5.14 to show that for every subset \( A \) of \( \mathbb{R}^d \) and any linear transformation \( T \) as above we have \( h_d(T(A)) = \sqrt{\det(T^* T)} h_d(A) \), where \( h_d \) is the \( d \)-dimensional Hausdorff measure considered either in \( \mathbb{R}^n \) or in \( \mathbb{R}^d \). See Folland [45, Proposition 11.21, pp. 352].

Actually, \( h_1 \) has the concrete interpretation of the length measure, e.g., the value \( h_1(\gamma) \) is the length of a rectifiable curve \( \gamma \) on \( \mathbb{R}^d \) and \( h_1(\gamma) = \infty \) if the curve \( \gamma \) is not rectifiable. In general, if \( M \) is a sufficiently regular \( n \)-dimensional
surface (e.g., $C^1$ submanifold) with $1 \leq n < d$ then the restriction of $h_n$ to $M$ is a multiple of the surface measure on $M$. Thus, often we normalize the Hausdorff measures so that $h_d$ agrees with the Lebesgue measure on $\mathbb{R}^d$. Indeed, it is not hard to show (e.g., see Folland [45, Theorem 11.25, pp. 353–354]) that if $f : V \to \mathbb{R}^d$ is a $C^1$ function on $V \subset \mathbb{R}^k$ such that the matrix $F(x) = (\partial_j f_i(x))$ has rank $k$ for every point $x$ then for every Borel subset $B$ of $V$ the image $f(B)$ is a Borel subset of $\mathbb{R}^d$ and

$$h_k(f(A)) = \int_A \sqrt{\det(F^*F)}dh_k,$$

i.e., the Hausdorff measure of the a $k$-dimensional $C^1$ submanifold of $\mathbb{R}^d$ can be computed in term of a Lebesgue integral in $\mathbb{R}^k$.

**Exercise 5.14.** Let $(X, d)$ be a metric space and $d_1$ be another metric equivalent to $d$, i.e., such that $ad \leq d_1 \leq bd$ for some positive constants $a$ and $b$. Prove that if $h_s$ and $h_s^1$ denote the Hausdorff measures corresponding to $d$ and $d_1$ then $a^s h_s(A) \leq h_s^1(A) \leq b^s h_s(A)$, for every subset $A$ of $X$. Briefly discuss the particular cases where $X = \mathbb{R}^n$ and either $d_1(x, y) = \max_i |x_i - x_i|$ or $d_1(x, y) = \sum_i |x_i - x_i|$. Can we easily compute $h_s^1(B^1)$, where $B^1$ is the unit ball in the $d_1$ metric? \[\square\]

To give an equivalent of Theorem 2.27 some notation is necessary. For a linear (or affine) transformation $\mathbb{R}^d$ into $\mathbb{R}^n$ we can use the following definition of Jacobian $J(T) = h_d(T(Q))/h_d(Q)$ or equivalently $J(T) = c_d h_d(T(Q))$, where $Q$ is the unit cube in $\mathbb{R}^d$ and $c_d$ the constant in Proposition 5.14. Note that $T(Q)$ is a parallelepiped in $\mathbb{R}^n$ and the $d$-dimensional Hausdorff measure $h_d$ can be used in either $\mathbb{R}^d$ (i.e., the Lebesgue measure) or $\mathbb{R}^n$. Certainly, if $T$ is not injective then $T(Q)$ is contained in a subspace of dimension strictly less than $d$, and thus, $h_d(T(Q)) = 0$, i.e., the Jacobian $J(T) = 0$. Alternatively, if $d = n$ then $J(T) = |\det(T)|$, and the invariance property of Theorem 2.27 can be applied. Actually, by means of the polar decomposition, the linear transformation $T = HS$ and $J(T) = J(S) = |\det(S)|$, where $S : \mathbb{R}^d \to \mathbb{R}^d$ is (linear) symmetric and $H : \mathbb{R}^d \to \mathbb{R}^n$ is (linear) orthogonal (or isometry), see also Exercise 5.13.

**Proposition 5.16** (invariance). Let $T$ be a linear transformation from $\mathbb{R}^d$ into $\mathbb{R}^n$, $d \leq n$. Then for every $A \subset \mathbb{R}^d$ we have $h_d(T(A)) = J(T)h_d(A)$, where $J(T)$ is the Jacobian of $T$, and $h_d$ is the Hausdorff outer measure considered either on $\mathbb{R}^n$ or on $\mathbb{R}^d$.

**Proof.** If $T$ is not injective then the Jacobian $J(T) = 0$ and $T(A)$ is contained in a subspace of dimension strictly less than $d$, which implies that $h_d(T(A)) = 0$ and the equality holds true.

If $T$ is injective then $T(\mathbb{R}^d)$ is a subspace of dimension $d$ in $\mathbb{R}^n$, the (direct) image preserves unions and intersections, and compact sets, which imply that the set function $A \mapsto h_d(T(A))$ defines a Borel measure $\mu$ on $\mathbb{R}^d$. Because $\mu$ and $\nu = J(T)h_d$ agree on the class of all cubes, we deduce that both measures
agree on the $\sigma$-algebra of Borel sets, and Remark 3.1 implies that both ($\mu$ and $\nu$) regular Borel outer measures are the same.

In view of Propositions 5.16 and 5.14, the Lebesgue measure $\ell_d$ on $\mathbb{R}^d$ could be (improperly) used as the Hausdorff/Lebesgue measure on $\mathbb{R}^n$ in the formula $\ell_d(T(A)) = J(T) \ell_d(A)$, for every $A \subset \mathbb{R}^d$. Actually, based on Exercise 5.11, it should be clear that in general, the Hausdorff measure $h_s$ on $\mathbb{R}^d$ is not $\sigma$-finite. However, because each $h_{s,\delta}$ is semi-finite and $h_s = \sup_{\delta > 0} h_{s,\delta}$, it is clear that the Hausdorff measure is semi-finite on $\mathbb{R}^d$ (see Exercise 2.15). Even more, if $K$ is a compact subset of $\mathbb{R}^d$ with $0 < h_s(K) \leq \infty$ then there exists another compact set $K_0 \subset K$ such that $0 < h_s(K_0) < \infty$, see Fededer [44, Section 2.10.47-48, pp. 204–206] and Rogers [96, Theorem 57, pp. 122].

Remark 5.17. Unless explicitly mentioned, it should be understood from the context that in all what follows the definition of the Hausdorff measure incorporate a coefficient $c_s = 2^{-s} \pi^{s/2}/\Gamma(s/2 + 1)$, i.e.,

$$h_{s,\delta}(A) = \inf \left\{ c_s \sum_n d^s(E_n) : A \subset \bigcup_n E_n, \ E_n \subset X, \ d(E_n) \leq \delta \right\}$$

and $h_s(A) = \lim_{\delta \to 0} h_{s,\delta}(A)$, to match the Lebesgue measure as suggested by Proposition 5.14. Note that the lower case $h$ (instead of the usual capital $H$) is used to denote the Hausdorff measure. Moreover, in most case, $h_s$ with $s = n$ integer $n = 1, \ldots, d - 1$ in $\mathbb{R}^d$ is actually denote by $\ell_n$ and refers to the (surface) Lebesgue measure of dimension $n$ in $\mathbb{R}^d$, even if the actually (general) definition is in term of the $n$-dimensional Hausdorff measure in $\mathbb{R}^d$.

The interested reader may consult, for instance, the books by Taylor [115], Evans and Gariepy [42], Lin and Yang [78], Mattila [79], Yeh [120, Chapter 7, pp. 675–826] or Billingsley [15, Section 3.19, pp. 247–258], among many others.

5.6 Area and Co-area Formulae

As mentioned early, the Lebesgue measure represents the volume in $\mathbb{R}^d$ while the length and surface area are given by the Hausdorff measure, except for a factor. Recall that on the Borel space $\mathbb{R}^d$ we denote by $\ell_n = c_n h_n$, where $c_n = 2^{-n} \pi^{n/2}/\Gamma(n/2 + 1)$, i.e., the $n$-dimensional surface measure, and $\ell_d$ is the Lebesgue measure. Also $\ell_0$ is the counting measure, the number of elements in the given set.

Let us recall the polar decomposition of a linear mapping $T : \mathbb{R}^d \to \mathbb{R}^n$ into an symmetric linear map $\mathbb{R}^{d\wedge n} \to \mathbb{R}^{d\wedge n}$ and an orthogonal linear map $H : \mathbb{R}^{d\wedge n} \to \mathbb{R}^{d\wedge n}$ such that $T = HS$ if $d \leq n$ and $T = SH^*$ if $d \geq n$. Thus, the Jacobian $J(T)$ is defined as $J(T) = |\det(S)|$, the determinant (with positive sign) of the symmetric (square) part of $T$. Next, based on Rademacher’s Theorem, the differential $Df$ of a given Lipschitz mapping $f : \mathbb{R}^d \to \mathbb{R}^n$ exits as a linear map $Df(x) : \mathbb{R}^d \to \mathbb{R}^n$, $\ell_d$-almost every $x$. Hence, the Jacobian of $f$ is defined as the Jacobian of its differential $Df$ (as a linear map), i.e., $J(f, x) = J(Df(x))$. 
Theorem 5.18. Let $f : \mathbb{R}^d \to \mathbb{R}^n$ be a Lipschitz function. Then for every $\ell_d$-measurable set $E \subset \mathbb{R}^d$ the mapping $y \mapsto \ell_d \mathbf{n}(E \cap f^{-1}(y))$ is $\ell_n$-measurable, we have the co-area formula
\[
\int_E J(f, x) \, dx = \int_{\mathbb{R}^n} \ell_{d-n}(E \cap f^{-1}(y)) \, dy, \quad \text{when} \quad d \geq n,
\]
and the area formula
\[
\int_E J(f, x) \, dx = \int_{\mathbb{R}^n} \ell_0(E \cap f^{-1}(y)) \, dy, \quad \text{when} \quad d \leq n,
\]
where $dx = d\ell_d(x)$ and $dy = d\ell_n(y)$, for $x$ in $\mathbb{R}^d$ and $y$ in $\mathbb{R}^n$. \qed

It is clear that the area formula is used for the length of a curve ($d = 1$, $n \geq 1$), surface area of a graph or surface area of a parametric hypersurface ($d \geq 1$, $n = d + 1$), and in general for submanifolds.

The both formulae generalize to change of variables, i.e., if $g : \mathbb{R}^d \to \mathbb{R}$ is an $\ell_d$-integrable function then
\[
\int_{\mathbb{R}^d} g(x) J(f, x) \, dx = \int_{\mathbb{R}^n} \left[ \sum_{x \in f^{-1}(y)} g(x) \right] \, dy, \quad \text{when} \quad d \leq n
\]
and, the restriction of $g$ to $f^{-1}(y)$, denoted by $g|_{f^{-1}(y)}$, is $\ell_{d-n}$-integrable for $\ell_n$-almost every $y$ and
\[
\int_{\mathbb{R}^d} g(x) J(f, x) \, dx = \int_{\mathbb{R}^n} dy \int_{f^{-1}(y)} g(x) \ell_{d-n}(dx), \quad \text{when} \quad d \geq n.
\]
Note that $f^{-1}(y)$ is a closed set in $\mathbb{R}^d$ for every $y \in \mathbb{R}^n$.

The co-area formula can be used to compute level sets and polar (or spherical) coordinates, e.g., if $g : \mathbb{R}^d \to \mathbb{R}$ is integrable then
\[
\int_{\mathbb{R}^d} g(x) \, dx = \int_0^\infty dr \int_{\partial B(0,r)} g(x) \ell_{d-1}(dx),
\]
where $\partial B(0,r)$ is the boundary (sphere) of the ball $B(0,r)$ with radius $r$ and center at the origin $0$, and again, we have
\[
\int_{\partial B(0,r)} g(x) \ell_{d-1}(dx) = \int_{S^{d-1}} g(rx') r^{d-1} \, dx',
\]
where $S^{d-1} = \partial B(0, 1)$ is the unit sphere in $\mathbb{R}^d$ and $dx' = \ell_{d-1}(dx')$. In spherical coordinates this means
\[
\int_{\Omega} f(x) \ell_d(dx) = \int_0^\infty dr \int_{\{x \in \Omega : |x| = r\}} f(x) \ell_{d-1}(dx) = \int_0^\infty r^{d-1} dr \int_{\{x' \in \mathbb{R}^d : |x'| = 1\}} f(rx') \mathbb{1}_{\{rx' \in \Omega\}} \ell_{d-1}(dx'). \tag{5.20}
\]
Moreover, the center of the spherical coordinates may be different from the origin 0. For instance, a prove of what was mentioned in this subsection can be found Evans and Gariepy [42] or Lin and Yang [78].

The interested reader may also check the Appendix C in Leoni [75, pp 543-579] for a quick refresh on Lebesgue and Hausdorff measure (and integration), in particular, if \( A \) and \( B \) are two measurable sets in \( \mathbb{R}^d \) such that \( A + B = \{ a + b : a \in A, \ b \in B \} \) is also measurable then Brunn-Minkowski’s inequality reads as

\[
(\ell_d(A))^{1/d} + (\ell_d(B))^{1/d} \leq (\ell_d(A + B))^{1/d}.
\] (5.21)

This estimate in turn can be used to deduce the isodiametric inequalities. Moreover, if we denote by \( \ell^*_d \) the Lebesgue outer measure in \( \mathbb{R}^d \) then the reader may find details (e.g., Stroock [112, Section 4.2, pp. 74-79]) on proving the so-called isodiametric inequality (see Remark 2.40)

\[
\ell^*_d(A) \leq \omega_d (r(A))^d, \quad \forall A \subset \mathbb{R}^d,
\]

where \( \omega_d = \pi^{d/2}/\Gamma(d/2 + 1) \) is the Lebesgue measure of the unit ball in \( \mathbb{R}^d \), and \( r(A) \) is the radius of \( A \), i.e., \( r(A) = \sup\{|x - y|/2 : x, y \in A\} \).

For a later use, the above co-area formulae can be summarized as

\[
\int_{\Omega} f(x)|\nabla \varrho(x)| \, dx = \int_{\mathbb{R}} ds \int_{\varrho^{-1}(s)} f(x) \ell_{d-1}(dx)
\] (5.22)

where \( \ell_{d-1} \) is the \((d - 1)\)-dimensional Hausdorff (Lebesgue) measure in \( \mathbb{R}^d \), \( \Omega \) is an open subset of \( \mathbb{R}^d \) and \( \varrho \) is a real-valued Lipschitz function defined on \( \Omega \). More general, if \( \varrho = (\varrho_1, \ldots, \varrho_n) \) is a Lipschitz function defined on \( \Omega \) with values in \( \mathbb{R}^n \), for some \( n = 1, \ldots, d - 1 \), then the formula (5.22) becomes

\[
\int_{\Omega} f(x)\sqrt{\nabla \varrho^* \nabla \varrho} \, dx = \int_{\mathbb{R}^n} ds \int_{\varrho^{-1}(s)} f(x) \ell_{d-n}(dx),
\] (5.23)

where the Jacobian \( J(\rho, x) = \sqrt{\nabla \varrho^* \nabla \varrho} \) is written in term of the \( n \times n \) square matrix \( \nabla \varrho^* \nabla \varrho = (\sum_{k=1}^d \partial_k \varrho_i \partial_k \rho_j) \).

Instead of actually proving of the above results, we prefer to follow Jones [65, section 15.J, pp. 494-505] and to give a number of steps to deduce Theorem 5.18 for a simple case when \( n = d \) and \( f \) is bi-Lipschitz, i.e., essentially, the formula of a change-of-variables in the \( d \)-dimensional Lebesgue integral for a bi-Lipschitz homeomorphism.

As early, denote by \( \ell \) and \( \ell^* \) the Lebesgue measure in \( \mathbb{R}^d \) and let \( f \) be a function from an open set \( \Omega \subset \mathbb{R}^d \) into \( \mathbb{R}^d \).

1. If \( f = (f_1, \ldots, f_d) \) is differentiable at a particular point \( x \) (i.e., there exists a matrix denoted by \( Df(x) = (\partial_j f_i(x)) \) such that \( |f_i(x + z) - f_i(x) - \sum_j z_j \partial_j f_i(x)| \to 0 \) as \( |z| \to 0 \) then every \( \varepsilon > 0 \) there exists \( \delta = \delta(\varepsilon, x) \) such that

\[
\ell^*(f(B)) \leq (|\det(Df(x))| + \varepsilon)\ell(B),
\]

for every ball \( B \) with center \( x \) and radius \( r \leq \delta \).
Proof. Since the Lebesgue measure is invariant under any translation and rotation (orthogonal change of variables), this estimate reduces to the case when \( x = 0 \) and \( f(x) = 0 \), i.e., for every \( \varepsilon' > 0 \) there exists \( \delta' > 0 \) such that

\[
|y| \leq \delta \quad \Rightarrow \quad |f(y) - Ty| \leq \varepsilon'|y|,
\]

where \( T = Df(0) \) is a diagonal matrix. Thus two cases should be considered, with a ball \( B = B_r \) centered at the origin and radius \( r \leq \delta' \). In any case, the set \( T(B_r) \) contained in a ball centered at the origin and radius \( Cr \), where \( C \) satisfies \( |Ty| \leq C|y| \) for every \( y \) in \( \mathbb{R}^d \).

First, if \( T \) is not invertible (i.e., \( \det(T) = 0 \)) then image \( T(B_r) \) contained in a \((d-1)\)-dimensional subspace, and therefore, it can be covered by a \( d \)-dimensional interval with arbitrary small size in one direction, i.e., there are \((d-1)\) sized have length bounded \( 2C_Tr \) and one size is smaller than \( 2r\varepsilon' \). This implies that

\[
\ell^*(T(B_r)) \leq 2^d C^{d-1} r^d \varepsilon',
\]

i.e., for any given \( \varepsilon > 0 \) we can choose a suitable \( \delta > 0 \) so that \( \ell^*(T(B_r)) \leq \varepsilon \ell(B_r) \), for every \( r \leq \delta \).

Next, if \( T \) is invertible and \( c > 0 \) satisfies \( |T^{-1}y| \leq c|y| \) then

\[
|y| \leq \delta \quad \Rightarrow \quad |f(y) - Ty| \leq \varepsilon'|y| \quad \Rightarrow \quad |T^{-1}f(y)| \leq (1 + c\varepsilon')|y|,
\]

which means that \( \ell^*(T^{-1}f(B)) \leq (1 + c\varepsilon)^d \ell(B) \). Again, the invariant property (Theorem 2.27) implies

\[
\ell^*(T^{-1}f(B)) = |\det(T)|\ell(f(B)),
\]

and the desired estimate follows. \( \square \)

(2) If \( f \) is a Lipschitz function from an open set \( \Omega \subset \mathbb{R}^d \) into \( \mathbb{R}^d \) then \( f \) preserves negligible and Lebesgue measurable sets.

Proof. Since \( f \) is Lipschitz continuous, the bound \( |f(x) - f(y)| \leq C|x - y| \) for every \( x, y \) in \( \Omega \) implies that the image of any ball \( B \) of radius \( r \) is contained into another ball of radius \( Cr \). Thus, any sequence \( \{B(x_i, r_i) : i \geq 1\} \) of small balls (with center \( x_i \) and radius \( r_i \)) covering a set \( N \) yields another cover \( \{B'(f(x_i), Cr_i) : i \geq 1\} \) by small balls of the image \( f(N) \). Hence \( f \) maps sets of zero Lebesgue measure into sets of zero Lebesgue measure.

Next, we check that a continuous function preserves \( \mathcal{F}_\sigma \)-sets. Indeed, a continuous function maps compact set into compact sets and because any function preserves union of sets and any \( \mathcal{F}_\sigma \)-set in \( \mathbb{R}^d \) is a countable union of compact sets, the assertion is verified.

Now, conclude by recalling that any measurable set is the union of a \( \mathcal{F}_\sigma \)-set and a negligible set. For instance, the reader may check Stroock [112, Section 2.2, pp. 30–33] for more details. \( \square \)
If \( f \) is differentiable at any point \( x \) in \( E \subset \Omega \) then
\[
\ell^*(f(E)) \leq \left( \sup_E J(f) \right) \ell^*(E)
\]  
(5.24)
where \( J(f) : x \mapsto J(f, x) = |\text{det}(Df(x))| \) denote the Jacobian of \( f \). This estimate yields Sard’s Theorem (i.e., the set where \( f \) is differentiable and the Jacobian vanishes has Lebesgue measure zero). Moreover, any negligible set of point where \( f \) is differentiable is mapped into a negligible set.

**Proof.** If desired, due to the monotone convergence, we may assume that \( E \) is bounded (or with finite measure) without any loss of generality, so that, for every \( \varepsilon > 0 \) there exists an open set \( G \) such that \( E \subset G \subset \Omega \) and \( \ell(G) \leq \ell^*(E) + \varepsilon < \infty \).

Now, by means of (1), for any given \( \varepsilon > 0 \) and for each \( x \) in \( E \) there exists \( \delta = \delta(x, \varepsilon) \) such that the ball \( B(x, 5\delta) \subset G \) and
\[
\ell^*(f(B(x, r))) \leq (\varepsilon + \sup_E J(f)) \ell^*(B(x, r)),
\]
for every \( r \leq \delta \). This form a Vitali cover of \( E \) and therefore, by Theorem 2.29, there exist a sequence \( \{B_i, i \geq 1\} \) of disjoint balls with center in points of \( E \) such that
\[
E \subset \left( \bigcup_{i < k} B_i \right) \cup \left( \bigcup_{i \geq k} B^5_i \right) \subset G
\]
where \( B^5_i \) is ball with the center as \( B_i \) but with radius 5 times the radius of \( B_i \). Hence,
\[
\ell^*(f(E)) \leq \sum_{i < k} \ell^*(f(B_i)) + \sum_{i \geq k} \ell^*(f(B^5_i)) \leq \left( \varepsilon + \sup_E J(f) \right) \left( \sum_{i < k} \ell(B_i) + \sum_{i \geq k} 5^d \ell(B_i) \right),
\]
Since the balls are disjoint and contained in the set \( G \) with finite measure, the remainder of the series vanishes, i.e., \( \sum_{i \geq k} \ell(B_i) \to 0 \) as \( k \to \infty \), we deduce
\[
\ell^*(f(E)) \leq \sum_i \ell^*(f(B_i)) \leq (\varepsilon + \sup_E J(f)) \sum_i \ell(B_i),
\]
i.e.,
\[
\ell^*(f(E)) \leq (\varepsilon + \sup_E J(f)) \ell(G) \leq (\varepsilon + \sup_E J(f)) \left( \ell^*(E) + \varepsilon \right),
\]
which proves estimate (5.24).

Finally, if \( N \) is a negligible set of differentiable points of \( f \) and \( \{B_i : i \in I\} \) is a cover of \( N \) with ball centered at some point in \( N \) then the image family \( \{f(B_i) : i \in I\} \) is a cover of \( f(N) \) and the above estimate holds for each ball \( B_i \). Hence, the set \( f(N) \) is also negligible. \( \square \)
If $f$ is measurable on $\Omega$ and differentiable at point of a measurable set $E$ then the Jacobian $J(f) = |\det(Df)|$ is measurable on $E$ and the following estimate

$$\ell^*(f(E)) \leq \int_E |\det(Df)(x)|dx$$

holds. In particular, if $f$ is differentiable at almost every point in $\Omega$, and $f$ preserves negligible and Lebesgue measurable subsets of $\Omega$ then the above estimate is valid for any measurable subset $E$ of $\Omega$, $f(E)$ is measurable, and

$$\int_{f(\Omega)} g(y)dy \leq \int_\Omega g(f(x))|\det(Df)(x)|dx,$$

for any nonnegative Borel measurable function $g$ in $\mathbb{R}^d$. 

**Proof.** The measurability of the Jacobian is clearly true. Because the function $f$ is continuous, it preserves $\mathcal{F}_\sigma$-sets. Hence, $f$ preserves measurable subset sets of $\Omega$, i.e. $f(E)$ is measurable.

Therefore, first assume that $\ell(E) < \infty$, and for any given $\varepsilon > 0$ define

$$E_k = \{x \in E : (k-1)\varepsilon \leq |\det(Df)(x)| < k\varepsilon\},$$

for any $k \geq 1$. Apply (3) to obtain $\ell(f(E_k)) \leq k\varepsilon \ell(E_k)$, and use the inclusion $f(E) \subset \bigcup_k f(E_k)$ to deduce

$$\ell(f(E)) \leq \sum_k (k-1)\varepsilon \ell(E_k) + \varepsilon \sum_k \ell(E_k),$$

which yields

$$\ell(f(E)) \leq \sum_k \int_{E_k} |\det(Df)(x)|dx + \varepsilon \ell(E),$$

i.e., estimate (5.25) is satisfied when $\ell(E) < \infty$.

If $\ell(E) = \infty$ then replace $E$ with $E^r = \{x \in E : |x| \leq r\}$ and use the monotone convergence to complete the argument.

Finally, remark that to replace the differentiability everywhere with only differentiability everywhere almost everywhere, we need to know, a priori, that $f$ maps negligible sets into negligible sets. Also, observe that inequality (5.26) reduces to inequality (5.25) if $g = 1_B$ with a Borel set $B \subset \Omega$. Thus, by linearity, inequality (5.26) remains valid for nonnegative simple functions and therefore, by the monotone convergence, we conclude.

If $f$ is bi-Lipschitz continuous (i.e., $c|x-y| \leq |f(x) - f(y)| \leq C|x-y|$ for every $x, y$ in $\Omega$ and some constants $C \geq c > 0$) and $g$ is a nonnegative Lebesgue measurable function in $f(\Omega)$ then the composition function $x \mapsto g(f(x))$ is Lebesgue measurable in $\Omega$ and

$$\int_{f(\Omega)} g(y)dy = \int_\Omega g(f(x))|\det(Df)(x)|dx.$$ 

In particular, if $g$ in integrable in $f(\Omega)$ then $x \mapsto g(f(x))$ is integrable in $\Omega$. 

(5) If $f$ is bi-Lipschitz continuous (i.e., $c|x-y| \leq |f(x) - f(y)| \leq C|x-y|$ for every $x, y$ in $\Omega$ and some constants $C \geq c > 0$) and $g$ is a nonnegative Lebesgue measurable function in $f(\Omega)$ then the composition function $x \mapsto g(f(x))$ is Lebesgue measurable in $\Omega$ and

$$\int_{f(\Omega)} g(y)dy = \int_\Omega g(f(x))|\det(Df)(x)|dx.$$ 

In particular, if $g$ in integrable in $f(\Omega)$ then $x \mapsto g(f(x))$ is integrable in $\Omega$. 

5.6. Area and Co-area Formulae

Proof. As mentioned early, Rademacher’s Theorem 5.19 (see below) implies that \( f \) is differentiable almost everywhere, and so the preceding steps (1),..., (4) can be used.

Because both, \( f \) and its inverse \( f^{-1} \) as Lipschitz continuous functions, and (1) implies that \( (g \circ f)^{-1}(B) = f^{-1}(g^{-1}(B)) \) is a Lebesgue measurable set for any Borel set \( B \) in \( \mathbb{R} \). Hence, \( g \circ f \) is measurable.

Apply the inequality (5.26) to get the \( \leq \) sign in (5.27). Next, re-apply inequality (5.26) with \( f^{-1} \) to obtain

\[
\int_{\Omega} h(x) dx \leq \int_{f(\Omega)} h(f^{-1}(y)) |\text{det}(Df^{-1})(y)| dy.
\]

Since \( \text{det}(Df^{-1})(y) = 1/ \text{det}(Df)(f^{-1}(y)) \), take \( h(x) = g(f(x))|\text{det}(Df)(x)| \) to deduce

\[
\int_{f(\Omega)} g(x) |\text{det}(Df)(x)| dx \leq \int_{f(\Omega)} g(y) dy,
\]

which yields the other sign \( \geq \), and the proof is completed.

The equality (5.27) can be written as

\[
\int_{f(\Omega)} h(f^{-1}(y)) dy = \int_{\Omega} h(x) |\text{det}(Df)(x)| dx,
\]

for any Lebesgue measurable function \( h \) in \( \Omega \).

(6) The function \( f \) is called \textit{locally bi-Lipschitz continuous} if for every closed ball \( B \subset \Omega \) there exist constants \( C \geq c > 0 \) such that

\[
c|x-y| \leq |f(x) - f(y)| \leq C|x-y|, \quad \forall x, y \in B.
\]

The \textit{multiplicity} \( m(f, \Omega) = m(y|f, \Omega) \) of \( y \) relative to \( f \) in \( \Omega \) is defined as the number of points (possible infinite or zero) \( x \) in \( \Omega \) such that \( f(x) = y \).

If \( f \) is locally bi-Lipschitz continuous and \( g \) is a nonnegative Lebesgue measurable function in \( f(\Omega) \) then the composition \( g \circ f : x \mapsto g(f(x)) \) and the multiplicity \( m(f, \Omega) : y \mapsto m(y|f, \Omega) \) are Lebesgue measurable functions in \( \Omega \), and

\[
\int_{f(\Omega)} g(y)m(y|f, \Omega) dy = \int_{\Omega} g(f(x)) |\text{det}(Df)(x)| dx,
\]

or equivalently (without using explicitly the multiplicity function)

\[
\int_{f(\Omega)} \sum_{x \in f^{-1}(y)} h(x) dy = \int_{\Omega} h(x) |\text{det}(Df)(x)| dx,
\]

for any nonnegative Lebesgue measurable function \( h \) in \( \Omega \).
Proof. First, note that a locally bi-Lipschitz continuous function is not necessarily continuous or one-to-one in $\Omega$, but it preserves negligible and Lebesgue measurable set and it is almost everywhere differentiable in view of Rademacher’s Theorem 5.19 (see below). Thus, the set critical points (i.e., where the function is not differentiable or the Jacobian vanishes) of a Lipschitz function is negligible, and closed if the gradient is continuous. Thus, the inverse function theorem implies that a continuously differentiable function is a locally bi-Lipschitz continuous function on the open $\Omega \setminus N$, with $N = \{ x : |\det(Df(x))| = 0 \}$, i.e., the above assertion (6) holds true for any continuously differentiable function $f$ on $\Omega$. Also, remark that the multiplicity $m(y|f, \Omega)$ can be expressed as the sum (possible infinite or empty) of Dirac measures $\sum_{x \in f^{-1}(y)} \delta_x(\Omega)$, i.e., the function $A \mapsto m(y|f, A)$ is a measure over $A \subset \Omega$, for any fixed $y$ in $\mathbb{R}^d$.

Since $f$ is bi-Lipschitz on any closed ball inside $\Omega$, for every $x$ in $\Omega \setminus N$ there is $\delta > 0$ such that for any ball $B(x, r)$ with center $x$ and radius $r$,
\begin{equation}
\ell(f(B(x, r))) = \int_{B(x, r)} |\det(Df(x))|dx, \quad \forall r \leq \delta \tag{5.31}
\end{equation}
Such a collection of balls is a fine (or Vitali) cover of $\Omega$ and thus (see Remark 2.36), there exists a countable sub-collection forming sequence $\{ B_i : i \geq 1 \}$ of disjoint balls such that
\begin{equation*}
\Omega \setminus N = \bigcup_{i} B_i,
\end{equation*}
for some negligible set $N$.

The function $f$ may not be one-to-one so that the union $\bigcup_i f(B_i)$ may be disjoint. Since $f$ is one-to-one in each ball $B_i$, the definition of the multiplicity yields the equality
\begin{equation*}
\sum_i 1_{f(B_i)} = m(f, \Omega \setminus N'),
\end{equation*}
and, because $f(N)$ is also a negligible set, we deduce from (5.31) the equality
\begin{equation*}
\int_{\mathbb{R}^d} m(y|f, E)dy = \int_{f(E)} \sum_i 1_{f(B_i)}(y)dy = \int_E |\det(Df(x))|dx,
\end{equation*}
with $E = \Omega$. Now, if $E$ is an open subset of $\Omega$ then we repeat the argument with $E$ instead of $\Omega$, i.e. (5.31) holds for any open set $E$. Since both expressions are measures, the equality (5.31) is valid for any measurable set $E$ of $\Omega$, and in the left-hand side, we can substitute $\mathbb{R}^d$ with $f(E)$ and $m(y|f, E)$ with $m(y|f, \Omega)$.

Hence, equality (5.29) holds for $g = 1_E$ and by linearity and monotonicity, it remains true for any nonnegative Lebesgue measurable function $g$. It is clear that equality (5.30) for $h$ is actually (5.29) for $h = g \circ f$. Conversely, if equality (5.30) holds then we cannot take $g = h \circ f^{-1}$, but instead we choose $h = 1_{B_i}$ to obtain (5.31), and so, equality (5.29) follows.\[Preliminary\]
After showing that a monotone function is differentiable almost everywhere, the almost differentiability of a Lipschitz continuous function in one variable is easily obtained, however, the same result in several dimensions is more delicate, e.g., Ziemer [124, Section 2.2, pp. 49–52].

**Proposition 5.19 (Rademacher).** If \( f \) is a Lipschitz continuous function on an open set \( \Omega \subset \mathbb{R}^d \) with values in \( \mathbb{R}^n \), i.e.,

\[
\sup_{x,y \in \Omega} \left\{ \frac{|f(x) - f(y)|}{|x - y|} \right\} < \infty,
\]

then

\[
\frac{f(x + z) - f(x) - z \cdot \nabla f(x)}{|z|} \to 0 \quad \text{as} \quad |z| \to 0, \quad a.e. x \in \Omega. \tag{5.32}
\]

i.e., \( f \) is differentiable for almost every \( x \) in \( \Omega \).

**Proof.** This is understood as Fréchet differentiable, i.e., there exists a negligible set \( N \) in \( \mathbb{R}^d \) such that for any \( x \) in \( \Omega \setminus N \) the gradient \( \nabla f(x) \) is defined and (5.32) is satisfied.

It is clear that we may assume \( n = 1 \) without loss of generality. If \( v \) is direction in \( \mathbb{R}^d \) with \( |v| = 1 \) and \( x \) is a point in \( \mathbb{R}^d \) then the function \( t \mapsto F_v(t) = f(x + tv) \) is a Lipschitz continuous functions of one-variable, and so, it is differentiable for almost every \( t \). The \( v \)-directional derivative of \( f \) at \( x \) is denoted by \( \nabla_v f(x) \) and it is equal to \( F'_v(0) \) wherever it exists.

Take a fixed \( v \) to show that \( \nabla_v f(x) \) is defined for almost every \( x \) in \( \Omega \). Indeed, let \( N_v \) be the set where \( v \)-directional derivative fails to exists, i.e.,

\[
N_v = \left\{ x \in \Omega : \limsup_{t \to 0} \frac{f(x + tv) - f(x)}{t} > \liminf_{t \to 0} \frac{f(x + tv) - f(x)}{t} \right\}.
\]

This is a Borel set and because the Lebesgue measure \( \ell_d \) is invariant under any orthogonal transformations \( \rho \), we have \( \ell_d(N_v) = \ell_d(\rho(N_v)) \). Since \( \rho(N_v) = N_{\rho(v)} \), we can choose any convenient direction \( v \) to establish the claim that \( \ell_d(N_v) = 0 \). Therefore, if \( v = (1,0,\ldots,0) \) then the fact that the function of one variable \( t \mapsto f(t, x_2, \ldots, x_d) \) is differentiable almost everywhere implies that any \( (x_2, \cdots, x_d) \)-section of \( N_v \) is a one-dimensional negligible set, i.e., \( \ell_1(\{x_1 : (x_1, x_2, \ldots, x_d) \in N_v \}) = 0 \), for every \( (x_2, \ldots, x_d) \), and then, Fubini theorem yields \( \ell_d(N_v) \). Therefore, the expressions \( \nabla_v f(x) \) and \( \nabla f(x) \) are defined for almost every \( x \) in \( \Omega \).

A second point is to show that \( \nabla_v f(x) = v \cdot \nabla f(x) \) for almost every \( x \) in \( \Omega \). To this end, take a continuously differentiable function \( \varphi \) in \( \Omega \) vanishing near the boundary \( \partial \Omega \) and use the definition as limit when \( t \to 0 \) and a one-dimensional change of variables to obtain

\[
\int_{\Omega} \nabla_v f(x) \varphi(x) dx = -\int_{\Omega} f(x) \nabla_v \varphi(x) dx,
\]

\[
-\int_{\Omega} f(x)(v \cdot \nabla \varphi(x)) dx = \int_{\Omega} (v \cdot \nabla f(x)) \varphi(x) dx.
\]
Since $\nabla_v \varphi(x) = v \cdot \nabla \varphi(x)$ and $\varphi$ is arbitrary, the desired claim is proved.

Now, writing $z = tv$ with $|v| = 1$, the differentiability condition becomes $D(x, v, t) = [f(x + tv) - f(x)]/t - v \cdot \nabla f(x) \to 0$ as $t \to 0$, uniformly in the directions $v$.

Next, given any $\varepsilon > 0$, use to precedent claims to find a negligible set $N$ in $\mathbb{R}^d$ and a finite $\varepsilon$-dense set of directions $\{v_1, \ldots, v_n\}$ such that $\nabla_{v_k} f(x) = v_k \cdot \nabla f(x)$ for every $k = 1, \ldots, n$ and any $x$ in $\Omega \setminus N$. Thus $\max_k |D(x, v_k, t)| \to 0$ as $t \to 0$, because there is a finite number of directions.

Remark that an $\varepsilon$-dense set of directions means that $\min_k |v - v_k| \leq \varepsilon$, for every $v$ in $\mathbb{R}^d$ with $|v| = 1$, and note that if $f$ is only absolutely continuous over any direction $v$ then all previous steps remain true. Finally, use the Lipschitz conditions to deduce the inequality

$$|D(x, v, t) - D(x, v_k, t)| \leq \left( \sup_{v, t} \left\{ \frac{|f(x + tv) - f(x)|}{t} \right\} + |\nabla f(x)| \right) |v - v_k|,$$

which implies that $|D(x, v, t)| \leq |D(x, v_k, t)| + C\varepsilon$, for a suitable constant $C$ independent of $x, v, v_k, t$ and $\varepsilon$. This completes the argument.
Chapter 6

Measures and Integrals

6.1 Signed Measures

Because measures can take infinite values, subtraction two measures is only allowed when at least one of them is finite. Thus a signed measure $\nu$ is a $\sigma$-additive set function on a measurable space $(\Omega, \mathcal{F})$ such that $\nu(\emptyset) = 0$. The $\sigma$-additivity implies that $\nu$ takes values in either $[-\infty, +\infty)$ or $(-\infty, +\infty]$; moreover, if $A = \sum_{i=1}^{\infty} A_i$ with $A_i \in \mathcal{F}$ and $|\nu(A)| < \infty$ then, by separating the positive and the negative terms, we deduce that the series $\sum_{i=1}^{\infty} \nu(A_i)$ is absolutely convergence. Also, for any $E \subset F$ measurable sets, the relation $\nu(F) = \nu(E) + \nu(F \setminus E)$ shows that if $|\nu(F)| < \infty$ then $|\nu(E)| < \infty$ (i.e., finite values can only be obtained by adding or subtraction real numbers. Hence it makes sense to say that a signed measure $\nu$ is finite if $|\nu(\Omega)| < \infty$, and similarly we define $\sigma$-finite signed measures.

Exercise 6.1. Regarding the above statements, first (a) prove that a series $\sum_{i=1}^{\infty} a_i$ of real numbers converges absolutely if and only if the series $\sum_{i=1}^{\infty} a_{\iota(i)}$ converges, for any bijective function $\iota$ between the positive integers. Next, let $\nu$ be a signed measure on $(\Omega, \mathcal{F})$ and $\{F_k\}$ be a sequence of disjoint sets in $\mathcal{F}$ such that $|\nu(\bigcup_k F_k)| < \infty$. Prove that (b) the series $\sum_k \nu(F_k)$ is absolutely convergence.

Hahn-Jordan Decomposition

The $\sigma$-additivity property applied to finite measures can be considered in a larger context, e.g., we may discuss measures with complex values (in $\mathbb{C}$) or with vector values in $\mathbb{R}^d$ or even more general with values in a topological vector space (usually a Banach space or a locally convex space). First we need to establish the Hahn-Jordan decomposition

Proposition 6.1. Let $\nu$ be a signed measure on $(\Omega, \mathcal{F})$. Then there exists a measurable set $P \in \mathcal{F}$ such that $\nu^+(F) = \nu(F \cap P)$ and $\nu^-(F) = -\nu(F \cap P^c)$ are measures satisfying $\nu(F) = \nu^+(F) - \nu^-(F)$, for every $F \in \mathcal{F}$. The set $P$
is not necessarily unique, but the positive and negative variations measures $\nu^+$ and $\nu^-$ are uniquely defined.

Proof. It suffices to consider the case where $-\infty < \nu(F) \leq \infty$, for every $F \in \mathcal{F}$. A set $B$ in $\mathcal{F}$ satisfying $\nu(F \cap B) \leq 0$, for every $F \in \mathcal{F}$, is called a negative set. Let $\{B_n\}$ be a minimizer sequence of negative sets such that

$$\lim_n \nu(B_n) = \inf \{ \nu(B) : B \text{ is a negative set} \} = \beta.$$ 

Hence the set $B_0 = \bigcup_n B_n$ is a negative set satisfying $\nu(B_0) \leq \nu(B_n)$, for every $n$, i.e., $\beta = \nu(B_0)$ is a real number. We complete the proof by showing that $A = B_0^c = \Omega \setminus B_0$ is a positive set, i.e., $\mu(F \cap A) \geq 0$, for every $F \in \mathcal{F}$.

Working by contradiction, suppose $F_0 \in \mathcal{F}$, $F_0 \subset A$ with $-\infty < \nu(F_0) < 0$. Then, $F_0$ cannot be a negative set since $B_0 \cup F_0$ would be a negative set with $\nu(B_0 \cup F_0) < \nu(B_0)$, contradicting the minimal character of $B_0$. Hence, there exists $B \in \mathcal{F}, B \subset F_0$, with $\infty > \nu(B) > 0$ and then $\nu(F_0 \setminus B) = \nu(F_0) - \nu(B) < 0$. Let $k_1$ be the smallest integer such that there exist a set $B = F_1$ satisfying $\nu(F_1) \geq 1/k_1$, i.e., for any other measurable set $F \subset F_0$ we have $\nu(F) \leq 1/(k_1 + 1)$. Now, applying the arguments used on $F_0$ to $F_0 \setminus F_1$, and iterating, we construct sequences $\{k_n\}$ and $\{F_n\} \subset \mathcal{F}$ such that $F_n \subset F_{n-2} \setminus F_{n-1}$, $\infty > \nu(F_n) > 1/k_n$, and

$$\nu(F_0 \setminus \bigcup_{i=1}^n F_i) = \nu(F_0) - \sum_{i=1}^n \nu(F_i) < 0, \quad \forall n \geq 2,$$

$$\nu(F) \leq \frac{1}{k_n - 1}, \quad \forall F \in \mathcal{F}, \ F \subset F_0 \setminus \bigcup_{i=1}^n F_i, \ n \geq 2.$$

Hence, $\sum_n 1/k_n \leq \sum_n \nu(F_n) = \nu(\bigcup_n F_n) < \infty$ and so $k_n \to 0$, which implies that $E_0 = F_0 \setminus \bigcup_{n=1}^\infty F_n$ is a negative set with

$$\nu(E_0) = \nu(F_0) - \sum_n \nu(F_n) \leq \nu(F_0) < 0,$$

which again, contradicts the minimal character of $B_0$. $\square$

A set $P$ as in Hahn-Jordan Decomposition (Proposition 6.1) is called a positive set for $\nu$. The measure $|\nu|(F) = \nu^+(F) + \nu^-(F)$, for any $F \in \mathcal{F}$, is called the variation of $\nu$, which can also be defined by

$$|\nu|(F) = \sup \left\{ \sum_{i=1}^n \nu(F_i) : F = \sum_{i=1}^n F_i, \ F_i \in \mathcal{F} \right\}, \quad \forall F \in \mathcal{F},$$

recall that the sum symbol $\sum$ also means a union of disjoint sets. Note that a signed measure $\nu$ is finite (i.e., $|\nu(\Omega)| < \infty$) if and only if $|\nu|$ is so, (i.e., $|\nu|(\Omega) < \infty$), and similarly for the concept of $\sigma$-finite.
Definition 6.2. Let \( \mu \) and \( \nu \) be two signed measures on a measurable space \((\Omega, \mathcal{F})\). The signed measure \( \nu \) is said to be **absolutely continuous** with respect to \( \mu \) and written \( \nu \ll \mu \) if for every \( F \in \mathcal{F} \) with \(|\mu|(F) = 0\) we also have \( \nu(F) = 0 \). On the contrary, these two measures \( \mu \) and \( \nu \) are called (mutually) **singular** and written \( \mu \perp \nu \) (or \( \nu \perp \mu \)) if there exists \( A \in \mathcal{F} \) such that \(|\mu|(A) = 0\) and \(|\nu|(\Omega \setminus A) = 0\).

It is clear that being singular is a symmetric property, while being absolutely continuous is not. Moreover, \( \mu \perp \nu \) if and only if there exits \( A \in \mathcal{F} \) such that for every \( F \in \mathcal{F} \) we have \( F \cap A = \emptyset \Rightarrow \mu(F) = 0 \) and \( F \subset A \Rightarrow \nu(F) = 0 \), i.e., \( \nu = 0 \) on \( A \) and \( \mu = 0 \) on \( \Omega \setminus A \). Similarly, \( \nu \ll \mu \) if an only if for every \( F \in \mathcal{F} \) such that \( \mu(E \cap F) = 0 \), for every \( E \in \mathcal{F} \), we have \( \nu(E \cap F) = 0 \), for every \( E \in \mathcal{F} \).

Exercise 6.2. Prove that (a) \( \nu \ll \mu \) if and only if (b) \( |\nu| \ll |\mu| \) if and only if (c) \( \nu^+ \ll \mu \) and \( \nu^- \ll \mu \).

**Radon-Nikodym Derivative**

If \( f \) is a quasi-integrable function in \((\Omega, \mathcal{F}, \mu)\), i.e., \( f = f^+ - f^- \) is measurable and either \( f^+ \) or \( f^- \) is integrable, then the expression

\[
F \mapsto \int_F f \, d\mu, \quad \forall F \in \mathcal{F}
\]

defines a signed measure which is absolutely continuous with respect to \( \mu \). The converse is precisely the Radon-Nikodym Theorem, and the Lebesgue decomposition completes the argument, namely, any \( \sigma \)-finite signed measure \( \nu \), on a \( \sigma \)-finite measure space \((\Omega, \mathcal{F}, \mu)\), can be written as \( \nu = \nu_a + \nu_s \), where \( \nu_a \ll \mu \) and \( \nu_s \perp \mu \).

**Theorem 6.3.** Let \((\Omega, \mathcal{F}, \mu)\) be a \( \sigma \)-finite measure space. Suppose that \( \nu \) is a \( \sigma \)-finite signed measure on \((\Omega, \mathcal{F})\), which is absolutely continuous with respect to \( \mu \). Then there exists a quasi-integrable function \( f \) such that

\[
\nu(F) = \int_F f \, d\mu, \quad \forall F \in \mathcal{F},
\]

where the function \( f \) is uniquely defined except in a set of \( \mu \)-measure zero.

**Proof.** First note that by means of the Hahn-Jordan decomposition, we can write \( \nu = \nu^+ - \nu^- \), which effectively reduces the problem to the case of a \( \sigma \)-finite measure \( \nu \). Now, we proceed in several steps:

(Step 1) Since \( \nu \) is \( \sigma \)-finite, the whole space \( \Omega \) can be written as a disjoint countable union \( \bigcup_n \Omega_n^\nu \) with \( |\nu(\Omega_n^\nu)| < \infty \). Next, because \( \mu \) is also \( \sigma \)-finite, each \( \Omega_n^\nu \) can be written as a disjoint countable union \( \bigcup_k \Omega_{n,k}^\nu \) with \( |\nu(\Omega_{n,k}^\nu)| < \infty \).
Hence, relabeling the double sequence, we have $\Omega = \bigcup_n \Omega_n$, with $\Omega_n \cap \Omega_m = \emptyset$ if $n \neq m$ and $|\nu_n(\Omega_n)| + \mu(\Omega_n) < \infty$, for every $n$. Therefore, it suffices to show the results for the case where $\nu$ and $\mu$ are finite measures.

(Step 2) If $G$ is the class of nonnegative $\mu$-integrable functions $g$ such that

$$\nu(F) \geq \int_F g \, d\mu, \quad \forall F \in \mathcal{F},$$

then there exists a function $f$ in $G$ such that

$$\int_\Omega f \, d\mu = \sup_{g \in G} \int_\Omega g \, d\mu.$$

Indeed, first note that if $g_1$ and $g_2$ belongs to $G$ then $g_1 \lor g_2$ also belongs to $G$. Thus, if $\{g_n\}$ is a maximizing sequence then $f_n = \max\{g_1, \ldots, g_n\}$ defines an increasing sequence in $G$ such that

$$\lim_n \int_\Omega f_n = \sup_{g \in G} \int_\Omega g \, d\mu.$$

The monotone convergence theorem ensures that $f = \lim_n f_n$ belongs to $G$ and provides a maximizer.

(Step 3) If $\lambda \neq 0$ is a measure absolutely continuous with respect to $\mu$ then there exists $\varepsilon > 0$ and $A \in \mathcal{F}$ with $\nu(A) > 0$ such that $\lambda(F \cap A) \geq \varepsilon \mu(F \cap A)$, for every $F \in \mathcal{F}$. Indeed, let $A_k$ the Hahn decomposition of the signed measure $\lambda_k = \lambda - (1/k)\mu$, i.e., $\lambda_k(F \cap A_k) \geq 0 \geq \lambda_k(F \setminus A_k)$, for every $F \in \mathcal{F}$. Set $A_0 = \bigcup_k A_k$ and $B_0 = \bigcap_k B_k$, with $B_k = \Omega \setminus A_k$. Since $0 \leq \lambda(B_0) \leq (1/k)\mu(B_0)$ we have $\lambda(B_0) = 0$, and because $\lambda$ is nonzero and $A_0 = \Omega \setminus B_0$, we deduce $\lambda(A_0) > 0$, i.e., there exists $k$ such that $\lambda(A_k) > 0$. Hence, we choose $A = A_k$ and $\varepsilon = 1/k$, for this particular $k$.

(Step 4) To complete the proof we show that the measure

$$\lambda(F) = \nu(F) - \int_F f \, d\mu, \quad \forall F \in \mathcal{F}$$

vanishes. To this purpose, assume $\lambda \neq 0$ and get a contradiction. Because $\nu \ll \mu$ implies $\lambda \ll \mu$, we can use (Step 3) to get a measurable set $A$ and a $\varepsilon > 0$ such that $\nu(A) > 0$ such that $\lambda(F \cap A) \geq \varepsilon \mu(F \cap A)$, for every $F \in \mathcal{F}$. Choose $h = f + \varepsilon 1_A$ to get

$$\int_F h \, d\mu = \int_F f \, d\mu + \varepsilon \mu(F \cap A) \leq \int_F f \, d\mu + \lambda(F \cap A) =$$

$$= \int_{F \setminus A} f \, d\mu + \nu(F \cap A) \leq \nu(F \setminus A) + \nu(F \cap A) = \nu(F),$$

which shows that $h$ belongs to the class $G$ and

$$\int_\Omega h \, d\mu = \int_\Omega f \, d\mu + \varepsilon \mu(A) > \int_\Omega f \, d\mu,$$

i.e., a contradiction. \qed
Sometimes, the function $f$ satisfying the conditions of Theorem 6.3 is denoted by $\frac{d\nu}{d\mu}$ and called the Radon-Nikodym derivative.

*Remark 6.4.* It is simple to show that for any $\mu$ and $\nu$ are two finite measures on $(\Omega, \mathcal{F})$ we have $\nu \ll \mu$ if and only if for every $\varepsilon > 0$ there exist $\delta > 0$ such that $F \in \mathcal{F}$ and $\mu(F) < \delta$ imply $\nu(F) < \varepsilon$. Indeed, by contradiction, suppose that for some $\varepsilon > 0$ there is a sequence $\{F_n\}$ of measurable sets such that $\mu(F_n) < 2^{-n}$ and $\nu(F_n) \geq \varepsilon$. If $F_0 = \bigcap_n \bigcup_{k \geq n} F_k$ then we deduce $\mu(F_0) \leq \sum_{k \geq n} 2^k = 2^{n-1}$ and $\nu(F_0) \geq \varepsilon$, i.e., $\mu(F_0) = 0$ and we obtain a contradiction. \hfill $\square$

**Lebesgue Decomposition**

A generalization of Radon-Nikodym arguments yields

**Theorem 6.5.** Let $\mu$ and $\nu$ be a two $\sigma$-finite signed measures on a measurable space $(\Omega, \mathcal{F})$. Then there exist two $\sigma$-finite signed measures $\nu_a$ and $\nu_s$ such that $\nu = \nu_a + \nu_s$, $\nu_a \ll \mu$ and $\nu_s \perp \mu$. Moreover the pair $\nu_a, \nu_s$ is uniquely determinate.

*Proof.* First, as in Theorem 6.3, and because $\nu \perp |\mu|$ (or $\nu \ll |\mu|$) if and only if $\nu \perp \mu$ (or $\nu \ll \mu$), and $\nu = \nu^+ - \nu^-$, we can reduce the discussion to the case where $\nu$ and $\mu$ are finite measures.

Since $\nu \ll \mu + \nu$, we apply Radon-Nikodym Theorem 6.3 to show the existence of an integrable function $f$ defined $(\mu + \nu)$-almost everywhere such that

$$\nu(F) = \int_F f \, d\mu + \int_F f \, d\nu, \quad \forall F \in \mathcal{F}. $$

Because $\nu(F) \leq \mu(F) + \nu(F)$ for every measurable $F$, we deduce $0 \leq f \leq 1$, $(\mu + \nu)$-almost everywhere, and so almost everywhere with respect to $\nu$ and $\mu$.

Define $A = \{f < 1\}$ and $B = \Omega \setminus A$, which are unique up to an $\mu + \nu$-negligible set. Thus

$$\nu(B) = \int_B f \, d\mu + \int_B f \, d\nu = \mu(B) + \nu(B), \quad \text{and} \quad \mu(B) < \infty,$$

which yields $\mu(B) = 0$. Now, defining $\nu_a(F) = \nu(F \cap A)$ and $\nu_s(F) = \nu(F \cap B)$, we have $\nu = \nu_a + \nu_s$ and $\nu_s \perp \mu$.

To check that $\nu_a \ll \mu$, suppose $\mu(F) = 0$ to have

$$\nu_a(F) = \nu(F \cap B) = \int_{F \cap A} f \, d\mu + \int_{F \cap A} f \, d\nu = \int_{F \cap A} f \, d\nu,$$

i.e.,

$$\int_{F \cap A} (1 - f) \, d\nu = 0 \quad \text{and} \quad 1 - f > 0 \text{ a.e.},$$

which implies $\nu(F \cap A) = 0$, i.e., $\nu_a(F) = 0$.

Certainly, the sets $A$ and $B$ are defined up to a negligible set with respect to $\mu + \nu$. However, if $\tilde{\nu}_a$ and $\tilde{\nu}_s$ is another decomposition with $\tilde{\nu}_a \ll \mu$ and $\tilde{\nu}_s \perp \mu$ then the signed measure $\lambda = \nu_a - \tilde{\nu}_a = \tilde{\nu}_s - \nu_s$ is both singular and absolutely continuous with respect to $\mu$, and so $\lambda = 0$. \hfill $\square$
Exercise 6.3. Give more details on the how to reduce the proof of Theorem 6.5 to the case where $\mu$ and $\nu$ are finite measures.

We may prove directly Hahn-Jordan decomposition and then the Radon-Nikodym Theorem and Lebesgue decomposition (e.g., as above or see Halmos [57, Chapter VI, pp. 117–136]) or we may deduce Radon-Nikodym Theorem from Riesz presentation (Corollary 6.14) and then show Hahn-Jordan and Lebesgue decomposition (e.g., Rudin [100, Chapter 6, pp. 124–144]).

A simple application of Radon-Nikodym Theorem is the following construction of the essential supremum and infimum, which is discussed later in Section 6.2. Let $\{f_i : i \in I\}$ be a family of real-valued measurable functions defined on a $\sigma$-finite measure space $(X, \mathcal{X}, \mu)$. An extended real-valued measurable function $g$ is called a measurable upper bound of the family $\{f_i : i \in I\}$ if for every $i$ there exists a null set $N_i$ such that $f_i(x) \leq g(x)$ for every $x$ in $X \setminus N_i$. Loosely phasing, the essential supremum $f^*$ is the smallest measurable upper bound, i.e., $f^*$ is a measurable upper bound and with respect to any other measurable upper bound $g$ we have $f^* \leq g$ almost everywhere. Similarly, we define the essential infimum by replacing $f_i$ with $-f_i$. It is simple to prove then the essential supremum (or infimum) is unique almost everywhere and unchangeable if each function of the family is modified almost everywhere. Certainly, this notion is only interesting when the family is uncountable, otherwise, we can take the pointwise supremum (or infimum), which is a measurable function.

Regarding the existence we proceed as follows: First, by means of the transformation $x \mapsto 1/2 + \arctan(x) / \pi$ as in Exercise 1.22, we reduce the problem to the case where the family is equi-bounded, i.e., to discuss the existence of the essential supremum, we may assume that $0 \leq f_i(x) \leq 1$, for every $i$ in $I$ and any $x$ in $X$, without any loss of generality. Second, for any measurable set $A$ consider the set function

$$\nu(A) = \sup \left\{ \sum_{k=1}^{n} \int_{A_k} f_{i_k}(x) \mu(dx) \right\},$$

where the supremum is taken over all finite measurable partitions $A = \sum_{k=1}^{n} A_k$ and any choice of indexes $i_k$ in $I$. It is clear that $\nu(A) \leq \mu(A)$, which shows that if $\{A_n\}$ is an increasing sequence of measurable sets then $\nu(\bigcup_{n \geq k} A_n) \leq \mu(\bigcup_{n \geq k} A_n)$. Moreover, it is not to hard to check that $\nu$ is finitely additive, and so, $\nu$ is an absolutely continuous measure with respect to $\mu$. Next, Radon-Nikodym Theorem 6.3 can be used to define $f^* = d\nu / d\mu$. Therefore, from the inequality

$$\int_{A} f_i \mu \leq \nu(A) = \int_{A} f^* \mu, \quad \forall i \in I, \forall A \in \mathcal{X},$$

and because the definition of $\nu$ implies that

$$\nu(A) \leq \int_{A} g \mu, \quad \forall A \in \mathcal{X},$$
for every measurable upper bound \( g \), we deduce that \( f^* \) is indeed the essential supremum of the family \( \{ f_i : i \in I \} \).

Finally, if the family is stable under the pairwise maximization (i.e., if \( i \) and \( j \) belong to \( I \) then there exists \( k \) in \( I \) such that \( \max\{ f_i(x), f_j(x) \} = f_k(x) \), for almost every \( x \)) then the construction of \( \nu \) proves that there exists sequence \( \{ i_n \} \subset I \) of indexes such that \( \{ f_{i_n} \} \) is an almost everywhere increasing sequence satisfying \( f_{i_n}(x) \to f^*(x) \) almost everywhere \( x \).

**Exercise 6.4.** With the above notation, fill in details for the previous assertions on the essential supremum and infimum. Compare with Exercise 1.23.

The concept of \( \sigma \)-additivity for set functions can be used in other situations, for instance, with complex-valued set functions. A \( \sigma \)-additive set function \( \eta : \mathcal{A} \to \mathbb{C} \) is called a complex measure, i.e., \( \mathcal{A} \) is a \( \sigma \)-algebra, \( \eta(\emptyset) = 0 \) and \( \eta(\bigcup_i A_i) = \sum_i \mu(A_i) \), for any disjoint sequence \( \{ A_i \} \) in \( \mathcal{A} \). Thus, \( \eta \) is a complex measure if and only if \( \eta(A) = \mu(A) + i\nu(A) \), where \( \mu \) and \( \nu \) are real-valued measures (i.e., finite measures). Thus, Radon-Nikodym Theorem 6.3 and the Lebesgue decomposition Theorem 6.5 can be re-stated for complex-valued measures.

**Exercise 6.5.** State the Radon-Nikodym Theorem 6.3 and the Lebesgue decomposition Theorem 6.5 for the case of complex-valued measures (and give details of the proof, if necessary).

For instance, the reader may take a look at Schilling [104, Chapter 19, pp. 202–225] and Taylor [114, Chapter 8, 348–378]. Certainly, the concept of differentiation of locally finite Borel measures is directly related to all the above, for instance, checking the book Mattila [80, Chapter 2, pp. 23–43] may prove very beneficial.

### 6.2 Essential Supremum

Now, we discuss briefly the supremum (or infimum) of a family of measurable functions, but instead of looking at the pointwise suprema, we use the following

**Definition 6.6.** Let \( \mathcal{G} \) be a (nonempty) family of (extended-valued) measurable functions in a measure space \( (\Omega, \mathcal{F}, \mu) \) and denote by \( \mathcal{G} \) the \( \sigma \)-algebra generated by \( \mathcal{G} \). A \( \mathcal{G} \)-measurable function \( g^* \) is called the essential supremum of the family \( \mathcal{G} \) if (1) \( g \leq g^* \) \( \text{a.e.} \), for every \( g \) in \( \mathcal{G} \) and (2) for any other \( \mathcal{G} \)-measurable function \( \bar{g} \) satisfying (1) with \( \bar{g} \) in lieu of \( g^* \), we must have \( \bar{g} \leq g^* \) \( \text{a.e.} \). Certainly the essential infimum is defined by using the family \( \mathcal{H} = \{-g : g \in \mathcal{G}\} \). On the other hand, we say that the family \( \mathcal{G} \) has a \( \sigma \)-finite support if there exits a monotone increasing sequence \( \{ G_i : i = 1, 2, \ldots \} \) of \( \mathcal{G} \)-measurable sets such that \( \mu(G_i) < \infty \), and \( \{ g \neq 0 : g \in \mathcal{G}\} \subset \bigcup_i G_i \).

If the essential supremum (infimum) exists then it must be unique almost everywhere, and then, it is usually denoted by \( \text{ess-sup}_{\mathcal{G}}\{g\} \) or \( \text{ess-sup}_{g \in \mathcal{G}}\{g\} \).
(\text{ess-inf}_G \{g\} \text{ or } \text{ess-inf}_{g \in G} \{g\}). It is clear that if two family of measurable functions are $G \subset H$ the same almost everywhere, in the sense that for each $g$ in $G$ there exists $h$ in $H$ such that $g = h$ a.e., then $\text{ess-sup}_G \{g\} = \text{ess-sup}_H \{h\}$. In other words, the essential supremum and essential infimum is a well defined operation on the equivalence classes of measurable functions $L^0(\Omega, \mathcal{F}, \mu)$, i.e., $G$ is actually a family of equivalence classes of measurable functions. The reader may want to take a look at Exercise 6.4 at the beginning of next Chapter.

A priori, for a countable family $G$ we can construct the essential supremum functions $g^*$ and $g_*$ by exhaustion, i.e., if $G = \{g_1, g_2, \ldots\}$ then the pointwise definitions $g^* = \lim_n \max_{i=1,2,\ldots,n} g_i$ and $g_* = \lim_n \min_{i=1,2,\ldots,n} g_i$ provided the essential supremum and infimum. Certainly, if the measure $\mu$ is $\sigma$-finite then any family $G$ has a $\sigma$-finite support. Similarly, since an integrable function must be zero outside of a $\sigma$-finite set, for a countable family $G$ of integrable functions we can construct a essential support as $\bigcup_{g \in G} \{g \neq 0\}$. Thus, in general, a essential support is loosely described as $\bigcup_{g \in G} \{g \neq 0\}$.

**Theorem 6.7.** Let $G$ be a (nonempty) family of (extended-valued) measurable functions in a measure space $(\Omega, \mathcal{F}, \mu)$ with $\sigma$-finite support, i.e., there exits a monotone increasing sequence $\{G_i : i = 1, 2, \ldots\}$ of $G$-measurable sets such that $\mu(G_i) < \infty$, and $\{g \neq 0 : g \in G\} \subset \bigcup_i G_i$. Then the essential supremum (infimum) function $g^*$ ($g_*$) of $G$ exits.

**Proof.** Naturally, only the case of the essential supremum needs to be discussed. First, by means of the transformation $g \mapsto \arctan(g)$ we may assume that the family $G$ is uniformly bounded with a $\sigma$-finite support. Moreover, we may also suppose that the family $G$ is closed (or stable) under the (pointwise) max operation, i.e., replace $G$ by the intersection of the families of $G$-measurable functions vanishing outside of the support of $G$, which are stable under the max operation (of finite many elements) and contains $G$. At this point, for each $i = 1, 2, \ldots$ we may construct a monotone increasing sequence $\{g_{i,1}^*, g_{i,2}^*, \ldots\}$ of functions in $G$ (with support in $G_i$) such that

$$
\lim_n \int_{G_i} g_{i,n}^* \, d\mu = \int_{G_i} \lim_n g_{i,n}^* \, d\mu = \sup_{g \in G} \int_{G_i} g \, d\mu < \infty, \quad \forall i.
$$

Now, for each fixed $i$, we verify that the pointwise (monotone) limit $g_i^* = \lim_n g_{i,n}^*$ is the essential supremum of the family $G_i = \{g \mathbb{1}_{G_i} : g \in G\}$. Indeed, to check condition (1) of Definition 6.6, for any $g$ in $G_i$ consider the sequence $\{h_{i,n}\} \subset G_i$, given by the pointwise maximum $h_{i,n} = \max\{g, g_{i,n}^*\}$. Since $g_{i,n}^*$ is monotone increasing in $n$, we have

$$
\int_{G_i} \max\{g, g_{i,n}\} \, d\mu = \lim_n \int_{G_i} \max\{g, g_{i,n}\} \, d\mu \leq \sup_{g \in G} \int_{G_i} g \, d\mu = \int_{G_i} \lim_n g_{i,n}^* \, d\mu.
$$

Because $\max\{g, g_i\} \geq g_i^* = \lim_n g_{i,n}^*$, we deduce $g \leq g_i^*$, a.e. On the other hand, $g_{i,n} = 0$ outside $G_i$ and so condition (2) is also satisfied.
Thus, for every $i$ we have established that (1) $g \mathbb{1}_{G_i} \leq g_i^*$ a.e., for every $g$ in $G$ and (2) if a $\mathcal{G}$ measurable function $h$ satisfies $g \mathbb{1}_{G_i} \leq h$ a.e., for every $g$ in $G$ then $h \leq g_i^*$. Finally, we obtain that pointwise maximum $g^* = \max_i g_i^*$ is indeed the desired essential supremum.

**Remark 6.8.** The result on the essential supremum (infimum) described in Theorem 6.7 can be presented as follows: if $\{f_i : i \in I\}$ is a family of (extended-valued) measurable functions with $\sigma$-finite support, then there exists a countable subset $J$ of indexes in $I$ such that the measurable function $x \mapsto \sup_{i \in J} f_i(x)$ satisfies (a) $f_i \leq f^*$ almost everywhere, for every $i$ in $I$, and (b) $f^* \leq f$ almost everywhere, for any other measurable function $f$ satisfying (a) with $f$ in lieu of $f^*$. Indeed, similar to Theorem 6.7, first assume (without any loss of generality) that each $f_i$ is $[0,1]$-valued measurable function vanishing outside $G_i = \bigcup_{k \geq 1} G_{i,k}$, increasing in $k$ and $\mu(G_{i,k}) < \infty$, for every $i,k$. Next, find a countable set of indexes $J_n^k$ such that

$$\sup_{n \geq 1} \int_{G_{i,k}} (\sup_{i \in J_n^k} f_i) \, d\mu = \sup_{N \in \mathcal{N}} \int_{G_{i,k}} (\sup_{i \in N} f_i) \, d\mu, \quad \forall k \geq 1,$$

where $\mathcal{N}$ is the family of all countable subset $N$ of indexes in $I$. Hence, use the countable set of indexes $J = \bigcup_{k,n} J_n^k$ to conclude.

Beside monotone properties (a) $\text{ess-sup}_{g \in \mathcal{G}} \{\Phi(g)\} = \Phi(\text{ess-sup}_{g \in \mathcal{G}} \{g\})$ for any measurable monotone increasing function $\Phi$, and (b) if $g \leq h$ a.e., for every $g \in \mathcal{G}$ and $h \in \mathcal{H}$ then $\text{ess-sup}_{g \in \mathcal{G}} \{g\} \leq \text{ess-sup}_{h \in \mathcal{H}} \{h\}$; by modifying a little the arguments of the above proof, we obtain

**Corollary 6.9.** Let $\mathcal{G}$ be a family of measurable functions, with $\sigma$-finite support, in a measure space $(\Omega, \mathcal{F}, \mu)$. Then there exists a sequence $\{g_i : i = 1, 2, \ldots\} \subset \mathcal{G}$ such that $\text{ess-sup}_{g \in \mathcal{G}} \{g\} = \lim_n \max_{1 \leq i \leq n} \{g_i\}$, and similarly for the essential infimum.

Therefore, the monotone convergence theorems for the integrals yield

**Corollary 6.10.** Let $\{f_i : i \in I\}$ be a family of measurable functions, with $\sigma$-finite support, in a measure space $(\Omega, \mathcal{F}, \mu)$. If there exists an integrable function $g$ such that for each $i$ in $I$ we have $f_i \geq g$ almost everywhere, then

$$-\infty < \sup_{i \in I} \left( \int_{\Omega} f_i \, d\mu \right) = \int_{\Omega} \left( \text{ess-sup}_{i \in I} f_i \right) \, d\mu \leq +\infty.$$

Similarly, if the function $\text{ess-inf}_{i \in I} \{f_i\}$ is integrable then

$$-\infty < \inf_{i \in I} \left( \int_{\Omega} f_i \, d\mu \right) = \int_{\Omega} \left( \text{ess-inf}_{i \in I} f_i \right) \, d\mu < +\infty,$$

provided some $f_i$ is also integrable.
On the other hand, given a family of measurable functions \( G \) and if the essential supremum \( \text{ess-sup}_{g \in G} \{ |g| \} \) exists then the essential support (unique up to a set of measure zero) exists and this ‘essential support’ is denoted by \( \text{ess-sup}_{g \in G} \{ |g| \} \neq 0 \). With this in mind, we can use the identities \( 1_{A \cap B} = \min\{1_A, 1_B\} \) and \( 1_{A \cup B} = \max\{1_A, 1_B\} \) to define the essential intersections and essential unions of any family \( \{ A_i : i \in I \} \) of measurable sets (actually, family of equivalence classes of measurable sets) as the essential infimum and the essential supremum, namely,

\[
\bigcap_{i \in I} A_i = \{ \text{ess-inf}_{i \in I} 1_{A_i} \neq 0 \} \quad \text{and} \quad \bigcup_{i \in I} A_i = \{ \text{ess-sup}_{i \in I} 1_{A_i} \neq 0 \},
\]

where both sets belongs to \( G \), the \( \sigma \)-algebra generated by either the family of functions \( \{ 1_{A_i} : i \in I \} \) or the family of sets \( \{ A_i : i \in I \} \). This yields a way of dealing with uncountable family of sets on a \( \sigma \)-finite measure space.

Let \( \{ f_i : i \in I \} \) be a (uncountable) family of measurable functions on a complete \( \sigma \)-finite measure space \( (\Omega, \mathcal{F}, \mu) \), and consider the pointwise essential suprema \( \overline{f}(x) = \text{ess-sup}_{i \in I} \{ f_i(x) \} \) and the essential supremum \( f^* = \text{ess-sup}_{i \in I} \{ f_i \} \). The pointwise suprema \( \sup_{i \in I} \{ f_i(x) \} \) is not well adapted for functions that may change in a set of measure zero (i.e., defined almost everywhere). Now, the equality \( \overline{f} = f^* \) a.e. is not expected, because \( \overline{f} \) would be necessarily measurable. However, Corollary 6.9 implies that \( \overline{f} \leq f^* \) a.e., and if the pointwise supremum \( \overline{f} \) is measurable then we must have \( f^* = \overline{f} \), almost everywhere. A similar argument applies to the pointwise infimum.

Similarly, given a (extended-valued) function (non-necessarily measurable) \( f \) with \( \sigma \)-finite support consider the family \( \overline{F} \) (or \( F \)) of all measurable functions \( g \) supported on the support of \( f \) and satisfying \( g \geq f \) a.e. (or \( g \leq f \) a.e., respectively). Define the upper (or lower) measurable envelop \( \overline{f} \) (or \( f \)) of \( f \) as essential infimum (or supremum) of \( \overline{F} \) (or \( F \)), i.e.,

\[
\overline{f} = \text{ess-inf}\{ \overline{F} \} = \text{ess-inf}\{ g \} \quad \text{and} \quad f = \text{ess-sup}\{ F \} = \text{ess-sup}\{ g \}.
\]

Invoking Corollary 6.9, there exist sequences \( \{ f_n \} \) and \( \{ \overline{f}_n \} \) of measurable functions satisfying \( f_n \leq f \leq \overline{f}_n \) a.e., and

\[
f = \lim_{n} \max_{k \leq n} f_k \quad \text{and} \quad \overline{f} = \lim_{n} \min_{k \leq n} \overline{f}_k.
\]

Thus, it is clear that \( f \leq f \leq \overline{f} \) a.e., and if the function \( f \) is measurable then \( \overline{f} = f = \overline{f} \), almost everywhere, since the essential extrema are always defined almost everywhere.

Going back to the pointwise essential infimum \( f_*(x) = \text{ess-inf}_{i \in I} \{ f_i(x) \} \) or supremum \( f^*(x) = \text{ess-sup}_{i \in I} \{ f_i(x) \} \), for a family of measurable functions \( \{ f_i : i \in I \} \) with \( \sigma \)-finite support, we may take \( f = f^* \) or \( f = f_* \) and consider the upper (or lower) measurable envelop, in short \( \text{ume}(f^*) \), \( \text{ume}(f_*) \), \( \text{lme}(f^*) \) and \( \text{lme}(f_*) \). The almost everywhere inequalities

\[
\text{lme}(f_*) \leq f_* \leq \text{ume}(f_*) \quad \text{and} \quad \text{lme}(f^*) \leq f^* \leq \text{ume}(f^*),
\]
follows from the previous discussion. Clearly, if dealing with a countable family then the essential supremum \( f^* \) and the essential infimum are measurable and the measurable envelops are not necessary. Moreover, these measurable envelops can be considered with respect to a sub \( \sigma \)-algebra.

Recall that \( L^0_\sigma(\Omega, \mu) \) denotes the real-valued (i.e., taking values in \([-\infty, +\infty]\) almost everywhere) measurable functions with \( \sigma \)-finite support, and \( L^0_\sigma(\Omega, \mu) \) the quotient space of their equivalence classes, under the \( \mu \) almost everywhere equality. In view of Theorem 6.7, for any family \( F \) of functions (actually, equivalence classes) in \( L^0_\sigma(\Omega, \mu) \) supported in a fixed \( \sigma \)-finite sets, the essential infimum \( \text{ess-inf}\{F\} \) and essential supremum \( \text{ess-sup}\{F\} \) are well defined.

Thus, for a given a sequence of measurable functions \( \{f_n\} \subset L^0_\sigma(\Omega, \mu) \), consider the family \( F(\{f_n\}) \) (or \( \overline{F(\{f_n\})} \), respectively) of all functions in \( L^0_\sigma(\Omega, \mu) \) with essential support in \( \text{ess-sup}(\{f_n\}) = \bigcup_n \text{ess-sup}(f_n) \) such that the limit \( \lim_n \mu(\{x \in A : f_n(x) < f(x)\}) = 0 \) (or \( \lim_n \mu(\{x \in A : f_n(x) > f(x)\}) = 0 \), respectively), for any measurable set \( A \) with finite measure \( \mu(A) < \infty \). The essential inferior and superior limits of the sequence \( \{f_n\} \) is defined as \( \text{ess-liminf}_n f_n = \text{ess-sup} \left( F(\{f_n\}) \right) \) and \( \text{ess-limsup}_n f_n = \text{ess-inf} \left( F(\{f_n\}) \right) \). If the measure \( \mu \) is finite, i.e., \( \mu(\Omega) < \infty \), then these conditions can be written in short as

\[
\text{ess-liminf}_n f_n = \text{ess-sup} \left\{ f \in F : \lim_n \mu(f_n < f) = 0 \right\} \quad \text{and} \\
\text{ess-limsup}_n f_n = \text{ess-inf} \left\{ f \in F : \lim_n \mu(f_n > f) = 0 \right\},
\]

where \( F \) is the family of all measurable functions.

**Proposition 6.11.** Let \( \{f_n\} \) be a sequence of elements in \( L^0_\sigma(\Omega, \mu) \). Then

\[
\liminf_n f_n \leq \text{ess-liminf}_n f_n \leq \text{ess-limsup}_n f_n \leq \limsup_n f_n, \quad \text{a.e.} \quad (6.1)
\]

Moreover, \( \text{ess-liminf}_n f_n = \text{ess-limsup}_n f_n = f \), almost everywhere, if and only if the sequence \( \{f_n\} \) converges to the function \( f \) in measure on any set with finite measure, i.e., if and only if \( \mu(\{x \in A : |f(x) - f_n(x)| \geq \varepsilon\}) \to 0 \), for every \( \varepsilon > 0 \) and any measurable \( A \) with \( \mu(A) < \infty \).

**Proof.** To show (6.1), if \( f \) is in \( F(\{f_n\}) \) then the relation

\[
\left\{ x \in A : \liminf_n f_n(x) < f(x) \right\} \subset \bigcup_{k \geq 1} \bigcap_{n \geq k} \left\{ x \in A : f_n(x) < f(x) \right\}
\]

yields \( \mu(\{x \in A : \liminf_n f_n(x) < f(x)\}) = 0 \) for every measurable set \( A \) with finite measure. Since \( \{f_n\} \) and \( f \) are supported in a \( \sigma \)-finite set, this implies \( \liminf_n f_n \geq f \), a.e., for every \( f \) in \( F(\{f_n\}) \). In view of Corollary 6.9, the essential supremum \( \text{ess-liminf}_n f_n \) is an almost everywhere pointwise limit of a finite maximum of function in \( F(\{f_n\}) \), and thus, the first inequality follows. Similarly, the last inequality \( \text{ess-limsup}_n f_n \leq \limsup_n f_n \), a.e., is obtained.
Finally, if \( f \) belongs to \( \mathcal{F}(\{f_n\}) \) and \( \overline{f} \) belongs to \( \overline{\mathcal{F}}(\{f_n\}) \) then
\[
\{ x \in A : \overline{f}(x) < f(x) \} \subset \{ x \in A : f_n(x) < f(x) \} \cup \\
\quad \cup \{ x \in A : f_n(x) > \overline{f}(x) \},
\]
which yields \( f \leq \overline{f} \), a.e. Hence, the \( \text{ess-limsup}_n f_n \leq \text{ess-limsup}_n f_n \), a.e., and the proof of the inequality (6.1) is completed.

Assume that \( f_n \rightarrow f \) in measure on any set with finite measure, and consider the functions \( f_\varepsilon = f - \varepsilon \mathbb{1}_F \) and \( f^\varepsilon = f + \varepsilon \mathbb{1}_F \), where \( F \) is the \( \sigma \)-finite measurable set containing the essential support of the sequence \( \{f_n\} \), \( F \supset \text{ess-sup}(\{f_n\}) \). The relation
\[
\{ x \in A : f_n(x) < f_\varepsilon(x) \} \cup \{ x \in A : f_n(x) > f^\varepsilon(x) \} \subset \\
\quad \subset \{ x \in A : |f_n(x) - f(x)| > \varepsilon \}
\]
implies that \( f_\varepsilon \) belongs to \( \mathcal{F} \) and \( f^\varepsilon \) belongs to \( \overline{\mathcal{F}} \). Hence, \( \text{ess-limsup}_n f_n \geq f_\varepsilon \) and \( \text{ess-limsup}_n f_n \leq f^\varepsilon \), and as \( \varepsilon \rightarrow 0 \), the equality \( \text{ess-limsup}_n f_n = \text{ess-limsup}_n f_n = f \) is deduced.

For the converse, suppose that \( \text{ess-limsup}_n f_n = \text{ess-limsup}_n f_n = f \), almost everywhere. Invoking Corollary 6.9, there exists two sequences \( \{\overline{f}_i\} \subset \overline{\mathcal{F}}(\{f_n\}) \) and \( \{f_i\} \subset \mathcal{F}(\{f_n\}) \) such that
\[
\lim_{k \to \infty} \max_{i \leq k} f_i(x) = \lim_{k \to \infty} \min_{i \leq k} \overline{f}_i(x) = f(x), \quad \text{a.e.} \tag{6.2}
\]
The relations
\[
\{ x \in A : f_n(x) > f(x) + \varepsilon \} \subset \{ x \in A : f(x) + \varepsilon < \min_{i \leq k} \overline{f}_i(x) \} \cup \\
\quad \cup \{ x \in A : f_n(x) > \overline{f}_1(x) \} \cup \ldots \cup \{ x \in A : f_n(x) > \overline{f}_k(x) \}
\]
and
\[
\{ x \in A : f_n(x) < f(x) - \varepsilon \} \subset \{ x \in A : f(x) - \varepsilon > \max_{i \leq k} f_i(x) \} \cup \\
\quad \cup \{ x \in A : f_n(x) < f_1(x) \} \cup \ldots \cup \{ x \in A : f_n(x) < f_k(x) \}
\]
imply that
\[
\lim_{n} \mu(\{ x \in A : |f_n(x) - f(x)| > \varepsilon \}) \leq \\
\quad \leq \mu(\{ x \in A : f(x) < \min_{i \leq k} \overline{f}_i(x) - \varepsilon \}) + \\
\quad \quad + \mu(\{ x \in A : f(x) > \max_{i \leq k} \overline{f}_i(x) + \varepsilon \}).
\]
In view of (6.2), as \( k \to \infty \), we deduce \( \lim_{n} \mu(\{ x \in A : |f_n(x) - f(x)| > \varepsilon \}) = 0 \), which prove the desired results.

**Exercise 6.6.** Based on Theorem 6.7, discuss and compare the statements in Exercises 1.22, 1.23, 4.23, 6.4 and 7.9. Consider also Remark 7.14.
6.3 Orthogonal Projection

Some of the properties valid in the Euclidean spaces \( \mathbb{R}^n \) or \( \mathbb{C}^n \) can be extended to some infinite dimensional spaces, such as \( L^2(\Omega, \mathcal{F}, \mu; \mathbb{R}^n) \) or \( L^2(\Omega, \mathcal{F}, \mu; \mathbb{C}^n) \). Perhaps, at this level, the reader should take a look at the beginning of the book Halmos [59] for a short introduction to Hilbert spaces.

Our interest is on the orthogonal projection and the representation of linear continuous functionals for the \( L^2 \) spaces, but there is not more effort in doing the arguments for a Hilbert space \( H \), a special class of Banach spaces, where the norm \( \| \cdot \| \) is given via a bilinear (or sesqui-linear, when working with complex-valued functions) continuous form \( (\cdot, \cdot) \), called scalar or inner product. For instance, for the \( L^2 \) space over the complex number, we have

\[
(f, g) = \int_{\Omega} f(x) \overline{g}(x) \mu(dx), \quad \forall f, g \in L^2(\Omega, \mathcal{F}, \mu; \mathbb{C}),
\]

and \( \|f\|^2 = (f, \overline{f}) \), where the notation \((\cdot, \cdot)\) is reserved for the duality, even when discussing real-valued functions \( f \) and the complex-conjugate operator \( f \mapsto \overline{f} \) is not used. This special form of the norm yields the so-called parallelogram equality \( \|f + g\|^2 + \|f - g\|^2 = 2\|f\|^2 + 2\|g\|^2 \), for every \( f, g \in H \), and the identity \( \|f + g\|^2 - \|f - g\|^2 = 4(f, g) \) allows the re-definition of the scalar product in term of the norm.

Actually recall that a Hilbert space is a vector space (on \( \mathbb{R} \) or \( \mathbb{C} \)) with a scalar (or inner) product satisfying:

a. \( (f, f) \geq 0 \), \( \forall f \in H \), and \( (f, f) = 0 \) only if \( f = 0 \);

b. \( (af + bg, h) = a(f, h) + b(g, h) \), \( \forall f, g, h \in H \) and \( a, b \in \mathbb{R} \) (or \( \mathbb{C} \));

c. \( (f, g) = (g, f) \), \( \forall f, g \in H \);

plus the completeness axiom: every Cauchy sequence \( \{f_n\} \subseteq H \), i.e., \( (f_n - f_m, f_n - f_m) \to 0 \) as \( n, m \to \infty \), is convergent, i.e., there exists \( f \in H \) such that \( (f_n - f, f_n - f) \to 0 \) as \( n, m \to \infty \). Hence, by considering the nonnegative quadratic \( r \mapsto \|f + rg\|^2 \) and using the linearity we deduce the Cauchy inequality,

\[
|(f, g)| \leq \|f\| \|g\|, \quad \forall f, g \in H,
\]

where the equality holds if and only if \( f \) and \( g \) are co-linear, i.e., \( f = cg \) or \( cf = g \) for some constant \( c \).

Two elements \( f, g \) in a Hilbert space \( H \) are called orthogonal if \( (f, g) = 0 \), and we may define the orthogonal complement of any nonempty subset \( V \subseteq H \) as \( V^\perp = \{h \in H : (h, v) = 0, \forall v \in V\} \). From the continuity and the linearity of the scalar product we deduce that \( V^\perp \) is a closed subspace of \( H \).

**Proposition 6.12 (Orthogonal Projection).** Let \( K \) be a closed convex set of \( H \). Then there exists a unique operator \( P : H \to K \) such that \( f \mapsto Pf \) satisfies

\[
(Pf - f, k - Pf) \geq 0, \quad \forall k \in K.
\]
Moreover, we have the estimate $\|Pf - Pg\| \leq \|f - g\|$ for every $f$ and $g$ in $H$; and if $K$ is a closed subspace then $P$ is linear and (6.3) becomes $(Pf - f, k) = 0$ for every $k$ in $K$.

Proof. First check the uniqueness. For any $g$ in $H$, $Pg$ satisfies

$$(Pg - g, k - Pg) \geq 0, \quad \forall k \in K.$$ 

Take $k = Pf$ and add (6.3) with $k = Pg$ to deduce $(f - g, Pf - Pg) \geq \|Pf - Pg\|^2$, which yields the estimate and the uniqueness. If $K$ is a closed subspace then $k - Pf \in K$ if and only if $k \in K$, i.e., (6.3) is equivalent to $(Pf - f, k) = 0$ for every $k \in K$ and the linearity of $P$ follows.

Next, for every fixed $f$ in $H$, consider the nonlinear functional $h \mapsto I(h) = (h - 2f, h)$ on $H$ and set $a = \inf \{I(h) : h \in K\}$. Since $I(h) \geq \|h\|^2 - 2\|f\| \|h\|$, we obtain $a \geq -\|f\|^2 > -\infty$, and so we can find a minimizing sequence $\{h_n\} \subset K$ such that $a \leq I(h_n) \leq a + n^{-1}$, for every $n \geq 1$. Because $K$ is convex, $h_{n,m} = (h_n + h_m)/2$ belongs to $K$ and we obtain

$$\|h_n\|^2 + \|h_m\|^2 - 2\|h_{n,m}\|^2 = I(h_n) + I(h_m) - 2I(h_{n,m}) \leq 1/n + 1/m,$$

after canceling the linear part of $I$. Hence, applying the parallelogram equality we have

$$\|h_n - h_m\|^2 = 2\|h_n\|^2 + 2\|h_m\|^2 - \|h_n - h_m\|^2 \leq 2/n + 2/m,$$

which proves that $\{h_n\}$ is a Cauchy sequence in $K$. The whole space $H$ is complete and $K$ is closed, therefore, there exists $h$ in $K$ such that $\|h_n - h\| \to 0$.

Now, for every $k$ in $K$ we have $h + \varepsilon(k - h)$ in $K$, for any $\varepsilon$ in $[0,1]$, and so $I(h + \varepsilon(k - h)) \geq I(h)$, i.e.,

$$2\varepsilon(h - f, k - h) + \varepsilon^2 \|k - h\|^2 \geq 0.$$ 

Thus, dividing by $\varepsilon$ and then vanishing $\varepsilon$, we get (6.3) with $Pf = h$. \qed

Sometimes, we write $P = P_K$ to emphasize the dependency on $K$. Also, $P_K$ is called the orthogonal projection over $K$. It is clear that $P_Kf = f$ for every $f$ in $K$, i.e., $P_K$ is idempotent. If $K$ is a closed subspace then $Pf - f$ belongs to $K^\perp$, i.e., $f = Pf + (f - Pf)$, which means $H = K \oplus K^\perp$. For any nonempty subset $V$ of $H$, we have defined its orthogonal complement $V^\perp = \{h \in H : (h, v) = 0, \forall v \in V\}$, but only when $V = K$ is a closed subspace we obtain $V = (V^\perp)^\perp$. Also, by writing $f = Pf + (f - Pf)$ we deduce $(Pf, g) = (Pf, Pg) = (f, Pg)$, for every $f, g \in H$, i.e., the projection is a symmetric operator.

If $(H, \|\cdot\|)$ is a Hilbert space then we denote by $H'$ its dual space, i.e., the space of all continuous linear functionals $T : H \to \mathbb{K}$, with $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$. We can check that $H'$ endowed with the dual norm

$$\|Tf\|_{H'} = \|Tf\|' = \sup \{|Tf| : \|f\| \leq 1\}$$

[

[Preliminary]

Menaldi

November 11, 2016

]
is a Banach space, and more detail is needed to see that $\| \cdot \|$ satisfies the parallelogram equality, and so, $H'$ is a Hilbert space.

Thus, if $f$ belongs to $H$ we can define $\Phi f : H \to \mathbb{R}$, $\Phi f(h) = (h, f)$, which results an element in $H'$. It is clear that the map $f \mapsto \Phi f$ is (sesqui-)linear from $H$ into $H'$, and Cauchy inequality shows that $\| \Phi f \|' = \| f \|$ for every $f$ in $H$.

**Theorem 6.13** (Riesz Representation). Let $H$ a Hilbert space. If $T : H \to \mathbb{K}$, with $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$, is a continuous linear functional then there exists $f$ in $H$ such that $T(h) = (h, f)$, for every $h$ in $H$. Moreover, the application $\Phi$ defined above is an isometry from $H$ onto its dual $H'$.

**Proof.** It is clear that only the fact that $\Phi$ is onto should be shown, i.e., given $T$ we can find $f$. To this purpose, denote by Ker$(T)$ the kernel or null space of $T$, i.e., all elements in $h \in H$ such that $T(h) = 0$. If Ker$(T) = H$ then $f = 0$ satisfies $\Phi(f) = T$, otherwise, there exits $g \neq 0$ in the orthogonal complement Ker$(T)^\perp$, and after diving by $T(g)$ if necessary, we may suppose $T(g) = 1$. Now, for any $h$ in $H$ we have $T(h - T(h)g) = 0$ and so $h - T(h)g$ belongs to Ker$(T)$, i.e., $(h - T(h)g, g) = 0$. This can written as $T(h)(g, g) = (h, g)$, for every $h$ in $H$. Hence, $f = g/(g, g)$ satisfies the desired condition. $\square$

Among other things this proves the

**Corollary 6.14.** Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $T : L^2 \to \mathbb{R}$ be a linear functional, which is continuous, i.e., for some constant $C > 0$,

$$|T(f)| \leq C \|f\|_2, \quad \forall f \in L^2.$$ 

Then there exists a unique function $g = g_T$ in $L^2$ such that

$$T(f) = \int_\Omega f g \, d\mu, \quad \forall f \in L^2,$$

and $\|T\|' = \|g\|_2$. $\square$

**Exercise 6.7.** Let $H$ be a Hilbert space with inner product $(\cdot, \cdot)$. First use Zorn's Lemma to show that any orthonormal set can be extended to an orthonormal basis $\{e_i : i \in I\}$, i.e., a maximal set of orthogonal vectors with unit length. Now, prove (1) that any element $x$ in $H$ can be written uniquely as $x = \sum_{i \in I} (x, e_i) e_i$, where only a countable number of $i$ have $(x, e_i) \neq 0$. Next, (2) verify that the cardinal of $I$ is invariant for any orthogonal basis. Finally, prove (3) that $H$ is isomorphic to the Hilbert space $l^2(I)$ of all functions $c : I \to \mathbb{R}$ (or complex valued) such that $\sum_{i \in I} |c_i|^2 < \infty$ (i.e., only a countable number of $c_i$ are nonzero and the series is convergent). $\square$

**Exercise 6.8.** If $(X, \mathcal{X}, \mu)$ is a measure space and $E$ belongs to $\mathcal{X}$ then we identify $L^2(E, \mu)$ with the subspace of $L^2(X, \mu)$ consisting of functions vanishing outside $E$, i.e., an element $f$ in $L^2(X, \mu)$ is in $L^2(E, \mu)$ if and only if $f = 0$ a.e. on $E^c$. Let $\{X_i\}$ be a sequence in $\mathcal{X}$ such that $X = \bigcup_i X_i$, and $\mu(X_i \cap X_j) = 0$ whenever $i \neq j$. Prove that (a) $\{L^2(X_i, \mu)\}$ is a sequence of mutually orthogonal
subspaces of $L^2(X,\mu)$ and (b) every $f$ in $L^2(X,\mu)$ can be written uniquely as $f = \sum_{i=1}^{\infty} f_i$, where $f_i$ belongs to $L^2(X_i,\mu)$ and the series converges in norm. Moreover, show that (c) if $L^2(X_i,\mu)$ is separable for every $i$ then so is $L^2(X,\mu)$. \[\square\]

6.4 Uniform Integrability

Let $(\Omega, \mathcal{F}, \mu)$ be a $\sigma$-finite measure space. On the vector space of integrable functions $\mathcal{L} = L^1$ we can define the semi-norm
\[
\|f\|_1 = \int_{\Omega} |f| \, d\mu,
\]
and using equivalence classes we obtain a norm and therefore $L^1 = \mathcal{L}/\sim$ or $L^1(\Omega, \mathcal{F}, \mu)$ is a normed space. Elements in $L^1$ are classes of equivalence, but we think of a function defined almost everywhere, and if necessary, we may complete the definition everywhere as along as the operations involving elements in $L^1$ does not depend on the particular extension used. Special attention is necessary to this point when dealing with measurable functions (or random variables or processes) in probability theory. Now, the inequality
\[
\varepsilon \mu(\{x : |f(x)| \geq \varepsilon\}) \leq \|f\|_1,
\]
shows that convergence in $L^1$ (also called in mean) implies convergence in measure. Note that
\[
\left| \int_{\Omega} (f - g) \, d\mu \right| \leq \int_{\Omega} |f - g| \, d\mu = \|f - g\|_1,
\]
if $f_n \to f$ in $L^1$ then the integral of $f_n$ converges to the integral of $f$, i.e, the integral is a continuous mapping from $L^1$ into $\mathbb{R}$.

In the construction of the integral we allow functions taking valued in the extended real numbers $\mathbb{R} = [-\infty, +\infty]$, but integrable functions are finite almost everywhere, i.e., the set $\{x \in \Omega : |f(x)| = \infty\}$ is negligible, the set $\{x \in \Omega : |f(x)| \geq \varepsilon\}$ is finite and thus, the support $\{x \in \Omega : |f(x)| > 0\}$ is $\sigma$-finite. Therefore, $L^1(\Omega, \mathcal{F}, \mu; \mathbb{R}) = L^1(\Omega, \mathcal{F}, \mu; \mathbb{R})$. Thus the space $L^0(\Omega, \mathcal{F}, \mu; \mathbb{R})$ of equivalence classes of measurable functions with values in $\mathbb{R}$ almost everywhere finite is $L^0(\Omega, \mathcal{F}, \mu; \mathbb{R})$, i.e., equivalence classes with the condition $\mu(\{|f| = \infty\}) = 0$.

Main Properties

The concept of uniform integrability applies to a set of measurable functions defined on a measure space, i.e., a subset of $L^0(\Omega, \mathcal{F}, \mu)$, but properly speaking, a subset of the space $L^0(\Omega, \mathcal{F}, \mu)$, namely, a subset of classes of equivalence classes of measurable functions.
Definition 6.15. Let \( \{ f_i : i \in I \} \) be a family of measurable functions almost everywhere finite (or elements in \( L^0 \)). If for every \( \varepsilon > 0 \) there exist \( \delta > 0 \) and a set \( A \in \mathcal{F} \) with \( \mu(A) < \infty \) such that for every \( F \in \mathcal{F} \) with \( \mu(F) < \delta \) we have

\[
\int_F f_i \, d\mu + \int_{A^c} |f_i| \, d\mu < \varepsilon, \quad \forall i \in I,
\]

then the family \( \{ f_i : i \in I \} \) is called \( \mu \)-equicontinuous. While, if for every \( \varepsilon > 0 \) there exist \( \delta > 0 \) and a set \( A \in \mathcal{F} \) with \( \mu(A) < \infty \) such that

\[
\int_{\{|f_i| \geq 1/\delta\}} |f_i| \, d\mu + \int_{A^c} |f_i| \, d\mu < \varepsilon, \quad \forall i \in I,
\]

then the family is called \( \mu \)-uniformly integrable. The words uniform integrability or uniformly integrable may be used when the reference measure \( \mu \) is clear from the context.

It is clear that if \( \mu(\Omega) < \infty \) then we can take \( A = \Omega \) and the above definition is greatly simplified. Both \( \mu \)-equicontinuous and \( \mu \)-uniformly integrable have in common the part relative to the set \( A \), namely, for every \( \varepsilon > 0 \) there exists a set \( A \in \mathcal{F} \) such that

\[
\mu(A) < \infty \quad \text{and} \quad \sup_{i \in I} \int_{A^c} |f_i| \, d\mu < \varepsilon. \quad (6.4)
\]

This condition is useful only when \( \mu(\Omega) = \infty \), it involves the behavior of the set \( \{|f_i| \leq \delta\} \), as \( \delta \rightarrow 0 \), and it could be called tightness.

On the other hand, if the family is almost everywhere equibounded, i.e., \( |f_i| \leq M \) almost everywhere, for every index \( i \) in \( I \), then \( \{|f_i| \geq 1/\delta\} \) is the empty set for \( \delta < 1/M \) and

\[
\int_F |f_i| \, d\mu \leq M \mu(F),
\]

proving that a part of \( \mu \)-equicontinuity and \( \mu \)-uniform integrability (except the tightness condition) is satisfied. Moreover, the condition on the set \( F \) could be called uniform or equi absolute continuity of the family of measures obtained from the integrals. Indeed, for an integrable function \( f \), the measure

\[
\mu_f(A) = \int_A |f| \, d\mu, \quad \forall A \in \mathcal{F},
\]

is absolutely continuous with respect to \( \mu \), then it should be clear that any finite family of integrable functions is a \( \mu \)-equicontinuous set. Also, the inequality

\[
\frac{1}{\delta} \mu(\{|f_i| \geq 1/\delta\}) \leq \int_{\{|f_i| \geq 1/\delta\}} |f_i| \, d\mu \leq \max_i \left\{ \int_{\Omega} |f_i| \, d\mu \right\}, \quad \forall \delta > 0,
\]

shows that any finite family of integrable functions is a \( \mu \)-uniformly integrable set. Furthermore, the following properties or comments hold true:
(1) If $\Omega = \{1, 2, \ldots\}$ and $\mu$ is the $\sigma$-finite measure $\mu(F) = \sum_{k=1}^{\infty} 1_{k \in F}$, for every $F \in 2^\Omega$, then there is no $F \in 2^\Omega$ such that $0 < \mu(F) < 1$, i.e., $\mu(F) < \delta < 1$ implies $F = \emptyset$ and therefore the condition on the uniform absolute continuity is always satisfied. Now, regarding condition (6.4), for any set $A \in 2^\Omega$ with $\mu(A) < n < \infty$ we have $A^c \supset \{k \geq n\}$. Thus, the sequence of functions $f_i : \Omega \to \mathbb{R}$, $f_i(k) = k^{-1/i} - (k + 1)^{-1/i}$ satisfies
\[
\int_{\{k \geq n\}} |f_i| \, d\mu = \sum_{k=n}^{\infty} f_i(k) = \lim_{k} (n^{-1/i} - (k + 1)^{-1/i}) = n^{-1/i} \geq n^{-1},
\]
and therefore, $\{f_i : i \geq 1\}$ fails to be $\mu$-equicontinuous (and $\mu$-uniformly integrable) because (6.4) is not satisfied.

(2) It is clear that if $\{f_i : i \in I\}$ is a $\mu$-equicontinuous family of functions then the equality
\[
\int_{F} |f_i| \, d\mu = \int_{F \cap \{f_i > 0\}} f_i \, d\mu - \int_{F \cap \{f_i < 0\}} f_i \, d\mu
\]
shows that the family $\{|f_i| : i \in I\}$ is also $\mu$-equicontinuous. Moreover, if $\{f_i : i \in I\}$ and $\{g_j : j \in J\}$ are two families of $\mu$-equicontinuous functions then for any constant $c$ the family $\{h_{i,j} = f_i + cg_i : i \in I, j \in J\}$ is also $\mu$-equicontinuous.

(3) If $\{f_i : i \in I\}$ is a family of $\mu$-uniformly integrable functions then the inequality
\[
\int_{F} |f_i| \, d\mu \leq \frac{\mu(F)}{\delta} + \int_{\{|f_i| \geq 1/\delta\}} |f_i| \, d\mu, \quad \forall F \in \mathcal{F}, \forall i \in I,
\]
shows that the family is also $\mu$-equicontinuous. Moreover, for $F = A$ with $\mu(A) < \infty$ as in the Definition 6.15, we deduce that $\sup_{i \in I} \|f_i\|_1 < \infty$.

(4) For a family $\{f_i : i \in I\}$ of $\mu$-equicontinuous functions, each member $f_i$ is an integrable function. Indeed if $F_n = \{|f_i| \geq n\}$ then $\bigcap_{n} F_n = \{|f| = \infty\}$, and for any set $A \in \mathcal{F}$ with $\mu(A) < \infty$ we have $\mu(F_n \cap A) \to 0$ as $n \to \infty$. Hence, take any $\varepsilon > 0$ and find $\delta > 0$ and $A \in F$ as above. Since $\Omega = F_n \cup (F_n^c \cap A^c) \cup (F_n^c \cap A)$, taking $n$ such that $\mu(F_n) < \delta$ and $F = F_n$ we deduce
\[
\int_{\Omega} |f_i| \, d\mu = \int_{F} |f_i| \, d\mu + \int_{F^c \cap A^c} |f_i| \, d\mu + \int_{F^c \cap A} |f_i| \, d\mu \leq 
\leq \int_{F} |f_i| \, d\mu + \int_{A^c} |f_i| \, d\mu + n\mu(A_i),
\]
i.e., each $f_i$ must be integrable.

(5) If $\{f_i : i \in I\}$ is a $\mu$-equicontinuous family of functions then we may have $\sup_i \|f_i\|_1 = \infty$. Indeed, if the measure $\mu$ is finite with an atom $A_1$ and
Then for any $\varepsilon > 0$ we can choose $A = \Omega$ and $\delta < \mu(A_1)$ to have
\[
0 = \int_F |f_i| \, d\mu + \int_{A^c} |f_i| \, d\mu \leq \delta \mu(\Omega) \leq \varepsilon, \quad \text{but} \quad \int_{\Omega} |f_i| \, d\mu = i,
\]
for every $F \in \mathcal{F}$ with $\mu(F) < \delta$. On the other hand, it is clear that if the set $A$ satisfying (6.4) can be decomposed into a finite number of measurable sets $A_1, \ldots, A_n$ such that
\[
\int_{A_k} |f_i| \, d\mu < \delta, \quad \forall k = 1, \ldots, n, \quad \forall i \in I,
\]
then $\sup \|f_i\|_1 < \infty$. Therefore, we deduce that if the family of measures induced by the functions $\{f_i : i \in I\}$ is uniformly absolutely $\mu$-continuous, i.e., for every $\varepsilon > 0$ there exist $\delta > 0$ such that for every $F \in \mathcal{F}$ with $\mu(F) < \delta$ we have
\[
\int_F f_i \, d\mu < \varepsilon, \quad \forall i \in I,
\]
and also, for every $\delta > 0$ there exist $A_1, \ldots, A_n$ in $\mathcal{F}$ such that $A = A_1 \cup \cdots \cup A_n$ has finite measure, $\mu(A) < \infty$, and
\[
\int_{A_k} |f_i| \, d\mu + \int_{A^c} |f_i| \, d\mu < \delta, \quad \forall k = 1, \ldots, n, \quad \forall i \in I,
\]
then $\{f_i : i \in I\}$ is $\mu$-uniformly integrable. In other words, if the measure $\mu$ is diffuse or non-atomic (i.e., for any set $A$ with $\mu(A) < \infty$ and for every $\delta > 0$ there is a decomposition of $A$ into a finite number of measurable sets, $A = A_1 \cup \cdots \cup A_n$ with $\mu(A_i) < \delta$, for every $i = 1, \ldots, n$), then any $\mu$-equicontinuous family $\{f_i : i \in I\}$ is also $\mu$-uniformly integrable.

(6) A family $\{f_i : i \in I\}$ of $\mu$-equicontinuous functions with $\sup_{i \in I} \|f_i\|_1 < \infty$ is $\mu$-uniformly integrable. Indeed, the inequality
\[
\mu(\{|f_i| \geq c\}) \leq \frac{1}{c} \int_{\Omega} |f_i| \, d\mu \leq \frac{1}{c} \sup_{i \in I} \|f_i\|_1,
\]
shows that for every $\delta > 0$ and any $i$ there exists $c$ sufficiently large so that the set $F_{i,c} = \{|f_i| \geq c\}$ satisfies $\mu(F_{i,c}) < \delta$. Now, for every $\varepsilon > 0$ there exists $\delta > 0$ such that
\[
F \in \mathcal{F} \quad \text{with} \quad \mu(F) < \delta \quad \implies \quad \int_F |f_i| \, d\mu \leq \varepsilon,
\]
and by taking $F = F_{i,c}$ we conclude. As a consequence we deduce that if $\{f_i : i \in I\}$ and $\{g_j : j \in J\}$ are two families of $\mu$-uniform integrable functions then for any constant $c$ the family $\{h_{i,j} = f_i + cg_j : i \in I, j \in J\}$ is also $\mu$-uniformly integrable.
Proposition 6.16. If $\delta$ and we conclude by taking $F$ such that $g$, function
Any family
\[ \int |f_i| \, d\mu \leq \int |g| \, d\mu \leq \frac{\varepsilon}{3}. \]
Next, if $A_n = \{ x \in \Omega : |g(x)| \geq 1/n \}$ then $1_{A_n} g \to 0$ almost everywhere as $n \to \infty$. Thus, Lebesgue dominate convergence Theorem 4.7 shows that
\[ \lim_n \int_{A_n} |g| \, d\mu = \lim_n \int_{\Omega} 1_{A_n} |g| \, d\mu = 0, \]
i.e., there exists $A = A_n$ such that
\[ \int_{A^c} |f_i| \, d\mu \leq \int_{A^c} |g| \, d\mu < \frac{\varepsilon}{3}, \quad \text{and} \quad \mu(A) \leq n \int_{\Omega} |g| \, d\mu, \]
and we conclude by taking $\delta \in (0, \delta_1]$ such that $\delta \mu(A) < \varepsilon/3$.

(8) Similarly to (7), any family $\{f_i : i \in I\}$ of measurable functions dominated
by a $\mu$-equicontinuous (or $\mu$-uniformly integrable) family $\{g_j : j \in J\}$ (i.e.,
for every $i$ there exists $j$ such that $|f_i| \leq g_j$ almost everywhere) results also
$\mu$-equicontinuous (or $\mu$-uniformly integrable).

(9) Let $r \to p(r), r > 0,$ be a nonnegative Borel measurable function such that
$p(r)/r \to \infty$ as $r \to \infty$, e.g., $p(r) = r^\alpha$ with $\alpha > 1$ or $p(r) = r \ln(1 + r)$. If $\sup_{i \in I} \|p(|f_i|)\|_1 = C < \infty$ then for every $\varepsilon > 0$ choose $\delta > 0$ such that
$p(r) \geq rC/\varepsilon$, for every $r \geq 1/\delta$. Thus
\[ \int_{\{|f_i| \geq 1/\delta\}} |f_i| \, d\mu \leq \frac{\varepsilon}{C} \|p(|f_i|)\|_1 \leq \varepsilon, \quad \forall i \in I. \]
Hence, if $\mu(\Omega) = \infty$ then we need only to add the condition (6.4), to deduce
that $\{f_i : i \in I\}$ is a $\mu$-uniformly integrable family.

Proposition 6.16. If $\{f_i : i \in I\}$ is $\mu$-equicontinuous then it is uniformly
$\sigma$-additive, i.e.,
\[ \forall \{B_n\} \subset \mathcal{F}, B_n \supset B_{n+1}, \cap_n B_n = \emptyset, \]
we have
\[ \lim_n \left( \sup_{i \in I} \int_{B_n} |f_i| \, d\mu \right) = 0. \tag{6.5} \]
Conversely, if either the index set $I$ is countable or the measure $\mu$ is $\sigma$-finite then
the uniform $\sigma$-additive condition (6.5) implies the $\mu$-equicontinuous condition.

Proof. Indeed, let $\{B_n\}$ be a decreasing sequence in $\mathcal{F}$ such that $\cap_n B_n = \emptyset$. From (6.4), for any $\varepsilon > 0$ there exists a measurable set $A$ with $\mu(A) < \infty$ such that
\[ \int_{B_n} |f_i| \, d\mu = \int_{B_n \cap A^c} |f_i| \, d\mu + \int_{B_n \cap A} |f_i| \, d\mu \leq \varepsilon + \int_{B_n \cap A} |f_i| \, d\mu. \]
Since $\mu(B_n \cap A) < \infty$ we have $\mu(B_n \cap A) \to 0$ and the $\mu$-equicontinuity (the condition on the set $F$) yields (6.5).

Conversely, if the index set $I$ is countable or the measure $\mu$ is $\sigma$-finite, the set $\bigcup_{i \in I} \{ f_i \neq 0 \}$ is contained in a $\sigma$-finite measurable set $E$, and so, there exists an increasing sequence of measurable sets $\{ E_k \}$ with $\mu(E_k) < \infty$ such that $E = \bigcup_k E_k$. Thus

$$\int_{E_k^c} |f_i| \, d\mu = \int_{E \setminus E_k} |f_i| \, d\mu \quad \text{and} \quad \lim_k \int_{E \setminus E_k} |f_i| \, d\mu = 0$$

where the limit is uniform in view of (6.5). Hence we deduce (6.4) with $A = E_k$ and $k$ sufficiently large. Therefore, if $\{ B_n : n \geq 1 \}$ is a sequence in $\mathcal{F}$ such that $\mu(B_n) < \infty$ and $\mu(A) = 0$ with $\bigcap_n B_n = A$, then $C_n = \bigcap_{k=1}^n (B_k \setminus A)$ forms a decreasing sequence satisfying $\bigcap_n C_n = \emptyset$ and the uniform $\sigma$-additivity (6.5) yields a contradiction with the $\mu$-equicontinuous condition.

Note that in the above Proposition 6.16, because the set $\{ f_i \neq 0 \}$ is $\sigma$-finite for every $i \in I$, the countability of the index set $I$ can be avoided if we assume that the $\sigma$-ring of all $\sigma$-finite measurable sets is countable generated. It is also clear that this condition is related to the separability of the Banach space $L^1(\Omega, \mathcal{F}, \mu)$. Another aspect of of the $\mu$-uniformly integrability is analyzed later, in Definition 6.24 and Theorem 6.25.

A family of measures $\{ \mu_i : i \in I \}$ is called uniform absolutely continuous on a measure space $(\Omega, \mathcal{F}, \mu)$ (or $\mu$-uniform absolutely continuous) if for every $\varepsilon > 0$ there exists $\delta > 0$ such that for every measurable set $F$ with $\mu(F) < \delta$ we have $\mu_i(F) < \varepsilon$, for every $i$ in $I$, see Remark 6.4. To mimic the $\mu$-equicontinuity of a family of integrable functions, we may add a tightness condition like: for every $\varepsilon > 0$ there exist measurable sets $A_i$ such that $\sup_{i \in I} \mu_i(A_i) < \infty$ and $\mu_i(A_i^c) < \varepsilon$, for every $i$ in $I$. Actually, given a family $(\Omega_i, \mathcal{F}_i, \mu_i)$ of measurable spaces and a family $\{ f_i : i \in I \}$ of measurable functions $f_i : \Omega_i \to \mathbb{R}$ almost everywhere finite, i.e., elements of $L^1(\Omega_i, \mathcal{F}_i, \mu_i)$, then we could say that they are equi-continuous if for every $\varepsilon > 0$ there exist $\delta > 0$ and sets $A_i$ in $\mathcal{F}_i$ such that $\sup_{i \in I} \mu_i(A_i) < \infty$ and for every set $F$ in $\mathcal{F}_i$ with $\mu_i(F_i) < \delta$ we have

$$\int_{F_i} f_i \, d\mu_i + \int_{A_i^c} |f_i| \, d\mu_i < \varepsilon, \quad \forall i \in I,$$

while the uniform integrability is stated with the condition

$$\int_{\{|f_i| \geq 1/\delta\}} |f_i| \, d\mu_i + \int_{A_i^c} |f_i| \, d\mu_i < \varepsilon, \quad \forall i \in I.$$

The reader can verify that most of the previous properties, (1) . . . , (9) above, remain true for this setting, where both the functions $f_i$ and the measures $\mu_i$ are indexed by $i$ in $I$. For instance, it is clear that property (7) make sense only when $\Omega_i = \Omega$ the same abstract space. Nevertheless, when comparing with the uniform $\sigma$-additivity property as in Proposition 6.16 we get some difficulties. In particular, if we are dealing with probability measures then we could
take $A_i = \Omega$ and virtually, this question does not occur. Similarly, if the abstract spaces $(\Omega_i, \mathcal{F}_i) = (\Omega, \mathcal{F})$ for every $i$ in $I$ then $\sup_{i \in I} \mu_i(A_i) < +\infty$ can be replaced by a more useful condition, namely, the family of finite measures \( \{\lambda_i(B) = \mu_i(B \cap A_i) : i \in I\} \) is uniformly $\sigma$-additive. Note that Vitali-Hahn-Saks Theorem 7.33 (later on) yields some light on this point, but the situation in general is complicated and some tools from Functional Analysis are really useful. Therefore, uniform absolutely continuity or uniform integrability or uniform $\sigma$-additivity for a family of measures is not completely discussed in these notes. Perhaps checking the viewpoint in Schilling [104, Chapter 16, pp. 163–175] may help.

**Mean Convergence**

When comparing the convergence almost everywhere (or in measure) with the mean convergence (i.e., in $L^1$) we encounter the following equivalence:

**Theorem 6.17 (Vitali).** Let \( \{f_n\} \) be a pointwise almost everywhere Cauchy sequence of integrable functions. Then \( \{f_n\} \) is a Cauchy sequence in $L^1$ if and only if \( \{f_n\} \) is $\mu$-equicontinuous.

**Proof.** First, for every $\varepsilon > 0$ there exist $A$ and $\delta > 0$ such that for any $F \in \mathcal{F}$ with $\mu(F) < \delta$ we have

\[
\int_F |f_n| \, d\mu + \int_{A^c} |f_n| \, d\mu + \delta \mu(A) < \frac{\varepsilon}{4}, \quad \forall n.
\]

Thus, the estimate

\[
\int_\Omega |f_n - f_k| \, d\mu \leq \int_{A^c} |f_n - f_k| \, d\mu + \int_{A \cap \{|f_n - f_k| \geq \delta\}} |f_n - f_k| \, d\mu + \delta \mu(A)
\]

shows that

\[
\int_\Omega |f_n - f_k| \, d\mu < \frac{\varepsilon}{2} + \int_{A \cap \{|f_n - f_k| \geq \delta\}} |f_n - f_k| \, d\mu.
\]

Since \( \{f_n\} \) is a almost everywhere Cauchy sequence and $\mu(A) < \infty$, there exists an index $n_\varepsilon$ such that $\mu(\{A \cap |f_n - f_k| \geq \delta\}) < \delta$, for every $n, k \geq n_\varepsilon$. Hence, taking $F = \{A \cap |f_n - f_k| \geq \delta\}$ we have

\[
\int_{A \cap \{|f_n - f_k| \geq \delta\}} |f_n - f_k| \, d\mu < \frac{\varepsilon}{2},
\]

i.e., $\|f_n - f_k\|_1 < \varepsilon$, for every $n_\varepsilon$.

Assuming that \( \{f_n\} \) is a Cauchy sequence in $L^1$, given $\varepsilon > 0$ there exists an index $n_\varepsilon$ such that $\|f_n - f_k\|_1 \leq \varepsilon/2$ for every $n, k \geq n_\varepsilon$. Thus for any $A \in \mathcal{F}$

\[
\int_{A^c} |f_n| \, d\mu \leq \int_{A^c} |f_{n_\varepsilon}| \, d\mu + \frac{\varepsilon}{2}, \quad \forall n \geq n_\varepsilon.
\]

(6.6)
Since each $f_i$ is integrable, for every $\delta > 0$ the set $F_{i,\delta} = \{|f_i| \geq \delta\}$ has finite measure,

$$\int_{F_{i,\delta}^c} |f_i| \, d\mu = \int_{\{0 < |f_i| < \delta\}} |f_i| \, d\mu \to 0 \text{ as } \delta \to 0,$$

and

$$\int_F |f_i| \, d\mu \to 0 \text{ as } \mu(F) \to 0,$$

for any fixed $i$. If $A_\delta = \bigcup_{i=1}^{n_\varepsilon} F_{i,\delta}$ then for every $k = 1, \ldots, n_\varepsilon$ we deduce

$$\int_{A_\delta^c} |f_k| \, d\mu \leq \int_{F_{k,\delta}^c} |f_k| \, d\mu \leq \max_{1 \leq i \leq n_\varepsilon} \int_{F_{i,\delta}^c} |f_i| \, d\mu \leq \frac{\varepsilon}{2},$$

provided $\delta$ is sufficiently small. Thus, there is $\delta$ such that for $A = A_\delta$ we have

$$\int_{A_\delta^c} |f_k| \, d\mu \leq \varepsilon, \quad \forall i \geq 1, \quad \text{and } \mu(A) < \infty.$$

Similarly,

$$\max_{1 \leq i \leq n_\varepsilon} \int_F |f_i| \, d\mu \to 0 \text{ as } \mu(F) \to 0,$$

and with $A^c = F$ in (6.6), we complete the proof of the $\mu$-equicontinuity. \(\square\)

Note that in the proof of the above result we have shown that if a Cauchy sequence in $L^0 \cap L^1$ (i.e., in measure) is $\mu$-equicontinuous then it is also a Cauchy sequence in $L^1$. Moreover, we may assume that the sequence $\{f_n\}$ of integrable functions is a Cauchy sequence in measure for every measurable set of finite measure, i.e., for every $\varepsilon > 0$ and every $A \in \mathcal{F}$ with $\mu(A) < \infty$ we have

$$\mu\left(\{x \in A : |f_n(x) - f_k(x)| \geq \varepsilon\}\right) \to 0 \text{ as } n, k \to \infty,$$

(6.7)

to deduce that $\{f_n\}$ is a Cauchy sequence in $L^1$ if and only if $\{f_n\}$ is $\mu$-equicontinuous. For instance, Lebesgue dominate convergence Theorem 4.7 or Proposition 7.1 can be restated as

$$\lim_n \int_{\Omega} f_n \, d\mu = \int_{\Omega} f \, d\mu,$$

for any $\mu$-equicontinuous sequence $\{f_n\}$ of measurable (necessarily integrable) functions which converges to a almost everywhere finite function $f$, in measure for every measurable set of finite measure.

Actually, it is a good exercise to revise the proof of Vitali Theorem 6.17 and to deduce the following generalization
Proposition 6.18. Let \( \{f_n\} \) be a Cauchy sequence of measurable functions, in measure for every measurable set of finite measure, i.e., (6.7). Then \( \{f_n\} \) is a \( p \)-Cauchy sequence, \( 0 < p < \infty \), i.e., for every \( \varepsilon > 0 \) there exists \( n_\varepsilon \) such that 
\[
\int_{\Omega} |f_n - f_m|^p \, d\mu < \varepsilon, \quad \forall n, m \geq n_\varepsilon,
\]
if and only if \( \{|f_n|^p\} \) is \( \mu \)-equicontinuous.

There are several application of Vitali Theorem 6.17, namely,

Corollary 6.19. Let \( \{f_n\} \) be a sequence of integrable functions which converges to an integrable function \( f \), in measure on every measurable set of finite measure. If 
\[
\lim_n \int_{\Omega} f_n^+ \, d\mu = \int_{\Omega} f^+ \, d\mu \quad \text{and} \quad \lim_n \int_{\Omega} f_n^- \, d\mu = \int_{\Omega} f^- \, d\mu \quad (6.8)
\]
then \( \{f_n\} \) is \( \mu \)-uniform integrable and \( f_n \to f \) in \( L^1 \).

Proof. From the elementary inequality 
\[
|a^+ - b^+| \vee |a^- - b^-| \leq |a - b| \leq |a^+ - b^+| + |a^- - b^-|, \quad \forall a, b \in \mathbb{R},
\]
we deduce that (1) \( f_n^+ \to f^+ \) and \( f_n^- \to f^- \) in measure (on every measurable set of finite measure), and that (2) \( f_n \to f \) in \( L^1 \) if and only if \( f_n^+ \to f^+ \) and \( f_n^- \to f^- \) in \( L^1 \). Hence we may assume that \( f_n \) and \( f \) are nonnegative, without any lost of generality.

Now, the dominate convergence implies that 
\[
\lim_n \int_{\Omega} (f_n \wedge f) \, d\mu = \int_{\Omega} f \, d\mu
\]
and by assumption
\[
\lim_n \int_{\Omega} (f_n + f) \, d\mu = 2 \int_{\Omega} f \, d\mu.
\]
Hence, the equality 
\[
|f_n - f| = f_n \vee f - f_n \wedge f = f_n + f - 2(f_n \wedge f),
\]
shows that \( \|f_n - f\|_1 \to 0 \).

Finally, the condition (6.8) implies that \( \sup_n \|f_n\|_1 < \infty \) and Vitali Theorem 6.17 (actually Proposition 6.18 with \( p = 1 \)) yields the \( \mu \)-equicontinuity of \( \{f_n\} \), and we deduce the \( \mu \)-uniform integrability condition.

Note that if 
\[
\lim_n \int_{\Omega} f_n \, d\mu = \int_{\Omega} f \, d\mu \quad \text{and} \quad \lim_n \int_{\Omega} |f_n| \, d\mu = \int_{\Omega} |f| \, d\mu
\]
then the relation \( a^\pm = (|a| \pm a)/2 \), for every real number \( a \), and the linearity of the integral show that (6.8) holds.
Proposition 6.20. If \( \{f_n\} \) is a sequence of \( \mu \)-uniformly integrable functions such that the negative part of the superior limit \( (\limsup_n f_n)^- \) is an integrable function then

\[
\limsup_n \int f_n \, d\mu \leq \int \limsup_n f_n \, d\mu. \tag{6.9}
\]

Proof. Since \( \{f_n\} \) are \( \mu \)-uniformly integrable functions, for a given \( \varepsilon > 0 \) there exists \( \delta > 0 \) and \( A \in \mathcal{F} \) with \( \mu(A) < \infty \) such that

\[
\int_{A \cap \{|f_n| > 1/\delta\}} |f_n| \, d\mu + \int_{A^c} |f_n| \, d\mu \leq \varepsilon, \quad \forall n,
\]

Now, decompose the integral

\[
\int_{A} f_n \, d\mu = \int_{A^c} f_n \, d\mu + \int_{A \cap \{|f_n| > 1/\delta\}} f_n \, d\mu + \int_{A \cap \{|f_n| \leq 1/\delta\}} f_n \, d\mu
\]

to check that for every \( n \) we have

\[
\int_{A} f_n \, d\mu \leq \varepsilon + \int g_n \, d\mu, \quad \text{where} \quad g_n = 1_{A \cap \{|f_n| \leq 1\}} f_n,
\]

with \( |g_n| \leq (1/\delta) \mathbb{1}_A \). Thus, by means of Lebesgue dominate convergence Theorem 4.7 we obtain

\[
\limsup_n \int f_n \, d\mu \leq \varepsilon + \int \limsup_n g_n \, d\mu.
\]

Hence, if \( \limsup_n f_n \geq 0 \) then \( \limsup_n g_n \leq \limsup_n f_n \) and we deduce (6.9). Otherwise, because \( (\limsup_n f_n)^- = g \) is an integrable function, we may replace \( f_n \) with \( f_n + g \) to obtain the desired inequality. \( \square \)

Let us comment on the above Proposition 6.20. First, for a measure space \((\Omega, \mathcal{F}, \mu)\), take a measurable set \( A \in \mathcal{F} \) with \( 0 < \mu(A) \leq 1 \) and find a finite partition \( A = \bigcup_{i=1}^{k} A_{k,i} \) with \( 0 < \mu(A_{k,i}) \leq 1/k \), for every \( i \). If \( \{a_k\} \) and \( \{b_k\} \) are two sequences of real numbers then we construct a sequence of functions \( \{f_n\} \) as follows: the sequence of integers \( \{1, 2, 3, \ldots, 10, 11, \ldots\} \) is grouped as \( \{(1); (2, 3); (4, 5, 6); (7, 8, 9, 10); \ldots\} \) where the \( k \) group has exactly \( k \) elements, i.e., for any \( n = 1, 2, \ldots \), we select first \( k = 1, 2, \ldots \), such that \((k-1)k/2 < n \leq k(k+1)/2\) and we write (uniquely) \( n = (k-1)k/2 + i \) with \( i = 1, 2, \ldots, k \) to define

\[
f_n(x) = \begin{cases} a_k & \text{if } x \in A \setminus A_{k,i}, \\ b_k & \text{if } x \in A_{k,i}. \end{cases}
\]

Now, we may construct a sequence of nonnegative function \( \{f_n\} \) with \( a_k = 0 \) and \( b_k = \sqrt{k} \), for every \( k \) so that

\[
\int_{\Omega} f_n \, d\mu = b_k \mu(A_{k,i}) \leq \frac{\sqrt{k}}{k} \leq \frac{2}{\sqrt{n}}.
\]
Because for every $x \in A$ there exist $i, k$ such that $x \in A_{k,i}$ and then $f_n(x) = b_k$, we deduce that $\limsup_n f_n(x) = \infty$, for every $x$ in $A$. Since the sequence $\{f_n\}$ converges (to 0) in $L^1$, this is an example of the strict inequality $0 < \infty$ in (6.9).

More general, if we choose $a_k = a$ and $b_k \to b$ with $\lim_k b_k/k = 0$ then

$$
\int_{\Omega} f_n \, d\mu = a\mu(A \setminus A_{k,i}) + b_k\mu(A_{k,i}) \to a\mu(A), \quad \text{as} \quad n \to \infty,
$$

while $\limsup_n f_n(x) = a \lor b$. This sequence $\{f_n\}$ is also $\mu$-uniformly integrable and the inequality (6.9) becomes $a\mu(A) \leq (a \lor b)\mu(A)$. For instance, if $a = -2$ and $b_k = -1$ we have the strict inequality $(-2)\mu(A) < (-1)\mu(A)$.

Another example, if the sequence $\{f_n\}$ admits a sub-sequence $\{f_{nk}\}$ convergence almost everywhere to some function $f$ then

$$
\limsup_n f_n \geq \limsup_k f_{nk} = f, \quad \text{a.e.}
$$

and therefore integrability of $(\limsup_n f_n)^-$ is guarantee. On the other hand, for a given $n = 1, 2, \ldots$, divide the interval $]0, 1]$ into $I_{k,n} = ](k-1)2^{-n}, k2^{-n}]$ with $k = 1, 2, \ldots, 2^n$ and define the functions $f_n(x) = (-1)^k$ for every $x$ in $I_{k,n}$ to check that

$$
|x : |f_n(x) - f_m(x)| \geq 1| = \frac{1}{2}, \quad \forall n \neq m,
$$

where $| \cdot |$ denotes the Lebesgue measure. Because $|f_n(x)| \leq 1$, this yields an example (in the Lebesgue space measure space $]0, 1]$) of an uniformly integrable sequence with no convergence (almost everywhere) sub-sequences, but with $\liminf_n f_n(x) = -1$ and $\limsup_n f_n(x) = 1$ for all $x$ in $]0, 1] \setminus \{k2^{-n} : k = 1, 2, \ldots, 2^n$, and $n = 1, 2, \ldots\}$. Again, in this case, the inequality (6.9) is satisfied strictly, $0 < 1$.

**Theorem 6.21.** Let $\{f_n\}$ be a sequence of $\mu$-uniformly integrable functions and consider the limits $\overline{f} = \limsup_n f_n$ and $\underline{f} = \liminf_n f_n$. Then

$$
\int_{\Omega} f \, d\mu \leq \liminf_n \int_{\Omega} f_n \, d\mu \leq \limsup_n \int_{\Omega} f_n \, d\mu \leq \int_{\Omega} \overline{f} \, d\mu,
$$

and necessarily, the positive part $(f)^+$ and the negative part $(\overline{f})^-$ are both integrable.

**Proof.** Since the sequence $\{f_n\}$ is $\mu$-integrable, we obtain that the numerical sequence $\{||f_n||_1\}$ is bounded, and in view of the above inequality (6.10), we deduce that $(f)^+$ and $(\overline{f})^-$ are both integrable.

Now, the point is to check that the extra assumption on the integrability of the limit $(\limsup_n f_n)^-$ is not necessary in Proposition 6.20.

Indeed, because $-f_n^- \leq f_n$ we obtain $-\liminf_n f_n^- = \limsup_n (-f_n^-) \leq \limsup_n f_n$ and therefore $(\limsup_n f_n)^- \leq \liminf_n f_n^-$. Hence, by Fatou lemma

---

1. a personal communication of N. Krylov
(see Theorem 4.6) we deduce
\[
\int \liminf_n f_n^- \, d\mu \leq \liminf_n \int f_n^- \, d\mu < \infty,
\]
which implies that \((\limsup_n f_n)^-\) is integrable.

Hence, we can apply Proposition 6.20 for the sequences \(\{f_n\}\) and \(\{-f_n\}\) to deduce the inequality (6.10).

Note that without assuming quasi-integrability for the limits \(f\) and \(\bar{f}\) (i.e., \((f)^+\) and \((\bar{f})^-\) are both integrable), the \(\mu\)-uniform integrability cannot be replaced by \(\mu\)-equicontinuity. Indeed, similar to example in (5) after Definition 6.15, for a finite measure \(\mu\) with two atoms \(A_1\) and \(A_2\), \(\Omega = A_1 \cup A_2\), and the sequence of functions with \(f_i = (i/\mu(A_1)) \mathbb{1}_{A_1}\) and \(f_i = (-i/\mu(A_2)) \mathbb{1}_{A_2}\), \(i \geq 1\) is \(\mu\)-equicontinous, but the limit \(f(A_k) = \lim_i f_i(A_k)\) is \(+\infty\) for \(k = 1\) and \(-\infty\) for \(k = 2\),
\[
\int \Omega \max(f_i) \, d\mu = 0,
\]
for any \(i \geq 1\), and \(f\) is not quasi-integrable.

**Convergence in Norm**

The following result, which is also an application of Vitali Theorem 6.17 makes a connection with \(p\)-integrable functions.

**Proposition 6.22.** Let \(\{f_n\}\) be a bounded sequence in \(L^p(\Omega, \mathcal{F}, \mu)\), for some \(0 < p < \infty\). If \(f_n\) converges to \(f\) in measure for every measurable set of finite measure then
\[
\lim_n \int \Omega \max(|f_n| - |f_n - f| - |f|^p) \, d\mu = 0.
\]

**Proof.** Firstly, for every \(\varepsilon > 0\) there exists a constant \(C_\varepsilon > 0\) such that for every numbers \(a\) and \(b\) we have
\[
\max(|a + b|^p - |b|^p) \leq \varepsilon |b|^p + C_\varepsilon |a|^p.
\]
Indeed, if \(0 < p \leq 1\) the the simple estimate \(|a + b|^p \leq |a|^p + |b|^p\) yields estimate (6.11). Now, for \(1 < p < \infty\), the function \(t \mapsto |t|^p\) is convex; and so \(|a + b|^p \leq (|a| + |b|)^p \leq (1 - \lambda)^{1-p}|a|^p + \lambda^{1-p} |b|^p\), for any \(\lambda\) in \((0, 1)\). Hence, by taking \(\lambda = (1 + \varepsilon)^{1/(1-p)}\) we deduce (6.11) with \(p > 1\).

Secondly, by assumption
\[
\int \Omega |f_n|^p \, d\mu \leq C < \infty, \quad \forall n,
\]
and since $|f_n - f|^p \leq 2^p(|f_n|^p + |f|^p)$, we obtain
\[
\int_{\Omega} |f_n - f|^p \, d\mu \leq 2^{p+1} C, \quad \forall n,
\]
for every $0 < p < \infty$.

Next, estimate (6.11) implies
\[
||f_n|^p - |f_n - f|^p - |f|^p| \leq ||f_n|^p - |f_n - f|^p| + |f|^p \leq \varepsilon |f_n - f|^p + (1 + C\varepsilon)|f|^p.
\]

Hence, by setting $g_n = (||f_n|^p - |f_n - f|^p - |f|^p| - \varepsilon |f_n - f|^p)^+$, we have $0 \leq g_n \leq (1 + C\varepsilon)|f|^p$ and so, Vitali Theorem 6.17 yields
\[
\lim_{n} \int_{\Omega} g_n \, d\mu = 0.
\]

Therefore
\[
\limsup_{n} \int_{\Omega} ||f_n|^p - |f_n - f|^p - |f|^p| \, d\mu \leq \varepsilon 2^{p+1} C,
\]
i.e., the desired result.

\[\square\]

• \textbf{Remark 6.23.} In particular, if $f_n$ converges to $f$ in measure for every measurable set of finite measure and $\|f_n\|_p \to \|f\|_p$ then $\|f_n - f\|_p \to 0$.

Also, we may generalize Definition 6.15 to $L^p$, with $1 \leq p < \infty$, as follows:

**Definition 6.24.** Let \(\{f_i : i \in I\}\) be a family of measurable functions almost everywhere finite (or elements in $L^0$). If for every $\varepsilon > 0$ there exist $\delta > 0$ and $A \in \mathcal{F}$ with $\mu(A) < \infty$ such that
\[
\int_{\{|f_i| \geq 1/\delta\}} |f_i|^p \, d\mu + \int_{A^\varepsilon} |f_i|^p \, d\mu < \varepsilon, \quad \forall i \in I,
\]
then the family is called \(\mu\)-uniformly integrable of order $p$, for $0 < p < \infty$.

\[\square\]

Actually, this means that a family \(\{f_i : i \in I\}\) of measurable functions almost everywhere finite is $\mu$-uniformly integrable (or $\mu$-equicontinuous) of order $p$ if and only if \(\{|f_i|^p : i \in I\}\) is $\mu$-uniformly integrable (or $\mu$-equiconitnuous).

**Theorem 6.25.** Let \(\{f_i : i \in I\}\) be a family of measurable functions almost everywhere finite in a measure space $(\Omega, \mathcal{F}, \mu)$. Then the following statements are equivalent:

1. \(\{f_i : i \in I\}\) are $\mu$-uniformly integrable of order $p$;
2. for any $\varepsilon > 0$ there exists a nonnegative $p$-integrable function $g$ such that
\[
\int_{\{|f_i| \geq g\}} |f_i|^p \, d\mu < \varepsilon, \quad \forall i \in I;
\]
(3) (a) there exists a constant $C > 0$ such that

$$
\int_{\Omega} |f_i|^p \, d\mu \leq C, \quad \forall i \in I,
$$

and (b) for every $\varepsilon > 0$ there exist a constant $\delta > 0$ and a nonnegative $p$-integrable function $h$ such that for every $F \in \mathcal{F}$

$$
\int_F h^p \, d\mu < \delta \quad \text{implies} \quad \int_F |f_i|^p \, d\mu < \varepsilon, \quad \forall i \in I.
$$

Proof. (1) $\Rightarrow$ (2): Given $\varepsilon > 0$ choose $\delta > 0$ and $A \in \mathcal{F}$ as in Definition 6.24 and set $g = (1/\delta) \mathbb{1}_A$. By means of the inequality

$$
1_{\{|f_i| \geq g\}} |f_i|^p \leq 1_{A^c} |f_i|^p + 1_{\{|f_i| \geq 1/\delta\}} |f_i|^p,
$$

we obtain

$$
\int_{\{|f_i| \geq g\}} |f_i|^p \, d\mu \leq \int_{A^c} |f_i|^p \, d\mu + \int_{\{|f_i| \geq 1/\delta\}} |f_i|^p \, d\mu \leq \varepsilon, \quad \forall i \in I.
$$

and because $\mu(A) < \infty$ the function $g$ is $p$-integrable.

(2) $\Rightarrow$ (3): For every nonnegative $p$-integrable function $g$ and every $F \in \mathcal{F}$ we have

$$
\int_F |f_i|^p \, d\mu = \int_{F \cap \{|f_i| \geq g\}} |f_i|^p \, d\mu + \int_{F \cap \{|f_i| < g\}} |f_i|^p \, d\mu \leq \int_{\{|f_i| \geq g\}} |f_i|^p \, d\mu + \int_F |g|^p \, d\mu.
$$

Hence, for $F = \Omega$ and $g$ as in (2) for $\varepsilon = 1$ we get (3) (a). Similarly, taking $g$ as in (2) for $\varepsilon/2$ and $h = g$, we deduce (3) (b).

(3) $\Rightarrow$ (1): Given $\varepsilon > 0$ choose $\delta > 0$ and $h \geq 0$ as in (3) (b). Define $A_r = \{h \leq r\}$ to check that

$$
r^p \mu(A_r) \leq \int_{A_r} h^p \, d\mu \leq \int_{\Omega} h^p \, d\mu < \infty,
$$

which means that $A_r$ has finite measure for every $r > 0$. Moreover, on the complement,

$$
\int_{A_r^c} h^p \, d\mu = \int_{\Omega} h^p \mathbb{1}_{h > r} \, d\mu \to 0 \quad \text{as} \quad r \to \infty.
$$

Hence, if $r$ is sufficiently large then take $A = A_r$ to deduce that the condition (3) (b) yields

$$
\int_{A_r^c} h^p \, d\mu < \delta \quad \text{implies} \quad \int_{A_r^c} |f_i|^p \, d\mu < \varepsilon, \quad \forall i \in I.
$$
i.e., one part Definition 6.24 of $\mu$-integrability of order $p$. Next, because $h$ is $p$-integrable, there exists $\delta' > 0$ such that

$$\mu(F) < \delta' \quad \text{implies} \quad \int_F h^p \, d\mu < \delta.$$ 

Thus, take $C$ as in (3) (a) to check the inequality

$$r\mu(\{|f_i| \geq r\}) \leq \int_{\{|f_i| \geq r\}} |f_i| \, d\mu \leq \int_{\Omega} |f_i| \, d\mu \leq C, \quad \forall i \in I.$$ 

Now, if $r$ sufficiently large so that $C/r \leq \delta'$ then $\mu(\{|f_i| \geq r\}) < \delta'$, and the condition (3) (b) with the set $F = \{|f_i| \geq r\}$ yields

$$\int_{\{|f_i| \geq r\}} |f_i| \, d\mu \leq \varepsilon,$$

proving the $\mu$-integrability of order $p$.

Alternatively, the proof may continue as follows:

(3) ⇒ (2): Given $\varepsilon > 0$ choose $\delta > 0$ and $h \geq 0$ as in (3) (b). If $C$ is as in (3) (a), then the inequality

$$a^p \int_{\{|f_i| \geq ah\}} |h|^p \, d\mu \leq \int_{\{|f_i| \geq ah\}} |f_i|^p \, d\mu \leq \int_{\Omega} |f_i|^p \, d\mu \leq C, \quad \forall a > 0,$$

shows we can select $a$ sufficiently large so that

$$\int_{\{|f_i| \geq ah\}} |h|^p \, d\mu \leq \frac{C}{a^p} \leq \delta.$$ 

Hence, the condition (3) (b) with $F = \{|f_i| \geq ah\}$ yields

$$\int_{\{|f_i| \geq ah\}} |f_i|^p \, d\mu \leq \varepsilon, \quad \forall i \in I,$$

i.e., we deduce (2) with $g = ah$.

(2) ⇒ (1): Given $\varepsilon > 0$ find $g$ as in (2). Thus, the inequality

$$\int_{\{|f_i| \geq 1/\delta\}} |f_i|^p \, d\mu = \int_{\{|f_i| \geq 1/\delta\} \cap \{|f_i| \geq g\}} |f_i|^p \, d\mu +$$

$$+ \int_{\{|f_i| \geq 1/\delta\} \cap \{|f_i| < g\}} |f_i|^p \, d\mu \leq$$

$$\leq \int_{\{|f_i| \geq g\}} |f_i|^p \, d\mu + \int_{\{g \geq 1/\delta\}} g^p \, d\mu.$$ 

proves that

$$\int_{\{|f_i| \geq 1/\delta\}} |f_i|^p \, d\mu \leq \varepsilon + \int_{\{g \geq 1/\delta\}} g^p \, d\mu, \quad \forall i \in I,$$
and since
\[
\lim_{\delta \to 0} \int_{\{g \geq 1/\delta\}} g^p \, d\mu = 0, \quad \forall g \in L^p,
\]
we can find \( \delta > 0 \) such that
\[
\int_{\{|f_i| \geq 1/\delta\}} |f_i|^p \, d\mu \leq 2\varepsilon, \quad \forall i \in I.
\]
Also the set \( A_r = \{ g \geq r \} \) has finite measure for every \( r > 0 \), and
\[
\int_{A_c^c} |f_i|^p \, d\mu = \int_{A_c^c \cap \{|f_i| \geq g\}} |f_i|^p \, d\mu + \int_{A_c^c \cap \{|f_i| < g\}} |f_i|^p \, d\mu \leq \int_{\{|f_i| \geq g\}} |f_i|^p \, d\mu + \int_{\{g < r\}} g^p \, d\mu
\]
ensures that there exists \( A = A_r \), for some \( r > 0 \), such that
\[
\int_{A_c} |f_i|^p \, d\mu \leq 2\varepsilon.
\]
Hence, the family \( \{f_i : i \in I\} \) is \( \mu \)-uniformly integrable of order \( p \).

\begin{itemize}
  \item \textbf{Remark 6.26.} Note that a measure space \( (\Omega, \mathcal{F}, \mu) \) is \( \sigma \)-finite if and only if there exists a strictly positive integrable function \( h \). Indeed, if \( \mu \) is \( \sigma \)-finite then there exists an increasing sequence \( \{\Omega_k\} \subset \mathcal{F} \) such that \( \Omega = \bigcup_k \Omega_k \) and \( 0 < \mu(\Omega_k) < \infty \). Thus, the function \( h = \sum_k (2^{-k}/\mu(\Omega_k)) \mathbb{1}_{\Omega_k} > 0 \) is integrable, for every \( p \). Conversely, if there exists a strictly positive integrable function \( h \) then the sets \( \Omega_k = \{h \geq 1/k\} \) satisfy the required condition. Moreover, if \( h > 0 \) and integrable then \( h^{1/p} \) is strictly positive and \( p \)-integrable.
  \end{itemize}

The following result applies for \( \sigma \)-finite measure spaces.

\begin{corollary}
Let \( h \) be a strictly positive \( p \)-integrable function on a measure space \( (\Omega, \mathcal{F}, \mu) \). Then we can revise the statements in Theorem 6.25 as follows:
\( (2) \) becomes: for every \( \varepsilon > 0 \) there exists \( \alpha > 0 \) such that
\[
\int_{\{|f_i| \geq \alpha h\}} |f_i|^p \, d\mu < \varepsilon, \quad \forall i \in I;
\]
and \( (3) \) (b) reads as: for every \( \varepsilon > 0 \) there exist a constant \( \delta > 0 \) such that for every \( F \in \mathcal{F} \)
\[
\int_F h^p \, d\mu < \delta \quad \text{implies} \quad \int_F |f_i|^p \, d\mu < \varepsilon, \quad \forall i \in I.
\]
The equivalence among properties (1), (2) and (3) remains true.
\end{corollary}
Proof. The inequalities

\[
\int_{\{|f_i| \geq \alpha h\}} |f_i|^p d\mu = \int_{\{|f_i| \geq \alpha h\} \cap \{|f| \geq g\}} |f_i|^p d\mu + \int_{\{|f_i| \geq \alpha h\} \cap \{|f| < g\}} |f_i|^p d\mu \leq \\
\leq \int_{\{|f| \geq g\}} |f_i|^p d\mu + \int_{\{|g| \geq \alpha h\}} |g|^p d\mu
\]

and

\[
\int_F |f_i|^p d\mu \leq \int_{\{|f_i| \geq \alpha h\}} |f_i|^p d\mu + \alpha^p \int_F |h|^p d\mu
\]

provide the equivalence.

Note that if \( \mu(\Omega) < \infty \) then we can take \( A = \Omega \) in the Definition 6.24, i.e., \( g = 1/\delta \) and \( h = 1 \) in conditions (2) and (3) of Theorem 6.25 and Corollary 6.27. For instance, the interested reader may consult the books by Bauer [10, Section 21, pp. 121–131], among others.

Also we have a practical criterium to check the \( \mu \)-uniformly integrability of order \( q \), compare with (9) in Section 6.4.

**Proposition 6.28.** Let \( \{f_i : i \in I\} \) be a family of measurable functions equi-bounded on \( L^p(\Omega, \mathcal{F}, \mu) \) and \( L^p(\Omega, \mathcal{F}, \mu) \), for some \( 0 < r < p < \infty \), i.e., there exists a constant \( C > 0 \) such that

\[
\left( \int_{\Omega} |f_i|^p d\mu \right) \leq C^p, \quad \left( \int_{\Omega} |f_i|^r d\mu \right) \leq C^r \quad \forall i \in I.
\]

Then for any \( q \) in the open interval \( (r, p) \) and for every \( \varepsilon > 0 \) there exist \( \delta > 0 \) and a measurable set \( A \) with \( \mu(A) < \infty \) such that

\[
\int_{\{|f_i| \geq 1/\delta\}} |f_i|^q d\mu + \int_{\Omega \setminus A} |f_i|^q d\mu < \varepsilon, \quad \forall i \in I,
\]

i.e., the family \( \{f_i : i \in I\} \) is \( \mu \)-uniformly integrable of order \( q \). \( \square \)

**Proof.** Indeed, because the family is bounded in \( L^p \), it has to be a family of measurable functions taking finite values almost everywhere and the set \( \{|f_i| \geq 1/\delta\} \) has finite \( \mu \)-measure for every \( \delta > 0 \) and \( i \) in \( I \).

Now, write \( q = sp \) with \( 0 < s < 1 \) to deduce

\[
|f_i|^q \frac{\delta^{-(1-s)r}}{\delta} \mathbb{1}_{\{|f_i| \geq 1/\delta\}} \leq |f_i|^{sp} |f_j|^{(1-s)r}, \quad \forall i, j \in I,
\]

and then apply Hölder inequality to obtain

\[
\delta^{-(1-s)r} \int_{\{|f_j| \geq 1/\delta\}} |f_i|^q d\mu \leq \left( \int_{\Omega} |f_i|^p d\mu \right)^s \left( \int_{\Omega} |f_j|^r d\mu \right)^{1-s} \leq C.
\]
Hence, for a given \( \varepsilon > 0 \) choose \( \delta > 0 \) so that \( C\delta^{(1-s)r} < \varepsilon/2 \). Next, fix \( j \) in \( I \) and take \( A = \{|f_j| \geq 1/\delta\} \), which has finite \( \mu \)-measure and satisfies

\[
\int_{\Omega \setminus A} |f_i|^q d\mu = \int_{\{|f_j| \geq 1/\delta\}} |f_i|^q d\mu \leq C\delta^{(1-s)r} < \varepsilon/2.
\]

Similarly, take \( i = j \) to deduce

\[
\int_{\{|f_i| \geq 1/\delta\}} |f_i|^q d\mu \leq C\delta^{(1-s)r} < \varepsilon/2,
\]

proving the \( \mu \)-uniform integrability of order \( q \).

To complete this section, we show a relation of totally bounded (or pre-compact) sets in \( L^p \) and uniformly integrable sets of order \( p \). Recall that a family of functions \( \{f_i : i \in I\} \) is a totally bounded subset of \( L^p(\Omega, F, \mu) \) if for every \( \varepsilon > 0 \) there exists a finite subset of indexes \( J \subset I \) such that for every \( i \) in \( I \) there exists \( j \) in \( J \) satisfying \( \|f_i - f_j\|_p < \varepsilon \), i.e., any element in \( \{f_i : i \in I\} \) is within a distance \( \varepsilon \) from the finite set \( \{f_j : j \in J\} \). Sometimes \( \{f_j : j \in J\} \) is called an \( \varepsilon \)-net relative to \( \{f_i : i \in I\} \). This concept of totally bounded sets is equivalent to pre-compact set on a complete metric space, in particular, this also applied to the topological vector space \( L^p(\Omega, F, \mu) \) with \( 0 < p < 1 \) and the distance \( \text{d}(f, g) = \|f - g\|_p \).

**Proposition 6.29.** Let \( \{f_i : i \in I\} \) be a totally bounded subset of \( L^p(\Omega, F, \mu) \), with \( 0 < p < \infty \). Then \( \{f_i : i \in I\} \) is \( \mu \)-uniformly integrable of order \( p \).

**Proof.** For a given \( \varepsilon > 0 \), denote by \( J_\varepsilon \subset I \) the finite subset of indexes obtained from the totally boundedness property. We assume \( 1 \leq p < \infty \) to able to use the triangular inequality \( \|f - g\|_p \leq \|f\|_p + \|g\|_p \). The case where \( 0 < p < 1 \) is treated analogously, by means of the inequality \( \|f - g\|_p \leq \|f\|_p^{p'} + \|g\|_p^{p'} \).

Since the finite family \( \{f_j : j \in J_\varepsilon\} \) is \( \mu \)-equicontinuous (also \( \mu \)-uniformly integrable) of order \( p \), for this same \( \varepsilon > 0 \) there exists \( \delta > 0 \) and \( A \) in \( F \) such that

\[ F \in F \text{ with } \mu(F) < \delta \quad \implies \quad \int_F |f_j|^p d\mu \leq \varepsilon, \quad \forall j \in J_\varepsilon, \]

and

\[ \int_{A^c} |f_j|^p d\mu < \varepsilon, \quad \forall j \in J_\varepsilon, \]

which combined with the inequality

\[
\inf \{\|f_i - f_j\|_p : j \in J_\varepsilon\} \leq \varepsilon, \quad \forall i \in I,
\]

show that \( \{|f_i|^p : i \in I\} \) is \( \mu \)-equicontinuous.
Now, we redo the argument to show that \( \mu \)-equicontinuity plus uniformly bounded is equivalent to \( \mu \)-uniformly integrability. Indeed, the family of functions \( \{f_i : i \in I\} \) is uniformly bounded in \( L^p \), namely
\[
\|f_i\|_p \leq \|f_i - f_j\|_p + \|f_j\|_p \leq \varepsilon + \sup \{\|f_j\|_p : j \in J_\varepsilon\} < \infty,
\]
and the inequality
\[
\mu(\{|f_i| \geq c\}) \leq \frac{1}{c^p} \int |f_i|^p d\mu \leq \frac{1}{c^p} \sup_{i \in I}\|f_i\|_p^p,
\]
shows that for every \( \delta > 0 \) there exists \( c > 0 \) (sufficiently large) so that the set \( F_{i,c} = \{|f_i| \geq c\} \) satisfies \( \mu(F_{i,c}) < \delta \), for every \( i \). Hence, by taking \( F = F_{i,c} \) within the \( \mu \)-equicontinuity condition for the whole family \( \{|f_i|^p : i \in I\} \), we deduce that \( \{f_i : i \in I\} \) is also \( \mu \)-uniformly integrable of order \( p \).

Certainly, the converse of Proposition 6.29 fails. For instance, the sequence \( \{f_n(x) = \sin nx\} \) on \( L^2(\pi, \pi) \) satisfies \( \|f_n\|_2 = 2\pi \) so that any \( L^2 \)-convergent subsequence would converge to a some function \( f \) with \( \|f\|_2 = 2\pi \). However, Riemann-Lebesgue Theorem 7.31 affirms that \( \langle f_n, g \rangle \to 0 \) for every \( g \) in \( L^1 \), which means that the sequence cannot be totally bounded in \( L^2(\pi, \pi) \). Nevertheless, because \( |f_n(x)| \leq 1 \), this sequence is \( \mu \)-uniformly integrable of any order \( p \).

An important role is played by the weak convergence in \( L^1 \), i.e., when \( \langle f_n, g \rangle \to \langle f, g \rangle \) for every \( g \) in \( L^\infty \). Actually, we have the Dunford-Pettis criterion: a set \( \{f_i : i \in I\} \) is sequentially weakly pre-compact in \( L^1(\Omega, \mathcal{F}, \mu) \) if and only if it is \( \mu \)-uniformly integrable (a partial proof for the case of a finite measure can be found in Meyer [81, Section II.2, T23, pp. 39-40]). However, any bounded set in \( L^p(\Omega, \mathcal{F}, \mu) \), \( 1 < p \leq \infty \), is weakly pre-compact. This is a general result (Alaoglu’s Theorem) valid for any reflexive Banach space, e.g., see Conway [29, Section V.3 and V.4, pp. 123–137].

**Exercise 6.9.** Consider the Lebesgue measure on the interval \((0, \infty)\) and define the functions \( f_i = \langle 1/i \rangle_{(i, 2i)} \) and \( g_i = 2^i \langle 1/(2i-1, 2i) \rangle \) for \( i \geq 1 \). Prove that (a) the sequence \( \{f_i : i \geq 1\} \) is uniformly integrable of any order \( p > 1 \), but not of order \( 0 < p \leq 1 \). On the contrary, show that (b) the sequence \( \{g_i : i \geq 1\} \) is uniformly integrable of any order \( 0 < p < 1 \), but the sequence is not equi-integrable of any order \( p \geq 1 \).

### 6.5 Representation Theorems

When discussing signed measures, it was clear that a signed measure cannot assume the values \(+\infty\) and \(-\infty\). However, a \( \sigma \)-finite signed measure \( \mu \) make sense, i.e., the measurable space \((\Omega, \mathcal{F})\) has a partition \( \Omega = \sum_k \Omega_k \) such that the restriction of \( \mu \) to \( \Omega_k \), denoted by \( \mu_k \), is a finite signed measure. This is essentially the situation of a linear functional on the space \( L^1(\Omega, \mathcal{F}, \mu) \).
There are the various versions of the so-called Riesz representation theorems. For instance, recall the definition of the Lebesgue spaces $L^p = L^p(\Omega, \mathcal{F}, \mu)$, for $1 \leq p < \infty$ and its dual, denoted by $(L^p)'$, the Banach space of linear continuous (or bounded) functional on $L^p$, endowed with the dual norm

$$\|g\|_p = \sup \left\{ \langle g, \varphi \rangle : \|\varphi\|_p \leq 1 \right\}, \quad \forall g \in (L^p)' ,$$

where $\langle \cdot, \cdot \rangle$ denote the duality pairing, i.e., $g$ acting on (or applied to) $\varphi$, and for the supremum, the functions $\varphi$ can be taken in $L^p$ or just a simple function, actually, $\varphi$ belonging to some dense subspace of $L^p$ is sufficient.

**Theorem 6.30.** For every $\sigma$-finite measure space $(\Omega, \mathcal{F}, \mu)$ and $p$ in $[1, \infty)$, the map $g \mapsto Tg$, defined by

$$\langle Tg, f \rangle = \int_{\Omega} gf \, d\mu,$$

gives a linear isometry from $(L^q, \| \cdot \|_q)$ onto the dual space of $(L^p, \| \cdot \|_p)$, with $1/p + 1/q = 1$.

**Proof.** First, Hölder inequality shows that $T$ maps $L^q$ into $(L^p)'$ with $\|Tg\|_p \leq \|g\|_q$. Moreover, by means of Proposition 4.32 and Remark 4.33 we have the equality, i.e., $\|Tg\|_p = \|g\|_q$, which proves that $T$ is an isometry.

To check that $T$ is onto, for any given element $g$ in the dual space $(L^p)'$ define

$$\nu_g(A) = \langle g, 1_A \rangle, \quad \forall A \in \mathcal{F}, \mu(A) < \infty.$$

Considering $\nu_g$ defined on measurable subsets $A \subset F$, for a fixed $F$ in $\mathcal{F}$ with $\mu(F) < \infty$, we have a signed measure $\nu_g$ on $F \subset \Omega$, which is absolutely continuous with respect to $\mu$. Thus Radon-Nikodym Theorem 6.3 yields an almost everywhere measurable function, still denoted by $g_F$, such that

$$\nu_g(A) = \int_A g_F \, d\mu, \quad \forall A \in \mathcal{F}, \ A \subset F.$$

By linearity, we have

$$\langle g, 1_F \varphi \rangle = \int_F g_F \varphi \, d\mu,$$

for any simple functions $\varphi$. Again, Proposition 4.32 and Remark 4.33 imply the equality $\|g 1_F\|_p = \|g_F\|_q$, where $g 1_F$ is the restriction of the functional $g$ to $F$, i.e., $\langle g 1_F, f \rangle = \langle g, 1_F f \rangle$.

Since for some sequence $\{f_n\}$ of functions in $L^p$ we have $\langle g, f_n \rangle \to \|g\|_p$, there exists a $\sigma$-finite measurable set $G$ (supporting all $f_n$) such that

$$\|g\|_p = \sup \left\{ \langle g, 1_G \varphi \rangle : \|\varphi\|_p \leq 1 \right\}, \quad (6.12)$$
and $G = \bigcup_n G_n$, for some monotone sequence $\{G_n\}$ of measurable sets with $\mu(G_n) < \infty$. Since $G_n \subset G_{n+1}$ implies $g_{G_n} = g_{G_{n+1}}$ on $G_n \setminus N_n$, with $\mu(N_n) = 0$, we can define a measurable function $g_\sigma$ such that $g_\sigma = 0$ outside of $G$ and $g_\sigma = g_{G_n}$ on $G_n \setminus N_n$, for every $n \geq 1$, i.e., we have

$$\langle g, 1_G \varphi \rangle = \int g_\sigma \varphi \, d\mu, \quad \text{for any simple function } \varphi$$

Now, apply Proposition 4.32 and Remark 4.33 to deduce that $\|g 1_G\|_p = \|g_\sigma\|_q$.

On the other hand, for any $\mu(F) < \infty$ with $F \cap G = \emptyset$ we must have $\nu_g(F) = 0$, i.e., $g_F = 0$ almost everywhere. Indeed, if $\nu_g(F) > 0$ then

$$\langle g, 1_F + 1_G \varphi \rangle = \langle g, 1_F \rangle + \langle g, 1_G \varphi \rangle$$

yields $\|g\|_p = \langle g, 1_F \rangle + \|g 1_G\|_p$, which contradict the equality (6.12). This proves that $g = g 1_G$ and $g = T(g_\sigma)$.

Recalling that a Banach space is called reflexive if it is isomorphic to its double dual, we deduce that $L^p(\Omega, \mathcal{F}, \mu)$ is reflexive for $1 < p < \infty$. On the other hand, if $L^1$ is separable and $L^\infty$ is not separable then $L^1$ cannot be reflexive, since it can be proved that if the dual space is separable so is the initial space.

Given a Hausdorff topological space $X$, denote by $C(X)$ the linear space of all real-valued continuous functions on $X$. The minimal $\sigma$-algebra $\mathcal{B}_a$ for which all continuous (and bounded) real functions are measurable is called the Baire $\sigma$-algebra. If $X$ is a metric space then $\mathcal{B}_a$ coincides with the Borel $\sigma$-algebra $\mathcal{B}$, but in general $\mathcal{B}_a \subset \mathcal{B}$. If $X$ is compact then $C(X)$ with the sup-norm, namely, $\|f\|_\infty = \sup_x |f(x)|$ becomes a Banach space. The dual space $C(X)'$, i.e., the space of all continuous linear functional $T : C(X) \to \mathbb{R}$, with the dual norm

$$\|T\|'_\infty = \sup \{ |T(f)| : \|f\|_\infty \leq 1 \}$$

is also a Banach space.

If $X$ is a compact Hausdorff space then denote by $M(X)$ the linear space of all finite signed measures on $(X, \mathcal{B}_a)$, i.e., $\mu$ belongs to $M(X)$ if and only if $\mu$ is a linear combination (real coefficients) of finite measures, actually it suffices $\mu = \mu_1 - \mu_2$ with $\mu_i$ measures. We can check that

$$\|\mu\| = \inf \{ \mu_1(X) + \mu_2(X) : \mu = \mu_1 - \mu_2 \}$$

defines a norm, which makes $M(X)$ a Banach space. Moreover, we can write $\|\mu\| = |\mu|(X)$, where $|\mu|(X) = \mu^+(X) + \mu^-(X)$ and $\mu = \mu^+ - \mu^-$, with $\mu^+$ and $\mu^-$ measures such that for some measurable set $A$ we have $\mu^+(A) = 0$ and $\mu^-(X \setminus A) = 0$.

**Theorem 6.31.** For every compact Hausdorff space $X$, the mapping $\mu \mapsto I_\mu$,

$$I_\mu(f) = \int_X f \, d\mu_1 - \int_X f \, d\mu_2, \quad \text{with } \mu = \mu_1 - \mu_2,$$

is a linear isometry from the space $(M(X), \|\cdot\|)$ onto $(C(X)', \|\cdot\|_\infty)$. \qed
For instance, the reader may consult the book by Dudley [37, Theorems 6.4.1 and 7.4.1, p. 208 and p. 239] for a complete proof of the above theorems. For the extension to locally compact spaces, e.g., see Bauer [10, Sections 28-29, pp. 170–188], among others.

- **Remark 6.32.** For locally compact space, the one-point compactification or Alexandroff compactification of $X$ yields the following version of Theorem 6.31: If $X$ is a locally compact Hausdorff space then the dual of the space $\left(C_* (X), \| \cdot \|_\infty \right)$ of all continuous functions vanishing at infinity (i.e., $f$ such that for every $\varepsilon > 0$ there exists a compact $K_\varepsilon$ satisfying $|f(x)| < \varepsilon$ for every $x \in X \setminus K_\varepsilon$) is the space $\left(M(X), \| \cdot \| \right)$ of all finite Borel (or Radon) measures on $X$. For instance the reader may check Malliavin [79, Section II.6, pp. 94–100].

For a locally compact (Hausdorff) space $X$ denote by $C_0 (X, \mathbb{R}^m)$ the linear space of all $\mathbb{R}^m$-valued continuous functions on $X$ with compact support, i.e., $f : X \rightarrow \mathbb{R}^m$ continuous and its $\text{supp}(f)$ (the closure of the set $\{ x \in X : f(x) \neq 0 \}$) is compact. Recall that a (outer) Radon measure on $X$ is a (signed) measure defined on the Borel $\sigma$-algebra which is finite for every compact subset of $X$, see Section 3.3.

**Theorem 6.33.** Let $T : C_0 (X, \mathbb{R}^m) \rightarrow \mathbb{R}$ be a linear functional satisfying

$$\|T\|_K = \sup \{ T(f) : f \in C_0 (X, \mathbb{R}^m), \ |f| \leq 1, \ \text{supp}(f) \subset K \} < \infty,$$

for every compact subset $K$ of $X$. Then $\mu$ defined by

$$\mu(U) = \sup \{ T(f) : f \in C_0 (X, \mathbb{R}^m), \ |f| \leq 1, \ \text{supp}(f) \subset U \},$$

for every open set $U$, is a Radon measure on $X$. Moreover, we have

$$T(f) = \int_X f \sigma \, d\mu, \ \forall f \in C_0 (X, \mathbb{R}^m),$$

where $\sigma : \mathbb{R}^m \rightarrow \mathbb{R}$ is a $\mu$-measurable function such that $|\sigma| = 1$.

For instance, a proof of this result (for $X = \mathbb{R}^n$) can be found in Evans and Gariepy [42, Section 1.8, pp. 49–54]. A simplified version (of this section and the previous one) is discussed in Stroock [112, Chapter 7, pp. 139–158]. In general, the reader may check Folland [45, Chapter 7, pp. 211–233] for a discussion on Radon measures and functional; and perhaps take a look at Kubrusly [74, Chapter 12, pp. 223–246] for some more details.
Chapter 7

Elements of Real Analysis

At this point we dispose of the basic arguments on measure theory, and we can push further the analysis. First, we study signed measures, next we discuss briefly some questions on real-valued functions, and finally some typical Banach spaces of functions. For instance, the interested reader may check the book Stein and Shakarchi [109] for most of the topic discussed below.

First, we give an improved version of the Lebesgue dominate convergence Theorem 4.7.

Proposition 7.1. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let $\{f_n\}$ be a sequence of measurable functions which convergence to $f$ in measure over any set of finite measure. Then

$$\lim_{n \to \infty} \int_{\Omega} f_n \, d\mu = \int_{\Omega} f \, d\mu,$$

provided there exists an integrable function $g$ satisfying $|f_n(x)| \leq g(x)$, almost everywhere in $x$, for any $n$.

Proof. Let us show that in the Lebesgue dominate convergence we can replace the (almost everywhere) pointwise convergence by the convergence in measure over any set of finite measure, i.e., for every $\varepsilon > 0$ and $A \in \mathcal{F}$ with $\mu(A) < \infty$ there exists an index $N = N(\varepsilon, A)$ such that $\mu(\{x \in A : |f_n(x) - f(x)| \geq \varepsilon\}) < \varepsilon$ for every $n \geq N$.

We argue by contradiction. If the limit does not hold then there exist $\varepsilon > 0$ and a subsequence $\{f_{n'}\}$ such that

$$\left| \int_{\Omega} f_{n'} \, d\mu - \int_{\Omega} f \, d\mu \right| \geq \varepsilon, \quad \forall n'.$$

However, because $f_{n'} \to f$ in measure on any finite set and the set $\{f_n \neq 0\} \cup \{f \neq 0\}$ is $\sigma$-finite, we can apply Theorem 4.17 to find a subsequence $\{f_{n''}\}$ such that $f_{n''} \to f$ (pointwise) almost everywhere. Hence, we may modify $f_{n''}$ in a negligible set (without changing its integral) to get $f_{n''} \to f$ pointwise, which yields a contradiction with Theorem 4.7. \qed
7.1 Differentiation and Approximation

We consider the Lebesgue measure \( \ell = dx \) in \( \mathbb{R}^d \) and Lebesgue measurable functions \( f : \mathbb{R}^d \to \mathbb{R} \), i.e., measurable with respect to the Lebesgue \( \sigma \)-algebra \( \mathcal{L} = \mathcal{B}^\ell \), the completion of the Borel \( \sigma \)-algebra \( \mathcal{B}(\mathbb{R}^d) \) relative to \( \ell \). Sometime, we may write \( \ell(A) = |A| \), meaning the Lebesgue outer measure of \( A \), for any \( A \subset \mathbb{R}^d \). In most of this section, we omit the word Lebesgue when using the properties: integrable, measurable, almost everywhere, etc, as long as the context clarifies the meaning.

Recall that if \( f : \mathbb{R}^d \to [-\infty, +\infty] \) is Lebesgue integrable then we can modify \( f \) in a set of measure zero so that \( f \) takes only real-values and remains Lebesgue integrable. Thus, contrary to continuous functions, an integrable function is better thought as a function defined almost everywhere. We denote by \( L^1 \) the space of all real-valued Lebesgue integrable functions, i.e., \( f \) is Lebesgue measurable and

\[
\|f\|_1 = \int_{\mathbb{R}^d} |f(x)| \, dx < \infty.
\]

Note that \( \| \cdot \|_1 \) satisfies all the property of a norm, except that \( \|f\|_1 = 0 \) only implies \( f = 0 \) almost every where, i.e., it is a semi-norm on \( L^1 \), the vector space of real-valued (Lebesgue) integrable functions in \( \mathbb{R}^d \). By considering functions defined almost everywhere we transform \( L^1 \) into the normed space \((L^1, \| \cdot \|_1)\), where the elements are equivalent classes and \( f = g \) in \( L^1 \) means \( \|f - g\|_1 = 0 \), i.e., \( f = g \) almost everywhere.

Similarly, we use the (vector) space \( L^\infty \) of all measurable functions bounded almost everywhere (so-called measurable essentially bounded functions), with the semi-norm \( \| \cdot \|_\infty \), defined by

\[
\|f\|_\infty = \inf \{ C \geq 0 : |f(x)| \leq C, \text{ a.e.} \},
\]

i.e., the infimum of all constants \( C > 0 \) such that \( \{|x \in \mathbb{R}^d : |f(x)| \geq C| \} = 0 \). Again, this yields the normed space \((L^\infty, \| \cdot \|_\infty)\).

Denote by \( \mathcal{C}^n \), with \( n = 0, 1, \ldots \), the collection of all real-valued functions continuously differentiable up to the order \( n \) in \( \mathbb{R}^d \), and also denote by \( \mathcal{C}_0^n \) is the sub-collection of functions in \( \mathcal{C}^n \) with compact support. Similarly, \( \mathcal{C}_b^n \) collection of all functions real-valued functions continuously bounded differentiable up to the order \( n \). It is clear that \( \mathcal{C}_0^n \subset L^1 \cap L^\infty \). Sometimes, it is convenient to use multi-indexes for derivatives, e.g., \( \alpha = (\alpha_1, \ldots, \alpha_d) \), \( \alpha_i = 0, 1, \ldots \), and \( |\alpha| = \alpha_1 + \cdots + \alpha_d \),

\[
\partial^\alpha f(x) = \frac{\partial^{|\alpha|} f(x)}{\partial_{x_1}^{\alpha_1} \cdots \partial_{x_d}^{\alpha_d}}, \quad \text{or} \quad \partial^\alpha = \partial_x^\alpha = \partial_{x_1}^{\alpha_1} \cdots \partial_{x_d}^{\alpha_d},
\]

to emphasize the variables and order of derivatives.

Recall, the support of a continuous function \( f \) is denoted by \( \text{supp}(f) \) and defined as the closure of the set \( \{ x \in \mathbb{R}^d : f(x) \neq 0 \} \), but for a measurable function \( f \) we define its support as the support of the induced measure, i.e., the
support of a Borel measure $\mu$ is the complement of the open set $\bigcup\{U \text{ open} : \mu(U) = 0\}$, and therefore the support of $f$ is the support of the measure

$$B \mapsto \int_B |f(x)| \, dx, \quad \forall B \in \mathcal{B}(\mathbb{R}^d),$$

i.e., $f = 0$ almost everywhere outside the $\text{supp}(f)$. Hence, denote by $\mathcal{L}^1_0$ the (vector) space of all integrable functions with compact support (i.e., vanishing outside of a ball).

**Approximation by Smooth Functions**

To approximate integrable function by smooth function, remark that by definition, for any integrable function $f$ there exists a sequence $\{f_k : k \geq 1\}$ of simple functions such that $\|f_k - f\|_1 \to 0$, which implies that the vector space of (integrable) simple functions is dense in $\mathcal{L}^1$, and in particular we deduce that $\mathcal{L}^1_0 \cap \mathcal{L}^\infty$ is dense in $\mathcal{L}^1$. Also we have

**Proposition 7.2.** Given a Lebesgue integrable function $f$ and a real number $\varepsilon > 0$ there exists a continuous functions $g$ such that

$$\int_{\mathbb{R}^d} |f(x) - g(x)| \, dx = \|f - g\|_1 < \varepsilon,$$

and $g$ vanishes outside of some ball, i.e., the space of continuous functions with compact supports $C^0_0$ is dense in $\mathcal{L}^1$.

**Proof.** Each real-valued measurable function $f$ can be written as $f = f^+ - f^-$, where $f^+$ and $f^-$ are nonnegative $m$-measurable functions. By Proposition 1.9, for any nonnegative measurable function $f^\pm$ there exists an increasing sequence $\{f^\pm_k : k \geq 1\}$ of simple functions such that $f^\pm_k(x) \to f^\pm(x)$, for almost everywhere $x$ in $\mathbb{R}^d$, as $k \to \infty$. Hence, by the monotone convergence, we obtain

$$\lim_k \int_{\mathbb{R}^d} |f^\pm_k(x) - f^\pm(x)| \, dx = 0,$$

whenever $f$ is integrable in $\mathbb{R}^d$. Now, for a fixed $k$, the simple function $f^\pm_k$ is a finite combination of expression of the form $c \mathbb{1}_E$, with $E$ a (Borel) measurable set of finite measure and $c$ a real number. For each $E$ and $\varepsilon > 0$ there exists an open set $U \supset E$ such that $m(U \setminus E) < \varepsilon$. Since $U$ is an open set in $\mathbb{R}^d$, see Remark 2.38, there exists an non-overlapping sequence $\{Q_i : i \geq 1\}$ of closed cubes such that $U = \bigcup_i Q_i$, which yields $m(U) = \sum_i m(Q_i)$, and

$$\lim_n \int_{\mathbb{R}^d} |\mathbb{1}_U(x) - \mathbb{1}_{F_n}(x)| \, dx = 0, \quad \text{with} \quad F_n = \bigcup_{i=1}^n Q_i.$$

Given $\varepsilon > 0$ and the cubes $F_n$, we can easily find a continuous function $g_{\varepsilon,n}$ such that

$$\int_{\mathbb{R}^d} |g_{\varepsilon,n}(x) - \mathbb{1}_{F_n}(x)| \, dx < \varepsilon.$$
Combining all, the desired approximation follows.

Alternatively, given an integrable function $f$, the dominated convergence implies that, for every $\varepsilon > 0$ there exists $r > 0$ such that the function $f_r(x) = 1_{\{|x| \leq r\}} 1_{\{|f(x)| \leq r\}} f(x)$ satisfies

$$\int_{\mathbb{R}^d} |f(x) - f_r(x)| \, dx \leq \frac{\varepsilon}{2}.$$

Now, for this $r > 0$, we apply Lusin’s Theorem 4.22 to obtain a closed set $C_r \subset B_r = \{x : |x| \leq r\}$ such that $f_r$ is continuous on $C_r$ and $m(B_r \setminus C_r) < \varepsilon/(5r)$. Next, essentially based on Tietze’s extension (see Proposition 0.2), $f_r$ can be extended to a continuous function $g_r$ on $\mathbb{R}^d$ satisfying the conditions: (a) $|g_r| \leq r$ on $\mathbb{R}^d$ and (b) the set $N_r = \{x \in \mathbb{R}^d : g_r(x) \neq f_r(x)\}$ is contained in some ball $B$ and $m(N_r) < \varepsilon/(4r)$. Hence

$$\int_{\mathbb{R}^d} |f(x) - g_r(x)| \, dx \leq \int_{\mathbb{R}^d} |f(x) - f_r(x)| \, dx +$$

$$+ \int_{B} |f_r(x) - g_r(x)| \, dx \leq \frac{\varepsilon}{2} + 2r m(N_r) \leq \varepsilon,$$

and $g = g_r$ is the desired function. \hfill $\Box$

The arguments used in proving Proposition 7.2 can be extended to a more general setting, e.g., replacing the Lebesgue measure $m$ on $\mathbb{R}^d$ by a Borel measure $\mu$ on a metric space $\Omega$. There are other arguments for approximation typical in $\mathbb{R}^d$, for instance, mollification and truncation.

Let us begin with the following results.

**Proposition 7.3.** If $f$ belong to $L^1$ then

$$\lim_{a \to 0} \int_{\mathbb{R}^d} |f(x + a) - f(x)| \, dx = 0,$$

i.e., the translation operator $\tau_a f = f(\cdot - a)$ is continuous in $L^1$.

**Proof.** Indeed, let us denote by $\mathcal{K}$ the collection of all functions $f$ in $L^1$ such that $\|f(\cdot + a) - f\|_1 \to 0$ as $a \to 0$. It is simple to check that $\mathcal{K}$ is a closed vector space, i.e.,

(a) if $\alpha, \beta \in \mathbb{R}$ and $f, g \in \mathcal{K}$ then $\alpha f + \beta g \in \mathcal{K}$,

(b) if $\{f_n : n \geq 1\} \subset \mathcal{K}$ and $\|f_n - f\|_1 \to 0$ then $f \in \mathcal{K}$.

Now, we use the same argument of Proposition 7.2 to successively approximate an integrable function by simple functions, next $c 1_A$ with $A$ measurable and $m(A) < \infty$ by $c 1_U$ with a bounded open set $U$, and then for every $\varepsilon > 0$ and $U$ we find a finite union of non-overlapping cubes $Q = \bigcup_{i=1}^n Q_i$ with $Q \subset U$ and $m(U \setminus Q) < \varepsilon$, to establish that the family of (simple) functions of the form $\sum_{i=1}^n a_i 1_{Q_i}$, where the cubes $Q_i$ have edges parallel to the axis, can approximate and integrable function in the $\| \cdot \|_1$ norm.
Since the characteristic function of a \(d\)-interval (or a cube) belongs to \(\mathcal{K}\), we deduce \(\mathcal{K} = \mathcal{L}^1\).

Alternatively, we may claim that any integrable function can be approximated in the \(\| \cdot \|_1\) norm by continuous functions with compact support (Proposition 7.2), which also belong to \(\mathcal{K}\).

**Exercise 7.1.** Let \(Q\) be the class of dyadic cubes in \(\mathbb{R}^d\), i.e., for \(d = 1\) we have the dyadic intervals \([k2^{-n}, (k + 1)2^{-n}]\), with \(n = \pm 1, \pm 2, \ldots\) and any integer \(k\). Consider the set \(D\) of all finite linear combinations of characteristic functions of cubes in \(Q\) and rational coefficients. Verify that (1) \(D\) is a countable set and prove (2) that \(D\) is dense in \(\mathcal{L}^1\), i.e., for any integrable function \(f\) and any \(\varepsilon > 0\) there is an element \(\varphi\) in \(D\) such that \(\| f - \varphi \|_1 < \varepsilon\). Moreover, by means of Weierstrauss approximation Theorem 0.3, (3) make an alternative argument to show that there is a countable family of truncated (i.e., multiplied by \(1\{\|x\| < r\}\)) polynomial functions which is dense in \(\mathcal{L}^1\), proving (again) that the space \(\mathcal{L}^1\) is separable.

For two integrable functions \(f\) and \(g\) we consider the convolution \(f \ast g\) given by the formula

\[
(f \ast g)(x) = \int_{\mathbb{R}^d} f(x - y) g(y) \, dy = \int_{\mathbb{R}^d} f(y) g(x - y) \, dy, \quad \forall x \in \mathbb{R}^d. \tag{7.1}
\]

It is clear that if either \(f\) or \(g\) is essentially bounded then \(x \mapsto (f \ast g)(x)\) is well defined and \(\| f \ast g \|_\infty \leq \| f \|_1 \| g \|_\infty\). Moreover, we can also check the inequality \(\| f \ast g \|_1 \leq \| f \|_1 \| g \|_1\), which means that the convolution \(f \ast g\) is defined almost everywhere, i.e., \(\mathcal{L}^1\) is a commutative algebra with the convolution product.

**Definition 7.4** (locally integrable). A function \(f : \mathbb{R}^d \to \mathbb{R}\) is called **locally integrable** if for every \(x\) in \(\mathbb{R}^d\) there exists an open neighborhood \(U_x\) such that \(f\) is integrable in \(U_x\), or equivalently, the restriction to any compact set in integrable. This class of functions is denoted by \(\mathcal{L}^1_{\text{loc}}\), and we say that a sequence of locally integrable functions \(\{f_n : n \geq 1\}\) converges to \(f\) **locally** in \(\mathcal{L}^1\) or in \(\mathcal{L}^1_{\text{loc}}\), if

\[
\lim_{n} \int_K |f_n(x) - f(x)| \, dx = 0,
\]

for every compact set \(K\) of \(\mathbb{R}^d\). Similarly, \(\mathcal{L}^\infty_{\text{loc}}\) is the space of **locally essentially bounded** functions, i.e., functions bounded almost everywhere on any compact set. We also have the spaces of equivalence classes \(L^1_{\text{loc}}\) and \(L^\infty_{\text{loc}}\).

Certainly, we mean \(f\) is Lebesgue measurable and \(f 1_K\) is in \(\mathcal{L}^1\). This concept of locally integrable can be used on locally compact spaces \(\Omega\) with a Borel measure \((\mu, \mathcal{B})\).

Also, recall that we say that a measurable function defined almost everywhere has compact support if it is equal to zero almost everywhere outside of a ball. The sub-vector space of \(\mathcal{L}^1\) (or \(L^1\)) of all functions with compact support
is denoted by \( \mathcal{L}^1_0 \) (or \( \mathcal{L}^1 \)), and similarly with \( \mathcal{L}^\infty \) (or \( \mathcal{L}^\infty_0 \)). The convolution \( f \ast g \) is also defined if \( f \) and \( g \) are only locally integrable and one of them has compact support, i.e., \( f \in \mathcal{L}^1_0 \) and \( g \in \mathcal{L}^1 \) or \( g \in \mathcal{L}^\infty \) implies \( f \ast g \in \mathcal{L}^1_0 \) or \( f \ast g \in \mathcal{L}^\infty_0 \), respectively.

In general, given two Lebesgue measurable functions \( f \) and \( g \), we say that the convolution \( f \ast g \) is defined if the functions inside the integrals in the expression (7.1) are integrable for almost every \( x \). Remark that the convolution operation commutes with the translation operator, i.e., \( \tau_a (f \ast g) = (\tau_a f) \ast g = f \ast (\tau_a g) \).

**Proposition 7.5.** Let \( f \) and \( g \) be two Lebesgue measurable functions in \( \mathbb{R}^d \).

(a) If \( f \) is integrable and \( g \) is essentially bounded then the convolution \( f \ast g \) is a bounded uniformly continuous function. Moreover, \( f \) is only locally integrable, \( g \) is only locally essentially bounded, and either \( f \) or \( g \) has a compact support then the convolution \( f \ast g \) is a continuous function.

(b) Denote by \( \partial_i f \) the partial derivative of \( f \) with respect to \( x_i \). If \( f \) is essentially bounded or integrable, \( g \) is integrable and the partial derivative \( \partial_i f \) is a bounded function then the \( i \)-partial derivative of the convolution \( f \ast g \) is a bounded uniformly continuous function and \( \partial_i (f \ast g) = (\partial_i f) \ast g \). Moreover, if \( f \) and \( g \) are only locally integrable, either \( f \) or \( g \) has a compact support, and the partial derivative \( \partial_i f \) is a locally bounded function then the \( i \)-partial derivative of the convolution \( f \ast g \) is continuous function and \( \partial_i (f \ast g) = (\partial_i f) \ast g \).

**Proof.** Consider the bound

\[
|(f \ast g)(x + a) - (f \ast g)(x)| \leq \int_{\mathbb{R}^d} |f(x + a - y) - f(x - y)| |g(y)| \, dy,
\]

where the integral is actually limited to the support of \( g \). Thus, if \( g \) is essentially bounded then Proposition 7.3 proves most of the above claim (a). For the local version of this claim, we remark that if \( f \) or \( g \) has a compact support then the integral is only on a compact set \( K \) (as long as \( x \) remain in another compact region) instead of \( \mathbb{R}^d \), and again, the continuity follows.

Next, by means of the Mean Value Theorem and the dominate convergence, we obtain \( \partial_i (f \ast g) = (\partial_i f) \ast g \) and in view of (a), we deduce the claim (b).

Certainly, we can iterate the property (b) to deduce that \( \partial^\alpha (f \ast g) = (\partial^\alpha f) \ast g \), for any multi-index \( \alpha \) with \( |\alpha| \leq n \), e.g., \( f \) belongs to \( \mathcal{C}^n \) and \( g \) is in \( \mathcal{L}^1 \).

Regarding the claim (b), we assume that the partial derivative \( \partial_i f \) exists a any point, so that the Mean Value Theorem can be applied, however, the expression \( (\partial_i f) \ast g \) is a continuous function even if \( \partial_i f \) is defined almost everywhere. Nevertheless, if we assume that \( \partial_i f \) is defined only almost everywhere then we may have a non-constant function with \( \partial_i f = 0 \) a.e. (like the Cantor function).

**Exercise 7.2.** Consider the various cases of Proposition 7.5 and verify the claims by doing more details, e.g., consider the case when \( g \) is only locally integrable. What if the \( \partial_i f \) exists everywhere, \( f \) and \( \partial_i f \) are locally integrable function, and \( g \) is essentially bounded with compact support?
Exercise 7.3. Let \( f \) and \( g \) be two nonnegative Lebesgue locally integrable functions in \( \mathbb{R}^d \). Prove (a) if \( \lim \inf_{|x| \to \infty} f(x)/g(x) > 0 \) and \( g \) is not integrable then \( f \) is also non integrable, in particular, \( \lim \inf_{|x| \to \infty} f(x)|x|^d = 0 \), for any integrable function \( f \), (b) if \( \lim \sup_{|x| \to \infty} f(x)/g(x) < \infty \) and \( g \) is integrable then \( f \) is also integrable. Finally, show that (c) if \( f \) is integrable and uniformly continuous in \( \mathbb{R}^d \) then \( \lim \sup_{|x| \to \infty} |f(x)| = 0 \). Given an example of a nonnegative integrable and continuous function on \([0, \infty)\) such that \( \lim \sup_{x \to \infty} f(x) = +\infty \).

Corollary 7.6. If \( k \) is an integrable kernel, i.e., an integrable function such that
\[
\int_{\mathbb{R}^d} k(x) \, dx = 1,
\]
and \( \{k_\varepsilon : \varepsilon > 0\} \) its corresponding mollifiers, i.e., \( k_\varepsilon(x) = \varepsilon^{-d} k(x/\varepsilon) \), for every \( x \) in \( \mathbb{R}^d \), then we have
\[
\lim_{\varepsilon \to 0} \left\| f \ast k_\varepsilon - f \right\|_1 = 0, \quad \forall f \in L^1.
\]
Moreover, if either \( f \) is essentially bounded in \( \mathbb{R}^d \) or the kernel \( k \) satisfies
\[
k(x) = \alpha(x)|x|^{-d}, \quad a.e. \ x \in \mathbb{R}^d, \quad \text{with} \ \alpha(x) \to 0 \ as \ |x| \to \infty,
\]
and \( f \) is uniformly continuous and bounded in a subset \( F \) of \( \mathbb{R}^d \), i.e., for every \( \varepsilon > 0 \) there exists \( \delta > 0 \) such that \( |x-y| < \delta \) and \( x \) in \( F \) imply \( |f(x) - f(y)| < \varepsilon \), and \( \sup_{x \in F} |f(x)| < \infty \), then \( (f \ast k_\varepsilon)(x) \to f(x) \) uniformly for \( x \) in \( F \), as \( \varepsilon \to 0 \).

Proof. First, a change of variables shows that \( \|k_\varepsilon\|_1 = \|k\|_1 \) and
\[
\int_{\mathbb{R}^d} k_\varepsilon(x) \, dx = \int_{\mathbb{R}^d} k(x) \, dx = 1.
\]
Thus, the inequality
\[
|f \ast k_\varepsilon(x) - f(x)| \leq \left| \int_{\mathbb{R}^d} [f(x-\varepsilon y) - f(x)] k(y) \, dy \right| \tag{7.2}
\]
and the dominate convergence show the second part of the claim.

To prove the first part, we remark that, for every \( \delta > 0 \), the expression
\[
\int_{|y|>\delta} |k_\varepsilon(y)| \, dy = \varepsilon^{-d} \int_{|y|>\delta/\varepsilon} |k(y/\varepsilon)| \, dy = \int_{|y|>\delta/\varepsilon} |k(y)| \, dy
\]
vanesishes as \( \varepsilon \to 0 \). Now, by means of the continuity of the translation, Proposition 7.3, given \( r > 0 \) there exists \( \delta > 0 \) such that
\[
\phi(y) = \int_{\mathbb{R}^d} |f(x-y) - f(x)| \, dx \leq r, \quad \text{if} \quad |y| \leq \delta,
\]
and clearly \(|\phi(y)| \leq 2\|f\|_1\), for every \(y\). Thus
\[
\|f * k_\varepsilon - f\|_1 \leq \int_{\mathbb{R}^d} \phi(y)|k_\varepsilon(y)|\,dy \leq r \int_{|y| \leq \delta} |k_\varepsilon(y)|\,dy + 2\|f\|_1 \int_{|y| > \delta} |k_\varepsilon(y)|\,dy,
\]
i.e., \(\lim_{\varepsilon \to 0} \|f * k_\varepsilon - f\|_1 \leq r\|k\|_1 = r\), and we conclude by letting \(r \to 0\).

Finally, if \(f\) is bounded and uniformly continuous then we can split the integral (7.2) in \(\mathbb{R}^d = \{|y| < \delta\} \cup \{|y| \geq \delta\}\) to get
\[
|f * k_\varepsilon(x) - f(x)| \leq 2\|f\|_\infty \int_{|y| \geq \delta} |k(y)|\,dy + \sup_{|y| < \delta} |f(x - \varepsilon y) - f(x)|
\]
and to deduce the uniform convergence on \(F\) when \(f\) is essentially bounded.

If the kernel is essentially bounded, and \(f\) is bounded and uniformly continuous on \(F\) then for every \(x\) in \(F\) and for any \(\varepsilon > 0\) there exists \(\delta > 0\) such that \(|f(x - y) - f(x)| < \varepsilon\) if \(|y| \leq \delta\). Thus
\[
|f * k_\varepsilon(x) - f(x)| \leq \varepsilon_1 \|k\|_1 + \int_{|y| > \delta} |f(x - y) - f(x)| |k_\varepsilon(y)|\,dy,
\]
where the last term is bounded by
\[
A + B = \int_{|y| > \delta} |f(x - y)| |k_\varepsilon(y)|\,dy + |f(x)| \int_{|y| > \delta} |k_\varepsilon(y)|\,dy.
\]
As prove above, the second term vanishes, i.e.,
\[
B \leq \left(\sup_{x \in F} |f(x)|\right) \int_{|y| > \delta/\varepsilon} |k(y)|\,dy,
\]
i.e., \(B \to 0\) as \(\varepsilon \to 0\), and
\[
A = \int_{|y| > \delta} |f(x - y)||\alpha(y/\varepsilon)||y|^{-d}\,dy \leq \delta^{-d} \sup_{|y| > \delta} \{||\alpha(y/\varepsilon)||\} \|f\|_1,
\]
which tends to 0 as \(\varepsilon \to 0\). □

**Exercise 7.4.** Prove a local version of Corollary 7.6, i.e., if the kernel \(k\) has a compact support then \(f * k_\varepsilon \to f\) in \(L^1_{\text{loc}}\), for every \(f\) in \(L^1_{\text{loc}}\), and locally uniformly if \(f\) belongs to \(L^\infty_{\text{loc}}\). □

There are a couple of well known kernels, for instance,
\[
p(x) = \frac{1}{\pi} \left(\frac{1}{1 + x^2}\right), \quad \forall x \in \mathbb{R}, \quad \text{Poisson kernel},
\]
\[
f(x) = \frac{1}{\pi} \left(\frac{\sin^2 x}{x^2}\right), \quad \forall x \in \mathbb{R}, \quad \text{Fejér kernel},
\]
\[
g(x) = \frac{1}{\sqrt{\pi}} \left(e^{-x^2}\right), \quad \forall x \in \mathbb{R}, \quad \text{Gauss-Weierstrauss kernel},
\]

and their multi-dimensional counterparts. Another typical case is the following kernel

\[ k(x) = \begin{cases} 
  c_r \exp \left[ - (r^2 - |x|^2)^{-1} \right] & \text{if } |x| < r, \\
  0 & \text{otherwise,}
\end{cases} \tag{7.3} \]

for any given constant \( r > 0 \) and a suitable \( c_r \) to meet the condition \( \|k\|_1 = 1 \). It is clear that the support of \( k \) is the ball centered at the origin with radius \( r \). Moreover, following standard theorems of calculus, we can verify that \( k \) is continuously differentiable of any order, i.e., \( k \) is a typical example of a function in \( \bigcap_{n=1}^{\infty} C^0_0(\mathbb{R}^d) = C^\infty(\mathbb{R}^d) \).

It is clear that if \( f \) is in \( L^1_{\text{loc}} \) and \( k \) is a kernel with compact support, e.g., like (7.3), then the convolution \( f \ast k_\varepsilon \) belongs to \( C^\infty \) and \( f \ast k_\varepsilon \to f \) in \( L^1_{\text{loc}} \).

Now we consider the space \( L^p(\mathbb{R}^d, \mathcal{B}, \mu) \), where \( \mu \) is a Radon measure, i.e., \( \mu(K) < \infty \) for every compact subset \( K \) of \( \mathbb{R}^d \). Assuming temporarily that the expression

\[ \|f\|_p = \int_{\mathbb{R}^d} |f|^p \, d\mu \]

is a norm (which is called \( p \)-norm, see later on this chapter), we have

**Proposition 7.7.** Any function \( f \) in a Radon measure space \( L^p(\mathbb{R}^d, \mathcal{B}, \mu) \) is limit in the \( p \)-norm of a sequence \( \{\varphi_n : n \geq 1\} \) of infinity differentiable functions with compact support.

**Proof.** By approximating \( f^+ \) and \( f^- \) separately, we may assume \( f \geq 0 \). Thus, writing \( f \) as a monotone limit of simple functions, we are reduced to the case of simple functions \( f = \sum_{i=1}^n a_i \mathbb{1}_{A_i} \), and so, to the a characteristic function \( \mathbb{1}_A \) of a measurable set \( A \) of finite measure. Next, by means of Proposition 2.17, we approximate \( \mathbb{1}_A \) by a sequence of semi-open \( d \)-interval \( [a, b] = [a_1, b_1] \times \cdots \times [a_d, b_d] \). Finally, we construct a sequence \( \{\phi_n : n \geq 1\} \) of smooth functions such that \( 0 \leq \phi_n \leq \mathbb{1}_I \) and \( \phi_n(x) \to \mathbb{1}_{[a, b]}(x) \) for every \( x \) in \( \mathbb{R}^d \).

**Exercise 7.5.** Let \( \{k_n\} \) be a sequence of infinite-differentiable functions such that \( k_n(x) = 1 \) if \( |x| < n \) and \( k_n(x) = 0 \) if \( |x| > n + 1 \), and let \( Q \) be the family of functions of the form \( qk_n \), where \( q \) is a polynomial with rational coefficients. By means of Weierstrass approximation Theorem 0.3, prove that \( Q \) is a dense family of \( L^p(\mathbb{R}^d, \mathcal{B}, \mu) \) under the \( p \)-norm, see last part of Exercise 7.1.

### 7.2 Partition of the Unity

Now, let \( K \) be a compact set in \( \mathbb{R}^d \) and \( U \) be an open set satisfying \( U \supset K \), with compact closure \( \overline{U} \). If \( B^1 = \{x \in \mathbb{R}^d : |x| \leq 1\} \) then for any \( \delta > 0 \) sufficiently small we have \( K \subset R = K + \delta B^1 \subset U \), and so we can select \( \varepsilon > 0 \) such that \( K + \varepsilon B^1 \subset R \) and \( R + \varepsilon B^1 \subset U \). Therefore, we may consider the convolution \( \mathbb{1}_R \ast k_\varepsilon \), where \( k \) is given by (7.3) with \( r = 1 \). Since the support of
\(k_\varepsilon\) satisfies supp\((k_\varepsilon) \subset B^1\) we deduce that there exists a function \(f = 1_R \ast k_\varepsilon\) with derivatives of any order such that \(f = 1\) on \(K\) and \(f = 0\) outside \(U\).

Another point is to use Corollary 7.6 with \(k\) given by (7.3) to show that for any \(f\) in \(L^1\) and \(\varepsilon > 0\) there exists a function \(g\) with derivatives of any order such that \(\|f - g\|_1 < \varepsilon\), i.e., the space \(C_0^\infty(\mathbb{R}^d) = \bigcap_{n=1}^\infty C_n^0(\mathbb{R}^d)\) is dense in \(L^1\).

**Theorem 7.8.** Let \(\{\Omega_\alpha : \alpha\}\) be an open cover of an open subset \(\Omega\) of \(\mathbb{R}^d\), i.e., \(\Omega_\alpha\) is open and \(\Omega = \bigcup_\alpha \Omega_\alpha\). Then there exists a smooth partition of the unity \(\{\chi_i : i \geq 1\}\) subordinate to \(\{\Omega_\alpha : \alpha\}\), i.e., (a) \(\chi_i\) belongs to \(C_0^\infty(\mathbb{R}^d)\), (b) for every \(i\) there exists \(\alpha = \alpha(i)\) such that \(\chi_i(x) = 0\) for every \(x\) in \(\Omega \setminus \Omega_\alpha\), namely, supp\((\chi_i) \subset \Omega_\alpha\), (c) \(0 \leq \chi_i(x) \leq 1\) and \(\sum_i \chi_i(x) = 1\), for every \(x\) in \(\Omega\), where the series is locally finite, namely, for any compact set \(K\) of \(\Omega\) the set of indices \(i\) such that the support of \(\chi_i\) intercept \(K\), supp\((\chi_i) \cap K \neq \emptyset\), is finite.

**Proof.** (1) First we show that there exits a locally finite subordinate open cover \(\{U_i : i \geq 1\}\) with compact closure \(U_i\), i.e., for any compact set \(K \subset \Omega\) the set of indices \(\{i \geq 1 : U_i \cap K \neq \emptyset\}\) is finite, and for every \(i \geq 1\) there exists \(\alpha(i)\) such that \(U_i \subset \Omega_{\alpha(i)}\).

Indeed, consider the compact sets

\[K_n = \{x \in \Omega : |x| \leq n \text{ and } d(x, \mathbb{R}^d \setminus \Omega) \geq 1/n\},\]

for \(n \geq 1\), where \(d(x, A) = \inf\{|x - y| : y \in A\}\). We have \(\Omega = \bigcup_n K_n\), \(K_{n-1} \subset K_n^\circ\), where \(K_n^\circ\) is the interior of \(K_n\). For \(n \geq 3\) define \(\Omega_{\alpha,n} = \Omega_\alpha \cap K_{n+1}^\circ \cap (\Omega \setminus K_{n-2}^\circ),\) and remark that \(\Omega_{\alpha,n}\) is a open cover of \(K_{n+1}^\circ \cap (\Omega \setminus K_{n-2}^\circ) \supset K_n \cap (\Omega \setminus K_{n-1}^\circ)\). On the other hand, for each \(x\) in \(K_n \cap (\Omega \setminus K_{n-1}^\circ)\) there exists an open set \(U_n(x)\) with closure \(\overline{U}_n(x)\) included in \(\Omega_{\alpha,n}\) for some \(\alpha(x)\). Hence, the family \(\{U_n(x) : x\}\) forms an open cover of the compact set \(K_n \cap (\Omega \setminus K_{n-1}^\circ)\) and so, there exists a finite subcover, i.e., \(x_1, \ldots, x_m, m = m(n)\) such that \(\{U_n(x_j) : j = 1, \ldots, m(n)\}\) cover \(K_n \cap (\Omega \setminus K_{n-1}^\circ)\), for every \(n \geq 3\). Thus, the family \(\{U_n(x_j) : j = 1, \ldots, m(n), n \geq 3\}\), now denoted by \(\{U_i : i \geq 1\}\), is countable and satisfies the required conditions.

(2) Next, we construct a continuous partition of the unity \(\{f_i : i \geq 1\}\) subordinate to \(\{U_i : i \geq 1\}\), and so, also subordinate to \(\{\Omega_\alpha : \alpha\}\). Indeed, we apply again the above argument (1), with \(\{U_i : i \geq 1\}\) instead of \(\{\Omega_\alpha : \alpha\}\), to obtain another locally finite subordinate cover \(\{V_i : i \geq 1\}\), which (after relabeling and deleting some \(U\)-open if necessarily) satisfies \(V_i \subset U_i \subset \Omega_\alpha, \alpha = \alpha(i)\), for every \(i \geq 1\). Now, we use Urysohn’s Lemma to get a continuous function \(g_i\) satisfying \(g_i(x) = 1\) for every \(x\) in \(V_i\) and \(g_i(x) = 0\) for any \(x\) in \(\mathbb{R}^d \setminus U_i\), i.e., supp\((g_i) \subset \overline{U}_i\). Since the covers are locally finite, for any compact \(K\) of \(\Omega\) there exists only finite many \(i\) such that \(U_i \cap K \neq \emptyset\) and so the finite sum \(g(x) = \sum_i g_i(x)\) defines a continuous function satisfying \(g(x) \geq 1,\) for every \(x\) in \(\Omega\). Hence, the family of continuous functions \(\{f_i : i \geq 1\}\), with \(f_i(x) = g_i(x)/g(x)\), is a partition of the unity subordinate to \(\{U_i : i \geq 1\}\), satisfying all the required conditions, except for the smoothness.

(3) To obtain a smooth partition we use the convolution with a smooth kernel \(k\) having compact support defined by (7.3) for \(r = 1\), as in Corollary 7.6.
with \( k_\varepsilon \). Indeed, again we apply (1) to get another locally finite subordinate cover \( \{ W_i : i \geq 1 \} \) which satisfies \( \overline{W}_i \subset V_i \subset \overline{V}_i \subset U_i \subset \Omega_\alpha, \alpha = \alpha(i) \), for every \( i \geq 1 \). If \( 2\varepsilon_i = \min\{d(\overline{V}_i, \Omega \setminus U_i), d(\overline{W}_i, \Omega \setminus V_i)\} \) then the convolution \( \varphi_i = 1_{V_i} * k_\varepsilon \) is an infinitely differentiable (smooth) function and, since \( \text{supp}(k_\varepsilon) \) is included in the ball centered at the origin with radius \( \varepsilon_i \), we have

\[
0 \leq \varphi_i \leq 1 \quad \text{in } \mathbb{R}^d, \quad \text{supp}(\varphi_i) \subset \overline{U}_i, \quad \text{and } \varphi_i = 1 \quad \text{on } \overline{W}_i.
\]

Moreover, the finite sum \( \varphi(x) = \sum_i \varphi_i(x) \) defines a smooth function satisfying \( \varphi(x) \geq 1 \), for every \( x \) in \( \Omega \). Hence, the family of smooth functions \( \{ \chi_i : i \geq 1 \} \), with \( \chi_i(x) = \varphi_i(x)/\varphi(x) \), is a partition of the unity subordinate to \( \{ U_i : i \geq 1 \} \), satisfying all the required conditions.

We note that in the above proof, we may go directly to (3) without using (2). However, (1) and (2) are valid for \( \sigma \)-compact locally compact Hausdorff topological spaces. Also, we may deduce (3) from (2) by using \( \varphi_i = g_i * k_\varepsilon \), with \( k \) as in (7.3) for \( r = 1 \) and \( 2\varepsilon_i = d(\overline{V}_i, \Omega \setminus \Omega_\alpha(i)) \). Indeed, we remark that \( g_i(x) > 0 \) implies \( \varphi_i(x) > 0 \) and then \( \varphi(x) = \sum_i \varphi_i(x) > 0 \), for every \( x \) in \( \Omega \). Alternatively, we may check that the functions \( g_i \) in (2) can be chosen infinitely differentiable, instead of just continuous. For instance, the reader may consult Folland [45, Section 4.5, pp. 132–136] and Malliavan [79, Section II.1, pp. 55–61].

- **Remark 7.9.** If \( \chi_0 \) is a smooth function defined on \( \mathbb{R}^d \) such that \( \chi(x) = 1 \) if \( |x| \leq 1 \) and \( \chi_0(x) = 0 \) if \( |x| \geq 3/2 \) then the sequence \( \{ \chi_i : i \geq 0 \} \) of smooth functions defined by

\[
\chi_i(x) = \chi_0(2^{-i}x) - \chi_0(2^{-i+1}x), \quad i = 1, 2, \ldots,
\]

satisfies \( \sum_{i=0}^\infty \chi_i(x) = 1 \) for every \( x \) in \( \mathbb{R}^d \) and is referred to as a dyadic resolution of the unity.

## 7.3 Lebesgue Points

For every locally integrable function \( f \) we define

\[
f^*(x) = \sup_{r>0} F(x,r), \quad F(x,r) = \frac{1}{|B(x,r)|} \int_{B(x,r)} |f(y)| \, dy, \tag{7.4}
\]

where \( |\cdot| \) denotes the Lebesgue measure and \( B(x,r) \) is the ball centered at \( x \) with radius \( r \), i.e., \( B(x,r) = x + rB^1 \), with the notation of Vitali’s Theorem 2.29. As mentioned in Remark 2.37, the ball may be considered in the Euclidean space \( \mathbb{R}^d \) or in \( \mathbb{R}^d \) with any (equivalent) norm, e.g., it could be a cube with center at \( x \) and sizes \( 2r \), moreover, we may take the cubes with edges parallel to the axis. For instance, a point here is the fact that the ratio of volume of a cube with sizes \( 2r \) and a ball of radius \( r \) is bounded both ways.

The function \( f^* \) is called Hardy-Littlewood maximal function of \( f \), a gauge of the size of the average of \( |f| \) around \( x \). We have (a) nonnegative \( f^*(x) \geq 0 \),
(b) sub-linear \((f + g)^* \leq f^* + g^*\), (c) positively homogeneous \((cf)^* = |c|f^*\), for every \(c\) in \(\mathbb{R}\), and (d) the bound \(f^*(x) \leq \|f\|_\infty\).

**Proposition 7.10.** Let \(f\) be an integrable function in \(\mathbb{R}^d\). Then \(f^*\) is a lower semi-continuous with values in \(\mathbb{R} \cup \{+\infty\}\) and

\[
|\{x \in \mathbb{R}^d : f^*(x) > c\}| \leq \frac{5^d}{c} \int_{\mathbb{R}^d} |f(y)| \, dy,
\]

for every \(c > 0\).

**Proof.** Since the boundary \(\partial B(x, r)\) is negligible the function \((x, r) \mapsto F(x, r)\) is continuous. Indeed, since \(\mathbb{1}_{B(x, r)}(y) \to \mathbb{1}_{B(x_0, r_0)}(y)\) as \((x, r) \to (x_0, r_0)\), for every \(y \in \mathbb{R}^d \setminus \partial B(x, r_0)\), i.e., \(\{y \in \mathbb{R}^d : |y - x| \neq r_0\}\), the dominate convergence yields the continuity. Hence, the function \(f^*\) is lower semi-continuous with values in \(\mathbb{R} \cup \{+\infty\}\) (i.e., also Borel measurable).

If \(E\) is a nonempty bounded measurable subset of \(\mathbb{R}^d\) then we can choose \(r_0 > 0\) such that \(E \subset B(0, r_0)\). For each \(x\) outside of \(B(0, 2r_0)\) define \(r_1 = r_1(x)\) the minimum \(r_1 > 0\) satisfying \(E \subset B(x, r_1)\), and also define \(r_2(x)\) the maximum \(r_2 > 0\) satisfying \(E \cap B(x, r_2) = \emptyset\). It is clear that \(0 \leq r_1(x) - r_2(x) \leq 2r_0\).

Now, if \(r \geq r_1\) then \(|E \cap B(x, r)|/|B(x, r)| \leq |E|/|B(x, r_1)|\) and if \(r \leq r_2\) then \(|E \cap B(x, r)| = 0\), i.e.,

\[
\frac{|E|}{|B(x, r_1)|} \leq (\mathbb{1}_E)^*(x) = \sup \left\{ \frac{|E \cap B(x, r)|}{|B(x, r)|} : r_2 \leq r \leq r_1 \right\} \leq \frac{|E|}{|B(x, r_2)|},
\]

for every \(x \in \mathbb{R}^d\) with \(|x| \geq 2r_0\). Hence, the inequality \(|x - y| \leq |y - x| \leq |x| + |y|\) implies \(|x_0 - r_0 \leq |x - y| \leq r_1\) and \(|x - y| \leq |x| + r_0\), for any \(y \in E\), i.e., we have \(|x - y| \leq r_1(x) \leq |x| + r_0\), and we deduce

\[
c_1|x|^{-d}|E| \leq (\mathbb{1}_E)^*(x) \leq c_2|x|^{-d}|E|, \quad \forall x \in \mathbb{R}^d, \ |x| \geq r.
\]

for some positive constants \(c_1, c_2\) and \(r\).

Thus, if \(f > 0\) in some set of positive measure then there exists a positive constant \(c_f\) such that \(f^*(x) \geq c_f|x|^{-d}\) for \(x\) sufficiently large, i.e., \(f^*\) is not integrable in \(\mathbb{R}^d\). However, if \(f\) is bounded with compact support then \(f^*(x) \leq C_f|x|^{-d}\) for every \(x \in \mathbb{R}^d\), which implies that the set \(E_c = \{x \in \mathbb{R}^d : f^*(x) > c\}\) has finite measure. Moreover, by definition of \(f^*\), for every \(x \in E_c\) there exists a ball \(B(x, r)\) such that

\[
c < \frac{1}{|B(x, r)|} \int_{B(x, r)} |f(y)| \, dy, \quad \text{i.e.} \quad |B(x, r)| < \frac{1}{c} \int_{B(x, r)} |f(y)| \, dy.
\]

Hence, by Vitali’s covering Corollary 2.33, for every \(\alpha > 5^d\) there exists a finite number of disjoint balls \(\{B(x_i, r_i) : i = 1, \ldots, n\}\) such that we have \(|E_c| \leq \alpha \sum_{i=1}^n |B(x_i, r_i)|\). Therefore,

\[
|E_c| \leq \alpha \sum_{i=1}^n \frac{1}{c} \int_{B(x_i, r_i)} |f(y)| \, dy \leq \frac{\alpha}{c} \int_{\mathbb{R}^d} |f(y)| \, dy,
\]

[Pre-]
and as \( \alpha \to 5^d \) we deduce (7.5), if \( f \) is bounded with compact support.

In general, we approximate \( |f| \) by an increasing sequence \( \{|f_n| : n \geq 1 \} \) of bounded functions with compact support (e.g., simple functions or continuous functions with compact support) to have

\[
\{ x \in \mathbb{R}^d : |f^*_n(x)| > c \} \leq \frac{5^d}{c} \int_{\mathbb{R}^d} |f_n(y)| \, dy \leq \frac{5^d}{c} \int_{\mathbb{R}^d} |f(y)| \, dy,
\]

which yields (7.5) as \( n \to \infty \). \( \square \)

**Remark 7.11.** We may use \( 3^d \) instead of \( 5^d \) in estimate (7.5), see Remark 2.34. \( \square \)

When a measurable function \( f \) satisfies (7.5) with some constant \( c_f \) replacing \( 5^d \) we say that \( f \) belongs to \( \text{weak-}L^1 \), see Exercise 4.19. Thus, Proposition 7.10 shows that (nonlinear) Hardy-Littlewood maximal application \( f \mapsto f^* \) maps \( L^1 \) into \( \text{weak-}L^1 \). It is also obvious that \( f \mapsto f^* \) maps \( L^\infty \) into itself, which allows the application of interpolation results used in singular integrals, e.g., see Folland [45, Section 6.5, pp. 200–208] and Stein [108, Chapter 1, pp. 3-25].

**Exercise 7.6.** With the notation of Proposition 7.10 prove that for any \( 1 < p < \infty \) and \( f \) in \( L^p(\mathbb{R}^d) \) we have that \( f^* \) is in \( L^p(\mathbb{R}^d) \) and \( \|f^*\|_p \leq C_p\|f\|_p \), i.e.,

\[
\int_{\mathbb{R}^d} |f^*(x)|^p \, dx \leq (C_p)^p \int_{\mathbb{R}^d} |f(x)|^p \, dx,
\]

with \((C_p)^p = 5^d 2^p p / (p - 1)\). We may proceed as follows: first use the distribution of \( f^* \), namely, \( m(f^*, r) = \ell(\{ x \in \mathbb{R}^d : f^*(x) > r \}) \), and estimate (7.5) to obtain the inequality

\[
m(f^*, r) \leq m(g^*_r, r/2) \leq \frac{5^d 2^d}{r} \|g_r\|_1,
\]

where \( g_r(x) = f(x) \mathbb{1}_{\{|f(x)| \geq r/2\}} \). Next, based on the formula

\[
\int_{\mathbb{R}^d} |f^*(x)|^p \, dx = p \int_0^\infty r^{p-1} m(f^*, r) \, dr,
\]

see Exercise 5.6, deduce the estimate

\[
\int_{\mathbb{R}^d} |f^*(x)|^p \, dx \leq p \int_0^\infty r^{p-1} \left( \frac{5^d 2^d}{r} \int_{\{2|f(x)| \geq r\}} |f(y)| \, dy \right) \, dr,
\]

which implies (7.6). See full details in Wheeden and Zygmund [119, Section 9.3, pp. 155–157]. \( \square \)

**Theorem 7.12.** Let \( f \) be a Lebesgue locally integrable function in \( \mathbb{R}^d \). Then almost every point is a Lebesgue point for \( f \), i.e., there exist a negligible \( N = N_f \), \( |N| = 0 \), such that

\[
\lim_{r \to 0} \frac{1}{|B(x, r)|} \int_{B(x, r)} |f(y) - f(x)| \, dy = 0, \quad \forall x \in \mathbb{R}^d \setminus N,
\]

where \( B(x, r) \) is the ball centered at \( x \) with radius \( r \).
Proof. First, we show a differentiability statement, namely,
\[ \lim_{r \to 0} F(x, r) = f(x), \quad F(x, r) = \frac{1}{|B(x, r)|} \int_{B(x, r)} f(y) \, dy, \tag{7.7} \]
for any integrable function \( f \) and almost every \( x \) in \( \mathbb{R}^d \). Indeed, let \( \{ f_n : n \geq 1 \} \) be a sequence of continuous functions with compact support (actually, continuous and integrable suffice) such that
\[ \int_{\mathbb{R}^d} |f_n(y) - f(y)| \, dy = \|f_n - f\|_1 \to 0. \]

If \( F_n(x, r) \) denotes the function as in (7.7) corresponding to \( f_n \) we have
\[ \limsup_{r \to 0} |F(x, r) - f(x)| \leq \limsup_{r \to 0} |F(x, r) - F_n(x, r)| + \limsup_{r \to 0} |F_n(x, r) - f_n(x)| + |f_n(x) - f(x)|, \]
for every \( n \geq 1 \) and any \( x \) in \( \mathbb{R}^d \). Since \( f_n \) is continuous, we obtain \( F_n(x, r) \to f_n(x) \) as \( r \to 0 \), for every fixed \( n \) and \( x \). On the other hand, we use Hardy-Littlewood maximal function to bound the first term \( |F(x, r) - F_n(x, r)| \leq (f - f_n)^*(x) \), i.e., we get
\[ \limsup_{r \to 0} |F(x, r) - f(x)| \leq (f - f_n)^*(x) + |f_n(x) - f(x)|, \quad \forall x, n. \]

Thus, given \( \varepsilon > 0 \), if
\[ E_\varepsilon = \left\{ x \in \mathbb{R}^d : \limsup_{r \to 0} |F(x, r) - f(x)| > \varepsilon \right\} \]
then
\[ E_\varepsilon \subset \left\{ x \in \mathbb{R}^d : 2(f - f_n)^*(x) > \varepsilon \right\} \cup \left\{ x \in \mathbb{R}^d : 2|f_n(x) - f(x)| > \varepsilon \right\}. \]

By means of Proposition 7.10 applied to \( f_n - f \) we get
\[ |E_\varepsilon| \leq \frac{5d2}{\varepsilon} \int_{\mathbb{R}^d} |f_n(x) - f(x)| \, dx + \frac{2}{\varepsilon} \int_{\mathbb{R}^d} |f(x) - f_n(x)| \, dx, \]
and as \( n \to \infty \), we obtain \( |E_\varepsilon| = 0 \), i.e., (7.7) holds.

Next, by replacing \( f \) by \( x \mapsto 1_{\{|x| < n\}} f(x) \), \( n \geq 1 \), we deduce that (7.7) remains true for any locally integrable \( f \).

Now, given \( f \), consider the countable family \( \{ g_q : q \in \mathbb{Q} \} \) of locally integrable functions \( g_q(x) = |f(x) - q| \), with \( \mathbb{Q} \) being the set of rational numbers. For each \( g_q \) there is a negligible subset \( N_q \subset \mathbb{R}^d \), where (7.7) does not holds for \( g_q \). Hence, for \( x \) in \( \mathbb{R}^d \setminus N \), with \( N = \bigcup_q N_q \), we have
\[ \lim_{r \to 0} \frac{1}{|B(x, r)|} \int_{B(x, r)} |f(y) - q| \, dy = |f(x) - q|, \quad \forall q \in \mathbb{Q}. \]

By taking a sequence of rational \( \{ q \} \) convergent to \( f(x) \) we obtain the desired result. \( \square \)
Besides using balls with another metric in $\mathbb{R}^d$, see Remark 2.37, it is clear that the balls $B(x, r)$ can be replaced by a family of measurable sets $E(x, r)$ having the properties: (a) the diameter of $E(x, r)$ vanishes as $r \to 0$, (b) there exists a constant $c > 0$ such that $|B(x, r)| \leq c|E(x, r)|$, for the smallest ball $B(x, r)$ containing $E(x, r)$. Indeed, the inequality

$$\frac{1}{|E(x, r)|} \int_{E(x, r)} |f(y) - f(x)| \, dy \leq \frac{c}{|B(x, r)|} \int_{B(x, r)} |f(y) - f(x)| \, dy$$

provides the desired conclusion.

There are other ways to establishing Theorem 7.12, e.g., one may use first Radon-Nikodym Theorem 6.3 to study derivative of set functions, for instance, see DiBenedetto [32, Section IV.8-11, pp. 184–193] or Rana [92, Section 9.2, pp. 322-331]. Also, based on the so-called Besicovtch’s covering arguments, a result similar to Theorem 7.12 is obtained for any Radon measure in $\mathbb{R}^d$, non necessarily the Lebesgue measure, e.g., Evans and Gariepy [42, Section 1.6, pp. 37–48] or Taylor [115, Chapter 11, pp. 139–156].

Exercise 7.7. Verify that if $E$ is a Lebesgue measurable set then almost every $x$ in $E$ is a point of density of $E$, i.e., we have $|E \cap B(x, r)|/|B(x, r)| \to 1$ as $r \to 0$. Similarly, almost every $x$ in $E^c$ is a point of dispersion of $E$, i.e., we have $|E \cap B(x, r)|/|B(x, r)| \to 0$ as $r \to 0$.

\[
\begin{align*}
\text{\bullet Remark 7.13.} & \quad \text{A real-valued function } f \text{ defined on an open interval } I \text{ is called approximately continuous at } x_0 \text{ in } I \text{ when there exists a (Lebesgue) measurable subset } E \text{ of } I \text{ such that } x_0 \text{ is a point of (Lebesgue) density of } E \text{ and } f/|E| \text{ is continuous at } x_0, \text{ i.e., (1) } \lim_{r \to 0} |E \cap B(x_0, r)|/|B(x_0, r)| = 1 \text{ and (2) } \lim_{x \to x_0, x \in E} f(x) = f(x_0). \text{ This definition can be adapted to a closed interval } I \text{ by using lateral or one-sided densities points and limits. If a function is approximately continuous at every point of } I \text{ then we say simply that } f \text{ is approximately continuous (on } I). \text{ There several interested properties on these functions, e.g., (a) a function is measurable if and only if it is approximatively continuous almost everywhere, (b) if } f \text{ and } g \text{ are approximatively continuous at } x_0 \text{ then the same if true of the functions } f + g, f - g, fg \text{ and of } f/g \text{ if } g(x_0) \neq 0. \text{ Note that if a function is equal almost everywhere to a continuous functions, or even less only continuous almost everywhere then it is approximately continuous. Besides the classic book Saks [102], for instance, the reader may want to check the books Bruckner [22, Section 2.2, pp. 15–20] or Gordon [54, Chapter 14, pp. 223–236].}
\end{align*}
\]

Exercise 7.8. Prove that if $f$ is $p$-locally integrable, i.e., $|f|^p$ is integrable on any compact subset of $\mathbb{R}^d$, $1 \leq p < \infty$, then

$$\lim_{r \to 0} \frac{1}{|B(x, r)|} \int_{B(x, r)} |f(y) - f(x)|^p \, dy = 0, \quad \forall x \in \mathbb{R}^d \setminus N,$$

or equivalently

$$\lim_{r \to 0} \int_{|y| \leq R} |f(x + ry) - f(x)|^p \, dy = 0, \quad \forall x \in \mathbb{R}^d \setminus N,$$
for some negligible set $N = N_f$, $|N| = 0$ and any radius $R > 0$. □

- **Remark 7.14.** For a locally essentially bounded function $f$ we can define

$$
\bar{f}(x, r) = \text{ess-sup}_{|y-x|<r} f(y) \quad \text{and} \quad \underline{f}(x, r) = \text{ess-inf}_{|y-x|<r} f(y)
$$

to have $f(x, r) \leq \bar{f}(x, r)$ almost everywhere in $x$ for every $r > 0$, but the values $\bar{f}(x, r)$ and $\underline{f}(x, r)$ are defined everywhere in $x$, and the limits

$$
\bar{f}(x) = \lim_{r \to 0} \text{ess-sup}_{|y-x|<r} f(y) \quad \text{and} \quad \underline{f}(x) = \lim_{r \to 0} \text{ess-inf}_{|y-x|<r} f(y)
$$

exist for every $x$. These are the superior and inferior essential limits, which can be used with functions defined almost everywhere, i.e., they are well defined on equivalence classes of functions. Moreover, for a one-dimensional $x$, we can consider the lateral superior and inferior essential limits. Since we are taken the supremum or infimum over a non-countable family, the measurability of either $\bar{f}(x, r)$ or $\underline{f}(x, r)$ may not be automatically ensured, e.g., restate Exercise 1.22 for essential limits and apply Proposition 2.3. Note that in view of the "essential" supremum or infimum, it is irrelevant to use a reduced ball like $0 < |y-x| < r$ instead of the full open ball $|y-x| < r$. □

**Exercise 7.9.** With the notation of Remark 7.14, consider a locally essentially bounded function $f$. Prove that (a) $\underline{f}(x, r) \leq f(x) \leq \bar{f}(x, r)$ at every Lebesgue point $x$ of $f$, for every $r > 0$. Next, (b) if $g(x) = \underline{f}(x) = \bar{f}(x)$ is finite for every $x$ in an open set $U$ then show that $f = g$ a.e. in $U$ and that $g$ is almost continuous in $U$, i.e., for any $x$ in $U$ and every $\varepsilon > 0$ there exist a null set $N$ and $\delta > 0$ such that $|y-x| < \delta$ and $y$ in $U \cap N^c$ imply $|g(y) - g(x)| < \varepsilon$, see Exercise 4.23. Therefore, a measurable function $f$ is called almost upper (lower) semi-continuous if $f = \bar{f}$ ($f = \underline{f}$) almost everywhere, and almost continuous if $f = \bar{f} = \underline{f}$ a.e. Finally, (c) verify that if $f$ is a continuous (or upper/lower semi-continuous) a.e. then $f$ is almost continuous (or almost upper/lower semi-continuous) a.e. Is there any function which is not continuous a.e, but nevertheless it is almost continuous a.e.? □

**Exercise 7.10.** Consider a locally integrable function $f$ defined on an open set $O$ of $\mathbb{R}^d$ and a bounded kernel $k$ with compact support, e.g., like (7.3). If $k_\varepsilon(x) = \varepsilon^{-d}k(x/\varepsilon)$ then prove that $(f \ast k_\varepsilon)(x) \to f(x)$ almost everywhere, indeed, for any Lebesgue point (Theorem 7.12) of $f$. □

**Exercise 7.11.** Let $g_1$ be a real-valued Lebesgue measurable function of two real variables, say $g_1(x, y)$ with $x, y$ in $\mathbb{R}$. Assume that $g_1$ is locally integrable in $\mathbb{R}^2$ and define the function

$$
g(x, y) = \int_0^x g_1(t, y) \, dt, \quad \forall x \in \mathbb{R},
$$

for almost every $y$ in $\mathbb{R}$. (1) Verify that $g$ is a locally integrable function, which is continuous in the first variable. Now, consider the set $A$ in $\mathbb{R}^2$ of all points
(x, y) such that \([g(x + h, y) - g(x, y)]/h \to g_1(x, y)\) as \(h \to 0\). (2) Prove that the complement \(N = A^c\) is a set of Lebesgue measure zero. Next, if \(f\) is a continuously differentiable function with compact support then (3) show that the convolution \(f \ast g\) is a continuously differentiable function and \(\partial_x (f \ast g) = (\partial_x f) \ast g\), where \(\partial_x\) denotes the partial derivative in the variable \(x\), and finally, by means of Fubini-Tonelli Theorem 4.14, (4) prove that \(\partial_x (f \ast g) = f \ast (\partial_x g)\), if \(f\) is a measurable (essentially) locally bounded function with compact support. Hint: regarding (4), first assume that

\[
g(x, y) = \int_{-\infty}^{x} g_1(t, y) \, dt, \quad \forall x, y \in \mathbb{R},
\]

to show that

\[
\int_{-\infty}^{b} (f \ast g_1)(x, y) \, dx = (f \ast g)(b, y), \quad \forall b, y \in \mathbb{R},
\]

and then, consider the general case. 

\[\square\]

7.4 Functions of one variable

Recall that a (real-valued) monotone function \(f(x)\) defined on an interval \((a, b)\) of \(\mathbb{R}\) can have only a countable number of discontinuities. By means of Vitali’s covering we show

**Proposition 7.15.** If \(f : (a, b) \to \mathbb{R}\) is a monotone increasing function then its derivative is a nonnegative function \(f'\) defined almost everywhere and

\[
\int_{a}^{b} f'(x) \, dx \leq f(b-) - f(a+),
\]

where \(f(x\pm)\) means the lateral limits at a point \(x\).

**Proof.** Only the main idea is given, for instance, full details are found in the books either Wheeden and Zygmund [119, Section 7.4, pp. 111–115] or Jones [65, Section 16.A, pp. 511–521] or Ovchinnikov [89, Chapter 4, 97–128]. The central argument is based on the four Dini’s derivatives

\[
D^\pm f(x) = \lim sup_{h \to 0^\pm} \frac{f(x + h) - f(x)}{h}, \quad \text{and}
\]

\[
D_\pm f(x) = \lim inf_{h \to 0^\pm} \frac{f(x + h) - f(x)}{h},
\]

where we show (with the help of Vitali’s covering) that the sets

\[
A_{r,s} = \{ x \in (a, b) : D_\pm f(x) > r > s > D^\mp f(x) \}\]
are negligible. Next, we define \( f_k(x) = [f(x + 1/k) - f(x)]k \) for \( k = 1, 2, \ldots \), which satisfies \( f_k(x) \to f'(x) \) for almost every \( x \), and so, \( f' \) is a measurable function. Since

\[
\int_a^{b-1/k} f_k(x)\,dx = k \int_{b-1/k}^b f(x)\,dx - k \int_a^{a+1/k} f(x)\,dx \leq f(b) - f(a),
\]

Fatou Lemma shows the desired inequality.

- **Remark 7.16.** It is perhaps important to mention that properties of derivative functions and reconstruction of continuous functions from its derivative are very delicate classic problems, e.g., Bruckner [22] and Gordon [54].
- **Remark 7.17.** Contrasting with results (a) the inverse of a monotone and continuous function is necessarily continuous (e.g., see Kirkwood [70, Theorem 4.16, pp. 101]), and (b) any monotone and continuous function maps Borel sets into Borel sets (e.g., see Burk [23, Proposition C.1, pp. 273–274]).

Differential inequalities are related questions, e.g.,

**Proposition 7.18.** Let \( f \) be a real-valued continuous function on \((a, b) \times [c, d]\), and let \( D \) denote one fixed choice of the four possible Dini derivatives \( D^+, D^-, D_+ \) or \( D_- \). Suppose that \( y \) and \( z \) are two continuous functions on \([a, b]\) with values in \([c, d]\) satisfying either \( Dy(x) \leq f(x, y(x)) \) and \( Dz(x) > f(x, z(x)) \), or \( Dy(x) < f(x, y(x)) \) and \( Dz(x) \geq f(x, z(x)) \), for any \( x \) in \((a, b)\). If \( y(a) < z(a) \) then \( y(x) < z(x) \), for every \( x \) in \((a, b)\).

**Proof.** Define \( x_* = \sup\{x \in (a, b) : y(x) < z(x)\} \) and \( x^* = \inf\{x \in (a, b) : y(x) \geq z(x)\} \). By contradiction, if the assertion is false then \( x_* < b, x^* > a, y(x) \geq z(x) \) for any \( x_* < x < b, y(x) < z(x) \) for any \( a < x < x^* \), and by continuity, \( y(x_*) = z(x_*) \) and \( y(x^*) = z(x^*) \). Therefore, for \( h > 0 \) sufficiently small, the inequalities

\[
\frac{y(x_* + h) - y(x_*)}{h} \geq \frac{z(x_* + h) - z(x_*)}{h}, \quad \text{and}
\]

\[
\frac{y(x^* - h) - y(x^*)}{h} \geq \frac{z(x^* - h) - z(x^*)}{h}
\]

imply \( Dy(x_*) \geq Dz(x_*) \) with \( D = D_+ \) or \( D = D^+ \), and \( Dy(x^*) \geq Dz(x^*) \) with \( D = D_- \) or \( D = D^- \). Hence, the differential inequalities satisfied by \( y \) and \( z \) yield either \( f(x_*, y(x_*)) > f(x_*, z(x_*)) \) or \( f(x^*, y(x^*)) > f(x^*, z(x^*)) \), which is a contradiction with the equality either \( y(x_*) = z(x_*) \) or \( y(x^*) = z(x^*) \).

There are several useful applications of Proposition 7.18, e.g.,

**Corollary 7.19.** Let \( I \) be an open interval and let \( D \) denote one fixed choice of the four possible Dini derivatives \( D^+, D^-, D_+ \) or \( D_- \). (1) A continuous function \( g \) is non-increasing on \( I \) if and only if \( Dg \leq 0 \) on \( I \). (2) If \( y \) and \( z \) are two continuous functions on \( I \) such that \( Dy \leq z \) on \( I \), then \( Dg \leq z \) on \( I \) for any other possible choice of a Dini derivative \( D \).
Proof. To prove (1), consider the functions \( y = g \) and \( z: x \mapsto \varepsilon + y(a) + \varepsilon(x - a) \), for a given \( a \) in \( I \) and \( \varepsilon > 0 \). These functions satisfy \( y(a) < \varepsilon + y(a) = z(a) \) and \( Dy \leq 0 \leq \varepsilon = z' \) on \( I_a = \{ x \in I : x \geq a \} \), i.e., the conditions of Proposition 7.18 with \( f = \varepsilon \), which imply that \( y(x) < z(x) = \varepsilon + y(a) + \varepsilon(x - a) \) for \( x \) in \( I_a \). Thus, take \( \varepsilon \to 0 \) to obtain \( y(x) \leq y(a) \) for every \( x \geq a \), with \( x \) and \( a \) in \( I \).

The converse is simpler, if \( g \) is non-increasing then \( g(x + h) - g(x) \leq 0 \) and \( g(x - h) - g(x) \geq 0 \), for any \( h > 0 \). Hence \( Dg(x) \leq 0 \), where again, \( D \) is any of the four possible Dini derivatives \( D^+, D^-, D_+ \) or \( D_- \).

To verify (2), remark that the function \( g \) defined for \( x \) in \( I \) by

\[
g: x \mapsto y(x) - \int_a^x z(t)dt, \quad \text{some } a \in I,
\]

satisfies \( Dg = Dy - z \leq 0 \). Thus, in view of (1), this implies that \( g \) is non-increasing, and again (1) yields that \( Dg = Dy - z \leq 0 \), for any other choice \( D \) of the four possible Dini derivatives \( D^+, D^-, D_+ \) or \( D_- \).

Recall that a function \( f \) has a bounded variation on the interval \( [a,b] \) if the following variation is finite

\[
\text{var}(f, [a,b]) = \sup \left\{ \sum_{i=1}^n |f(x_i) - f(x_{i-1})| : a = x_0 < x_1 < \cdots < x_n = b \right\},
\]

where the supremum is taken over all partitions of the interval \( [a,b] \) and \( \text{var} \) is referred to as the variation operator, and similarly, we define the positive \( \text{var}^+(f, [a,b]) \) and the negative \( \text{var}^-(f, [a,b]) \) variation operators exchanging the absolute value \( |\cdot| \) with the positive part \( \lfloor \cdot \rfloor^+ \) and the negative part \( \lfloor \cdot \rfloor^- \) of a real number in the above definition. Sometimes, we use the shorter notation \( V_f(x) = \text{var}(f, [a,x]) \). It is clear that a function with bounded variation is necessarily bounded (may be discontinuous) and that the set of bounded variation functions forms an algebra, i.e., if \( f \) and \( g \) have bounded variation and \( a, b \) are constants then \( af + bg \) and \( fg \) have also bounded variations. Moreover, if \( \psi \) is a Lipschitz continuous function in \( [a,b] \) (i.e., there is a constant \( M > 0 \) such that \( |\psi(x) - \psi(y)| \leq M|x - y| \), for every \( x, y \) in \( [a,b] \) then \( \text{var}(\psi \circ f, [a,b]) \leq M\text{var}(f, [a,b]) \). For instance, the reader interested in differential equation with bounded variation functions may take a look at the appendix in the book Miller and Rubinovich [82, Appendix 377–423].

**Exercise 7.12.** With the above notation verify that (1) \( \text{var}, \text{var}^\pm \) are additive functions on intervals and sub-additive on the functions, e.g., if \( a < c < b \) then \( \text{var}(f, [a,c]) + \text{var}(f, [c,b]) = \text{var}(f, [a,b]) \) and \( \text{var}(f + \alpha g, [a,b]) \leq \text{var}(f, [a,b]) + |\alpha| \text{var}(g, [a,b]), \) for any real constant \( \alpha \). Next, assuming that \( f \) is a function with bounded variation, show that (2) \( \text{var}(f, [a,b]) = \text{var}^+(f, [a,b]) + \text{var}^-(f, [a,b]) \); (3) \( f \) can be written as the difference of two monotone functions, namely, \( f(x) = \text{var}(f, [a,x]) - (f(x) - \text{var}(f, [a,x])) \) or (4) \( f(x) - f(a) = \text{var}^+(f, [a,x]) - \text{var}^-(f, [a,x]) \). Moreover, if \( \pi = \{ a = x_0 < x_1 < \cdots < x_n = b \} \) is a partition of \( [a,b] \) with mesh (or norm) \( |\pi| = \max_i \{(x_i - x_{i-1})\} \) then prove that
\[
(5) \quad \text{var}(f, \pi) = \sum_{i=1}^{n} |f(x_i) - f(x_{i-1})| \text{ converges to var}(f, [a, b]) \text{ as } |\pi| \to 0, \text{ provided } f \text{ is also continuous. Furthermore, if } f \text{ is also right-continuous then prove (6) the variation functions } x \mapsto \text{var}(f, [a, x]) \text{ and } x \mapsto \text{var}^\pm(f, [a, x]) \text{ are right-continuous and var}(f, [a, x]) \to 0 \text{ as } x \to a^+ \text{. Hint: for instance, check Gordon [54, Chapter 4, pp. 49–68].}
\]

Based on the previous Exercise 7.12, any function with bounded variation can be written as the difference of two monotone functions, thus, Proposition 7.15 proves that the derivative \( f' \) of a bounded variation function \( f \) exists almost everywhere and that \( f' \) is integrable in \( (a, b) \).

As it is known, the Cantor-Lebesgue function (see Exercise 2.25) is an example of an increasing continuous \( f \) satisfying \( f(1) = 1, f(0) = 0 \), and with derivative \( f'(x) = 0 \), almost every \( x \) in \( (0, 1) \). On the other hand, a function \( f \) on \( [a, b] \) is called absolutely continuous, if for any \( \varepsilon > 0 \) there exists \( \delta > 0 \) such that
\[
\sum_{i=1}^{n} |f(b_i) - f(a_i)| < \varepsilon \quad \text{if} \quad \sum_{i=1}^{n} (b_i - a_i) < \delta, \quad a_{i+1} \geq b_i > a_i, \quad n \geq 1.
\]

Also, a function \( f \) with bounded variation and such that \( f'(x) = 0 \), for almost every \( x \) in \( [a, b] \), is said to be singular function on \( (a, b) \). We can show that

1.- An absolutely continuous function is continuous and has bounded variation;
2.- If \( f \) is both absolutely continuous and singular on \( [a, b] \) then \( f \) is constant in \( [a, b] \);
3.- A function \( f \) is absolutely continuous on \( [a, b] \) if and only if \( f' \) exists almost everywhere, \( f' \) is integrable on \( [a, b] \) and
\[
\int_{a}^{x} f'(y) dy = f(x) - f(a), \quad \forall x \in [a, b];
\]
4.- If \( f \) is a function of bounded variation on \( [a, b] \) then the function
\[
g(x) = f(x) - \int_{a}^{x} f'(y) dy, \quad \forall x \in [a, b]
\]
is singular;
5.- If \( f \) is an absolutely continuous function then
\[
\text{var}(f, [a, b]) = \int_{a}^{b} |f'(x)| dx \quad \text{and} \quad \text{var}^\pm(f, [a, b]) = \int_{a}^{b} (f'(x))^\pm dx,
\]
are the total, positive and negative variations;
6.- If \( f \) is (a) continuous on \( [a, b] \), (b) \( f' \) exists almost everywhere, (c) \( f' \) is integrable, and (d) \( f \) maps null sets (i.e., set of Lebesgue measure zero) into null sets, then \( f \) is an absolutely continuous function on \( [a, b] \);
7.- If \( f \) is Lipschitz continuous on \( [a, b] \), i.e., there exists a constant \( M > 0 \)

such that \(|f(x) - f(y)| \leq M|x - y|\), for every \(x, y \in [a, b]\), then \(f\) is absolutely continuous on \([a, b]\).

Remark that Assertion 3 above can be regarded as a Lebesgue version of the Fundamental Theorem of Calculus, e.g., see Ovchinnikov [89, Sections 4.4-5, 114–128] or Rana [92, Section 6.3, pp. 191–207].

- **Remark 7.20.** A non-negative finitely additive set function \(\nu\) defined on an algebra \(\mathcal{A}\) (of subsets of \(\Omega\)) is called **purely finitely additive** if for any (\(\sigma\)-additive) pre-measure \(\mu\) satisfying \(0 \leq \mu \leq \nu\) on \(\mathcal{A}\) follows \(\mu = 0\). The so-called Hewitt-Yosida Theorem affirms that any finitely additive set function \(\nu\) (defined on an algebra) can be decomposed in a unique sum \(\nu = \nu_\sigma + \nu_p\) of a \(\sigma\)-additive measure \(\nu_\sigma\) and a purely finitely additive set function \(\nu_p\). A typical example of a purely finitely additive set function is \(\nu_c\) defined on the algebra \(\mathcal{A} \subset 2^{[0,1)}\), generated by the semi-ring (or semi-algebra) \(\mathcal{S}\) of all the subsets \([a, b)\) with \(0 \leq a < b \leq 1\), by \(A \mapsto \lambda_c(A)\), where \(\lambda_c(A) = 1\), whenever there exists \(\varepsilon > 0\) such that \([c - \varepsilon, t) \subset A\) and \(\lambda_c(A) = 0\), otherwise. To check this, note that \(\mu \leq \nu_c\) then \(\mu([c, 1)) \leq \nu_c([c, 1)) = 1\) and also \(\mu([a, c - \varepsilon)) = 0\) whenever \(0 \leq a < c - \varepsilon\); which implies that \(\mu([0, c/2) = 0\) and \(\mu([c(2k - 1)/(2k), c(2k + 1)/(2k + 2)) = 0\), for any \(k = 1, 2, \ldots\), and therefore, if \(\mu\) is \(\sigma\)-additive then \(\mu([0, t)) = 0\). Moreover, it can be proved that any finitely additive set function \(\nu\) defined on the algebra \(\mathcal{A}\) and having only values in \([0,1)\) has the form \(\nu_c\) above or \(\delta_c\) (the Dirac measure), for some point \(c\) in \([0,1]\). Actually, it can also be proved that any purely finitely additive set function \(\nu\) has the form \(\nu = \sum_{i=1}^{\infty} a_i \nu_{c_i}\), for some sequences \(\{c_i\} \subset [0,1)\) and \(\{a_i\} \subset [0, \infty)\) with \(\sum_i a_i < \infty\). For full details, the reader may consult the book Swartz [113, Sections 2.4–2.6, pp. 41–61].

**Exercise 7.13.** If \(f\) is a function Lebesgue integrable in \((0, a)\), with \(a > 0\) and
\[
g(x) = \int_x^a \frac{f(t)}{t} \, dt, \quad \forall x \in (0, a),
\]
then \(g\) is Lebesgue integrable in \((0, a)\). When the equality
\[
\int_0^a g(x) \, dx = \int_0^a f(x) \, dx
\]
is valid?

**Exercise 7.14.** If \(f\) and \(g\) are two absolutely continuous functions on \([a, b]\) then prove that the following integration-by-parts formula
\[
\int_a^b f'(x)g(x) \, dx + \int_a^b f(x)g'(x) \, dx = f(b)g(b) - f(a)g(a)
\]
is meaningful and correct.

**Exercise 7.15.** Let \(f\) be a right-continuous increasing function and \(m_f\) be its corresponding Lebesgue-Stieltjes measure on \(\mathbb{R}\). First prove that (a) the function \(f\) is absolutely continuous if and only if the measure \(m_f\) is absolutely continuous with respect to the Lebesgue measure \(\ell\). Next, (b) extend these results to functions with bounded variation.
**Exercise 7.16.** Let \( f \) be a monotone increasing function on \([0, \infty)\), and denote by \( f(a-) \) and \( f(b+) \), \( a > 0 \), the lateral limits from the left and from the right. Verify (1) that \( f(x+) = f(x-) \) for any \( x \) except for a countable number of points, and that the functions \( x \mapsto f_-(x) = f(x-) \) with \( f_-(0) = 0 \) and \( x \mapsto f_+(x) = f(x+) \) with \( f_+(0) = f(0) \) are monotone increasing functions, \( f_+ \) is continuous from the right and \( f_- \) is continuous from the left. Prove (2) that \( f_- \leq f \leq f_+ \) and if \( f_- l(0) = 0 \) and \( f_- l(x) = \sum_{0 \leq y < x} (f(y) - f(y)) \forall x > 0 \), then show that \( f_l \) and \( f_- r = f_+ - f_l \) are continuous from the left and \( f_r = f - f_l \) and \( f_+ r \) are continuous from the right, and all four functions are monotone increasing. 

**Exercise 7.17.** A function \( f \) has no discontinuities of the second class at a point \( x \) of \( \mathbb{R} \) if the limits from the left and from the right exists and are finite whenever they are defined (e.g., if a point \( x \) is isolated from the left then the limit from the left can not be defined). Show that a function \( f \) defined on a compact subset \( D \) of \( \mathbb{R} \) has no discontinuities of the second class if and only if there exists a sequence \( \{f_k\} \) of piecewise functions defined on \( \mathbb{R} \) such that \( f_k \) converges to \( f \) uniformly on \( D \). Question: A uniform limit of a function with no discontinuities of the second class is again a function with no discontinuities of the second class? 

Certainly, if \( f \) is a function of bounded variation in \([a,b]\) then \( f \) defines a measure \( \mu_f \) on \([a,b]\), which is absolutely continuous (or singular) with respect to the Lebesgue measure if and only if \( f \) is absolutely continuous (or singular) in \([a,b]\). 

The previous results can be used with convex functions. Indeed, if \( f \) is convex in \((a,b)\) then

\[
\frac{f(x_2) - f(x_1)}{x_2 - x_1} \geq \frac{f(x) - f(x_1)}{x - x_1}, \quad \forall a < x_1 < x < x_2 < b,
\]

which implies that the lateral derivatives of \( f \) exist at each point and \( D^+ f = D_+ f \) and \( D^- f = D_- f \) are monotone increasing (right or left continuous) functions (and so, \( f' \) can only be discontinuous at a countable number of points), and the estimate

\[
D^+ f(y) \leq \frac{f(x) - f(y)}{x - y} \leq D^- f(x), \quad \forall a < y < x < b.
\]

holds. Thus, the second derivative \( f'' \) exists almost everywhere and \( f''(x) \geq 0 \). However, the measure associated with the derivative (i.e., associated with \( x \mapsto f'(x+) \), the right-hand limit) is not necessarily absolutely continuous with
respect to the Lebesgue measure, and then, the condition $f''(x) \geq 0$ almost everywhere becomes a sufficient condition for convexity only if $f'$ is absolutely continuous, e.g., if $g$ is a primitive of the Cantor-Lebesgue function then $g$ is not a linear function (i.e., convex and concave at the same time) and $g'' = 0$ almost everywhere.

- Remark 7.21. If $f$ is a real-valued continuously differentiable function on an open interval $I$ then the function $g : x \mapsto |f(x)|$ has lateral derivatives (i.e., from the right $g'_+$ and from the left $g'_-$) and $-|f'| \leq g'_- \leq g'_+ \leq |f'|$ in $I$. Indeed, the chain rule shows that the function $g$ is differentiable at any $x_0$ where $f(x_0) \neq 0$, and $g'(x_0) = f'(x_0)$ if $f(x_0) > 0$ or $g'(x_0) = -f'(x_0)$ if $f(x_0) < 0$. Thus to calculate the right-hand derivative at any $x_0$ with $g(x_0) = f(x_0) = 0$ write

$$
\frac{g(x) - g(x_0)}{x - x_0} = \left| \frac{f(x) - f(x_0)}{x - x_0} \right|, \quad x > x_0,
$$

to get $g'_+(x_0) = |f'(x_0)|$, and similarly for the left-hand derivative $g'_-(x_0) = -|f'(x_0)|$. Actually, an analogous argument can be used when the absolute value is replaced by a convex function $\Phi$. Indeed, if $f'(x_0) \neq 0$ then $f$ is either increasing or decreasing at $x_0$ and the limit of ratio $[\Phi(f(x_0 + h)) - \Phi(f(x_0))]/[f(x_0 + h) - f(x_0)]$ exists, i.e., $g'_+(x_0) = \Phi'_+(f(x_0))f'(x_0)$ and $g'_-(x_0) = \Phi'_-(f(x_0))f'(x_0)$ if $f'(x_0) > 0$, while $g'_+(x_0) = \Phi'_-(f(x_0))f'(x_0)$ and $g'_-(x_0) = \Phi'_+(f(x_0))f'(x_0)$ if $f'(x_0) < 0$. If $f'(x_0) = 0$ then $g'(x_0) = 0$, because the ratio of $\Phi$ is bounded.

Exercise 7.18. Complete the details on the above statements on convex functions, in particular, show that if $f$ is convex in $(a, b)$, i.e., $f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y)$, for every $x, y$ in $\Omega$ and $t \in [0, 1]$, then (1) $f$ is necessarily continuous on $(a, b)$, (2) the left-hand derivative $f'_+(t) = \lim_{h \to 0^+} f(t + h) - f(t)/h$ exists and is right-continuous and increasing at any $t$, and (3) if $f$ is absolutely continuous and the derivative $f'$ is continuous on the complement of a countable set of points. Moreover, (4) show that any slope $m = m_x$ satisfying $f'_-(x) \leq m \leq f'_+(x)$ provides a supporting line $x$ in $(a, b)$, i.e., $f(y) \geq f(x) + m(y - x)$ for every $y$ in $(a, b)$. Finally, consider a function $f$ defined only on the dyadic points of $(a, b)$, i.e., for $x = a + k2^{-n}(b - a)$, $k = 1, \ldots, n - 1$, $n = 1, 2, \ldots$, such that $2f((x + y)/2) \leq f(x) + f(y)$ for any dyadic point $x$ and $y$ and prove that (5) $f$ can be uniquely extended to a convex function defined on the whole interval $(a, b)$.

Exercise 7.19. Prove Jensen’s inequality, i.e., if $\Phi : \mathbb{R} \to \mathbb{R}$ is a convex function and $\psi$ is a real-valued integrable function in a probability space $(\Omega, \mathcal{F}, P)$ then

$$
\Phi\left( \int_\Omega \psi(\omega)P(d\omega) \right) \leq \int_\Omega \Phi(\psi(\omega))P(d\omega).
$$

What can be said when $\phi$ is not necessarily integrable? In particular, deduce that if $f$ and $k$ are real-valued measurable functions on a measure space $(X, \mathcal{X}, \mu)$ such $fk$ is integrable and $k \geq 0$ is a kernel (i.e., an integrable function with integral equals to 1) then

$$
\Phi\left( \int_X fk \, d\mu \right) \leq \int_X \Phi(f)k \, d\mu,
$$

Exercise 7.20. Use Proposition 7.15 to prove that (a) if \{f_k\} is a sequence of monotone increasing functions defined on the interval \([a,b]\) such that the numerical series \(g(x) = \sum_k f_k(x)\) converges for every \(x\) in \([a,b]\) then we have \(g'(x) = \sum_k f'_k(x)\), for almost every \(x\) in \([a,b]\). Next, show that (b) the derivative of the variation satisfies \(V_f' (x) = |f'(x)|\), for almost every \(x\) in \((a,b)\).

Exercise 7.21. Prove, as much as possible, the above claims 1, \ldots, 7, and then check the references below for full details.

For a more comprehensive study, the reader to check the first three chapters of the book Leoni [75, Chapters 2 and 3, pp. 3–113] regarding monotone, bounded variation and absolutely continuous functions. In particular, the reader should note the concept of functions of bounded variation may take various forms.

A (right-continuous) function \(\alpha : \mathbb{R} \to \mathbb{R}\) is said to be of finite variation if \(\alpha\) has bounded variation on any bounded interval, or equivalently, it can be expressed as the different of two (right-continuous) monotone increasing functions, \(\alpha = \alpha_1 - \alpha_2\) where \(\alpha_i : \mathbb{R} \to \mathbb{R}\) is a (right-continuous) non-decreasing function, for \(i = 1, 2\). In this case, the integral

\[
\int_{[a,b]} f(s) \, d\alpha(s) = \int_{[a,b]} f(s) \, d\alpha_1(s) - \int_{[a,b]} f(s) \, d\alpha_2(s),
\]

for any bounded Borel function \(f\), actually, \(d\alpha\) is the (unique) Lebesgue-Stieltjes signed-measure associated with \(\alpha\).

Exercise 7.22. Prove the integration by part formula, namely, if \(\alpha\) and \(\beta\) are two right-continuous function of finite variation then

\[
\int_{[a,b]} \beta(s) \, d\alpha(s) + \int_{[a,b]} \alpha_-(s) \, d\beta(s) = \alpha(b)\beta(b) - \alpha(a)\beta(a),
\]

for every real numbers \(b > a\) and where \(\alpha_-(s) = \alpha(s-)\) is the left-continuous version obtained from \(\alpha\).

The following result is called Kunita-Watanabe inequality

Exercise 7.23. Let \(\alpha\) be a right-continuous function of finite variation on \([0, +\infty[\) and let \(a\) and \(b\) be two non-decreasing right-continuous functions on \([0, +\infty[\). If

\[
|\alpha(t) - \alpha(s)| \leq \sqrt{a(t) - a(s)} \sqrt{b(t) - b(s)}, \quad \forall t > s \geq 0,
\]

then for any Borel functions \(f\) and \(g\) on \(\mathbb{R}\) we have

\[
\int_{[0, +\infty[} |f(s)| |g(s)| \, d\alpha(s) \leq \left( \int_{[0, +\infty[} |f(s)|^2 \, d\alpha(s) \right)^{1/2} \left( \int_{[0, +\infty[} |g(s)|^2 \, db(s) \right)^{1/2},
\]

provided \(k \geq 0\) is a kernel, i.e., with integral equal to 1. Hint: verify first that because \(\Phi\) is convex then for every \(t_0\) there exists a slope \(\alpha(t_0)\) such that \(\Phi(t) \geq \Phi(t_0) + \alpha(t_0)(t - t_0)\), for every \(t\) in \(\mathbb{R}\).
where $|d\alpha(t)| = \text{var}(\alpha, [0,t])$ is the variation function associated with $\alpha$. Hint: If $v(t) = r^2a(t) + 2ra(t) + b(t)$ then consider the quadratic form $r \mapsto [v(t) - v(s)]$ to deduce that the Radon-Nikodym derivatives $d\nu/d\mu = r^2a'(t) + 2ra'(t) + b'(t)$ is nonnegative, where $d\mu$ is the Lebesgue-Stieltjes measure associated with the function $\mu(t) = \text{var}(\alpha, [0,t]) + a(t) + b(t)$. Hence, the inequality $|\alpha'| \leq \sqrt{a'b'}$ and Hölder inequality (for the case $p = q = 2$, i.e., Cauchy inequality) complete the argument.

For more details on this section, the reader may consult other books, e.g., Folland [45, Chapter 3, pp. 85–111], Gordon [54, Chapter 4, pp. 49–68], Jones [65, Chapter 16, pp. 511–580], Leoni [75, Part I, pp. 3–228], Riesz and Nagy [94, Chapters I and II, pp. 3–104] and Wheeden and Zygmund [119, Chapter 7, pp. 98–124]. Perhaps the reader may take a look at Pugh [91, Chapter 6, pp 363–417] for historic-type context.

For one variable, we may define first sets of measure zero (eventually cover some rudiments of measure theory in $\mathbb{R}$) and then introduce the Lebesgue integral as the antiderivative, by means of Proposition 7.15 and a short version of Fubini for series of functions. For instance, a *Lebesgue primitive* of a given almost everywhere defined function $f : \mathbb{R} \to [0, \infty)$, is an increasing and bounded below function $F : \mathbb{R} \to \mathbb{R}$ such that $F' = f$ almost everywhere, e.g., see Bridges [19, Chapter I.2, pp. 79–122]. Also, see Dshalalow [36, Chapter 9, pp. 517–550] and Taylor [114, Chapter 9, pp. 379–422].

### 7.5 Lebesgue Spaces

Perhaps the most typical measures are the Lebesgue measure in $\mathbb{R}^d$ and the counting measure in $\mathbb{N}$, with the corresponding $L^p = L^p(\mathbb{R}^d)$ space of Lebesgue almost everywhere measurable real-valued functions with norm

$$
\|f\|_p = \int_{\mathbb{R}^d} |f(x)|^p dx < \infty
$$

and $\ell^p = \ell^p(\mathbb{R})$ space of real-valued (or complex-valued or $\mathbb{R}^d$-valued) sequences $x = \{x_n\}$ such that

$$
\|a\|_p = \|\{a_n\}\|_p = \sum_{n=1}^{\infty} |a_n|^p < \infty,
$$

with $1 \leq p < \infty$. Also $L^\infty = L^\infty(\mathbb{R}^d)$ is the space of all Lebesgue essentially bounded (i.e., almost everywhere measurable and bounded outside of a negligible set) real-valued functions, namely

$$
\|f\|_\infty = \inf \{C > 0 : |f(x)| \leq C, \text{ a.e.}\},
$$

where the infimum is $\infty$ if the function is not bounded outside a negligible set. Similarly, $\ell^\infty = \ell^\infty(\mathbb{R}^d)$ is the space of all bounded real-valued sequences with
the norm
\[\|a\|_\infty = \|\{a_n\}\|_\infty = \sup \{|a_n| : n \geq 1\}.\]

Certainly, we have the spaces \(L^p(A)\) for any measurable non-negligible subset \(A \subset \mathbb{R}^d\) (of particular interest is the case when \(A = \Omega\) an open set), \(L^p(\mathbb{R}^d; \mathbb{C})\) or \(L^p(\mathbb{R}^d; \mathbb{R}^n)\) (functions with complex values or with values in \(\mathbb{R}^n\)), \(\ell^p(\mathbb{R})\) or \(\ell^p(\mathbb{C})\) (sequences with complex values or with values in \(\mathbb{R}^n\), \(n \geq 1\)). Moreover, if \((0,1] \times \cdots \times(0,1] = (0,1)^d\) is an \(d\)-dimensional interval, and so, with finite measure, i.e., the convolution can be considered in counting measure is a Haar measure on \((\mathbb{Z}, +)\). Actually, we may replace \(\mathbb{Z}\) by any countable set, or even any set of indexes \(I\), where real-valued “sequences” means functions \(a : I \to \mathbb{R}\) with countable support, i.e., such that \(\{i \in I : a_i \neq 0\}\) is finite or countable.

As we have seen, these are Banach spaces, which are separable if \(1 \leq p < \infty\). If \(A\) have a finite measure then \(L^p(A) \subset L^q(A)\) for any \(1 \leq p < q \leq \infty\), and on the contrary, \(\ell^q \subset \ell^p\). In general, \(L^p \cap L^q\) is a subspace of \(L^r\) for any \(1 \leq p \leq r \leq q \leq \infty\).

Recall the convolution defined on \(\mathbb{R}^d\) by the expression

\[(f \ast g)(x) = \int_{\mathbb{R}^d} f(x - y) g(y) \, dy = \int_{\mathbb{R}^d} f(y) g(x - y) \, dy, \tag{7.8}\]

which is defined almost everywhere for any \(f\) (or \(g\)) in \(L^1\) and \(g\) (or \(f\)) in \(L^\infty\). To define the convolution we use the topological group structure \((\mathbb{R}^d, +)\). In general, if \((\Omega, +)\) is a locally compact (abelian) group then a translation-invariant Radon measure on \(\Omega\) is called a Haar measure, and there is one and only one (up to a multiplicative constant) Haar measure, e.g. see Folland \[45, Section 11.1, pp. 339–348\] or Cohn \[28, Chapter 9, pp. 297-327\]. For instance, the Lebesgue measure is a Haar measure on \((\mathbb{R}^d, +)\) with the Euclidean topology and the counting measure is a Haar measure on \((\mathbb{Z}, +)\) or \((\mathbb{R}^d, +)\) with the discrete topology. Thus,

\[(a \ast b)_n = \sum_k a_{n-k} b_k = \sum_k a_k b_{n-k} \tag{7.9}\]

is a discrete version of (7.8). We are more interested in the continuous case.

If \(f\) and \(g\) have support in \(\mathbb{R}^d_+ = [0, \infty)^d\), then we have

\[(f \ast g)(x) = \int_{(0,x)} f(x - y) g(y) \, dy = \int_{(0,x)} f(y) g(x - y) \, dy,\]

where \((0,x) = (0,x_1) \times \cdots \times (0,x_d)\) is a bounded \(d\)-dimensional interval, and so, with finite measure, i.e., the convolution can be considered in \(L^1_{\text{loc}}\).

**Proposition 7.22** (Young Inequality). *If \(f\) belongs to \(L^p(\mathbb{R}^d)\) and \(g\) belongs to \(L^q(\mathbb{R}^d)\) then \(f \ast g\) belongs to \(L^r(\mathbb{R}^d)\) and

\[\|f \ast g\|_r \leq \|f\|_p \|g\|_q,\]

provided \(1 \leq p, q, r \leq \infty\) and \(1/p + 1/q - 1/r = 1\).
Proof. We integrate in $y$ the expression

$$|f(x - y)g(y)| = (|f(x - y)|^{p/r}|g(y)|^{q/r}) \times$$

$$\times (|f(x - y)|^{p(1/r-1)}) \times (|g(y)|^{q(1/q-1/r)})$$

and we apply H"older inequality as in Remark 4.29 with the exponents $r, p_1$ and $q_1$ satisfying $1/p_1 = 1/p - 1/r$ and $1/q_1 = 1/q - 1/r$, to obtain

$$|(f \star g)(x)|^r \leq \left((|f|^p \ast |g|^q)(x)\right)(\|f\|_p^{r-p}\|g\|_q^{r-q}).$$

Hence, integrating in $x$ we deduce

$$\|f \star g\|_r^r \leq \|f\|_p^p \|g\|_q^q \|f\|_p^{r-p} \|g\|_q^{r-q} = \|f\|_p^p \|g\|_q^q,$$

i.e., the desired estimate, for $p, q, r$ finite.

Analogously, we treat the limiting cases when some of the exponents are infinite. \hfill \Box

Most of the properties proved in Section 7.1 valid for $L^1$ can be extended to $L^p$, with $1 \leq p < \infty$.

(a) The translation is continuous in $L^p(\mathbb{R}^d)$, i.e., if $\tau_a f(\cdot) = f(\cdot + a)$ then $\|\tau_a f - f\|_p \to 0$ as $a \to 0$, for every $f$ in $L^p$. Indeed, we argue similar to Proposition 7.3. \hfill \Box

(b) The space $C_0^0$ of all continuous functions on $\mathbb{R}^d$ with compact support is dense in $L^p$, i.e., for every $\varepsilon > 0$ and $f$ in $L^p(\mathbb{R}^d)$ there exists $g_\varepsilon$ in $C_0^0(\mathbb{R}^d)$ such that $\|f - g_\varepsilon\|_p < \varepsilon$. Indeed, essentially the same arguments as those of Proposition 7.2. \hfill \Box

(c) The kernel convolution converges in $L^p$, i.e., with the notation of Corollary 7.6, $\|f \ast k_\varepsilon - f\|_p \to 0$ as $\varepsilon \to 0$, for every $f$ in $L^p$. Indeed, we apply H"older inequality to the right-hand term of

$$|(f \ast k_\varepsilon)(x) - f(x)| \leq \int_{\mathbb{R}^d} (|f(x - y) - f(x)| |k_\varepsilon(y)|^{1/p}) (|k_\varepsilon(y)|^{1/q}) dy,$$

with $1/p + 1/q = 1$, and we integrate in $dx$ to obtain

$$\|f \ast k_\varepsilon - f\|_p^p \leq \|k_\varepsilon\|_1^{p/q} \int_{\mathbb{R}^d} dx \int_{\mathbb{R}^d} |f(x - y) - f(x)|^p |k_\varepsilon(y)| dy.$$
The continuity of the translation \((a)\) shows that \(\phi(y) \to 0\) as \(|y| \to 0\), and so, for every \(\varepsilon_1 > 0\) there exists \(\delta > 0\) such that

\[
\int_{\{|y|<\delta\}} \phi(y) |k_\varepsilon(y)| dy \leq \varepsilon_1 \int_{\{|y|<\delta\}} |k_\varepsilon(y)| dy \leq \varepsilon_1 \|k\|_1 \leq \varepsilon_1.
\]

Since \(\phi\) is bounded, i.e., \(\|\phi(y)\|_\infty \leq (2\|f\|_p)^p\), we obtain

\[
\int_{\{|y|\geq\delta\}} \phi(y) |k_\varepsilon(y)| dy = \int_{\{|y|\geq\delta/\varepsilon\}} \phi(\varepsilon y) |k(y)| dy \leq (2\|f\|_p)^p \int_{\{|y|\geq\delta/\varepsilon\}} |k(y)| dy,
\]

where the right-hand side tends to 0 as \(\varepsilon \to 0\). This proves that \(f \ast k_\varepsilon \to f\) in \(L^p\), as \(\varepsilon \to 0\).

Based on these properties we have

**Proposition 7.23.** Let \(\Omega\) be an open subset of \(\mathbb{R}^d\) and \(C_0^\infty(\Omega)\) be the space of all real-valued functions having derivatives of any order and compact supports. Then \(C_0^\infty(\Omega)\) is dense in \(L^p(\Omega)\), for any \(1 \leq p < \infty\).

**Proof.** We may apply directly Proposition 7.7, or we could use the following argument.

It is clear that we can find a sequence \(\{\Omega_n : n \geq 1\}\) of open sets with compact closure satisfying

\[
\overline{\Omega}_n \subset \Omega_{n+1} \subset \overline{\Omega}_{n+1} \subset \Omega, \quad \forall n \quad \text{and} \quad \bigcup_n \Omega_n = \bigcup_n \overline{\Omega}_n = \Omega.
\]

By means of the dominate convergence we check that

\[
\int_{\Omega} |\mathbbm{1}_{\Omega_n}(x)f(x) - f(x)|^p dx \to 0,
\]

i.e., \(\|\mathbbm{1}_{\Omega_n}f - f\|_p \to 0\) as \(n \to \infty\). Hence, we are reduced to approximate functions with compact supports.

Therefore, let \(f\) be a function in \(L^p(\Omega)\) which vanishes outside of some compact set \(K = K_f \subset \Omega\). It is then clear that there exists a continuous function \(k\) with compact support inside \(\Omega\) such that \(k = 1\) on \(K\) and \(0 \leq k \leq 1\) on \(\Omega\). Now, as in Proposition 7.2, for every \(\varepsilon > 0\) there exists a continuous function \(g_\varepsilon\) with compact support such that

\[
\int_{\mathbb{R}^d} |\mathbbm{1}_K(x)f(x) - g_\varepsilon(x)|^p dx < \varepsilon,
\]

which implies that \(\|f - kg_\varepsilon\|_p < \varepsilon\). Actually, by means of a convolution with a smooth kernel, we can choose \(kg_\varepsilon\) in \(C_0^\infty(\Omega)\). \(\square\)

**Proposition 7.24.** If \(1 \leq p < \infty\) and \(A\) is a measurable subset of \(\mathbb{R}^d\) then \(L^p(A)\) is separable Banach space.
Proof. It is clear that only the case $A = \mathbb{R}^d$ needs consideration. Indeed, for any function in $L^p(A)$ can be extended by zero to be obtain an element in $L^p(\mathbb{R}^d)$ and backward, any function $f$ in $L^p(\mathbb{R}^d)$ becomes a function in $L^p(A)$ by setting $g = 1_A f$, which is a continuous linear transformation.

There are several ways to check that $L^p = L^p(\mathbb{R}^d)$ is separable. For instance, we may consider functions of the form $p(x)1_B$ where $p$ are polynomials with rational coefficients and $B$ are closed balls centered at the origin of radius $1/n$, for $n = 1, 2, \ldots$.

Alternatively, we may consider simple functions of the form $\sum_{j=1}^n a_j 1_{A_j}$, where $a_i$ are rational numbers and $\{A_j\}$ are disjoint $d$-intervals with rational extremes, i.e., of the form $\prod_{i=1}^d [\alpha_i, \beta_i]$, with $\alpha_i$ and $\beta_i$ rational numbers. It is clear that any simple function can be approximate in the $L^p$-norm with simple functions of the above form.

• Remark 7.25. It is clear some of the arguments used in the above Proposition 7.24 can be applied to any Radon measure $(\mathcal{F}, \mu)$ in $\mathbb{R}^d$, so that $L^p(\mathbb{R}^d, \mathcal{F}, \mu)$ is a separable Banach space.

The particular case $L^2(A)$ or $L^2(A; \mathbb{C})$ is a real or complex separable Hilbert space with the scalar or inner product

$(f, g) = \int_A f(x)\overline{g(x)} \, dx$ or $(f, g) = \int_A f(x)\overline{g(x)} \, dx$,

where $\overline{g}$ means the complex-conjugate. We denote by $\| \cdot \| = \| \cdot \|_2$ the corresponding norm.

The following definitions apply to any Hilbert space, but we focus in $L^2$. A family of functions $\{\varphi_i : i \in I\}$ is orthogonal if $(\varphi_i, \varphi_j) = 0$ for every $i \neq j$; it is orthonormal if also $\|\varphi_i\| = 1$, for every $i$; and it is called complete if the only function orthogonal to any $\varphi_i$ is the zero, i.e., if $(f, \varphi_i) = 0$ for every $i$ implies $f = 0$. The (finite) linear combinations of elements in the family is called the span, and if this family $\{\varphi_i : i \in I\}$ of functions is orthogonal then it is called an orthogonal basis if its span is dense in $L^2$. As seen below, for an orthogonal set, being complete or being a basis is an equivalent concept.

Proposition 7.26. There exists a complete orthonormal basis for $L^2$. Moreover, any orthonormal basis is countable and complete.

Proof. If $\{\varphi_i : i \in I\}$ is an orthonormal basis then

$\|\varphi_i - \varphi_j\|^2 = (\varphi_i - \varphi_j, \overline{\varphi_i - \varphi_j}) = \|\varphi_i\|^2 + \|\varphi_j\|^2 = 2$,

for any $i \neq j$. Because $L^2$ is separable, the set of indices $I$ can be at most countable.

If $\{\varphi_i : i \geq 1\}$ is a orthogonal basis and $(f, \varphi_i) = 0$ for every $i$ then $(f, \varphi) = 0$ for any $\varphi$ linear combination of elements in the basis, and so

$\|f\|^2 = (f, \overline{\varphi} - \varphi) \leq \|f\| \|f - \varphi\|$.
Since linear combinations are dense in $L^2$, the quantity $\|f - \varphi\|$ can be made arbitrary small, which implies that $f = 0$, i.e., $\{\varphi_i : i \geq 1\}$ is complete.

Finally, we apply the Gram-Schmidt procedure to a countable dense set $\{\phi_i : i \geq 1\}$ to obtain an orthonormal family $\{\varphi_i : i \geq 1\}$, which is a basis by construction. Thus, we get a complete orthonormal family or system or basis.

It is clear that we have proved that any separable Hilbert space has a (countable) complete orthonormal basis $\{\varphi_i : i \geq 1\}$.

**Exercise 7.24.** With the Lebesgue measure on $]0, +\infty[$, consider the Haar-type functions $f_i(s) = 1_{2i-1 < s \leq 2i} - 1_{2(i-1) < s \leq 2i - 1}$ and $f_{i,n}(s) = 2^{-n/2}f_i(s2^n)$, for $i = 1, \ldots, 4^n$, $n \geq 0$. First, show that if $n \geq m$ then $f_{i,n}f_{j,m} = 0$ except for $i$ within $(j-1)2^{n-m} + 1$ and $j2^{n-m}$. Secondly, show that $\{f_{i,n}\}$ is an orthonormal system in $L^2([0, \infty[)$. Thirdly, prove that $\{f_{i,n}\}$ can be completed to be a basis by adding the functions $\tilde{f}_i(s) = \delta_{i,0}(s) = 1_{(i-1) < s \leq i}$, for $i = 1, 2, \ldots$, for instance, show that $1/2\{\tilde{f}_{i,0}\} \equiv 1/2\{f_{i,0}\}$ yields $\{\tilde{f}_{i,1}(s) = 1_{i-1 < s \leq i}\}$, and $1/2\{\tilde{f}_{i,1}\} \equiv 1/2\{f_{i,1}\}$ yields $\{\tilde{f}_{i,2}(s) = 1_{i-1 < 4s \leq i-1}\}$ and so on. Finally, discuss a similar construction on $L^2([a, b[)$, for $b > a$.

Recall that $\ell^2(\mathbb{R})$ or $\ell^2(\mathbb{C})$ is the space of all real-valued or complex-valued sequences $a = \{a_i : i \geq 1\}$ such that $\|a\|_2 = \left(\sum_{i \geq 1} |a_i|^2\right)^{1/2}$ is finite, which is a separable Hilbert space with the scalar or inner product $(a, b) = \sum_{i \geq 1} a_i\bar{b}_i$.

**Proposition 7.27.** Let $\{e_i : i \geq 1\}$ be a complete orthonormal basis in a separable Hilbert space $H$, e.g., $H = L^2$, with norm $\| \cdot \|$ and inner product $(\cdot, \cdot)$. Then for any given element $h$ in $H$ the series $h_n = \sum_{i=1}^n (h, \bar{e}_i)e_i$, $n \geq 1$, converges to $h$ and Parseval’s formula

$$\|h\|^2 = \sum_{i=1}^\infty |(h, e_i)|^2, \quad \forall h \in H,$$

holds. Moreover, the mapping $T : H \to \ell^2$ defined by $T(h) = \{(h, \bar{e}_i) : i \geq 1\}$ is a linear isometry.

**Proof.** By means of the linearity of the inner product we have

$$\|h_n - h_m\|^2 = \sum_{i=m+1}^n |(h, e_i)|^2 \quad \text{and} \quad \|h - h_n\|^2 = \|f\|^2 - \sum_{i=1}^n |(h, e_i)|^2,$$

which proves that the sequence of partial sum $\{h_n : n \geq 1\}$ is convergent to some function $g$ in $L^2$. Since $h - g$ is orthogonal to any $e_i$, we deduce that $h = g$, $\|h_n - h\| \to 0$ and Parseval’s formula holds.

It is clear that $T$ is linear and that $T^{-1}(a) = \sum_{i \geq 1} a_i e_i$. Also, the parallelogram identity $\|h + g\|^2 + \|h - g\|^2 = 2[(h, \bar{g}) + (g, h)]$ shows that

$$(T(h), T(g)) = \sum_{i \geq 1} (h, e_i)(g, e_i), \quad \forall h, g \in H,$$

i.e., $T$ preserves the inner product.  \[\square\]
Remark 7.28. Based on the Proposition 4.32 and Corollary 6.14, we deduce that the dual space of $L^p(\Omega, \mathcal{F}, \mu)$ with $1 \leq p < \infty$ (i.e., the space of continuous linear functional on $L^p$) is isomorphic to $L^q(\Omega, \mathcal{F}, \mu)$ with $1/p + 1/q = 1$. 

Perhaps the reader may want to take a look at the book Lieb and Loss [77, Chapters 1 and 2, pp. 1–77] for a concrete review on the previous material. Also, plenty of exercises can be found in the book by Gelbaum [49].

7.6 Trigonometric Series

Now we take a closed look at the Hilbert space $L^2(-\pi, \pi) = L^2([a, b], \mathbb{C})$ with the Lebesgue measure and complex-valued functions.

The sequence $e^{ikx}$, $k = 0, \pm 1, \pm 2, \ldots$ is typical orthogonal system in $L^2(-\pi, \pi)$, i.e.,

$$\int_{-\pi}^{\pi} e^{ikx} e^{i\ell x} dx = 0, \quad \text{if } k \neq \ell,$$

and

$$\int_{-\pi}^{\pi} |e^{ikx}|^2 dx = 2\pi, \quad \forall k,$$

where the $\overline{}$ means the conjugate of a complex number, namely, $e^{i\ell x} = e^{-i\ell x}$. Note that there is not change if the open interval $(-\pi, \pi)$ is replaced by $(a, a + 2\pi)$, or in general, a change of scale (i.e., $x = y/r$) transforms $L^2(-\pi, \pi)$ into $L^2([a, b])$, for any arbitrary nonempty open interval $(a, b)$ and even adapt the notation to a multidimensional situation.

The Fourier coefficients and the complex-valued trigonometric series

$$c_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{ikx} dx \quad \text{and} \quad f \sim \sum_{k=-\infty}^{+\infty} c_k e^{ikx} \quad (7.10)$$

are defined for a given complex-valued function $f$ in $L^1(-\pi, \pi)$, note that the series has a double summation, i.e., $k$ ranging over the integer numbers. By looking at the real and the imaginary parts independent, the expressions $(e^{ikx} + e^{-ikx})/2$ and $(e^{ikx} - e^{-ikx})/(2i)$ yield

$$\frac{1}{2}, \cos x, \sin x, \ldots, \cos kx, \sin kx, \ldots,$$

which is an equivalent real-valued orthogonal system. Thus, the (real) Fourier coefficients and the trigonometric series

$$\begin{cases}
\quad a_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos kx \, dx \quad \text{and} \quad b_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin kx \, dx, \\
\quad a_0 = 2c_0, \quad a_k = c_k + c_{-k}, \quad b_k = i(c_k - c_{-k}), \quad 2c_k = a_k + ib_k, \\
\quad \text{and} \quad f \sim \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx),
\end{cases} \quad (7.11)$$

are defined for a given complex-valued function $f$ in $L^1(-\pi, \pi)$. Note that the series has a double summation, i.e., $k$ ranging over the integer numbers.

By looking at the real and the imaginary parts independent, the expressions $(e^{ikx} + e^{-ikx})/2$ and $(e^{ikx} - e^{-ikx})/(2i)$ yield

$$\frac{1}{2}, \cos x, \sin x, \ldots, \cos kx, \sin kx, \ldots,$$
can be used with real-valued functions, to avoid the use of complex numbers. It should be clear that
\[ \sum_{k=-n}^{n} c_k e^{ikx} = \frac{a_0}{2} + \sum_{k=1}^{n} (a_k \cos kx + b_k \sin kx). \]

Usually, the complex values \( c_k[f] \), and the real values \( a_k[f] \) and \( b_k[f] \), given by (7.10) and (7.11), are called the Fourier (cosine and sine) coefficients of the function \( f \). In both cases, the terms of the Fourier series are harmonic oscillations, and its study is referred to as harmonic analysis.

All these Fourier series are called trigonometric series, a partial sum is called a trigonometric polynomial. If a partial sum \( \sum_{k=1}^{n} c_k e^{ikx} \) of order \( n \) is multiplied by \( e^{inx} \) and regarded as a power polynomial \( p(z) = \sum_{k=1}^{n} c_k e^{i(k+n)x} \) in the variable \( z = e^{ix} \) of degree at most \( 2n \), then it must have no more than \( 2n \) zeros. Therefore, a partial sum of order \( n \) with more than \( 2n \) zeros within the interval \( (-\pi, \pi] \) must vanish.

If the function \( f \) is symmetric (or even) then all sine coefficients vanishes, i.e., \( b_k[f] = 0 \), for instance, a particular choice of \( f(x) = 1 \) for \( x \) in \((-r, r) \subset (-\pi, \pi)\) yields
\[ 1_{(-r, r)} \sim 2r \frac{1}{\pi} \left\{ \frac{1}{2} + \sum_{k=1}^{\infty} \frac{\sin kr}{kr} \cos kx \right\} = \frac{r}{\pi} \left\{ 1 + \sum_{k=-\infty}^{+\infty} \frac{\sin kr}{kr} e^{ikx} \right\}. \]

Similarly, if the \( f \) is antisymmetric (odd) then its Fourier series is a cosine series, i.e., \( a_k[f] = 0 \), for instance, a particular choice of \( f(x) = x \) for any \( x \) in \( (-\pi, \pi) \) yields
\[ x \sim \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \sin kx = \frac{1}{2} \sum_{|k|=1}^{\infty} \frac{(-1)^{k+1}}{k} e^{ikx}, \]
i.e., the summation on the last series ranges over all nonzero integers. It should be clear that the trigonometric series associated with a function \( f \) defined on \( (-\pi, \pi) \) produces may or may not be convergent, but if it converges then the series yields a \( 2\pi \)-periodic function defined on \( \mathbb{R} \). In this sense, a \( 2\pi \)-periodic function \( f \) defined on \( \mathbb{R} \) could be used as our initial function \( f \) defined on \( (-\pi, \pi) \), and conversely.

**Proposition 7.29.** If \( f \) is an absolutely continuous function on \([ -\pi, \pi ]\) with \( f(-\pi) = f(\pi) \) then the termwise differentiation of the trigonometric series associated to \( f \) is the trigonometric series associated with \( f' \). Moreover, if \( g \) is an integrable function defined on \( (-\pi, \pi) \) and \( G \) satisfies \( G' = g \) almost everywhere in \( (-\pi, \pi) \) then the function \( x \mapsto F(x) = G(x) - c_0[g]x \) is absolutely continuous on \([ -\pi, \pi ]\) and \( F(-\pi) = F(\pi) \).

**Proof.** First, note that the condition \( f(-\pi) = f(\pi) \) allows us to extend the definition of \( f \) to the whole real line \( \mathbb{R} \) as a periodic absolutely continuous
function, and its derivative, \( f' \), is also a periodic function almost everywhere defined on \( \mathbb{R} \). Similarly, because of \( F(-\pi) = F(\pi) \), the function \( F \) is a periodic absolutely continuous function on the real line \( \mathbb{R} \), and the first part can be applied to \( F \) to deduce that \( c_k[F] = c_k[g]/(ik) \) for any \( k \neq 0 \).

Since
\[
\int_{-\pi}^{\pi} f'(x) \, dx = f(\pi) - f(-\pi) = 0,
\]
the first Fourier coefficient \( c_0[f'] = 0 \). If \( k \neq 0 \) then an integration by parts and the equality \( e^{ik\pi} = e^{-ik\pi} \) yield
\[
\int_{-\pi}^{\pi} f'(x) e^{ikx} \, dx = ik \int_{-\pi}^{\pi} f(x) e^{ikx} \, dx,
\]
i.e., \( c_k[f'] = ik c_k[f] \), which show the validity of the first part of the claim.

Finally, the equality
\[
F(\pi) - F(-\pi) = \int_{-\pi}^{\pi} g(x) \, dx - c_0[g](2\pi) = 0
\]
complete the proof.

The orthogonality of the trigonometric system yields Bessel’s inequality
\[
\sum_{k=-\infty}^{+\infty} |c_k[f]|^2 \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x)|^2 \, dx
\]
as a consequence of Propositions 7.26 and 7.27. Actually, we have

**Theorem 7.30.** The trigonometric system is complete, i.e., if an integrable function on \((-\pi, \pi)\) has all its Fourier coefficients zero then the function is equal to zero almost everywhere. Moreover, Parseval’s equality
\[
\frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) \overline{g(x)} \, dx = \sum_{k=-\infty}^{+\infty} c_k[f] \overline{c_k[g]}, \quad \forall f, g \in L^2(\mathbb{R})
\]
holds.

**Proof.** First, assume that \( f \) is a real-valued continuous function on \([-\pi, \pi]\) such that \( c_k[f] = 0 \) for every \( k \). If \( |f| \) attains a nonzero maximum value at some point \( x^* \), i.e., \( M = f(x^*) > 0 \), and by continuity, there exists a small interval \( I_\delta = ((x^* - \delta) \cap (-\pi), (x^* + \delta) \cap \pi) \), with \( \delta > 0 \), such that \( f(x) > M/2 \) for every \( x \) in \( I_\delta \). The trigonometric polynomial \( p(x) = 1 + \cos(x - x_0) - \cos \delta \) satisfies (a) \( p(x) > 1 \) for any \( x \) in \( I_\delta \) and (b) \( |p(x)| \leq 1 \) for every \( x \) in \( I_\delta^c = (-\pi, \pi) \setminus I_\delta \). Because all Fourier coefficients vanishes, we must have
\[
\int_{-\pi}^{\pi} f(x)(p(x))^n \, dx = 0, \quad \forall n = 1, 2, \ldots
\]
However, \((-\pi,\pi) = I_\delta^c \cup I_\delta\), and on the complement \(I_\delta^c\),
\[
\left| \int_{I_\delta^c} f(x)(p(x))^n \, dx \right| \leq 2\pi M, \quad \forall n,
\]
while setting \(a = \sup_{I_{\delta/2}} p(x) > 1\) and noting that \(f(x) \geq M/2 > 0\) for any \(x\) in \(I_\delta \supset I_{\delta/2}\) we deduce
\[
\int_{I_\delta \cap [-\pi,\pi]} f(x)(p(x))^n \, dx \geq \int_{I_{\delta/2} \cap [-\pi,\pi]} f(x)(p(x))^n \, dx \geq \frac{M \delta}{2^2} a^n,
\]
All these imply the contradiction
\[
\int_{-\pi}^{\pi} f(x)(p(x))^n \, dx \to +\infty,
\]
which prove the result under the extra assumption that \(f\) is a real-valued continuous function on \([-\pi,\pi]\).

If \(f\) is a complex-valued continuous function then the condition \(c_k[f] = 0\) for every \(k\) implies that \(c_k[F] = 0\) for every \(k\), and then, the previous argument can be used on the real part and the imaginary part of \(f\).

Finally, if \(f\) is only integrable then the hypothesis that \(c_0[f] = 0\) implies that
\[
F(x) = \int_{-\pi}^{x} f(y) \, dy, \quad \forall x \in [-\pi,\pi]
\]
is an absolutely continuous function satisfying \(F(-\pi) = F(\pi)\), to which Proposition 7.29 can be applied to deduce that \(c_k[f] = ikc_k[F]\), for any \(k\). Thus, apply all the previous arguments to the function \(F\) to obtain \(c_k[F] = 0\) for any \(k\), which yields the completeness of the trigonometric system.

Parseval’s equality follows from Propositions 7.27 and the relation
\[
\Re(f, g) = \frac{1}{2} \left[ \|f + g\|_2^2 - \|f\|_2^2 - \|g\|_2^2 \right]
\]
between the inner product and the norm in \(L^2\).

There several consequences of the completeness of the trigonometric system, for instance, the equality \(f = \sum_{k=-\infty}^{+\infty} c_k[f]e^{ikx}\) almost everywhere for any function \(f\) in \(L^2([-\pi,\pi])\) implies that if \(f\) is continuous and the Fourier series converges uniformly then the sum of the series must be \(f\). Also, the relation \(c_k[f] = c_k[f']/(ik)\) and Bessel and Cauchy-Schwarz inequalities imply
\[
\sum_k |c_k[f]| \leq \left( \sum_k |c_k[f']|^2 \right) \left( \sum_k |ik|^{-2} \right) < \infty,
\]
which means that if \(f\) is an absolutely continuous function on \([-\pi,\pi]\) such that \(f(-\pi) = f(\pi)\) and it derivative \(f'\) is square-integrable (in particular, if \(f\) is continuously differentiable) then the Fourier series of \(f\) converges absolutely and uniformly to \(f\).
7.6. Trigonometric Series

Theorem 7.31. The Fourier coefficients \( c_k[f] \) of any integrable function \( f \) on \([-\pi, \pi]\) tend to zero as \(|k|\) goes to infinite. Hence, also \( a_k[f] \to 0 \) and \( b_k[f] \to 0 \) as \( k \to \infty \). Moreover, if \( f \) satisfies \( f(-\pi) = f(\pi) \) and it is Hölder continuous, i.e.,

\[
|f(x) - f(y)| \leq M|x - y|^\alpha, \quad \forall x, y \in [-\pi, \pi]
\]

for some constant \( M > 0 \) and \( 0 < \alpha < 1 \), then \( |c_k[f]|k^\alpha \leq M\pi^\alpha \), for every \( k \).

Furthermore, if \( f \) is a bounded variation function on \([-\pi, \pi]\) then \( 2\pi |c_k[f]|k \leq |f(\pi) - f(-\pi)| + \var(f, [-\pi, \pi]) \), for every \( k \).

Proof. First, in view of Bessel’s inequality, if \( f \) belongs to \( L^2([-\pi, \pi]) \) then the series with \( |c_k[f]|^2 \) converges and therefore \( c_k[f] \to 0 \) as \(|k| \to \infty \). Because the sequence \( f_n = f1_{|f| \leq n} \) converges to \( f \) as \( n \to \infty \), in \( L^1([-\pi, \pi]) \), and \( |f_n|^2 \leq n|f| \), for every \( \varepsilon > 0 \) there exists \( g \) in \( L^2([-\pi, \pi]) \) and \( h \) with \( \|h\|_1 \leq \varepsilon \) such that \( f = g + h \). Thus \( c_k[f] = c_k[g] + c_k[h] \), with \( |c_k[h]| \leq \|h\|_1 \leq \varepsilon \) and \( c_k[g] \to 0 \) as \(|k| \to \infty \). Hence, \( \limsup_k |c_k[f]| \leq \varepsilon \), i.e., \( c_k[f] \to 0 \) as \(|k| \to \infty \).

Actually, an estimate can be obtained. Indeed, if \( f \) is extended by periodicity to the whole real line then, the change of variable \( x = y + \pi/k \) yields

\[
c_k[f] = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)e^{ikx} \, dx = -\frac{1}{2\pi} \int_{-\pi}^{\pi} f(y + \pi/k)e^{iky} \, dy.
\]

Hence, from the semi-sum follows

\[
c_k[f] = \frac{1}{4\pi} \int_{-\pi}^{\pi} f(x) - f(x + \pi/k) \, e^{ikx} \, dx
\]

\[
|c_k[f]| \leq \frac{1}{4\pi} \int_{-\pi}^{\pi} |f(x) - f(x + \pi/k)| \, dx,
\]

and the continuity of the translation, Proposition 7.3, shows again that \( c_k[f] \to 0 \) as \(|k| \to \infty \), for any integrable function \( f \). Moreover, the expression

\[
w_1(f, \delta) = \sup_{|y| \leq \delta} \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x) - f(x + y)| \, dx
\]

gives the estimate \( |c_k[f]| \leq w_1(f, \pi/k) \).

Therefore, if \( f \) is Hölder continuous on the interval \([-\pi, \pi + \delta]\) then \( w_1(f, \delta) \leq M\delta^\alpha \), which yields \( |c_k[f]| \leq M\pi^\alpha k^{-\alpha} \) as desired.

Similarly, if \( f \) is a function of bounded variation then

\[
2\pi ik c_k[f] = \int_{-\pi}^{\pi} f(x)ike^{ikx} \, dx = f(x)e^{ikx}|_{-\pi}^{\pi} - \int_{-\pi}^{\pi} e^{ikx} \, df(x),
\]

which yields

\[
2\pi k |c_k[f]| \leq |f(\pi) - f(-\pi)| + \var(f, [-\pi, \pi])
\]
as desired.

\[\square\]
Note that $f$ is absolutely continuous and $f(-\pi) = f(\pi)$ then
the relation $c_k[f] = c_k[f']/(ik)$ implies that $c_k[f] k \to 0$ as $|k| \to \infty$.

\textbf{Remark 7.32.} The above arguments can be adapted to show that

$$\lim_{n \to \infty} \int_{-\pi}^{\pi} f(x) \sin nx \, dx = \lim_{n \to \infty} \int_{-\pi}^{\pi} f(x) \cos nx \, dx = 0, \quad \forall f \in L^1(\mathbb{R}). \quad (7.13)$$

Indeed, if $f$ is continuously differentiable with support in $[a,b]$ then an integra-

tion by parts yields

$$\int_{-\pi}^{\pi} f(x) \sin nx \, dx = -f(x) \frac{\cos nx}{n} \bigg|_{-\pi}^{\pi} + \frac{1}{n} \int_{-\pi}^{\pi} f'(x) \cos nx \, dx,$$

which proves (7.13). Approximating $f$ by smooth functions the desired result
is obtained. \hfill \square

\textbf{Exercise 7.25.} The Dirichlet kernel of index $n$ is the function

$$D_n(x) = \frac{1}{2} + \cos x + \cdots + \cos nx$$

Prove (1) that $D_n(x) = n + 1/2$ if $x = 0, \pm 2\pi, \pm 4\pi, \ldots$, and

$$D_n(x) = \left[ \sin \left( (n + 1/2)x \right) \right] / [2 \sin(x/2)]$$

otherwise, and (2)

$$\int_{-\pi}^{\pi} D_n(x) \, dx = \pi, \quad \forall n = 0, 1, 2, \ldots,$$

and (3) if $f$ belongs to $L^1([-\pi,\pi])$ then

$$S_n(f, x) = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x \pm y) D_n(y) \, dy, \quad \forall n = 0, 1, 2, \ldots,$$

where either sign $+$ or $-$ can be chosen, and $S_n(f, x)$ is the Fourier sum asso-
ciated with $f$, i.e., $S_n(f, x) = a_0[f]/2 + \sum_{k=1}^{n} (a_k[f] \cos kx + b_k[f] \sin kx)$ \hfill \square

If a function $f$ is continuous on $[-\pi,\pi]$ and $f(-\pi) = f(\pi)$ then its Fourier
series may be unbounded at some point. In general, the pointwise convergence
of the Fourier series is a delicate question.

For instance, the interested reader should take a look first at the book Whee-
den and Zygmund [119, Chapter 12, pp. 211–263] and Zaanen [123], and later
at the classic book Zygmund [125].

\section*{7.7 Some Complements}

Notice: This section is not self-contained (in this book, because more function
analysis is necessary, in particular, a Baire category argument is used), but it
seems to complete well some questions discussed early.

The following theorem is perhaps the most important result relative to the
theory of set functions, and a proof can be found in Dunford and Schwartz [40,
Vol 1, Section III.7, pp. 155–164] or Yosida [122, Section II.2, pp. 70–72]. An
additive set function $\lambda : \mathcal{F} \to \mathbb{R}^d$ is called $\mu$-continuous if for every $\varepsilon > 0$
there exists $\delta > 0$ such that for every $F \in \mathcal{F}$ with $\mu(F) < \delta$ we have $|\lambda(F)| < \varepsilon$, see
Definition 6.2, Remark 6.4 and Definition 6.15.
Theorem 7.33 (Vitali-Hahn-Saks). Let \((\Omega, \mathcal{F}, \mu)\) be a measure space and \(\{\lambda_n\}\) be a sequence of \(\mu\)-continuous additive set functions \(\mathbb{R}^d\)-valued. If the limit \(\lim_n \lambda_n(A)\) exists and is finite for every \(A\) in \(\mathcal{F}\) then \(\{\lambda_n\}\) is \(\mu\)-uniformly continuous, i.e., for every \(\varepsilon > 0\) there exists \(\delta > 0\) such that for every \(F\) in \(\mathcal{F}\) with \(\mu(F) < \delta\) we have \(\sup_n |\lambda_n(F)| < \varepsilon\).

Proof. First, recall that the set \(L^1(\Omega, \mathcal{F}, \mu)\) of all real-valued (or complex-valued) integrable functions is a Banach space and the subset \(\mathcal{F}_0(\mu)\) of all functions integrable functions \(f = \mathbb{1}_F\), with \(F\) in \(\mathcal{F}\) is a closed (just use the fact that from any convergence sequence in \(L^1\) we can a convergent subsequence almost everywhere), but certainly, \(\mathcal{F}_0(\mu)\) is not a vector subspace. Thus \(\mathcal{F}_0(\mu)\) is a complete metric space and the space of simple functions (almost measurable functions taking a finite number of values) \(\mathcal{S}(\Omega, \mathcal{F}, \mu)\) is the linear vector space generated by \(\mathcal{F}_0(\mu)\) is a dense subspace in \(L^1(\Omega, \mathcal{F}, \mu)\). Moreover, the complete metric space \(\mathcal{F}_0(\mu)\) can be also regarded as the sets in \(\mathcal{F}\) with finite \(\mu\)-measure and identified \(\mu\)-almost everywhere, where the distance is given by \(d(A, B) = \mu(A \cup B \setminus A \cap B)\). If \(\mathcal{F}(\mu)\) denotes the elements in \(\mathcal{F}\) with the \(\mu\)-almost everywhere equality and the distance \(d(A, B) = \arctan(\mu(A \cup B \setminus A \cap B))\), then \(\mathcal{F}(\mu)\) is also a complete metric space. Since \(\lambda_n\) is \(\mu\)-continuous, we may consider \(\lambda_n\) as a \(\mathbb{R}^d\)-valued continuous function on \(F(\mu)\). Thus

\[
\mathcal{F}_{n,\varepsilon} = \{A \in \mathcal{F}(\mu) : \sup_{k \geq 1} |\lambda_n(A) - \lambda_{n+k}(A)| \leq \varepsilon\}, \quad \forall n \geq 1, \forall \varepsilon > 0,
\]

is a closed subset of \(\mathcal{F}(\mu)\) and because the limit \(\lim_n \lambda(A)\) exists and is finite, we have the equality \(\mathcal{F}(\mu) = \bigcup_n \mathcal{F}_{n,\varepsilon}\), for every \(\varepsilon > 0\).

Any complete metric space is a second category set, in particular \(\mathcal{F}(\mu)\) is a second category set and thus, at least one \(\mathcal{F}_{m,\varepsilon}\) must has nonempty interior (see Baire category arguments later on!). Hence, there exists \(\delta > 0\) and \(A_0\) in \(\mathcal{F}(\mu)\) such that

\[
d(A, A_0) < \delta \quad \text{implies} \quad \sup_{k \geq 1} |\lambda_m(A) - \lambda_{m+k}(A)| \leq \varepsilon.
\]

Thus, for any \(A\) in \(\mathcal{F}(\mu)\) with \(\mu(A) < \delta\) we take \(A_1 = A \cup A_0\) and \(A_2 = A_0 \setminus A \cap A_0\) to have \(A = A_1 \setminus A_2\) and therefore

\[
|\lambda_n(A)| \leq |\lambda_m(A)| + |\lambda_m(A) - \lambda_n(A)| \leq \\
\leq |\lambda_m(A)| + |\lambda_m(A_1) - \lambda_n(A_1)| + |\lambda_m(A_2) - \lambda_n(A_2)| \leq \\
\leq |\lambda_m(A)| + 2\varepsilon, \quad \forall n \geq m,
\]

which shows that the sequence \(\{\lambda_n\}\) is \(\mu\)-uniformly continuous. \(\square\)

Corollary 7.34. Let \(\{f_n\}\) be a bounded sequence in \(L^1(\Omega, \mathcal{F}, \mu)\) such that the limit

\[
I_n(A) = \int_A f_n \, d\mu, \quad \lim_n I_n(A) = I(A), \quad \forall A \in \mathcal{F}
\]
exists and is finite. Then $I$ is $\sigma$-additive real-valued set function and the finite measures

$$A \mapsto \nu_n(A) = \int_A |f_n| \, d\mu, \quad \forall A \in \mathcal{F}$$

are uniformly $\sigma$-additive, i.e., if $A_k \in \mathcal{F}$, $A_k \subseteq A_{k-1}$ and $\bigcap_k A_k = \emptyset$ then $\sup_n \nu_n(A_k) \to 0$ as $k \to \infty$. Moreover, for every $\varepsilon > 0$ there exists $A \in \mathcal{F}$ with $\mu(A) < \infty$ such that $\nu_n(A^c) < \varepsilon$.

**Proof.** Define the finite measure

$$\lambda(A) = \sum_{n=1}^{\infty} 2^{-n} \lambda_n(A) \quad \text{with} \quad \lambda_n(A) = \frac{1}{\|f_n\|_1} \int_A |f_n| \, d\mu,$$

to get that $\lambda(A_k) \to 0$ as $k \to \infty$.

Since $I_n$ is $\lambda$-continuous and $\lambda$ is finite, Vitali-Hahn-Saks Theorem 7.33 implies that $\{I_n\}$ are uniformly $\sigma$-additive and thus, $I$ is $\sigma$-additive. Actually, for every $\varepsilon > 0$ there exists $\delta > 0$ such that $\lambda(E) < \delta$ implies $|I_n(E)| < \varepsilon$, for any $n$. Now, consider

$$I_n^+(A) = \int_A f_n^+ \, d\mu, \quad I_n^-(A) = \int_A f_n^- \, d\mu, \quad \forall A \in \mathcal{F}, \forall n.$$

Since $\lambda(A_k \cap B) \leq \lambda(A_k)$ for every $B \in \mathcal{F}$ we have $I_n(A_k \cap B) < \varepsilon$ if $\lambda(A_k) < \delta$, and in particular for $B = 1_{\{f_n > 0\}}$ we deduce $I_n^+(A_k) < \varepsilon$. Similarly, we obtain $I_n^-(A_k) < \varepsilon$, and therefore $\nu_n(A) = I_n^+(A_k) + I_n^-(A_k) < 2\varepsilon$ if $\lambda(A_k) < \delta$. Hence, $\{\nu_n\}$ are uniformly $\sigma$-additive.

Finally, because each $f_n$ is integrable, the set $E = \bigcup_n \{f_n \neq 0\}$ is $\sigma$-finite, i.e., $E = \bigcup_k E_k$ with $E_k \subseteq E_{k+1}$, $\mu(E_k) < \infty$ and also $\nu_n(E^c) = 0$. Therefore $\nu_n(E_k^c) = \nu_n(E \setminus E_k) < \varepsilon$ if $\lambda(E \setminus E_k) < \delta$, which must hold for $k$ sufficiently large since $\bigcap_k (E \setminus E_k) = \emptyset$. Thus, we choose $A = E_k$ to conclude. \qed

**Weak Convergence**

Comparing with Definition 6.15, we see that the condition on the set $A$ with finite measure is assured when the sequence $\{f_n\}$ is weakly convergent. Thus, we make some comments on weak convergence.

**Definition 7.35.** A sequence $\{f_n\}$ in $L^p(\Omega, \mathcal{F}, \mu)$, $1 \leq p \leq \infty$, converges weakly to $f$ if

$$\lim_n \int_\Omega f_n g \, d\mu = \int_\Omega f g \, d\mu, \quad \forall g \in L^q(\Omega, \mathcal{F}, \mu),$$

where $1/p + 1/q = 1$, and with brackets this is written as $\langle f_n, g \rangle \to \langle f, g \rangle$, even sometimes, the notation $f_n \rightharpoonup f$ is used. In this context, the convergence in norm (i.e., when $\|f_n - f\|_p \to 0$) is called strong convergence and usually denoted by $f_n \to f$. \qed
Recall that $q = \infty$ when $p = 1$ and that any function (actually equivalence class) $f(x)$ belonging to the Banach space $L^\infty(\Omega, \mathcal{F}, \mu)$ includes the condition of $\sigma$-finite non-zero range, i.e., the set $\{x \in \Omega : |f(x)| \neq 0\}$ is a countable union of sets with $\mu$-finite measure. Also note that technically, the weak convergence defined above for $L^\infty$ is actually called *weak* convergence, in the general context of Banach and dual spaces. On the other hand, as expected, by means of Hölder inequality,

$$|\langle f_n - f, g \rangle| \leq \|f_n - f\|_p \|g\|_q,$$

we show that weak convergence implies strong convergence.

**Proposition 7.36.** If $\{f_n\}$ is a sequence in $L^p(\Omega, \mathcal{F}, \mu)$ weakly convergent to $f$ then

$$\|f\|_p \leq \liminf_n \|f_n\|_p,$$

for any $1 \leq p \leq \infty$, i.e., the norm $\|\cdot\|_p$ is a weakly lower semi-continuous function.

**Proof.** Assume first $1 \leq p < \infty$. Since the function $g = |f|^{p/q}\text{sign}(f)$ belongs to $L^q$, with $1/p + 1/q = 1$, we have $\langle f_n, g \rangle \to \langle f, g \rangle = \|f\|_p^p$. However, Hölder inequality implies

$$|\langle f_n, g \rangle| \leq \|f_n\|_p \|g\|_p = \|f_n\|_p \|f\|_p^{p/q}$$

and (7.14) for $p < \infty$.

For $p = \infty$, we may assume that $\|f\|_p > 0$ and that $f$ vanishes outside of a set of $\sigma$-finite measure, namely, $\bigcup_k \Omega_k$ with $\Omega_k \subset \Omega_{k+1}$ and $0 < \mu(\Omega_k) < \infty$. Thus, for any $\varepsilon$ in the interval $(0, \|f\|_\infty)$, the set $\Omega_{k,\varepsilon} = \{x \in \Omega_k : |f(x)| \geq \|f\|_\infty - \varepsilon\}$ must have a positive measure for $k$ sufficiently large. Therefore, define $g = \text{sign}(f)1_{\Omega_{k,\varepsilon}}$ to have

$$\langle f_n, g \rangle \to \langle f, g \rangle \geq (\|f\|_\infty - \varepsilon)\mu(\Omega_{k,\varepsilon}).$$

Again, Hölder inequality yields

$$|\langle f_n, g \rangle| \leq \|f_n\|_p \mu(\Omega_{k,\varepsilon}), \quad \text{with} \quad 0 < \mu(\Omega_{k,\varepsilon}) < \infty,$$

and, we deduce

$$\liminf_n \|f_n\|_\infty \geq \|f\|_\infty - \varepsilon,$$

i.e., (7.14).

**Remark 7.37.** Related to Remark 6.23 we have the following result: if $f_n \rightharpoonup f$ weakly in $L^p(\Omega, \mathcal{F}, \mu)$ with $1 < p < \infty$ and $\|f_n\|_p \to \|f\|_p$ then $\|f_n - f\|_p \to 0$ as $n \to \infty$. This assertion fails for $p = 1$ or $p = \infty$, e.g., see DiBenedetto[32, Section V.11, pp. 236–238].

[PRELIMINARY]
It is clear that Banach-Steinhaus Theorem (or uniformly boundedness principle), as seen later on, proves that any weakly convergence sequence is bounded. The converse (i.e., that a bounded sequence contains a convergent subsequence) holds true for any dual space of Banach space (Alaoglu’s Theorem, e.g., see Conway [29, Section V.3 and V.4, pp. 123–137]). The proof is similar to the one given below, valid for any separable reflexive space (recall that this fails for $L^1$). On the other hand, a so-called density argument shows that if a bounded sequence $\{x_n\}$ in a normed space $X$ (say $L^p$) is such that $f(x_n) \to f(x)$ for every $f$ in a dense set $D$ of the dual space $X'$ then $x_n \to x$, weakly in $X$.

- **Remark 7.38.** Consider the space $\ell^p$ of all sequence $x = \{x_k\}$ with finite sum $\|x\|_p^p = \sum |x_k|^p < \infty$, i.e., the space $L^p$ with the discrete measure $\mu(A) = \sum 1_{k \in A}$, $A \subset \mathbb{N}$. On this space $\ell_p$ with $1 < p < \infty$, the weak convergence can be characterized as follows: a sequence $x^{(n)}$ converges weakly to $x$ if and only if (a) it is bounded, i.e., there exists a constant $C$ such that $\|x^{(n)}\|_p \leq C$ for every $n$ and (b) each coordinate converges, i.e., for every $k$, $x_k^{(n)} \to x_k$ as $n \to \infty$, e.g., see Bachman and Narici [7, Section 14.1, pp. 231–238].

A $\sigma$-algebra $\mathcal{F}$ is called $\mu$-separable if the algebra $\mathcal{F}_0 = \{F \in \mathcal{F} : \mu(F) < \infty\}$ is the completion of a countable generated algebra, i.e., there exists a countable subset $Q$ of $\mathcal{F}_0$ such that for any set $F$ in $\mathcal{F}_0$ there is a sequence $\{F_n\}$ in $Q$ satisfying $\mu(F \setminus F_n) \cup (F_n \setminus F) \to 0$. In this case, the space $L^q(\Omega, \mathcal{F}, \mu)$ is separable for any $1 \leq q < \infty$. Certainly, this includes the case where $\Omega$ is a Polish space (i.e., a separable and complete metrizable space) and $\mu$ is a $\sigma$-finite regular Borel measure.

**Proposition 7.39.** Let $\mathcal{F}$ be $\mu$-separable and $\{f_n : n \geq 1\}$ be a bounded sequence in $L^p(\Omega, \mathcal{F}, \mu)$ with $1 < p \leq \infty$. Then there exists a weakly convergent subsequence $\{f_{n_k} : k \geq 1\}$.

**Proof.** Essentially, this is the Cantor diagonal argument. The conjugate of the exponent $p$ is $q$, $1/p + 1/q = 1$, with $1 \leq q < \infty$. Thus, let $\{g_i : i \geq 1\}$ be a dense sequence in $L^q(\Omega, \mathcal{F}, \mu)$. Since the numerical sequence $\{(f_n, g_i) : n \geq 1\}$ is bounded for each $i$, by means of Cantor diagonal procedure, we construct a subsequence $\{f_{n_k} : k \geq 1\}$ such that the numerical sequence $\{(f_{n_k}, g_i) : n \geq 1\}$ is convergent for every $i \geq 1$.

Since $\{g_i : i \geq 1\}$ is dense, for every $g$ in $L^q(\Omega, \mathcal{F}, \mu)$ and for any $\varepsilon > 0$ there exists $g_i$ such that $\|g - g_i\| < \varepsilon$. Hence the inequality

$$|\langle f_{n_k} - f_{n_h}, g \rangle| \leq |\langle f_{n_k} - f_{n_h}, g_i \rangle| + \|g - g_i\| \sup_n \|f_n\|$$

shows that the numerical sequence $\{(f_{n_k}, g) : n \geq 1\}$ converges for every $g$ and defines a linear functional on $L^p(\Omega, \mathcal{F}, \mu)$. By Riesz representation Theorem 6.5, there exists $f$ in $L^p(\Omega, \mathcal{F}, \mu)$ such that $f_{n_k} \to f$ weakly. □

- **Remark 7.40.** Referring to Vitali-Hahn-Saks Theorem 7.33, we can prove that a sequence $\{f_n : n \geq 1\}$ in $L^1(\Omega, \mathcal{F}, \mu)$ is weakly compact if and only if it is...
bounded and the integrals
\[ I_n(A) = \int_A f_n d\mu, \quad n \geq 1, \]
are uniformly \(\sigma\)-additive, e.g., see Dunford and Schwartz\[40, \text{Theorem IV.8.9, pp. 292–296}\]. Certainly, if \(\{f_n : n \geq 1\}\) is \(\mu\)-uniformly integrable (see Definition 6.15) then \(\{I_n : n \geq 1\}\) is uniformly \(\sigma\)-additive, see Proposition 6.16. Hence, form Corollary 7.34 we deduce that sequentially weakly compact in \(L^1\) is equivalent to \(\mu\)-uniformly integrable, i.e., the Dunford-Pettis criterium. \(\square\)

## Totally Bounded Sets

Recall that a subset \(\{f_i : i \in I\}\) of a metric space \((X, d)\) is totally bounded if for every \(\varepsilon > 0\) there exists a finite subset of indexes \(J \subset I\) such that for every \(i\) in \(I\) there exists \(j\) in \(J\) satisfying \(d(f_i, f_j) < \varepsilon\), i.e., any element in \(\{f_i : i \in I\}\) is within a distance \(\varepsilon\) from the finite set \(\{f_j : j \in J\}\). Sometimes \(\{f_j : j \in J\}\) is called an \(\varepsilon\)-net relative to \(\{f_i : i \in I\}\). It is clear that a Cauchy sequence is a totally bounded set, and conversely, any totally bounded set contains a Cauchy sequence. Indeed, based on the existence of \(\varepsilon\)-nets, we can construct (by induction) a sequence \(\{f_n : n \geq 1\}\) (of the given totally bounded set) such that \(d(f_{n-1}, f_n) < 2^{-n}\), for any \(n \geq 2\), which is a Cauchy sequence. In a metric space, compactness is equivalent to sequentially compactness, and then, a totally bounded sets is equivalent to pre-compact (i.e., closure compact) set on a complete metric space, in particular, this also applied to the \(L^p\)-spaces \(L^p(\Omega, F, \mu)\) with \(0 < p < 1\) and the distance \(d(f, g) = \|f - g\|_p\) and to the Banach spaces \(L^p(\Omega, F, \mu)\) with \(1 \leq p \leq \infty\).

The following characterization of pre-compact (or totally bounded) sets in \(L^p(\Omega)\) is sometime referred to as Fréchet-Kolmogorov Theorem, e.g., Yosida \[122, \text{Section X.1, pp. 274–277}\] and DiBenedetto\[32, \text{Section V.22, pp. 260–262}\]. This applies to \(L^p(\Omega) = L^p(\Omega, F, \mu)\), where \(\Omega\) an open subset of \(\mathbb{R}^d\) and \(\mu\) is the Lebesgue measure. We use the notation
\[
\tau_h f = f(\cdot + h), \quad \forall h \in \mathbb{R}^d, \quad h\text{-translations},
\]
\[
\|f\|_{p,A} = \left( \int_A |f(x)|^p \, dx \right)^{1/p}, \quad \forall A \subset \mathbb{R}^d, \text{ measurable},
\]
\[
\Omega_\delta = \{ x \in \Omega : d(x, \partial \Omega) > \delta, \ |x| < 1/\delta \}, \quad \forall \delta > 0,
\]
where \(d(x, \partial \Omega) = \inf \{|x - y| : y \in \partial \Omega|\}\) is the distance from the point \(x\) to the boundary \(\partial \Omega\) of \(\Omega\). Also \(\overline{\Omega}_\delta\) denotes the closure of the open set \(\Omega_\delta\).

**Theorem 7.41.** A uniformly bounded \(\{f_i : i \in I\}\) subset of functions in \(L^p(\Omega)\), with \(1 \leq p < \infty\) and \(\Omega\) an open subset of \(\mathbb{R}^d\), is pre-compact or totally bounded if and only if for every \(\varepsilon > 0\) there exists \(\delta > 0\) such that
\[
\|\tau_h f_i - f_i\|_{p, \Omega_\delta} \leq \varepsilon \quad \text{and} \quad \|f_i\|_{p, \Omega \setminus \Omega_\delta} \leq \varepsilon, \quad \forall i \in I, \quad (7.15)
\]
and for every translation \(\tau_h\) with \(|h| \leq \delta\).
Thus, estimate (7.16) yields the inequality

\( |(f \ast \kappa_\eta)(x) - f(x)| \leq \int_{|y| \leq \eta} |(\tau_y f)(x) - f(x)| \, dy, \)

which yields the estimate

\[
\|f \ast \kappa_\eta - f\|_{P,\Omega_\delta} \leq \sup_{|h| \leq \eta} \|\tau_h f - f\|_{P,\Omega_\delta}, \quad \forall \eta \leq \delta, \tag{7.16}
\]

and similarly,

\[
|(f \ast \kappa_\eta)(x)| \leq \|\kappa_\eta\|_q \|f\|_{P,\Omega_\delta}, \quad \forall x \in \overline{\Omega_{2\delta}}, \forall \eta < \delta, \tag{7.17}
\]

for any \(1 \leq p \leq \infty\), with \(1/p + 1/q = 1\). Note that \(\|\kappa_\eta\|_q\) is unbounded as \(\eta \to 0\), for any \(q > 1\) or equivalently, \(1 \leq p < \infty\).

Also, recall the continuity of the translations in \(L^1(\mathbb{R}^d)\) proved in Proposition 7.3, which is easily extended to \(L^p(\mathbb{R}^d)\) with the tools in Section 7.5. Therefore, an extension by zero of functions in \(L^p(\Omega)\) shows that

\[
\lim_{h \to 0} \|\tau_h f - f\|_{P,\Omega_\delta} = 0, \quad \forall f \in L^p(\Omega) \tag{7.18}
\]

and \(1 \leq p < \infty\).

Suppose \(\{f_i : i \in I\}\) is a uniformly bounded set in \(L^p\) satisfying (7.15). Then for any \(\varepsilon' > 0\) we will construct an \(\varepsilon'\)-net, proving that \(\{f_i : i \in I\}\) is totally bounded. Indeed, for a fixed small \(\eta > 0\), consider the family of functions \(\{f_i \ast \kappa_\eta : i \in I\}\), as defined on \(\overline{\Omega_{2\eta}}\), and apply estimate (7.17) to \(f_i\) and \(\tau_h f_i - f_i\) to obtain, for every \(x\) in \(\overline{\Omega_{2\eta}}\) the inequalities

\[
|(f_i \ast \kappa_\eta)(x)| \leq \|\kappa_\eta\|_q \|f_i\|_{P,\Omega_\eta},\]

\[
|\tau_h f_i \ast \kappa_\eta(x) - (f_i \ast \kappa_\eta)(x)| \leq \|\kappa_\eta\|_q \|\tau_h f_i - f_i\|_{P,\Omega_\eta} \quad \forall |h| < \eta.
\]

Together with condition (7.15), this shows that \(\{f_i \ast \kappa_\eta : i \in I\}\) is uniformly bounded and equi-continuous set of continuous functions on \(\overline{\Omega_{2\eta}}\). Hence, Arzelà-Ascoli Theorem (see later on) implies that \(\{f_i \ast \kappa_\eta : i \in I\}\) is pre-compact, and therefore, there exists an \(\varepsilon''\)-net of continuous functions defined on \(\overline{\Omega_{2\eta}}\), which is denoted by \(\{g_j : j \in J\}\), with \(g_j = f_j \ast \kappa_\eta\) and \(J\) a finite subset of indexes. Thus, estimate 7.16 yields the inequality

\[
\|f_i - f_j\|_p \leq \|f_i - f_j\|_{P,\Omega_\eta} + \|f_i - g_j\|_{P,\Omega_\eta} + \|f_j \ast \kappa_\eta - f_j\|_{P,\Omega_\eta} \leq \\
\leq 2 \sup_i \|f_i\|_{P,\Omega_\eta} + \|f_i - g_j\|_{P,\Omega_{2\eta}} + \sup_{|h| \leq \eta} \|\tau_h f_j - f_j\|_{P,\Omega_\eta}.
\]
with $0 < \eta \leq \delta/2$. Hence, by means of condition (7.15), we deduce that \( \{f_j : j \in J\} \) is an \( \varepsilon' \)-net for \( \{f_i : i \in I\} \) with

To prove the converse, suppose that \( \{f_i : i \in I\} \) is a totally bounded set in \( L^p(\Omega) \), i.e., for every \( \varepsilon' > 0 \) there exists an finite set of indexes \( J \) such that \( \{f_j : j \in J\} \) is an \( \varepsilon' \)-net, namely,

\[
\min_{j \in J} \|f_i - f_j\|_{p,\Omega} \leq \varepsilon', \quad \forall i \in I.
\]

This shows that \( \{f_i : i \in I\} \) is uniformly bounded and that for every \( i \in I \) there exists \( j \) in \( J \) such that

\[
\|f_i\|_{p,\Omega \setminus \Omega_\delta} \leq \|f_j\|_{p,\Omega \setminus \Omega_\delta} + \varepsilon', \quad \forall \delta > 0
\]

and

\[
\|\tau_h f_i - f_i\|_{p,\Omega_\delta} \leq \|\tau_h f_i - \tau_h f_j\|_{p,\Omega_\delta} + \|\tau_h f_j - f_j\|_{p,\Omega_\delta} + \|f_j - f_i\|_{p,\Omega_\delta} \leq 2\varepsilon' + \|\tau_h f_j - f_j\|_{p,\Omega_\delta}, \quad \forall \delta > 0.
\]

Because \( J \) is finite and the translations are continuous, see property (7.18), there is \( \delta > 0 \) such that

\[
\max_{j \in J} \|f_j\|_{p,\Omega \setminus \Omega_\delta} \leq \varepsilon' \quad \text{and} \quad \max_{j \in J} \|\tau_h f_j - f_j\|_{p,\Omega_\delta} \leq \varepsilon', \quad \forall |h| \leq \delta.
\]

Hence \( \|f_i\|_{p,\Omega \setminus \Omega_\delta} \leq 2\varepsilon' \) and \( \|\tau_h f_i - f_i\|_{p,\Omega_\delta} \leq 3\varepsilon' \), for every \( i \in I \). This shows condition (7.15) with \( \varepsilon = 3\varepsilon' \).

Since any function in \( L^p(\Omega) \) can be extended by zero to a function in \( L^p(\mathbb{R}^d) \), we can rephrase the previous results in \( L^p(\mathbb{R}^d) \) as follows:

**Proposition 7.42.** A family \( \{f_i : i \in I\} \) of functions in \( L^p(\mathbb{R}^d) \), with \( 1 \leq p < \infty \), is totally bounded or pre-compact if and only if (1) there exists a constant \( C > 0 \) such that \( \|f_i\|_p \leq C \), for every \( i \in I \); (2) for every \( \varepsilon > 0 \) there exists \( n \) such that \( \|1_{|x| \geq r} f_i(x)\|_p \leq \varepsilon \) for every \( r > n \) and for every \( i \in I \); and (3) for every \( \varepsilon > 0 \) there exists \( \delta > 0 \) such that \( \|\tau_h f_i - f_i\|_p < \delta \) for every \( |h| < \delta \) and for every \( i \in I \)

**Proof.** Since a totally bounded set is necessarily bounded, this result is a consequence of the previous Theorem 7.41 for \( \Omega = \mathbb{R}^d \). However, it is worthwhile to remark some arguments used.

For instance, to verify (2) we get first an \( \varepsilon/2 \)-net \( \{f_j : j \in J\} \) with \( J \) a finite subset of indexes of \( I \). Because \( |f_j|^p \) is integrable, the dominate convergence Lebesgue theorem shows that \( 1_{|x| \leq r} f_j(x) \to f_j(x) \) in \( L^p \) for each \( j \) and so, for every \( \varepsilon > 0 \) there exists \( r > 0 \) such that \( \|1_{|x| \geq r} f_j(x)\|_p < \varepsilon/2 \) for every \( j \) in \( J \). However, for each \( i \in I \) there exists \( j \) such that \( \|f_i - f_j\|_p < \varepsilon/2 \), and we conclude.

To verify (3), we can compute \( \|\tau_h f - f\|_p \) to show that \( \|\tau_h f - f\|_p \to 0 \) as \( h \to 0 \) for every \( f = 1_A \) where \( A \) is a \( d \)-interval in \( \mathbb{R}^d \). Next, by linearity, this remains true for any finite valued function \( f \) and finally, but density, this holds true for any function \( f \) in \( L^p \).

Remark 7.43. It is clear that (7.15) of Theorem 7.41 or Proposition 7.42 can be restated as follows: for every \( \varepsilon > 0 \) there exists \( \delta > 0 \) such that
\[
\| \tau_{r,k}f_i - f_i \|_{p,\Omega_\delta} \leq \varepsilon \quad \text{and} \quad \| f_i \|_{p,\Omega \setminus \Omega_\delta} \leq \varepsilon, \quad \forall i \in I,
\]
(7.19)
and for every one-dimensional translation \( \tau_{r,k} \) with \( |r| \leq \delta, \ k = 1, \ldots, d \), i.e., with
\[
(\tau_{r,k}f)(x_1, \ldots, x_k, \ldots, x_d) = f(x_1, \ldots, x_k + r, \ldots, x_d).
\]
Indeed, it suffices to note that for any translation \( \tau_h \) with \( h = (h_1, \ldots, h_d) \) we have
\[
\tau_h f - f = (\tau_{h_1,1} f_1 - f_1) + (\tau_{h_2,2} f_2 - f_2) + \cdots + (\tau_{h_d,d} f_d - f_d),
\]
with \( f_1 = f \) and \( f_k = \tau_{h_k,k} f_{k-1}, \) for \( k = 2, \ldots, d \).

Exercise 7.26. If \( A \) is a totally bounded set of a normed space \((X, \| \cdot \|)\) then prove that the convex hull (or convex envelope) \( \text{co}(A) \) of \( A \) (i.e., the smallest convex set containing \( A \)) is also totally bounded. In particular, the closed convex hull of a compact set of a Banach space is also compact. Hint: Use the following argument (1) if \( F \subset X \) is a finite set then the convex hull \( \text{co}(F) \) of \( F \) is a totally bounded set. Next, let \( A \) be a totally bounded subset of \( X \) and let \( B_1 \) be an open balls containing the origin. By using the previous result, (2) find a finite set \( F \) such that \( A \subset F + B_1 \) and deduce that \( \text{co}(A) \) lies inside \( K + B_1 \) for some totally bounded set \( K \). Now, take any two open balls \( B_1 \) and \( B \) containing the origin and satisfying \( B_1 + B_1 \subset B \). Finally, because \( K \) is totally bounded, (3) find another finite \( E \) such that \( \text{co}(A) \subset (E + B_1) + B_1 \subset E + B \), and deduce that \( \text{co}(A) \) is indeed totally bounded.

Exercise 7.27. Banach-Saks Theorem states that if \( \{f_n\} \) is a weakly convergence sequence to \( f \) in \( L^p(\Omega, \mathcal{F}, \mu), \ 1 \leq p < \infty \) then there exists a subsequence \( \{f_{n_k}\} \) such that the arithmetic means \( g_k = (f_n_1 + \cdots + f_{n_k})/k \) strongly converges to \( f \), i.e., \( \|g_k - f\|_p \to 0 \). Prove this result for a Hilbert space \( H \) with scalar product \( \langle \cdot, \cdot \rangle \) and norm \( \| \cdot \| \), in particular for \( p = 2 \). Hint: First reduce the problem to the case where \( f = 0 \), and \( \|f_n\| \leq 1 \) for every \( n \geq 1 \). Next, construct a subsequence satisfying \( |(f_{n_i}, f_{n_{k+1}})| \leq 1/k \), for every \( i = 1, \ldots, k \), and deduce that \( \|g_k\|^2 \leq 3/k \), see Riesz and Nagy [94, Section 38, pp. 80–81.].

All exercises are re-listed here, but now, most of them have a (possible) solution. Certainly, this is not for the first reading. This part is meant to be read after having struggled (a little) with the exercises. Sometimes, there are many ways of solving problems, and depending of what was developed “in the theory”, solving the exercises could have alternative ways. In any way, some exercises are trivial while other are not simple. It is clear that what we may call “Exercises” in one textbook could be called “Propositions” in others. Most exercises correspond to Chapters 1, 2, 4 and 7, which are the center of the material.

Most of the various references mentioned in previous chapters contain many exercises, moreover, there are several books dedicated to ‘just exercises’, e.g., Gasiński and Papageorgiou [48], Yeh [121], among others.

(1.1) Classes of Sets

Exercise 1.1. Prove that the algebra $A$ (ring) generated by a $S$ semi-algebra (semi-ring) is the class of finite disjoint unions of sets in $S$, i.e., $A \in A$ if and only if $A = \sum_{i=1}^{n} A_i \in S$. Hint: prove first that the class of finite disjoint union of sets in $S$ is stable under the formation of finite unions.

Proof. Since $S$ is a semi-ring, the classes $F = \{\sum_{i=1}^{n} A_i : A_i \in S\}$ and $\{\bigcup_{i=1}^{n} A_i : A_i \in S\}$ are exactly the same. Indeed, define $B_i = A_1, B_i = A_i \setminus (A_1 \cup \cdots \cup A_{i-1})$ to get $\bigcup_{i=1}^{n} A_i = \sum_{i=1}^{n} B_i$, and each $B_i$ is a disjoint unions of elements in $S$. Thus, $F$ is stable under the formation of finite intersections.

Now, let us show that the class $F$ is stable under the formation of finite intersections. Indeed, by means of the set indicator (or characteristic) function, if $F = F_1 \cap F_2$ and $F_j = \sum_{i=1}^{n(j)} A_{i,j}$ then

$$1_F = 1_{F_1} 1_{F_2} = \left( \sum_{i=1}^{n(1)} 1_{A_{i,1}} \right) \left( \sum_{i=1}^{n(2)} 1_{A_{i,2}} \right) = \sum_{i=1}^{n(1)} \sum_{k=1}^{n(2)} 1_{A_{i,1}} 1_{A_{k,2}},$$

which shows that $F$ is a finite disjoint union of the sets $A_{i,1} \cap A_{k,2}$ in $S$. 

253
Next, because differences of sets in $S$ can be written as a finite disjoint union of the sets in $S$, we conclude that $\mathcal{F}$ is also stable under the formation of differences. \hfill \Box

**Exercise 1.2.** First show that any $\sigma$-algebra is a $\sigma$-ring and that a $\sigma$-ring is stable under the formation of countable intersections. Next, prove that an algebra $A$ (a ring $\mathcal{R}$) is a $\sigma$-algebra (a $\sigma$-ring) if and only if $A$ ($\mathcal{R}$) is stable under the formation countable increasing unions.

*Proof.* If $A$ is a $(\sigma)$-algebra of subsets of $\Omega$ then the equality $A \setminus B = A \cap \Omega \setminus B$ shows that $A$ is also a $(\sigma)$-ring. Also from the equalities $A = \bigcup_i A_i$ and $\bigcap_i A_i = A \setminus \bigcup_i (A \setminus A_i)$ we deduce that any $\sigma$-ring is stable under the formation of countable intersections. Since any countable union $A = \bigcup_{i \geq 0} A_i$ can be written as the countable increasing union $A = \bigcup_{i \geq 0} B_i$, with $B_0 = A_0$, $B_i = B_{i-1} \cup A_i$, this implies that any $A$ (or ring $\mathcal{R}$), which is stable under the formation countable increasing unions (or decreasing intersections), is indeed a $\sigma$-algebra (or $\sigma$-ring). \hfill \Box

**Exercise 1.3.** Since only countable operations are involved, we can convince oneself that if $\mathcal{K}$ has the cardinality of the continuum (or greater) then $\sigma(\mathcal{K})$ preserves its cardinality. Make a formal argument to show the validity of the previous statement, e.g., see Rama [92, Section 4.5, pp. 110-112] where transfinite induction is used.

*Proof.* First, construct the class $\mathcal{K}^* = \{K \in 2^\Omega : \text{ either } K \text{ or } \Omega \setminus K \text{ is in } \mathcal{K}\}$, from a given class $\emptyset \subseteq \mathcal{K} \subseteq 2^\Omega$. Next, by induction define the increasing sequence of classes $\{\mathcal{A}_n\}$ as follows: $\mathcal{A}_0 = \mathcal{K}^*$ and $\mathcal{A}_n$ is the class of all finite unions of sets in $\mathcal{A}_{n-1}^*$. By construction, if $A$ and $B$ belong to $\mathcal{A}_n$ then $\Omega \setminus A$ and $A \cup B$ belong to $\mathcal{A}_{n+1}$, which prove that $\mathcal{A} = \cup \mathcal{A}_n$ is an algebra. Moreover, an algebra $\mathcal{B}$ containing $\mathcal{K}$ should contain any class $\mathcal{A}_n$, i.e., $\mathcal{A}$ is indeed the algebra generated by $\mathcal{K}$. This proves that if $\mathcal{K}$ is countable then so is the algebra generated by $\mathcal{K}$.

The same argument can be applied using transfinite induction, i.e., for any ordinal $\alpha$ we define class $\mathcal{A}_\alpha$ of all finite unions of sets in the class $(\cup_{\beta < \alpha} \mathcal{A}_\beta)^*$. Next, if $\omega_0$ denotes the first uncountable ordinal then we show as above that union $\cup_{\beta < \omega_0} \mathcal{A}_\alpha$ is the $\sigma$-algebra generated by $\mathcal{K}$. Thus again, if $\mathcal{K}$ is countable then so is the $\sigma$-algebra generated by $\mathcal{K}$. \hfill \Box

**Exercise 1.4.** A subset $\mathcal{D}$ of $2^\Omega$ containing the empty set is called a Dynkin class if $\mathcal{D}$ is stable under the formation of complements and countable disjoint unions. Prove that a Dynkin class $\mathcal{D}$ is also a $\lambda$-class. How about the converse? (e.g., see Bauer [9, Section I.2]).

*Proof.* First, if $\mathcal{D}$ is a Dynkin class then $\Omega \in \mathcal{D}$; and for any $A \subset B$, both in $\mathcal{D}$, we have $B \setminus A = B \cap A^c = (B^c \cup A)^c$, which proves that a Dynkin class is also stable under the formation of proper differences. Hence, for an increasing sequence $\{A_i\}$ in $\mathcal{D}$ we have $\bigcup_i A_i = \bigcup_i B_i$, with $B_1 = A_1$ and $B_n = A_n \setminus A_{n-1}$. Thus, $\mathcal{D}$ is a Dynkin class if and only if $\mathcal{D}$ is a $\lambda$-class. \hfill \Box
Exercise 1.5. Given a family \( \{ F_{i,j} : i \in I_j, j \in J \} \) of subsets of \( \Omega \), verify the distributive formula

\[
\bigcup_{j \in J} \bigcap_{i \in I_j} F_{i,j} = \bigcap_{k \in K} \bigcup_{j \in J} F^k_j \quad \text{and} \quad \bigcap_{j \in J} \bigcup_{i \in I_j} F_{i,j} = \bigcup_{k \in K} \bigcap_{j \in J} F^k_j,
\]

where \( K = \prod_{j \in J} I_j \), i.e., \( \{ i_j : j \in J \} \), and \( F^k_j = F_{i,j} \). It is clear that if \( J \) is finite and each \( I_j \) is countable then \( K \) is a countable set, however, if for instance, \( I_j = \{ 0, 1 \} \) for every \( j \) in an infinite set of indexes \( J \) then \( K = \{ 0, 1 \}^J \) is not a countable set of indexes.

Proof. First, only one of the two identities should be proved, the other follows by means of formation of complements and the relation \( (A \cap B)^c = A^c \cup B^c \).

With the use of the indicator (or characteristic) function of a set, \( 1_A(x) > 0 \) if \( x \in A \) and \( 1_A(x) = 0 \) if \( x \in \Omega \setminus A \), and the algebraic identity

\[
\prod_{j \in J} \sum_{i \in I_j} F_{i,j} = \sum_{k \in K} \prod_{j \in J} F^k_j
\]

we deduce the identity \( \bigcap_{j \in J} \bigcup_{i \in I_j} F_{i,j} = \bigcup_{k \in K} \bigcap_{j \in J} F^k_j \).

Exercise 1.6. If \( \mathcal{E} \subset 2^\Omega \) with \( \emptyset \in \mathcal{E} \) then define \( \mathcal{E}_\sigma \) as the class of all countable unions of sets in \( \mathcal{E} \) and define \( \mathcal{E}_{\sigma^\delta} \) as the class of all countable intersections of sets in \( \mathcal{E}_\sigma \). Verify that (1) \( \mathcal{E}_\sigma \) (or \( \mathcal{E}_{\sigma^\delta} \)) is stable under the formation of countable unions (or countable intersections). Prove that (2) if \( \mathcal{E} \) is stable under the formation of finite intersections then so is \( \mathcal{E}_\sigma \). Deduce (3) that if \( \mathcal{E} \) is stable under the formation of finite unions and finite intersections then (a) for any \( E \) in \( \mathcal{E}_\sigma \) there exists a monotone increasing sequence \( \{ E_n \} \subset \mathcal{E}_\sigma \), \( E_n \subset E_{n+1} \), such that \( E = \bigcup_n E_n = \lim_n E_n \) and (b) for any \( F \) in \( \mathcal{E}_{\sigma^\delta} \) there exists a monotone decreasing sequence \( \{ F_n \} \subset \mathcal{E}_\sigma \), \( E_n \supset F_{n+1} \), such that \( F = \bigcap_n F_n = \lim_n F_n \).

Proof. (1) It is clear the a countable union (or product) of countable set is again a countable set, thus, if \( A_j = \bigcup_{i \in I} E_{i,j} \) with \( E_{i,j} \in \mathcal{E} \) then \( A = \bigcup_{j \in J} A_j = \bigcup_{i \in I, j \in J} E_{i,j} \), i.e., \( \mathcal{E}_\sigma \) is stable under the formation of countable unions.

In a similar manner, if \( B_j = \bigcup_{i \in I} F_{i,j} \) with \( F_{i,j} \in \mathcal{E}_\sigma \) then \( B = \bigcap_{j \in J} B_j = \bigcap_{i \in I, j \in J} E_{i,j} \), i.e., \( \mathcal{E}_\sigma \) is stable under the formation of countable intersections.

(2) If \( A = \bigcup_{j \in J} A_j \) and \( B = \bigcup_{j \in J} B_j \) with \( A_j \) and \( B_j \) in \( \mathcal{E} \) then \( A \cap B = \bigcup_{i,j} (A_i \cap B_j) \), which proves that \( \mathcal{E}_\sigma \) is stable under the formation of finite intersection if \( \mathcal{E} \) is so.

(3) If \( E \) belongs to \( \mathcal{E}_\sigma \) then \( E = \bigcup_{i \geq 1} A_i \) with \( A_i \) in \( \mathcal{E} \). Since the set \( E_n = \bigcup_{1 \leq i \leq n} A_i \) also belongs to \( \mathcal{E} \), we have \( E_n \subset E_{n+1} \) and \( E = \bigcup_n E_n = \lim_n E_n \).

Similarly, if \( F \) belongs to \( \mathcal{E}_{\sigma^\delta} \) then \( F = \bigcap_{i \geq 1} B_i \) with \( B_i \) in \( \mathcal{E}_\sigma \). In view of part (2), the set \( F_n = \bigcap_{1 \leq i \leq n} B_i \) also belongs to \( \mathcal{E}_\sigma \), we obtain \( F_n \supset F_{n+1} \) and \( F = \bigcap_n F_n = \lim_n F_n \).

Exercise 1.7. Let \( \mathcal{E} \subset 2^X \) be a semi-ring such that \( X \) is a countable union of elements in \( \mathcal{E} \) and consider the class of subsets in \( 2^X \) defined by \( \mathcal{F} = \{ \bigcup_{k=1}^\infty E_k : E_k \in \mathcal{E} \} \), recall that \( \sum \) means disjoint union of sets. (1) Modify the arguments
of Exercise 1.1 to prove that $\mathcal{F}$ is stable under the formation of countable unions. (2) Show that $\mathcal{F}$ is stable under the formation of finite intersection. Why the distributive formula of Exercise 1.5 cannot be used to show that $\mathcal{F}$ is stable under countable intersections? (3) Show that if $A$ belongs to $\mathcal{F}$ and $B$ belongs to the ring generated by $\mathcal{E}$ then the difference $A \setminus B$ is $\mathcal{F}$.

Proof. (1) It is clear that the class $\bar{\mathcal{F}} = \{ \bigcup_{k=1}^{\infty} E_k : E_k \in \mathcal{E} \}$ is stable under the formation of countable unions. Indeed, if $A = \bigcup_{i \geq 1} A_i$ and $B = \bigcup_{j \geq 1} B_j$ then $A \cup B = \bigcup_{n \geq 1} C_n$, where $C_{2k-1} = A_k$ and $C_{2k} = B_k$. Thus, the same arguments in Exercise 1.1 show that $\bar{\mathcal{F}}$ is actually $\mathcal{F}$. Indeed, if $A = \bigcup_{k=1}^{\infty} A_k$ with $A_k$ in $\mathcal{E}$ then $A = \sum_{k=1}^{\infty} B_k$, where $B_1 = A_1$, $B_2 = A_2 \setminus A_1$, $B_3 = (A_3 \setminus A_2) \setminus A_1$, and in general, $C_k = A_k$, $C_{k-1} = C_k \setminus A_{k-1}$, $\ldots$, $C_1 = C_2 \setminus A_1$, and $B_k = C_1$. Since $\mathcal{E}$ is a semi-ring, each $B_k$ is a finite disjoint union of sets in $\mathcal{E}$, and we deduce that $\mathcal{F} = \bar{\mathcal{F}}$.

(2) If $A_n = \bigcup_{i \geq 1} E_{i,n}$ (with $E_{i,n}$ in $\mathcal{E}$) and $B = \bigcap_{1 \leq n \leq N} A_n$ then for $K = \{1, 2, \ldots \}^N$ (i.e., $k = \{k_n, 1 \leq n \leq N\}$ with $k_n$ in $\{1, 2, \ldots \}$) and $C_k = \bigcap_{1 \leq n \leq N} E_{k_n,n}$, we obtain $B = \bigcup_{k \in K} C_k$. Because the set of indexes $K$ is countable and $\mathcal{E}$ is stable under the formation of finite intersections, the set $C_k$ belongs to $\mathcal{E}$, and we deduce that $B$ belongs to $\mathcal{F}$, i.e., $\mathcal{F}$ is stable under the formation of finite intersection.

The distributive formula of Exercise 1.5 can only be used to show that $\mathcal{F}$ is stable under finite intersections, a countable intersection may produce an uncountable union.

(3) Similarly, $\mathcal{F}$ may not be stable under the formation of differences, and so, $\mathcal{F}$ is not necessarily a $\sigma$-ring. However, if $A$ belongs to $\mathcal{F}$ and $B$ belongs to the ring generated by the semi-ring $\mathcal{E}$ then the difference $A \setminus B$ belongs to $\mathcal{F}$. Indeed, if $A = \bigcup_{i \geq 1} A_i$ and $B = \bigcup_{j \geq 1} B_j$ then $A \setminus B = \bigcup_n C_{n,k}$, where

$$C_{n,b} = A_n \setminus (B_1 \cup \cdots \cup B_k) = (\cdots (A_n \setminus B_1) \setminus B_2 \cdots) \setminus B_k,$$

i.e., $C_{1,1} = A_1 \setminus B_1$, $C_{1,2} = C_{1,1} \setminus B_2$, $\ldots$, $C_{1,k} = C_{1,k-1} \setminus B_k$. \hfill $\Box$

Exercise 1.8. A (nonempty) class $\mathcal{L}$ of subsets of $2^X$ is called a lattice if it is stable under finite intersections and finite unions (which may or may not contain the empty set $\emptyset$). Verify that (a) any intersection of lattices is a lattice; (b) if $\mathcal{L}$ is a lattice then the complement class $\{ A : A^c \in \mathcal{L} \}$ is also a lattice; (c) if $X = \mathbb{R}$ and $\mathcal{L}_c$ is the class of all bounded closed intervals $[a,b]$, $-\infty < a \leq b < \infty$, and the empty set $\emptyset$, then the smallest lattice class containing $\mathcal{L}_c$ is the class of all closed sets in $\mathbb{R}$; (d) if $X = \mathbb{R}$ and $\mathcal{L}_o$ is the class of all bounded open intervals $(a,b)$, $-\infty < a < b < \infty$, and the empty set $\emptyset$, then the smallest lattice class containing $\mathcal{L}_o$ is the class of all open sets in $\mathbb{R}$. (e) How about the equivalent of (c) and (d) for $X = \mathbb{R}^d$? Finally, note that a $\sigma$-lattice is a class stable under countable intersections and countable unions, show that a $\sigma$-lattice is a monotone class, and give an example of a monotone class (with an infinite number of elements) which is not a lattice.
Moreover, if an oval

Proof. It is clear the validity of (a) and (b). To prove (c) or (d), we need to use the fact that any closed (open) set is a finite union of its closed (open) connected components, and that any connected closed (open) set in $\mathbb{R}$ is an interval. For (e), we may have closed (open) $d$-intervals in $\mathbb{R}^d$, but the connected components are not easily identified.

By definition, we deduce that a $\sigma$-lattice is also a monotone class. If $\{A_i\}$ is an increasing sequence with $A = \bigcup_i A_i$ nonempty, and $B$ is another set such that the inclusions $B \cup A_1 \subset B \cup A$ and $A \subset B \cup A$ are both strict, then the class $\mathcal{M} = \{\emptyset, A, B, (B \cup A_1), A_i : i \geq 1\}$ is a monotone class, which is not a lattice, $A \cup B$ does not belongs to $\mathcal{M}$. The $\sigma$-lattice $\mathcal{L}$ generated by $\mathcal{M}$ is the class $\mathcal{L} = \{\emptyset, A, B, (B \cup A), (B \cup A_1), (B \cap A), (B \cap A_i), A_i : i \geq 1\}$, which is not a ring, if $B \cap A \neq \emptyset$ then $B \setminus A$ does not belongs to $\mathcal{L}$.

Exercise 1.9. A (nonempty) class $\mathcal{E}$ of subsets of $2^\Omega$ is called an oval if for any elements $U, A, V$ in $\mathcal{E}$ we have $U|A|V := (U \cap A^c) \cup (V \cap A)$ in $\mathcal{E}$ (which may or may not contain the empty set $\emptyset$). Prove that $\mathcal{E}$ is a ring if and only if $\mathcal{E}$ is oval and $\emptyset$ belongs $\mathcal{E}$.

Proof. Since $U|A|V = (U \setminus A) \cup (V \cap A)$, it is clear that a ring is also an oval. Moreover, if an oval $\mathcal{E}$ contains $\emptyset$ then using the identities $U|A|\emptyset = (U \setminus A)$, $\emptyset|A|V = (V \cap A)$ and $U|V|V = (U \cup V)$ we deduce that $\mathcal{E}$ is also a ring.

Actually, Exercises 1.8 and 1.9 are part of more advanced topics, the interested reader may consult König [72, Section 1.1, pp. 1-10] to understand the context.

(1.2) Borel Sets and Topology

Exercise 1.10. Let $I$ be a family of indexes and $\mathcal{E}_i$ be a class (of subsets of $\Omega_i$) generating a $\sigma$-algebra $\mathcal{F}_i$. Verify that the product $\sigma$-algebra $\prod_{i \in I} \mathcal{F}_i$ can be generated by cylinder sets of the form $\prod_{i \in I} E_i$, where $E_i \in \mathcal{E}_i$, $i \in I$ and $E_i = \Omega_i$, $i \notin J$ with $J \subset I$, finite. Now, for a finite product $I = \{1, \ldots, n\}$ consider the $\sigma$-algebra $\mathcal{A}$ generated by the hyper-rectangles of the form $E_1 \times \cdots \times E_n$ with $E_i \in \mathcal{E}_i$. Discuss why in proving that $\mathcal{A}$ is actually the product $\sigma$-algebra $\prod_{i=1}^n \mathcal{F}_i$, we may need the following assumption: for every $i = 1, \ldots, n$ there exists a monotone increasing sequence $\{E_{i,k} : k \geq 1\}$ such that $\Omega_i = \bigcup_k E_{i,k}$.

Proof. Let $\Omega = \prod_{i \in I} \Omega_i$, for a given index $i$, define the projection $p_i : \Omega \rightarrow \Omega_i$ as $p_i((\omega_i : i \in I)) = \omega_i$. A cylinder is a set of the form $\bigcap_{i \in I} p_i^{-1}(F_i) = \prod_{i \in I} F_i$, where $F_i \in \mathcal{F}_i$, $i \in I$ and $F_i = \Omega_i$, $i \notin J$ with $J \subset I$, finite. This means that the product $\sigma$-algebra $\mathcal{F} = \prod_{i \in I} \mathcal{F}_i$ is the smaller $\sigma$-algebra for which each projection $p_i$ is measurable, and because $p_i$ is measurable if and only if $p_i^{-1}(E_i) \in \mathcal{F}$ for any $E_i \in \mathcal{E}$, we deduce that product $\sigma$-algebra $\mathcal{F}$ is generated by cylinder sets of the form $E_i \in \mathcal{E}_i$, $i \in I$ and $E_i = \Omega_i$, $i \notin J$ with $J \subset I$, finite.

For a finite product $\mathcal{F}_1 \times \cdots \times \mathcal{F}_n$, we may consider the hyper-rectangles of the form $E_1 \times \cdots \times E_n$ with $E_i \in \mathcal{E}_i$. To see that a set of the form $p_j^{-1}(E_j)$ belongs to
the \( \sigma \)-algebra generated by all hyper-rectangle, we write \( \Omega_i = \bigcup_{k \geq 1} E_{i,k} \) for \( i \neq j \) to get \( p_j^{-1}(E_j) = \bigcup_k E_{1,k} \times E_{j-1,k} \times E_{j+1,k} \times E_{n,k} \). Note that if \( \mathcal{F}_1 = \{ \emptyset, \Omega_1 \} \), \( \mathcal{E}_i = \{ \emptyset \} \), and \( \mathcal{F}_2 = \mathcal{E}_2 \) with at least four sets, then the \( \sigma \)-algebra generated by the hyper-rectangle is the trivial one \( \{ \emptyset, \Omega_1 \times \Omega_2 \} \), while the product \( \sigma \)-algebra \( \mathcal{F} \) (generated by the cylinders) is \( \{ \emptyset, \Omega_1 \times F_2 : F_2 \in \mathcal{F}_2 \} \). \( \square \)

**Exercise 1.11.** Show that the Borel \( \sigma \)-algebra \( \mathcal{B}(\mathbb{R}^d) \) of the \( d \)-dimensional Euclidean space \( \mathbb{R}^d \), which is generated by all open sets, can also be generated by (1) all closed sets, (2) all bounded open rectangles of the form \( \{ x \in \mathbb{R}^d : a_i < x_i < b_i, \forall i \} \), with \( a_i, b_i \in \mathbb{R} \), (3) all unbounded open rectangles of the form \( \{ x \in \mathbb{R}^d : x_i < b_i, \forall i \} \), with \( b_i \in \mathbb{R} \), (4) all unbounded open rectangles with rational extremes, i.e., \( \{ x \in \mathbb{R}^d : x_i < b_i, \forall i \} \), with \( b_i \in \mathbb{Q} \), rational, (5) in the previous expressions we may replace the sign \( < \) by any of the signs \( \leq \) or \( > \) or \( \geq \) (i.e., replace open by closed or semi-closed) and the results remain true. Moreover, give an argument indicating that \( \mathcal{B}(\mathbb{R}^d) \) has the cardinality of the continuum (e.g., see Rama [92, Section 4.5, pp. 110-112]). Finally, prove that \( \mathcal{B}(\mathbb{R}^n \times \mathbb{R}^m) = \mathcal{B}(\mathbb{R}^n) \times \mathcal{B}(\mathbb{R}^m) \).

**Proof.** Since closed sets are complement of open sets, the assertion (1) follows. To check (2), we mention that the set of all bounded open rectangles (also called \( d \)-intervals) forms a basis for the topology in \( \mathbb{R}^d \), moreover, the endpoints of the bounded open \( d \)-intervals could be taken to be rational numbers, i.e., (4) also holds true. From \( d \)-dimensional version of expressions like \( (a,b) = (-\infty,b) \setminus (-\infty,a) \) and \( (-\infty,a] = \bigcap_{k \geq 1} (-\infty,a+1/k) \), we deduce assertions (3) and (5).

As in Exercise 1.3, if we realize that the cardinal of the set of all ordinal \( \alpha < \beta \) is actually the same as the cardinal of \( \beta \) then we see that the Borel \( \sigma \)-algebra is expressible as a continuum union (over the ordinal \( \alpha < \omega_0 \)) of sets with continuum cardinal.

To check that \( \mathcal{B}(\mathbb{R}^n \times \mathbb{R}^m) = \mathcal{B}(\mathbb{R}^n) \times \mathcal{B}(\mathbb{R}^m) \), we need to point out that a countable basis on \( \mathbb{R}^n \times \mathbb{R}^m \) can be constructed as a product of two countable basis (where we may include the whole space), one in \( \mathbb{R}^n \) and another in \( \mathbb{R}^m \). \( \square \)

**Exercise 1.12.** If \( \{ \mathcal{S}_i : i \in I \} \) is a family of semi-algebras (semi-rings) then define \( \mathcal{S} \) as the class of sets of the form \( \prod_{i \in I} A_i \), with \( A_i \in \mathcal{S}_i \) and \( A_i = \Omega_i, i \notin J \), for some finite non-empty index \( J \subset I \). Prove that \( \mathcal{S} \) is also a semi-algebra (semi-ring). **Hint:** For instance, make use of the equality

\[
(A_1 \times \cdots \times A_n \times \Omega_{n+1} \times \cdots)^c = (A_1 \times \cdots \times A_n)^c \times \Omega_{n+1} \times \cdots,
\]

to reduce the question to the case when the index \( I \) is finite. Next, for the case of \( \Omega_1 \times \Omega_2 \), we have the equalities \( (A_1 \times A_2)^c = A_1^c \times \Omega_2 + A_1 \times A_2^c \), \( (A_1 \times A_2) \cap (B_1 \times B_2) = (A_1 \cap B_1) \times (A_2 \cap B_2) \), \( (A_1 \times A_2) \setminus (B_1 \times B_2) = (A_1 \setminus B_1) \times A_2 + (A_1 \cap B_1) \times (A_2 \setminus B_2) \).

**Proof.** Note that the equality

\[
(A_1 \times \cdots \times A_n \times \Omega_{n+1} \times \cdots)^c = (A_1 \times \cdots \times A_n)^c \times \Omega_{n+1} \times \cdots,
\]
are measurable.

Thus, the function \( \mathbb{1}_A \) is measurable if and only if the set \( A \) is measurable. Also show that if \( V \) is an additive group of measurable functions (i.e., any function in \( V \) is measurable, the function identically zero belongs to \( V \), and \( f \pm g \) is in \( V \) for every \( f, g \) in \( V \)) then the family of sets \( \mathcal{L} = \{ A \subset \Omega : \mathbb{1}_A \in V \} \) is a \( \ell \)-class (or additive class) in \( 2^{\Omega} \).

**Proof.** Since the function \( \mathbb{1}_A \) takes only two values, the pre-image is

\[
(\mathbb{1}_A)^{-1}(B) = \begin{cases} 
\Omega & \text{if } 0, 1 \in B, \\
A & \text{if } 1 \in B \text{ and } 0 \notin B, \\
\Omega \setminus A & \text{if } 0 \notin B \text{ and } 1 \in B, \\
\emptyset & \text{if } 0, 1 \notin B.
\end{cases}
\]

Thus, the function \( \mathbb{1}_A \) is measurable if and only if the set \( A \) is measurable.

Since \( \mathbb{1}_A + \mathbb{1}_B = \mathbb{1}_{A \cup B} \) if \( A \cap B = \emptyset \) and \( \mathbb{1}_A - \mathbb{1}_B = \mathbb{1}_{A \setminus B} \) if \( A \supset B \), we deduce that \( \mathcal{L} \) is a \( \ell \)-class.

**Exercise 1.14.** Let \( \{ f_n \} \) be a sequence of measurable functions with extended real values and set \( \overline{f}(x) = \sup_n f_n(x) \) and \( \underline{f}(x) = \inf_n f_n(x) \). Discuss why the expressions \( \overline{f}^{-1}([\infty, a]) = \bigcap_n f_n^{-1}([\infty, a]) \) and \( \underline{f}^{-1}([a, +\infty]) = \bigcap_n f_n^{-1}([a, +\infty]) \) for every \( a \in \mathbb{R} \) show that \( \overline{f} \) and \( \underline{f} \) are measurable.

**Proof.** Certainly, we need to mention that \( \sigma \)-algebras are stable under the formation of countable intersections. Next, the Borel \( \sigma \)-algebra \( \mathcal{B}(\mathbb{R}) \) for the extended real numbers \( \mathbb{R} = [-\infty, +\infty] \) is the \( \mathcal{B}(\mathbb{R}) \) and unions of either \( -\infty \) or \( +\infty \) or both any some element in \( \mathcal{B}(\mathbb{R}) \). Moreover, because \( [-\infty, a] = [-\infty, a] \setminus \bigcap_{n \geq 1} [-\infty, -n] \) and \( [a, +\infty) = [a, +\infty] \setminus \bigcap_{n \geq 1} [n, +\infty] \), we deduce that the class of intervals either \( [-\infty, a] \) or \( [a, +\infty] \), for \( a \) in \( \mathbb{R} \) (or rational numbers), generates Borel \( \sigma \)-algebra \( \mathcal{B}(\mathbb{R}) \).

**Exercise 1.15.** Let \( \{ f_n : n = 1, 2, \ldots \} \) be a sequence of measurable functions from a measurable space \( (X, \mathcal{X}) \) into another measurable space \( (Y, \mathcal{Y}) \). Prove that if \( (Y, d) \) is also a complete metric space and \( \mathcal{Y} \) is the corresponding
Borel σ-algebra then the set \( \{ x \in X : f_n(x) \text{ convergent} \} \) is a measurable set. Hint: may need to use the fact that \( \{ f_n(x) \} \) is convergent if and only if \( \sup_{n \geq k} d(f_n(x), f_k(x)) \to 0 \) as \( k \to \infty \).

**Proof.** In a complete metric space, Cauchy sequences and convergent sequences are equivalent, i.e., \( \{ f_n(x) \} \) is convergent if and only if \( \sup_{n \geq k} d(f_n(x), f_k(x)) \to 0 \) as \( k \to \infty \). Since the function \( g(x) = \inf_n \sup_{n \geq k} d(f_n(x), f_k(x)) \) is measurable, and the set \( A = \{ x \in X : f_n(x) \text{ is convergent} \} \) is the pre-image of \( g^{-1}(\{0\}) \), we deduce that \( A \) is measurable. \( \square \)

**Exercise 1.16.** Discuss the difference between the Borel σ-algebras \( \mathcal{B}(\prod_{i \in I} \Omega_i) \) in the product space and \( \prod_{i \in I} \mathcal{B}(\Omega_i) \), when \( \Omega_i \) is a topological space and \( I \) is a countable or uncountable set of index.

**Proof.** The point is that the product topology is generated by open cylinder sets and the product σ-algebra \( \prod_{i \in I} \mathcal{B}(\Omega_i) \) could be also generated by open cylinder sets. However, in a topology, any kind of union of open sets is an open set, but only a countable union of Borel sets is a Borel set. Thus, the product topology could contain open sets that are not “Borel product” sets, even for a finite product of space.

If the topology of each \( \Omega_i \) has a countable basis and the set of indexes \( I \) is countable, then the Borel σ-algebras \( \mathcal{B}(\prod_{i \in I} \Omega_i) \) in the product space and \( \prod_{i \in I} \mathcal{B}(\Omega_i) \), are both generated the countable number of cylinder sets (here, use the fact that the set of all subsets with a finite number of elements of a countable set is again a countable set), namely, \( \prod_{i \in I} A_i \) with \( A_i = \Omega_i \) for every \( i \) in \( I \setminus J \) and \( A_j \) in the chosen countable basis for \( \Omega_j \), for every \( j \) in \( J \), for some finite family of indexes \( J \).

If the set of indexes \( I \) is uncountable, e.g., \( \mathbb{R}^T \) (set of all functions from \( T \) into \( \mathbb{R} \)) with \( T \) uncountable, we see that \( \mathcal{B}(\prod_{i \in I} \Omega_i) \) could strictly larger than \( \prod_{i \in I} \mathcal{B}(\Omega_i) \). Indeed, if we take an uncountable family \( \{ A_i : i \in J \} \) of closed sets \( A_i \) in \( \Omega_i \) with \( A_i \neq \Omega_i \) for any \( i \) in \( J \) then the set \( \prod_{i \in I} A_i \) with \( B_i = A_i \) for \( i \) in \( J \) and \( B_i = \Omega_i \) for \( i \) in \( I \setminus J \) does not belong to the product σ-algebra \( \prod_{i \in I} \mathcal{B}(\Omega_i) \), but \( \prod_{i \in I} B_i = \bigcap_{i \in J} (A_i \times \prod_{i \notin J} \Omega_i) \) is a closed set in the product topology. In particular, a singleton (i.e., a set of only one element) is closed in \( \mathbb{R}^T \), but it is not in the product σ-algebra \( \mathcal{B}(\mathbb{R}) \).

Similarly, if \( \Omega_1 \) has only uncountable basis, e.g., \( \{ B_{1,k} : k \in K \} \) with \( K \) an uncountable set of indexes, then cylinder sets of the form \( B_{1,k} \times B_2 \) with \( k \) in \( K \) and \( B_2 \) open set in \( \Omega_2 \) is a basis for the product topology in \( \Omega_1 \times \Omega_2 \), but it is not a basis for the product Borel σ-algebra \( \mathcal{B}(\Omega_1 \times \Omega_2) \). \( \square \)

**Exercise 1.17.** Let \( (\Omega, \mathcal{F}), (E_i, \mathcal{E}_i), i \in I \) be measurable spaces, and set \( E = \prod_i E_i, \mathcal{E} = \prod_i \mathcal{E}_i \), and \( \pi_i : E \to E_i \) the projection. Prove that \( f : \Omega \to E \) is \( (\mathcal{F}, \mathcal{E}) \)-measurable if and only if \( \pi_i \circ f \) is \( (\mathcal{F}, \mathcal{E}_i) \)-measurable for every \( i \in I \).

**Proof.** Any measurable cylinder set has the form \( C = \prod_{i \in J} A_i \), where \( A_i \) are sets in \( \mathcal{E}_i \) and \( A_i = \Omega_i \) for every \( i \) in \( I \setminus J \), for some finite set of indexes \( J \). Thus, by means of the projection functions, we have \( C = \bigcap_{i \in J} \pi_i^{-1}(A_i) \).
Since the class of measurable cylinder sets generates the product $\sigma$-algebra $E$, a function $f : \Omega \to E$ is measurable if and only if the pre-image of a measurable cylinder is a measurable set in $\Omega$. Also, the composed function $\pi_i \circ f : \Omega \to E_i$ is measurable if and only if $(\pi_i \circ f)^{-1}(A_i) = f^{-1}(\pi_i^{-1}(A_i))$ is measurable in $\Omega$ for every $A_i$ in $E_i$. Because the pre-image (though any function $f$) preserves intersections, we conclude.

**Exercise 1.18.** Given an example of a non-measurable real-valued function $f$ on a given measurable space $(\Omega, F)$ with $F \neq 2^\Omega$ such that $f^2$ is measurable. Prove that an (extended) real-valued $f$ is measurable if and only if its positive part $f^+ = \max\{f, 0\}$ and its negative part $f^- = -\min\{f, 0\}$ are both measurable.

**Proof.** For instance, if $A$ is a non measurable set then the function $f = 1_A - 1_{\Omega \setminus A}$ is non measurable, $f^{-1}(0, +\infty)) = A$. However, $f^2(x) = 1$ for every $x$ in $\Omega$, and a constant functions is always measurable.

It is clear that if $f^-$ and $f^+$ are both measurable then $f = f^+ - f^-$ is also measurable. On the other hand, if $f$ is measurable, because the constant function $g(x) = 0$ is also measurable, the max or min of measurable functions is also measurable, we deduce that $f^-$ and $f^+$ are both measurable. Alternatively, first we check that $f$ is measurable then $|f|$ is also measurable, as the composition with the measurable function absolute value $x \mapsto |x|$ from $(\mathbb{R}, B(\mathbb{R}))$ into itself (or with the extended real numbers $\mathbb{R}$). Next, we write $f^- = (|f| - f)/2$ and $f^+ = (|f| + f)/2$ to conclude. Note that, we may have $|f|$ measurable, even when $f$ is not measurable.

**Exercise 1.19.** Let $f$ be a function between two measurable space $(X, \mathcal{X})$ and $(Y, \mathcal{Y})$. Prove that (1) $f^{-1}(\mathcal{Y}) = \{f^{-1}(B) \in 2^X : B \in \mathcal{Y}\}$ is a $\sigma$-algebra in $X$ and (2) $\{B \in 2^Y : f^{-1}(B) \in \mathcal{X}\}$ is a $\sigma$-algebra in $Y$.

**Proof.** The assertions follow from the fact that pre-image preserves the set operations like the formation of any intersections, any unions, any differences and complements, in particular, $f^{-1}(\bigcup_n B_n) = \bigcup_n f^{-1}(B_n)$, $f^{-1}(Y) = X$, and $f^{-1}(A \setminus B) = f^{-1}(A) \setminus f^{-1}(B)$.

**(1.4) Some Examples**

**Exercise 1.20.** Give an example of a topological space $X$ where the Borel, the Baire $\sigma$-algebra are all distinct.

**Proof.** If $X$ is a product of uncountable many compact Hausdorff spaces each having more than one point, then a singleton (a set of only one element) is closed and hence a Borel set, but it is not a Baire set. Actually, it is recommended read more on Baire sets and functions, e.g., check the first pages of Dudley [37, Section 7.1, pp. 222–228].
(1.5) Various Tools

**Exercise 1.21.** Let $\mathcal{S}$ be a semi-ring of measurable sets in $(\Omega, \mathcal{F})$ such that $\sigma(\mathcal{S}) = \mathcal{F}$ and there exists a sequence $S_i$ in $\mathcal{S}$ satisfying $\Omega = \bigcup_i S_i$. Denote by $\mathcal{S}$ the vector space generated by all functions of the form $\mathbb{1}_A$ with $A$ in $\mathcal{S}$. Besides having a finite number of values, what other property is needed to give a characterization of a function in $\mathcal{S}$? Consider the following questions:

1. Verify that if $\varphi, \phi \in \mathcal{S}$ then $\max\{\varphi, \phi\} \in \mathcal{S}$ (i.e., $\mathcal{S}$ a lattice) and $\varphi \phi \in \mathcal{S}$.

2. Now, define $\bar{\mathcal{S}}$ as the semi-space of extended real-valued functions which can be expressed as the pointwise limit of a monotone increasing sequence of functions in $\mathcal{S}$. Show that if $f(x) = \lim_n f_n(x)$, with $f_n(x) \leq f_{n+1}(x)$, for every $x$ in $\Omega$ and $f_n$ in $\bar{\mathcal{S}}$, then $f$ belongs to $\bar{\mathcal{S}}$. Verify that the function $1 = 1_\Omega$ may not belongs to $\mathcal{S}$, but $1$ belongs to $\bar{\mathcal{S}}$. Moreover, if $u : \mathbb{R}^2 \to \mathbb{R}$ is a continuous function, nondecreasing in each variable and with $u(0,0) = 0$, then show that $x \mapsto u(f(x), g(x))$ belongs to $\bar{\mathcal{S}}$ for every $f$ and $g$ in $\bar{\mathcal{S}}$. Therefore, $\bar{\mathcal{S}}$ is a lattice but not necessarily a vector space.

**Proof.** Certainly, a real-valued function $f$ on $\Omega$ belongs to $\mathcal{S}$ if and only if $f$ takes a finite number of values and the pre-image of any singleton $f^{-1}\{\{a\}\}$, for any $a \neq 0$, belongs to the ring generated by $\mathcal{S}$ (i.e., it is a disjoint unions of sets in $\mathcal{S}$). In other words, $f = \sum_{i=1}^n c_i \mathbb{1}_{A_i}$ is a finite linear combination of $\mathbb{1}_A$ with $A$ in $\mathcal{S}$, moreover, with $A_i$ disjoint sets belong to $\mathcal{S}$. Note that the null function $0 = \mathbb{1}_\emptyset$ belongs to $\mathcal{S}$.

1. To check that $\max\{\varphi, \phi\}$ and $\varphi \phi$ are in $\mathcal{S}$, we write $\varphi = \sum_{i=1}^n a_i \mathbb{1}_{A_i}$ and $\phi = \sum_{j=1}^m b_j \mathbb{1}_{B_j}$ with $\{A_i\}$ and $\{B_j\}$ disjoint finite sequences of sets in $\mathcal{S}$, and $\{a_i\}$ and $\{b_j\}$ are finite sequences of real numbers. Since

$$
\max\{a \mathbb{1}_{A_i}(x), b \mathbb{1}_{B_j}(x)\} = \begin{cases} 
  a \lor b & \text{if } x \in A \cap B, \\
  a \lor 0 & \text{if } x \notin A \cup B, \\
  b \lor 0 & \text{if } x \in B \setminus A, \\
  0 & \text{if } x \notin A \cup B,
\end{cases}
$$

we have

$$
\max\{\varphi, \phi\} = \sum_{i,j} \left[ (a_i \lor b_j) \mathbb{1}_{A_i \cap B_j} + (a_i \lor 0) \mathbb{1}_{A_i \setminus B_j} + (b_j \lor 0) \mathbb{1}_{B_j \setminus A_i} \right],
$$

which proves that $\max\{\varphi, \phi\}$ belongs in $\mathcal{S}$. Similarly, the equality

$$
\varphi \phi = \sum_{i,j} (a_i b_j) \mathbb{1}_{A_i \cap B_j}
$$

shows $\varphi \phi$ is in $\mathcal{S}$. Therefore $\mathcal{S}$ is a vector lattice. Actually, for any function $u : \mathbb{R}^2 \to \mathbb{R}$, with $u(0,0) = 0$, we have

$$
u(\varphi, \phi) = \sum_{i,j} \left[ u(a_i, b_j) \mathbb{1}_{A_i \cap B_j} + u(a_i, 0) \mathbb{1}_{A_i \setminus B_j} + u(b_j, 0) \mathbb{1}_{B_j \setminus A_i} \right],
$$
where the sum in $i, j$ reduces to just one term when applied to any point $x$. Hence, $u(\varphi, \phi)$ belongs to $S$.

(2) Indeed, for each $f_n$ in $\bar{S}$, there exists an increasing sequence $\{g_{n,k} : k \geq 1\}$ in $S$ such that $f_n(x) = \lim_k g_{n,k}(x)$, for every $x$ in $\Omega$. Thus, $g_n = \max\{g_{n,1}, \ldots, g_{n,n}\}$ defines an increasing sequence of functions in $S$ satisfying $f_m \leq \lim_n g_n \leq f$, for every $m$. This shows that $\{g_n\}$ converges pointwise to $f$, i.e., $f$ belongs to $\bar{S}$.

Similarly, for a given function $u : \mathbb{R}^2 \rightarrow \mathbb{R}$ as above, if $f$ and $g$ belong to $\bar{S}$, we can define $h_k = u(f_k, g_k)$ to construct a sequence in $S$. Since $u$ is increasing and continuous, as in the previous paragraph, we deduce that $u(f, g)$ belongs to $\bar{S}$. Thus, $\bar{S}$ is a lattice, i.e., $f$ and $g$ in $\bar{S}$ implies $\max\{f, g\}$ in $\bar{S}$. However, $af + bg$ is in $\bar{S}$ when $a, b \geq 0$, but not in general.

Finally, it is clear that $\Omega$ may or may not belong to the semi-ring $S$, but because $\Omega = \bigcup S_i$ for some sequence of sets in $S$, we deduce that $1_\Omega = \lim_n 1_{A_n}$, with $A_n = \bigcup_{i=1}^n S_i$, always belongs to $\bar{S}$. □

**Exercise 1.22.** Let $(X, X')$ and $(Y, Y')$ be two measurable spaces and let $f : X \times Y \rightarrow [0, +\infty]$ be a measurable function. Consider the functions $f_\ast(x) = \inf_{y \in Y} f(x, y)$ and $f^\ast(x) = \sup_{y \in Y} f(x, y)$ and assume that for every simple (measurable) function $f$ the functions $f_\ast$ and $f^\ast$ are also measurable. By means of the approximation by simple functions given in Proposition 1.9:

(1) Prove that $f_\ast$ and $f^\ast$ are measurable functions.

(2) Extend this result to a real-valued function $f$.

(3) Now, if $g : \mathbb{R}^d \rightarrow \mathbb{R}$ is a locally bounded (Borel) measurable function, then show that $\overline{g}(x) = \lim_{r \rightarrow 0} \sup_{0 < |y - x| < r} g(y)$ and $\underline{g}(x) = \lim_{r \rightarrow 0} \inf_{0 < |y - x| < r} g(y)$ are measurable functions.

(4) Finally, give some comments on the assumption about the measurability of $f_\ast$ and $f^\ast$ for a simple function $f$.

Hint: apply the transformation arctan to obtain a bounded function, i.e., $f_\ast(x) = \tan\left(\inf_{y \in Y} \arctan f(x, y)\right)$ and $f^\ast(x) = \tan\left(\sup_{y \in Y} \arctan f(x, y)\right)$. Regarding (4), the brief comments in next Chapter about analytic sets and universal comprehension hold an answer, see Proposition 2.3.

**Proof.** (1) Since $f^\ast = -(f)_\ast$, we need to consider only one of them, e.g., $f^\ast$. Moreover, the transformation $t \mapsto \arctan$ and its inverse allow us to reduce the problem to the case of a bounded function. Now, by means of Proposition 1.9, there is a sequence $\{f_n\}$ of measurable simple functions such that $f_n \rightarrow f$ uniformly in $X \times Y$. Hence, the inequality

$$|f^\ast(x) - f^\ast_n(x)| \leq \sup_{y \in Y} |f(x, y) - f_n(x, y)|,$$

implies that $f^\ast_n \rightarrow f$ uniformly in $X$. By assumption, $f^\ast_n$ is measurable, and so is $f^\ast$.

(2) If $f$ is a real-valued functions then we can use the relation $f^\ast = a + (f - a)^\ast$ for any constant $a$ to reduce the problem to the case where $F$ is nonnegative.
(3) By taking a sequence in $r$, we only need to show that the functions
\( \sup_{0<|y-x|<r} g(y) \) and \( \inf_{0<|y-x|<r} g(y) \) are measurable. Moreover, as above, we need to consider only one of them (e.g., the expression with the sup), for a bounded nonnegative function $g$. Thus, by continuity, \((x,y)\mathbb{R}^d \times \mathbb{R}^d : 0 <|x-y|<r \} \) is an open set; and therefore, the function $f(x,y) = g(y)$ if $0 <|y-x|<r$ and $f(x,y) = 0$ otherwise is measurable. Hence, applying (2), we deduce that $f^* = \sup_{0<|y-x|<r} g(y)$ is indeed measurable.

Alternatively, we can use the explicit expression $f(x,y) = g(y)1_{|y-x|<r}$ to approximate $g$ by a sequence of simple functions, to reduce the problem to the case where $g$ assume only a finite number of values, i.e., we only have to establish the measurability for functions $f(x,y)$ of the form $a1_{|y-x|<r}$, or equivalently of the form $\varphi(x-y)$, with smooth functions $\varphi$ vanishing outside of a compact set. Hence, $\overline{g}$ and $g$ are measurable, independently of the assumption that $f^*$ and $f_*$ are measurable for every measurable simple function $f$.

(4) If $f = \sum_{i=1}^{n} a_i 1_{A_i}$ or equivalently $f = \sup_{1\leq i \leq n} a_i 1_{A_i}$, for a disjoint finite sequence \( \{A_i\} \) of measurable sets and constants $a_i \geq 0$, then $f^*(x) = \sup_{1\leq i \leq n} \sup_{y \in Y} a_i 1_{A_i(x,y)}$, so that measurability of $f^*$ is actually reduced to the measurability for the particular case $f = 1_{A}$, for any measurable set in the product $\sigma$-algebra $\mathcal{X} \times \mathcal{Y}$. Now, applying Exercise 1.7, take a product measurable set $A$ (i.e., in $\mathcal{X} \times \mathcal{Y}$) that can be written as a disjoint union of non-empty cylinder sets (rectangles in this case), i.e., $A = \sum_{i=1}^{\infty} X_i \times Y_i$, where $X_i$ and $Y_i$ are non-empty measurable sets. Again, we can write $(1_{A})^*(x) = \sup_{i\geq 1} \sup_{y \in Y} 1_{X_i \times Y_i}(x,y) = \sup_{i \geq 1} 1_{X_i}(x)$, which proves that $(1_{A})^*$ is measurable, i.e., the assumption on the measurability of $f^*$ and $f_*$ for the special case of simple functions (with sets as above) is always satisfied.

Exercise 1.23. Let \( \{f_t : t \in T\} \) be a family of measurable functions from $(\Omega, \mathcal{F})$ into a Borel space $(E, \mathcal{E})$. Assume that the set of indexes $T$ has a (sequential) topology and there exists a countable subset of indexes $Q \subset T$ such that for every $x$ in $\Omega$ and any $t$ in $T$ there is a sequence \( \{t_n\} \) of indexes in $Q$ such that $t_n \to t$ and $f_{t_n}(x) \to f_t(x)$.

(a) Prove that for metric spaces, the above condition is equivalent to the following statement: the sets \( \{x \in \Omega : f_t(x) \in C, \forall t \in O\} \) and \( \{x \in \Omega : f_t(x) \in C, \forall t \in O \cap Q\} \) are equal, for every closed set $C$ in $E$ and any open set $O$ in $T$.

(b) If the mapping $t \mapsto f_t(x)$ is continuous for every fixed $x$ in $\Omega$, then prove that any countable dense set $Q$ in $T$ satisfies the above condition.

(c) If $E = [-\infty, +\infty]$ then for the family \( \{f_t : t \in T\} \) of extended real-valued functions, we have

\[
\begin{align*}
    f^*(x) &= \sup_{t \in T} f_t(x) = \sup_{t \in Q} f_t(x), \\
    f_*(x) &= \inf_{t \in T} f_t(x) = \inf_{t \in Q} f_t(x),
\end{align*}
\]

which prove that $f^*$ and $f_*$ are measurable functions.

Proof. (a) It is clear that $A = \{x \in \Omega : f_t(x) \in C, \forall t \in O \cap Q\} \subset B = \{x \in \Omega : f_t(x) \in C, \forall t \in O\}$, and to check the converse, take $x$ in $B$ and $t$ in $O \setminus Q$. 

Because there exists a sequence \( \{t_n\} \) of indexes in \( Q \) such that \( t_n \to t \) and \( f_{t_n}(x) \to f_t(x) \), and because \( O \) is open and \( C \) is closed, we deduce that \( t_n \) is in \( O \) (for \( n \) sufficiently large), and \( f_t(x) \) belongs to \( C \), i.e., \( x \) must be in \( A \). On the other hand, note that we can express the above sets as

\[
A = \bigcap_{s \in O \cap Q} \{x \in \Omega : f_s(x) \in C\} \quad \text{and} \quad B = \bigcap_{s \in O} \{x \in \Omega : f_s(x) \in C\}.
\]

Thus, if there exist an open set \( O \), a closed set \( C \) and some \( x \) in \( B \setminus A \) then there is an index \( t \) in \( O \setminus Q \) such that \( f_t(x) \) does not belongs to \( C \), while \( f_s(x) \) belongs to \( C \), for every \( s \) in \( O \cap Q \). Hence, there can not exits a sequence \( \{t_n\} \) in \( Q \) such that \( f_{t_n}(x) \to f_t(x) \).

Actually, we require only that the topology on \( E \) and \( T \) be sequential, in particular metric spaces.

(b) If the mapping \( t \mapsto f_t(x) \) is continuous, for every fixed \( x \) in \( \Omega \), and \( Q \) is a dense set of indexes, then for every \( t \) there exits a sequence \( \{t_n\} \) in \( Q \) such that \( t_n \to t \). Hence, by continuity, we obtain \( f_{t_n}(x) \to f_t(x) \).

(c) For instance, we use the equalities \( \{x : f^*(x) > a\} = \bigcup_{t \in Q} \{x : f_t(x) > a\} \) and \( \{x : f_*(x) \geq a\} = \bigcap_{t \in Q} \{x : f_t(x) \geq a\} \), for every real number \( a \), to check that \( f^* \) and \( f_* \) are measurable. \( \square \)

**Exercise 1.24.** Let \( \mathcal{P} = \{\Omega_1, \ldots, \Omega_n\} \) be a finite partition of \( \Omega \), i.e., \( \Omega = \sum_{i=1}^n \Omega_i \), with \( \Omega_i \neq \emptyset \). Describe the algebra \( \mathcal{A} \) generated by the finite partition \( \mathcal{P} \) and prove that each \( \Omega_i \) is an atom of \( \mathcal{A} \). How about if \( \mathcal{P} \) is a countable or uncountable partition?

**Proof.** Because \( \Omega_i \setminus \Omega_j \) and \( \Omega_i \cap \Omega_j \) is either \( \emptyset \) or \( \Omega_i \), we deduce that a set \( A \) belongs to the algebra \( \mathcal{A} \) generated by the partition \( \mathcal{P} \) if and only if \( A = \emptyset \) or \( A = \sum_{i \in I} \Omega_i \) for some subset of indexes \( I \) of \( \{1, \ldots, n\} \).

Recall that an atom \( A \) of \( \mathcal{A} \) is an elements \( A \in \mathcal{A} \) such that \( A \neq \emptyset \), and any \( B \subset A \) with \( B \in \mathcal{A} \) results \( B = \emptyset \) or \( B = A \). Thus, it is clear that each \( \Omega_i \) is an atom in \( \mathcal{A} \).

The class \( \mathcal{A} \) is an finite algebra, and therefore, \( \mathcal{A} \) is also a \( \sigma \)-algebra. It is almost similar for the case of a countable partition \( \mathcal{P} \), \( \Omega = \sum_{i=1}^\infty \Omega_i \), with \( \Omega_i \neq \emptyset \), a set \( A \) belongs to the algebra \( \mathcal{A} \) generated by the partition \( \mathcal{P} \) if and only if \( A = \emptyset \) or \( A = \sum_{i \in I} \Omega_i \) for some finite subset of indexes \( I \) of \( \{1, 2, \ldots\} \). However, the algebra \( \mathcal{A} \) is infinite, and the elements of the \( \sigma \)-algebra \( \sigma(\mathcal{P}) \) have the same form, but with a countable subset of indexes \( I \).

Finally, if the set of indexes for the partition \( \mathcal{P} \) is uncountable then the elements of the ring \( \mathcal{A} \) generated by \( \mathcal{P} \) have the same form, with a finite subset of indexes \( I \), while \( \sigma \)-ring uses a countable subset of indexes \( I \). To get the algebra or \( \sigma \)-algebra, we need to add subset of indexes \( I \) with either finite or countable complement. \( \square \)
(2.1) Abstract Measures

**Exercise 2.1.** Let \((\Omega, \mathcal{A}, \mu)\) be a measure space and \(\{A_i : i \geq 1\}\) be a sequence in \(\mathcal{A}\). We write \(\lim \inf_k A_k = \bigcup_{n} \bigcap_{i \geq n} A_i\) and \(\lim \sup_k A_k = \bigcap_{n} \bigcup_{i \geq n} A_i\). Prove (a) \(\mu(\lim \inf_k A_k) \leq \lim \inf_k \mu(A_k)\) and (b) if \(\mu(\bigcup_{i \geq n} A_i) < \infty\) for some \(n\) then \(\mu(\lim \sup_k A_k) \geq \lim \sup_k \mu(A_k)\).

**Proof.** Consider the monotone sequences

\[ B_k = \bigcap_{i \geq k} A_i = \bigcup_{n \leq k} \bigcap_{i \geq n} A_i \quad \text{and} \quad C_k = \bigcup_{i \geq k} A_i = \bigcap_{n \leq k} \bigcup_{i \geq n} A_i. \]

Since \(B_k \subset A_i \subset C_k\), for every \(i \geq k\), the monotonicity of the measure \(\mu\) implies \(\mu(B_k) \leq \inf_{i \geq k} \mu(A_i)\) and \(\mu(C_k) \geq \sup_{i \geq k} \mu(A_i)\). Hence

\[ \mu(\lim \inf_k A_k) = \lim_k \mu(B_k) \quad \text{and} \quad \mu(\lim \sup_k A_k) = \lim_k \mu(C_k), \]

which yield (a) and (b).

Note that, using the indicator or characteristic set function, we have

\[ \mathbb{1}_A = \lim \inf_{n \geq k} \mathbb{1}_{A_i} = \lim_k \mathbb{1}_{A_k} \quad \text{and} \quad \mathbb{1}_{\overline{A}} = \lim \sup_{n \geq k} \mathbb{1}_{A_i} = \lim_k \mathbb{1}_{A_k}, \]

where \(A = \lim \inf_k A_k\) and \(\overline{A} = \lim \sup_k A_k\).

**Exercise 2.2.** Let \((\Omega, \mathcal{A}, \mu)\) be a measure space. Verify that (a) if \(A\) and \(B\) belong to \(\mathcal{A}\) then \(\mu(A) + \mu(B) = \mu(A \cup B) + \mu(A \cap B)\). Also, prove that (b) if \(\{A_i : i \geq 1\}\) is a sequence of sets in \(\mathcal{A}\) satisfying \(\mu(A_i \cap A_j) = 0\) for \(i \neq j\) then \(\mu(\bigcup_i A_i) = \sum_i \mu(A_i)\).

**Proof.** To check (a) note that if \(\mu(A \cap B) < \infty\) then \(B = A \cup (B \setminus A)\) implies \(\mu(B \setminus A) = \mu(B) - \mu(A \cap B)\). Thus, from the equality \(A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B)\) we obtain

\[ \mu(A \cup B) = \mu(A \setminus B) + \mu(B \setminus A) + \mu(A \cap B) = \mu(A) + \mu(B) - \mu(A \cap B). \]
Hence \( \mu(A) + \mu(B) = \mu(A \cup B) + \mu(A \cap B) \), which is trivially valid when \( \mu(A \cup B) = \infty \).

For (b), note that a sequence satisfying the condition \( \mu(A_i \cap A_j) = 0 \) for \( i \neq j \) is usually referred to as a \( \mu \)-non overlapping. Now, by means of the monotone continuity from below, it suffices to show the above equality for a finite number of sets, i.e., for \( 1 \leq i \leq n \), \( \mu(\bigcup_{i \leq n} A_i) = \sum_{i \leq n} \mu(A_i) \). Moreover, for \( n = 2 \) the equality follows from part (a), thus we proceed by induction. Indeed, assuming the equality is valid for \( n \), define \( A = \bigcup_{i \leq n} A_i \) and \( B = A_{n+1} \). Since \( A \cap B = \bigcup_{i \leq n} (A_i \cap A_{n+1}) \) we have \( \mu(A \cap B) = 0 \) and from part (1), we obtain \( \mu(A) + \mu(B) = \mu(A \cup B) \). Therefore

\[
\mu(\bigcup_{i \leq n+1} A_i) = \mu(A \cup B) = \mu(B) + \mu(A) = \\
= \mu(A_{n+1}) + \sum_{i \leq n} \mu(A_i) = \sum_{i \leq n+1} \mu(A_i),
\]

i.e., equality is also valid for \( n + 1 \). \(\square\)

**Exercise 2.3.** Let \( \{\mu_n\} \) be a sequence of measures. Prove (a) if \( \{c_n\} \) is a sequence of nonnegative numbers then \( \mu = \sum_n c_n \mu_n \) is a measure; (a) if the sequence \( \{\mu_n\} \) is increasing then \( \mu = \lim \mu_n \) is a measure; (c) if the sequence \( \{\mu_n\} \) is decreasing and \( \mu_1 \) finite then \( \mu = \lim \mu_n \) is a measure.

**Proof.** To check (a), recall that for any double sequence \( \{a_{i,j}\} \) of positive real numbers (where the symbol \( +\infty \) may be include), we show that \( \sum_i \sum_j a_{i,j} = \sum_j \sum_i a_{i,j} = \sum_{i,j} a_{i,j} \). Hence, the \( \sigma \)-additivity of \( \mu \) follows immediately.

Essentially the same argument is used for (b), i.e., for any double monotone increasing sequence \( \{a_{i,j}\} \) of positive real numbers (where the symbol \( +\infty \) may be include), we can verify that the iterated limits are equal to the double limit, namely, \( \lim_i \lim_j a_{i,j} = \lim_j \lim_i a_{i,j} = \lim_{i,j} a_{i,j} \).

Finally, it is also clear that (c) can be reduced to (b), when \( \mu_1 \) is a finite measure. \(\square\)

**Exercise 2.4.** Let \( (\tilde{A}, \tilde{\mu}) \) be the completion of \( (A, \mu) \) as above. Verify that \( \tilde{A} \) is a \( \sigma \)-algebra and that \( \tilde{\mu} \) is indeed well defined. Moreover, show that if \( \tilde{A} \in \tilde{A} \) then there exist \( A, B \in A \) such that \( A \subset \tilde{A} \subset B \) and \( \mu(A) = \mu(B) \). Furthermore, if \( \tilde{A} \subset \Omega \) and there exist \( A, B \in A \) such that \( A \subset \tilde{A} \subset B \) with \( \mu(A) = \mu(B) < \infty \) then \( \tilde{A} \in \tilde{A} \).

**Proof.** If \( \{A_i, N_i\} \) is a sequence of sets in the completion \( \sigma \)-algebra \( \tilde{A} \) then there exist a sequence \( \{A_i, N_i\} \) of sets in \( A \) such that \( A_i \subset \tilde{A}, \tilde{A} \setminus A_i \subset N_i \), and \( \mu(N_i) = 0 \). Hence \( A = \bigcup_i A_i \subset \bigcup_i \tilde{A} = \tilde{A}, \tilde{A} \setminus A \subset \bigcup_i N_i = N \), and \( \mu(N) = \sum_i \mu(N_i) = 0 \), i.e., \( \tilde{A} \) is closed under the formation of countable unions. Moreover, if \( A \) belongs to \( \tilde{A} \) and \( A \subset \tilde{A}, \tilde{A} \setminus A \subset N \), and \( \mu(N) = 0 \), then \( \tilde{A} \subset A \cup N \) and the complement \( (A \cup N)^c \subset \tilde{A}^c, \tilde{A}^c \setminus (A \cup N)^c = \tilde{A}^c \setminus (A \cup N) \subset N \), i.e., the complement \( \tilde{A}^c \) also belongs to \( \tilde{A} \).
Certainly, if $A_1 \subset \bar{A}$, $\bar{A} \setminus A_1 \subset N_1$, and also $A_2 \subset \bar{A}$, $\bar{A} \setminus A_2 \subset N_2$, i.e., $\bar{A} = A_1 \cup (\bar{A} \setminus A_1) \subset A_1 \cup N_1$, $i=1,2$. This implies that $(A_1 \setminus A_2) \cup (A_2 \setminus A_1) \subset N_1 \cup N_2$, and because $\mu(N_1) = \mu(N_2) = 0$, we deduce $\mu(A_1) = \mu(A_2)$.

As above, if $A$ belongs to $\bar{A}$ then $A \subset \bar{A}$, $\bar{A} \setminus A \subset $, and $\mu(N) = 0$, which implies $\bar{A} = A \cup (\bar{A} \setminus A) \subset A \cup N = B$, with $A \subset \bar{A} \subset B$ and $\mu(B) \leq \mu(A)+\mu(N)$, i.e., $\mu(A) = \mu(B)$. Conversely, if there exist $A$ and $B$ in $\mathcal{A}$ such that $A \subset \bar{A} \subset B$ with $\mu(A) = \mu(B) < \infty$ then define $N = B \setminus A$ to show that $\bar{A}$ belongs to the completion $\sigma$-algebra $\mathcal{A}$.

\[\square\]

\section*{(2.2) Caratheodory’s Arguments}

\begin{exercise}
With the notation of Proposition 2.6, if the initial set measure $\mu(E) = \sum_{\omega_i \in E} r_i$, for some sequences $\{\omega_i\} \subset \Omega$ and $\{r_i\} \subset [0, \infty]$ then the same weighted counting expression holds $\mu^*(E)$, for any $E \in \mathcal{E}$, i.e.,

$$\mu^*(E) = \sum_{\omega_i \in E} r_i = \sum_{i=1}^{\infty} r_i \mathbf{1}_{\omega_i \in E}, \quad \forall E \in \mathcal{E},$$

where $\mathbf{1}_{\omega_i \in E} = 1$ if $\omega_i$ is in $E$ and vanishes otherwise. What can be said about $\mu^*(A)$, for any $A \subset 2^\Omega$?

\end{exercise}

\begin{proof}
Denote by $\bar{\mu}(A)$ the weighted counting expression on the right, for any $A \subset \Omega$. If $\{E_n\} \subset \mathcal{E}$ is a sequence covering $A$ and $\omega_i \in A$ then $\omega_i$ belongs to $\bigcup_n E_n$, i.e., $\omega_i$ belongs to some $E_n$, i.e.,

$$\sum_n \sum_{i=1}^{\infty} r_i \mathbf{1}_{\omega_i \in E_n} = \sum_{i=1}^{\infty} r_i \mathbf{1}_{\omega_i \in \bigcup_n E_n} \geq \sum_{i=1}^{\infty} r_i \mathbf{1}_{\omega_i \in A},$$

which yields $\mu^*(A) \geq \bar{\mu}(A)$. Because any set $E \in \mathcal{E}$ can be covered by itself, we deduce $\mu^*(E) \geq \bar{\mu}(E)$, for every $E \in \mathcal{E}$.

In general, we may not have $\mu^*(A) = \bar{\mu}(A)$. For instance, if there exist two distinct points $\omega_1$ and $\omega_2$ (with $f_1$ and $f_2$ non-zero) which cannot be separated by $\mathcal{E}$ (i.e., for $E$ in $\mathcal{E}$ is such that $E$ contains either both $\omega_1$ and $\omega_2$ or none of them) then for $A = \{\omega_1\}$ we have $\mu^*(A) = f_1 + f_2$ and $\bar{\mu}(A) = f_1$.

Actually, Theorem 2.9 and Proposition 2.15 will prove that $\mu^*(A) = \bar{\mu}(A)$, for any $\mu^*$-measurable set $A$, which includes the $\sigma$-algebra generated by $\mathcal{E}$.

\end{proof}

\begin{exercise}
Denote by $\mu^*(\cdot, \mathcal{E})$ the outer measure (2.1). Assume that $\mu$ is initially defined on $\mathcal{E}' \supset \mathcal{E}$. Verify (1) that $\mu^*(\cdot, \mathcal{E}') \leq \mu^*(\cdot, \mathcal{E})$. Next, prove (2) that if for every $\varepsilon > 0$ and $E'$ in $\mathcal{E}'$ there exists $E$ in $\mathcal{E}$ such that $E \supset E'$ and $\mu(E') \leq \mu(E) + \varepsilon$ then $\mu^*(\cdot, \mathcal{E}') = \mu^*(\cdot, \mathcal{E})$.

\end{exercise}

\begin{proof}
For (1) note that the infimum is taken on a larger family, thus $\mu^*(\cdot, \mathcal{E}') \leq \mu^*(\cdot, \mathcal{E})$ follows from the definition.

\[\square\]
To check (2), take $\varepsilon > 0$ and a sequence $\{E'_n : n \geq 1\} \subset \mathcal{E}'$ such that $A \subset \bigcup_n E'_n$. For each $E'_n$ there exists $E_n$ in $\mathcal{E}$ such that $E_n \supset E'_n$ and $\mu(E_n) \leq \varepsilon 2^{-n} + \mu(E'_n)$. Hence, $\bigcup_n E_n \supset A$ and

$$\sum_n \mu(E_n) \leq \sum_n \varepsilon 2^{-n} + \sum_n \mu(E'_n) = \varepsilon + \sum_n \mu(E'_n),$$

which implies that $\mu^*(\cdot, \mathcal{E}) \leq \mu^*(\cdot, \mathcal{E}')$. \hfill $\Box$

**Exercise 2.7.** Let $\mu^*$ be the outer measures induced by a set function $\mu : \mathcal{E} \to [0, +\infty]$, as in Proposition 2.6. Recall that $\mathcal{E}_\sigma$ denotes the class of countable unions of sets in $\mathcal{E}$, and also that $\mathcal{E}_{\sigma\delta}$ denotes the class of countable intersections of sets in $\mathcal{E}_\sigma$. Prove that (1) for every $A \subset \Omega$ and any $\varepsilon > 0$ there exists a set $E$ in $\mathcal{E}_\sigma$ such that $A \subset E$ and $\mu^*(E) \leq \mu(A) + \varepsilon$. Deduce that (2) for every $A \subset \Omega$ there exists a set $F$ in $\mathcal{E}_{\sigma\delta}$ such that $A \subset F$ and $\mu^*(F) = \mu^*(A)$.

**Proof.** From the definition

$$\mu^*(A) = \inf \left\{ \sum_{n=1}^\infty \mu(E_n) : E_n \in \mathcal{E}, A \subset \bigcup_{n=1}^\infty E_n \right\}, \quad \forall A \subset \Omega,$$

of infimum, for every $\varepsilon > 0$ there exists a sequence $\{E_n\} \subset \mathcal{E}$ such that $\sum_n \mu^*(E_n) \leq \mu(A) + \varepsilon$. Since $E = \bigcup_n E_n$ belongs to $\mathcal{E}_\sigma$, $\mu^*(E_n) \leq \mu(E_n)$ and $\mu^*$ is sub $\sigma$-additive, we obtain

$$\mu^*(E) \leq \sum_n \mu^*(E_n) \leq \sum_n \mu(E_n) \leq \mu^*(A) + \varepsilon,$$

proving (1).

Now, for every $\varepsilon = 1/k$, $k = 1, 2, \ldots$, we use part (1) to find a set $E_k$ in $\mathcal{E}_\sigma$ such that $\mu^*(E_k) \leq \mu(A) + 1/k$. Hence, the set $F = \bigcap_k E_k$ belongs to $\mathcal{E}_{\sigma\delta}$, $A \subset F$ and

$$\mu^*(A) \leq \mu^*(F) \leq \mu^*(E_k) \leq \mu(A) + \frac{1}{k}, \quad \forall k = 1, 2, \ldots,$$

which yields (2) as $k \to \infty$. \hfill $\Box$

**Exercise 2.8.** Let $\mu^*_i (i = 1, 2)$ be the outer measures induced by the initial set functions $\mu_i : \mathcal{E} \to [0, +\infty]$, as in Proposition 2.6, and assume that the class $\mathcal{E}$ is stable under the formation of finite unions and finite intersections, and that every set in $\mathcal{E}$ is $\mu^*_i$-measurable (for $i = 1, 2$). Prove that if $\mu^*_1(E) \leq \mu^*_2(E)$ for every $E$ in $\mathcal{E}$ then $\mu^*_1(A) \leq \mu^*_2(A)$, for every $A \subset \Omega$.

**Proof.** First, recall Exercise 1.6. If $E = \bigcup_i E_i$ and $F = \bigcup_i F_i$ with $E_i$ and $F_i$ in $\mathcal{E}$ then $E \cap F = \bigcup_{i,j} (E_i \cap F_j)$, and because $\mathcal{E}$ is stable under the formation of finite intersections, the set $E_i \cap F_j$ is also in $\mathcal{E}$. Hence, $\mathcal{E}_\sigma$ is stable under the formation of finite intersections.

Recall that $\mathcal{E}_\sigma$ is stable under the formation of countable unions and $\mathcal{E}_{\sigma\delta}$ is stable under the formation of countable intersections.
Now, in view of Exercise 2.7, part (2), for every $A \subset \Omega$ there exists a set $F_i$ in $\mathcal{E}_{\sigma \delta}$ such that $A \subset F_i$ and $\mu^*_1(F_i) = \mu^*_1(A)$. The set $F = F_1 \cap F_2$ is also in $\mathcal{E}_{\sigma \delta}$, $A \subset F$ and $\mu^*_1(F) = \mu^*_1(A)$. Thus, we are reduced to show that $\mu^*_1(A) \leq \mu^*_2(A)$, for every $F$ in $\mathcal{E}_{\sigma \delta}$.

Since $\mathcal{E}$ is stable under the formation of finite unions, for every $E$ in $\mathcal{E}_\delta$ there exists a monotone increasing sequence $\{E_n\} \subset \mathcal{E}$ such that $E = \bigcup_n E_n \equiv \lim_n E_n$. In view of the $\sigma$-additive of $\mu^*_1$ on the $\mu^*_1$-measurable sets $E_n$ and $E$, the monotone continuity from below shows that

$$\mu^*_1(E) = \lim_n \mu^*_1(E_n) \leq \lim_n \mu^*_2(E_n) = \mu^*_2(E),$$

i.e., $\mu^*_1 \leq \mu^*_2$ on the class $\mathcal{E}_\sigma$.

Because $\mathcal{E}_\sigma$ is stable under the formation of finite intersections, for any $F$ in $\mathcal{E}_{\sigma \delta}$ there exists a monotone decreasing sequence $\{F_n\} \subset \mathcal{E}_\sigma$ such that $F = \bigcap_n F_n = \lim_n E_n$. Hence, almost the same argument as above yields

$$\mu^*_1(F) = \lim_n \mu^*_1(F_n) \leq \lim_n \mu^*_2(F_n) = \mu^*_2(F).$$

Indeed, we deduce from Exercise 2.7, part (1), that for every $F$ with $\mu^*_2(F) < \infty$ there exists a set $F''$ in $\mathcal{E}_\sigma$ such that $F \subset F''$ and $\mu^*_2(F'') < \infty$. Each set $F_n = F_n \cap F''$ is in $\mathcal{E}_\sigma$,

$$\mu_1(F_n) \leq \mu_2(F_n) \leq \mu_2(F'') < \infty, \quad \forall n,$$

and the monotone decreasing sequence $\{F_n\}$ satisfies $\lim_n F_n = F$. Hence, the monotone continuity from above can be applied to deduce that $\mu^*_1 \leq \mu^*_2$ on the class $\mathcal{E}_{\sigma \delta}$, which completes the proof.

**Exercise 2.9.** Let $\mu_i \ (i \geq 1)$ be initial set function $\mu_i : \mathcal{E} \to [0, +\infty]$ as in Proposition 2.6, where $\mathcal{E}$ is now a ring. Assume that all set initial functions are additive, and all but one are $\sigma$-additive, i.e., $\mu_1$ is additive and $\mu_i \ (i \geq 2)$ are $\sigma$-additive. Prove that $(\sum_i \mu_i)^* = \sum_i \mu_i^*$.

**Proof.** Indeed, the inequality $(\sum_i \mu_i)^* \geq \sum_i \mu_i^*$ follows from the definition of the infimum, so that only the converse inequality $(\sum_i \mu_i)^* \leq \sum_i \mu_i^*$ needs a discussion. Moreover, since $\mathcal{E}$ is a ring and all $\mu_i$ are additive, by Remark 2.8, the infimum defining $\mu_i$ implies that only disjoint covering should be considered. Thus, for any set $E$ in $\mathcal{E}$ with $\mu^*_1(E) < \infty$ and any $\varepsilon > 0$ there exists a sequence $\{E_n\}$ of disjoint sets in $\mathcal{E}$ such that $E = \sum_n E_n$ and $\sum_n \mu_1(E_n) \leq \mu^*_1(E) + \varepsilon$. Since $\mu_i \ (i \geq 2)$ is $\sigma$-additive we have $\sum_n \mu_i(E_n) = \mu_i(E) = \mu_i^*(E)$.

$$(\sum_i \mu_i)^*(E) \leq \sum_n \left[ \sum_i \mu_i(E_n) \right] \leq \varepsilon + \sum_i \mu_i^*(E),$$

which shows that $(\sum_i \mu_i)^*(E) \leq \sum_i \mu_i^*(E)$, for every $E$ in $\mathcal{E}$.

As in Exercise 2.8, by the monotone continuity of measures, we deduce that $(\sum_i \mu_i)^*(F) \leq \sum_i \mu_i^*(F)$, for every $F$ in $\mathcal{E}_{\sigma \delta}$.  

Now, by means of Exercise 2.7, part (2), for every $A \subset \Omega$ there exist sets $F_i$, $i = 0, 1, \ldots$, in $\mathcal{E}_{\sigma \delta}$ such that $A \subset F_i$, $(\sum_{n} \mu_i) (A) = (\sum_{n} \mu_i) (F_0)$ and $\mu^*_i (A) = \mu^*_i (F_i)$. The set $F = \bigcup_{i \geq 0} F_i$ also belongs to $\mathcal{E}_{\sigma \delta}$, and we have $A \subset F$, $(\sum_{n} \mu_i) (A) = (\sum_{n} \mu_i) (F)$ and $\mu^*_i (A) = \mu^*_i (F)$. Hence, $(\sum_{n} \mu_i) (A) \leq \sum_{n} \mu^*_i (A)$, for every $A \subset \Omega$, and we conclude the proof.

Exercise 2.10. Let $\mu^*$ be the outer measure on $\Omega$ induced by an additive set function $\mu$ defined on an algebra (actually, a semi-ring suffices) $A = \mathcal{E}$, given by (2.1). Denote by $A_\sigma$ the class of countable unions of sets in $A$, and by $A_{\sigma \delta}$ the class of countable intersections of sets in $A_\sigma$. (1) Prove that for every subset $E \subset \Omega$ and every $\epsilon > 0$ there exists a set $A$ in $A_\sigma$ such that $E \subset A$ and $\mu^*(A) \leq \mu^*(E) + \epsilon$, and (2) deduce that for some $B$ in $A_{\sigma \delta}$ such that $E \subset B$ and $\mu^*(E) = \mu^*(B)$.

Now, by means of Exercise 2.7, part (2), for every $B$ there exist a sequence $\{\mathcal{E}_i\}$ such that $E \subset \mathcal{E}_i$ and $\mu^*(E_\epsilon) < \infty$, for every $i$. (3) Show that a $\sigma$-finite set $E$ is $\mu^*$-measurable if and only if there exists $B$ in $A_{\sigma \delta}$ such that $E \subset B$ and $\mu^*(B \cap E) = 0$. Finally, if $\mu^*(\Omega) < \infty$ then (4) prove that $E$ is $\mu^*$-measurable if and only if $\mu^*(E) = \mu(\Omega) - \mu^*(\Omega \setminus E)$.

Proof. Note that (1) and (2) are particular cases of Exercise 2.7, and essentially, we repeat the same arguments.

(1) By the definition

$$
\mu^*(E) = \inf \left\{ \sum_{n=1}^{\infty} \mu(E_n) : E_n \in \mathcal{E}, E \subset \bigcup_{n=1}^{\infty} E_n \right\}, \quad \forall E \subset \Omega,
$$

for every $\epsilon > 0$ there exist a sequence $\{E_n\} \subset \mathcal{E}$ such that $E \subset \bigcup_{n=1}^{\infty} E_n$ and $\sum_{n=1}^{\infty} \mu(E_n) \leq \mu^*(E) + \epsilon$. Moreover, $\mu^*(E) \leq \mu(E)$, for every $E \in \mathcal{E}$. Because $\mathcal{E}$ is an algebra, in the proof of Theorem 2.9, we have seen that if $\mu$ is additive (not necessarily $\sigma$-additive) then the elements of $\mathcal{E}$ are $\mu^*$-measurable, which implies that every sets in the $\sigma$-algebra $\sigma(\mathcal{E})$ generated by $\mathcal{E}$ is $\mu^*$-measurable. Furthermore, the outer measure $\mu^*$ restricted to the $\sigma(\mathcal{E})$ is a measure and (if $\mu$ is $\sigma$-additive then $\mu^*$ agrees with $\mu$ on $\mathcal{E}$, which is not needed!). Hence, the set $A = \bigcup_n E_n$ belongs to $A_\sigma \subset \sigma(\mathcal{E})$, $E \subset A$ and

$$
\mu^*(A) \leq \sum_{n=1}^{\infty} \mu^*(E_n) \leq \sum_{n=1}^{\infty} \mu(E_n) \leq \mu^*(E) + \epsilon.
$$

Moreover, if $\mu$ is $\sigma$-additive and $A$ also belongs to $\mathcal{E}$ then $\mu^*(A) = \mu(A)$.

(2) Now, take $\epsilon = 1/n$ and apply (1) the construct a sequence $\{A_n\} \subset A_\sigma$ such that $E \subset A_n$ and $\mu^*(A_n) \leq \mu^*(E) + 1/n$. Therefore, $B = \bigcap_n A_n \supset A$ and $\mu^*(B) \leq \mu^*(A_n) \leq \mu^*(E) + 1/n$, i.e., $\mu^*(B) = \mu^*(E)$.

(3) From the definition, we deduce that any subset $N \subset \Omega$ with $\mu^*(N) = 0$ is $\mu^*$-measurable, and because $\mu$ is additive on the algebra $\mathcal{E}$, any set in $\sigma(\mathcal{E})$ is $\mu^*$-measurable. Thus, given any subset $E \subset \Omega$, if there exists $B$ in $A_{\sigma \delta} \subset \sigma(\mathcal{E})$ such that $E \subset B$ and $\mu^*(B \cap E) = 0$ then $B$ and $N = B \setminus E$ are $\mu^*$-measurable, and so is $E = B \setminus N$. 

Conversely, if $E$ is $\sigma$-finite then apply (2) (or (1)) to the initial sequence \(F_i \subset 2^\Omega\) we obtain a sequence \(\{E_n\} \subset \mathcal{A}_\delta\) (or \(\{E_n\} \subset \mathcal{A}_\sigma\)) such that \(E \subset \bigcup_n E_n\) and \(\mu^*(E_n) < \infty\), for every \(n \geq 1\). Now, apply (1) to the set \(E \cap E\) to obtain a set \(A_{n,k}\) in \(\mathcal{A}_\sigma\) with \(E \cap E \subset A_{n,k}\) and \(\mu^*(A_{n,k}) \leq \mu^*(E) + 2^{-n}/k\), \(k, n \geq 1\). Since \(\mu^*(E \cap E) < \infty\) and
\[
\mu^*(A_{n,k}) = \mu^*(E \cap E) + \mu^*(A_{n,k} \setminus (E \cap E)),
\]
we have
\[
\mu^*(A_{n,k} \setminus E) \leq \mu^*(A_{n,k} \setminus (E \cap E)) \leq 2^{-n}/k, \quad \forall k, n \geq 1.
\]
Therefore, the set \(A_k = \bigcap_n A_{n,k}\) belongs to \(\mathcal{A}_\sigma\), \(E \subset A_k\) and
\[
\mu^*(A_k \setminus E) \leq \sum_n \mu^*(A_{n,k} \setminus (E \cap E)) \leq \sum_n \frac{2^{-n}}{k} \leq \frac{1}{k}.
\]
Hence, if \(B = \bigcap_k A_k\) then \(B\) is in \(\mathcal{A}_\sigma\), \(E \subset B\) and \(\mu^*(B \setminus E) = 0\).

(4) Since \(\mu^*(\Omega) < \infty\), if \(E\) is \(\mu^*\)-measurable then the additivity of \(\mu^*\) on measurable sets implies that \(\mu^*(E) = \mu(\Omega) - \mu^*(\Omega \setminus E)\).

For the converse, apply (2) to \(E\) and \(E^c\) to obtain two sets \(A\) and \(B\) in \(\mathcal{A}_\sigma\) such that \(E \subset A, E^c \subset B, \mu^*(E) = \mu^*(A)\), and \(\mu^*(E^c) = \mu^*(B)\). Because \(\Omega = A \cup B\) and \(\mu^*\) is additive, we obtain \(\mu^*(\Omega) + 2\mu^*(A \cap B) = \mu^*(A) + \mu(B)\). However, by assumption, we have
\[
\mu^*(A) = \mu^*(E) = \mu(\Omega) - \mu^*(E^c) = \mu(\Omega) - \mu^*(B),
\]
and then we deduce \(\mu^*(A \cap B) = 0\). Since \(A \setminus E = A \cap E^c \subset A \cap B\), the set \(N = A \setminus E\) has outer measure zero, and so, it is \(\mu^*\)-measurable. Hence \(E = A \setminus N\) is also measurable.

Note that only the fact that the initial set function \(\mu\) is additive on a semiring \(\mathcal{E}\) is actually necessary in the above arguments for (1),\ldots,(4).

\[\Box\]

**Exercise 2.11.** Let \(\mu^*\) be an outer measure on \(\Omega\) and \(\{\Omega_i : i \geq 1\}\) be a sequence of disjoint \(\mu^*\)-measurable sets. Prove that \(\mu^*(E \cap (\bigcup_i \Omega_i)) = \sum_i \mu^*(E \cap \Omega_i)\), for every \(E \in 2^\Omega\).

**Proof.** In view of the \(\sigma\)-additivity of \(\mu^*\) we have \(\mu^*(E \cap (\bigcup_i \Omega_i)) \leq \sum_i \mu^*(E \cap \Omega_i)\), and to verify the converse inequality, we need show only that
\[
\sum_{i=1}^n \mu^*(E \cap \Omega_i) = \mu^*(E \cap (\bigcup_{i \leq n} \Omega_i)) \leq \mu^*(E \cap (\bigcup_i \Omega_i)), \quad \forall n \geq 1,
\]
which can be proved by induction as follows.

Apply the \(\mu^*\)-measurability (Caratheodory’s) condition to the sets \(\Omega_n\) with the generic set \(E_n = E \cap (\bigcup_{i \leq n} \Omega_i)\) to obtain \(\mu^*(E_n) = \mu^*(E \cap \Omega_n) + \mu^*(E \cap \Omega_n^c)\). Since \(E_n \cap \Omega_n = E \cap \Omega_n\) and \(E_n \cap \Omega_n^c = E_{n-1}\) we deduce
\[
\mu^*(E \cap (\bigcup_{i \leq n} \Omega_i)) = \mu^*(E_n) = \mu^*(E \cap \Omega_n) + \mu^*(E_{n-1}) = \\
= \mu^*(E \cap \Omega_n) + \mu^*(E \cap \Omega_{n-1}) + \mu^*(E_{n-2}) = \ldots = \\
= \mu^*(E \cap \Omega_n) + \mu^*(E \cap \Omega_{n-1}) + \cdots + \mu^*(E \cap \Omega_1),
\]
Exercise 2.12. Revise the definitions and proofs of this section to use outer measures defined on hereditary $\sigma$-rings. The class of $\mu^*$-measurable set becomes a hereditary $\sigma$-ring in Theorem 2.5. The class $\mathcal{E}$ needs not to cover the whole space $\Omega$ in Propositions 2.6. Theorem 2.9 is practically undisturbed if $\sigma$-ring is replaced by $\sigma$-ring, e.g., see Halmos [57, section 10, pp 41-48]. Referring to Propositions 2.6, as a typical application, consider the hereditary $\sigma$-ring of all set covered by a sequence of sets in $\mathcal{E}$ with $\mu$-finite value (i.e., $\sigma$-finite sets relative to $\mathcal{E}$), to construct $\sigma$-finite measures defined on $\sigma$-rings. Consider the alternative way of beginning with $\mu$ define on a class $\mathcal{E}$ such that $\Omega$ belongs to $\mathcal{E}$ and $\mu(E) < \infty$ for every $E \neq \Omega$.

Proof. A class $\mathcal{C}$ is called hereditary, if every subset of a set in $\mathcal{C}$ also belongs to $\mathcal{C}$, i.e., $A \subset B$ with $B \in \mathcal{C}$ implies $A \in \mathcal{C}$. Certainly, for the trivial case when the whole space $\Omega$ belongs to $\mathcal{C}$ we have $\mathcal{C} = 2^\Omega$. For instance, if $\mathcal{I}$ is the class of all bounded $d$-intervals in $\mathbb{R}^d$ then the hereditary class generated by $\mathcal{I}$ is the class of all bounded sets. In general, if $\mathcal{R}_0$ is the $\sigma$-ring generated by a class $\mathcal{E}$, then the hereditary class generated by $\mathcal{R}_0$ is the class $\mathcal{R}$ of all $\sigma$-covered sets relative to $\mathcal{E}$, i.e., the class of all sets $A$ such that there exits a sequence $\{E_n\} \subset \mathcal{E}$ such that $A \subset \bigcup_n E_n$. This class $\mathcal{R} = \mathcal{R}_H(\mathcal{E})$ is a $\sigma$-ring with the hereditary property, so-called hereditary $\sigma$-ring. Certainly, if there exits a sequence $\{E_n\}$ in $\mathcal{E}$ such that $\Omega = \bigcup_n E_n$ then $\mathcal{R}_0$ is the $\sigma$-algebra generated by $\mathcal{E}$ and $\mathcal{R}_H(\mathcal{E}) = 2^\Omega$.

Referring to Propositions 2.6, we may use finite set functions, i.e., to assume that $\mu: \mathcal{E} \to [0, +\infty)$ be such that $\emptyset \in \mathcal{E}$ and $\mu(\emptyset) = 0$. Then $\mu^*(A)$ is defined, a priori, for any set $A$ in the hereditary $\sigma$-ring $\mathcal{R}_H(\mathcal{E})$ generated by $\mathcal{E}$. Sometime, we may want to define $\mu^*(A) = +\infty$ when $A$ cannot be covered by a sequence of sets in $\mathcal{E}$, i.e., effectively defining $\mu^*$ on the whole class $2^\Omega$. Alternatively, we may allow $\mu$ to take the $+\infty$ value on some sets in the initial class $\mathcal{E}$, and pay attention to the hereditary $\sigma$-ring of all $\sigma$-finite sets relative to $(\mathcal{E}, \mu)$, i.e., the class of all sets $A \subset \bigcup_n E_n$, for some sequence $\{E_n\} \subset \mathcal{E}$ with $\mu(E_n) < \infty$.

The condition of $\mu^*$-measurability or Carathéodory’s condition for a set $A \in \mathcal{R}_H(\mathcal{E})$ can be re-written as $\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \setminus A)$, for every $E \in \mathcal{R}_H(\mathcal{E})$. Thus the set of all $\mu^*$-measurable sets form a $\sigma$-ring and measures are considered defined on $\sigma$-rings, but outer measure are defined on hereditary $\sigma$-rings. Theorem 2.9 holds as follows.

If $\mu$ is a measure on a ring $\mathcal{E}$ and $\mu^*$ is defined as above then (a) the restriction $\mu^*|\mathcal{E} = \mu$ and (b) every set in the $\sigma$-ring $\mathcal{R}$ generated by $\mathcal{E}$ is $\mu^*$-measurable and $\bar{\mu} = \mu^*|\mathcal{R}$ is a measure. Moreover, if $\bar{\mu}$ is $\sigma$-finite (i.e., for any $A$ in $\mathcal{R}$ there exists a sequence $\{A_n\} \subset \mathcal{R}$ such that $\bigcup_{n=1}^{\infty} A_n = A$ with $\bar{\mu}(A_n) < \infty$) then $\bar{\mu}$ is uniquely determinate on $\mathcal{R}$, i.e., if $\nu$ is another measure on $\mathcal{R}$ such that $\nu|\mathcal{E} = \mu$ then $\nu = \bar{\mu}$.

Perhaps, the relevance of studying measures on $\sigma$-rings (instead of measures on $\sigma$-algebras) resides on the importance of the $\sigma$-ring of $\sigma$-finite sets. As discussed later, we will be able to integrate (with a finite value) only those functions that vanish outsize of a $\sigma$-finite set. It is also clear that hereditary $\sigma$-rings play the role of the whole class $2^\Omega$. □
From a Semi-Ring

**Exercise 2.13.** Complete the proof of Proposition 2.11, i.e., verify that the extension of \( \mu \) to \( \mathcal{R}_0 \) is well defined and indeed is \( \sigma \)-additive.

*Proof.* Indeed, if \( A \) belongs to the ring \( \mathcal{R}_0 \) generated by the semi-ring \( \mathcal{E} \) then \( A \) is a disjoint union of sets in \( \mathcal{E} \), but this union may not be unique, i.e., \( A = \bigcup_{i=1}^n E_i = \bigcup_{j=m}^n F_j \) with \( E_i \) and \( F_j \) in \( \mathcal{E} \), for every \( i, j \). Since \( E_i = \bigcup_{j=1}^m E_i \cap F_j \) and \( \mu \) is additive on \( \mathcal{E} \) we have \( \mu(E_i) = \sum_{j=1}^m \mu(E_i \cap F_j) \). By symmetry, we also \( \mu(F_j) = \sum_{i=1}^n \mu(F_j \cap E_i) \), and therefore

\[
\sum_{i=1}^n \mu(E_i) = \sum_{i=1}^m \left( \sum_{j=1}^n \mu(E_i \cap F_j) \right) = \sum_{j=1}^m \left( \sum_{i=1}^n \mu(F_j \cap E_i) \right) = \sum_{i=1}^m \mu(F_i),
\]

i.e., \( \mu \) is well defined on \( \mathcal{R}_0 \).

If \( \{A_n\} \) is a disjoint sequence of sets in the ring \( \mathcal{R}_0 \) such that \( \sum_n A_n = A \) belongs to \( \mathcal{R}_0 \) then each \( n \geq 1 \) we can write \( A_n = \bigcup_{i \in I_n} E_{n,i} \) and also \( A = \bigcup_{i \in I} E_i \), with \( E_{n,i} \) and \( E_i \) in \( \mathcal{E} \), for finite indexes \( I_n \) and \( I \). By definition \( \mu(A) = \sum_{i \in I} \mu(E_i) \), and because \( A_{n,i} = A \cap E_i = \bigcup_{j \in I_{n,i}} (E_{n,j} \cap E_i) \) belongs to \( \mathcal{E} \) and is expressed as countable disjoint union of sets in \( \mathcal{E} \), the \( \sigma \)-additivity of \( \mu \) on \( \mathcal{E} \) implies

\[
\mu(A_{n,i}) = \sum_{j \in I_{n,i}} \mu(E_{n,j} \cap E_i).
\]

Adding in \( i \in I \), and because \( E_{n,j} = E_{n,j} \cap A = \bigcup_{i \in I} E_{n,j} \cap E_i \), we obtain

\[
\mu(A) = \sum_{i \in I} \left( \sum_{j \in I_{n,i}} \mu(E_{n,j} \cap E_i) \right) = \sum_{j \in I_{n,i}} \left( \sum_{i \in I} \mu(E_{n,j} \cap E_i) \right) = \sum_{j \in I_{n,i}} \mu(E_{n,j}) = \sum_{n} \left( \sum_{j \in I_{n,i}} \mu(E_{n,j}) \right) = \sum_{n} \mu(A_n),
\]

which proves the \( \sigma \)-additivity of the extension of \( \mu \) to \( \mathcal{R}_0 \).

Certainly, if the initial set function \( \mu \) assume infinite values for some sets in \( \mathcal{E} \) then we may reduce to the finite case by considering \( \mu \) defined only on the semi-ring

\[
\mathcal{E}_0 = \{E \in \mathcal{E} : \mu(E) < \infty\}
\]

of finite sets. \( \square \)

**Exercise 2.14.** Let \((X, \mathcal{X}, \mu)\) be a \( \sigma \)-finite measure space and \((Y, \mathcal{Y})\) be a measurable space. By means of a monotone argument prove that the function \( y \mapsto \mu(A_y) \) is measurable from \((Y, \mathcal{Y})\) into the Borel space \([0, \infty]\), for every \( A \) in the product \( \sigma \)-algebra \( \mathcal{X} \times \mathcal{Y} \) with section \( A_y = \{x \in X : (x, y) \in A\} \). Hint: the equality \((A \cup B)_x = A_x \cup B_x\) could be of some use here.
Proof. To verify that the section operation preserves unions, let us compare the set \( B(y) = \bigcup_{i \in I} (A_i)_y \) and the set \( A_y \) with \( A = \bigcup_{i \in I} A_i \). A point \( x \) belongs to \( B(y) \) if and only if there exists \( i \) in \( I \) such that \( x \) belongs to \( (A_i)_y \), i.e., if and only if \((x, y)\) belongs to \( A_i \), for some \( i \) in \( I \). Since, a point \( x \) belongs to \( A_y \) if and only if there exits \( i \) in \( I \) such that \((x, y)\) belongs to \( A_i \), we deduce the equality \( \bigcup_{i \in I} (A_i)_y = \bigcup_{i \in I} A_i \). The same argument holds true for the intersections, i.e., we have \( \bigcap_{i \in I} (A_i)_y = \bigcap_{i \in I} A_i \). Similarly, sections are preserved by the formation of differences, indeed, \( x \) belongs to \( A_y \setminus B_y \) if and only if \( x \) belongs to \( A_y \) and does not belong to \( B_y \), i.e., if and only if \((x, y)\) belongs to \((A \setminus B)\). In particular, for the complement set operation, \((A^c)_y = (A_y)^c\).

Consider the class \( \mathcal{C} \) of sets \( A \) in product \( \sigma \)-algebra \( \mathcal{X} \times \mathcal{Y} \) with the property that all sections are measurable, i.e., \( A_y \) belongs to \( \mathcal{X} \), for every \( y \) in \( \mathcal{Y} \). In view of the above properties on sections, this class is a \( \sigma \)-algebra. Since \( y \)-sections of a rectangle set \( E \times F \), with \( E \) in \( \mathcal{X} \) and \( F \) in \( \mathcal{Y} \), are either \( \emptyset \) or \( E \), we deduce that \( \mathcal{C} \) contains all rectangle, i.e., we deduce that \( \mathcal{C} = \mathcal{X} \times \mathcal{Y} \). This means that all sections of sets in the product \( \sigma \)-algebra \( \mathcal{X} \times \mathcal{Y} \) are measurable sets.

Because \((X, \mathcal{X}, \mu)\) is a \( \sigma \)-finite measure space, there exists an increasing sequence \( \{X_k\} \subset \mathcal{X} \) with \( \mu(X_k) < \infty \), for every \( k \geq 1 \). Consider the class \( \mathcal{A} \) of sets \( A \) in \( \mathcal{X} \times \mathcal{Y} \) for which the function \( y \mapsto \mu(X_k \cap A_y) \) is measurable, for any \( k \geq 1 \). To verify that \( \mathcal{A} \) \( \lambda \)-class, i.e., it is stable under the formation of countable monotone unions, monotone differences and it contains the whole space \( X \times Y \), we recall that \( \sigma \)-additivity of \( \mu \) implies the monotone continuity from below and that because we are dealing with finite values, we have \( \mu(X_k \cap (A_y \setminus B_y)) = \mu(X_k \cap A_y) - \mu(X_k \cap B_y) \). Hence, a monotone argument (Proposition 1.6) shows that \( \mathcal{A} \) is indeed the whole product \( \sigma \)-algebra \( \mathcal{X} \times \mathcal{Y} \), i.e., each function \( y \mapsto \mu(X_k \cap A_y) \) is measurable, for any \( k \geq 1 \). Therefore, the limit \( y \mapsto \mu(A_y) = \lim_k \mu(X_k \cap A_y) \) is also measurable.

An alternative argument, is to use Exercise 1.7 to write any set \( A \) in the product \( \sigma \)-algebra \( \mathcal{X} \times \mathcal{Y} \) as a countable disjoint union of rectangles, i.e, \( A = \sum_n E_n \times F_n \) with \( E_n \) in \( \mathcal{X} \) and \( F_n \) in \( \mathcal{Y} \). Since the sections satisfy \( A_y = \bigcap_n (E_n \times F_n)_y \) and \( \mu \) is \( \sigma \)-additive we have

\[
\mu(A_y) = \sum_n \mu(E_n \times F_n)_y = \sum_n \mu(E_n) 1_{\{y \in F_n\}},
\]

which prove that the function \( y \mapsto \mu(A_y) \) is measurable, without assuming that \( \mu \) is \( \sigma \)-finite.

**Exercise 2.15.** First (1) show that any \( \sigma \)-finite measure is semi-finite, and give an example of a semi-finite measure which is not \( \sigma \)-finite. Next, (2) prove that if \( \mu \) is a semi-finite measure then for every measurable set \( A \) with \( \mu(A) = \infty \) and any real number \( c > 0 \) there exits a measurable set \( F \) such that \( F \subset A \) and \( c < \mu(F) < \infty \). Now, if \( \mu \) is a measure (non necessarily semi-finite) then the finite part \( \mu_f \) is defined by the expression

\[
\mu_f(A) = \sup\{\mu(F) : F \in \mathcal{A}, F \subset A, \mu(F) < \infty\}.
\]
Finally, (3) prove that $\mu_f$ is a semi-finite measure, and that if $\mu$ is semi-finite then $\mu = \mu_f$. Moreover, (4) show that there exists a measure $\nu$ (non necessarily unique) which assumes only the values 0 and $\infty$ such that $\mu = \mu_f + \nu$.

**Proof.** Recall that a measure $\mu$ on (a $\sigma$-algebra) $A$ is called semi-finite if for every $A$ in $A$ with $\mu(A) = \infty$ there exists $F$ in $A$ satisfying $F \subset A$ and $0 < \mu(F) < \infty$.

(1) If $\mu$ is a $\sigma$-finite measure then there exists an increasing sequence $\{F_n\}$ of measurable sets such that $\Omega = \bigcup_n F_n$ and $\mu(F_n) < \infty$. Now, for any set measurable $A$ with $\mu(A) = \infty$, the $\sigma$-additivity (in this case, the monotone continuity from below) implies that $\mu(A) = \lim_n \mu(A \cap F_n)$, and thus, there exists a set $F = A \cap F_n$ such that $F \subset A$ and $0 < \mu(F) < \infty$.

If we take $\Omega = \mathbb{R}$, $A = 2^\Omega$ and $\mu(A) = 0$ the number of points in $A$ then $\mu$ the only sets with finite measure are the set with a finite number of elements. Thus, $\mu$ is not $\sigma$-finite because $\mathbb{R}$ is not countable. However, $\mu$ is semi-finite. Indeed, any $A$ with $\mu(A) = \infty$ is necessary nonempty, and therefore, there exist a point $a$ in $A$. Thus, we can take $F = \{a\} \subset A$ and $\mu(F) = 1$.

(2) Let us assume that $\mu$ is semi-finite and show that if $A$ is measurable and $\mu(A) = \infty$ then there exists a sequence $\{F_n\}$ of subsets of $A$ with $\mu(F_n) < \infty$, for every $n \geq 1$, and such that $\mu(F_n) \to \infty$, i.e., the supremum $\alpha = \sup \{\mu(F) : F \subset A, \mu(F) < \infty\}$ is infinite. Indeed, because $\mu$ is semi-finite we have $\alpha > 0$ and thus, there exists a sequence $\{F_n\}$, $F_n \subset A$, with $0 < \mu(F_n) \to \alpha$. The monotone continuity from below implies that $\mu(F) = \alpha$, with $F = \bigcup_n F_n$. Now, to see that $\mu(F) = \infty$ we argue by contradiction, if $\mu(F) < \infty$ then $\mu(A \setminus F) = \mu(A) - \mu(F) = \infty$ and therefore, again, because $\mu$ is semi-finite, there exists a set $E \subset A \setminus F$ with $\mu(E) > 0$, and necessarily $\mu(E \cup F) \leq \alpha$. Since $\mu(E \cup F) = \mu(E) + \mu(F)$ and $\mu(F) = \alpha$, we have the contradiction $\mu(E) = 0$.

(3) Remark that the definition makes sense even if $\mu$ is not semi-finite, we always have $\mu(\emptyset) = 0$. From the definition we can see that $\mu_f(A) = \mu(A)$ for any measurable set $A$ with $\mu(A) < \infty$. Thus $\mu_f(A) \leq \mu(A)$ for every measurable set $A$, and certainly $\mu_f$ is monotone, i.e., $A \subset B$ implies $\mu_f(A) \leq \mu_f(B)$, and $\mu_f(\emptyset) = 0$.

To verify that $\mu_f$ is $\sigma$-additive, take a disjoint sequence $\{A_n : n \geq 1\}$ of measurable sets, $A = \sum_n A_n$. By the definition of sup, for every $c < \mu_f(A)$ there exists a measurable set $F_c \subset A$ such that $c < \mu(F_c) < \infty$. Since

$$\mu_f(F_c) = \mu(F_c) = \sum_n \mu(F_c \cap A_n) = \sum_n \mu_f(F_c \cap A_n) \leq \sum_n \mu_f(A_n),$$

we deduce that $\mu_f(A) \leq \sum_n \mu_f(A_n)$. For the converse inequality, we need to consider only the case when $\mu_f(A) < \infty$, which implies that all $\mu_f(A_n) < \infty$, and thus, for every $\varepsilon > 0$ and any $n \geq 1$ there exits a measurable set $F_n \subset A_n$ such that $\mu_f(A_n) - \varepsilon 2^{-n} \leq \mu(F_n)$. Because the series converges, adding in $n \geq 1$ we obtain

$$\sum_n \mu_f(A_n) \leq \varepsilon + \sum_n \mu(F_n) \leq 2\varepsilon + \sum_{n \leq k} \mu(F_n),$$
for some \( k \) sufficiently large. For this \( k \), the additivity property of \( \mu \) implies that

\[
\mu \left( \bigcup_{n \leq k} F_n \right) = \sum_{n \leq k} \mu(F_n) < \infty.
\]

Hence, the definition of \( \mu_f \) yields \( \mu \left( \bigcup_{n \leq k} F_n \right) \leq \mu_f(A) \), and as \( \varepsilon \to 0 \), we deduce \( \sum_n \mu_f(A_n) \leq \mu_f(A) \).

Next, to check that \( \mu_f \) is semi-finite, take a measurable set \( A \) with \( \mu_f(A) = \infty \), and by the definition of sup, for every \( c > 0 \) there exists \( F_c \subset A \) such that \( c \leq \mu_f(F_c) < \infty \), and because \( \mu_f(F_c) = \mu(F_c) \) we deduce that \( \mu_f \) is semi-finite.

Similarly, if \( \mu \) is semi-finite and \( \mu(A) = \infty \) then by (2), there exists a sequence \( \{F_n\} \) of subsets of \( A \) with \( \mu(F_n) < \infty \), for every \( n \geq 1 \), and such that \( \mu(F_n) \to \infty \). Hence \( \mu_f(A) = \infty \), i.e., we deduce that \( \mu = \mu_f \).

(4) Note that \( \mu_f \leq \mu \) and if \( \mu_f(A) < \mu(A) = +\infty \) then \( \mu_f(A) = 0 \). Let \( \mathcal{R} \) be the \( \sigma \)-rings class of all \( \sigma \)-finite sets relative to \( \mu \), i.e., \( R \) belongs to \( \mathcal{R} \) if and only there exits a sequence \( \{R_n\} \) of measurable sets with \( \mu(R_n) < \infty \), for every \( n \geq 1 \) and \( R = \sum_n R_n \). If \( R \) belongs to \( \mathcal{R} \) then \( \mu(R) = \mu_f(R) \), and \( \mu(A) = \infty \) for every \( A \) which is not in \( \mathcal{R} \).

Define the set function \( \nu(A) = 0 \) if \( A \) belongs to \( \mathcal{R} \) and \( \nu(A) = \infty \) otherwise. It is clear that \( \mu = \mu_f + \nu \), and to check that \( \nu \) is \( \sigma \)-additive, let \( \{A_n : n \geq 1\} \) be a disjoint sequence of measurable sets, \( A = \sum_n A_n \). If each \( A_n \) is in \( \mathcal{R} \) then also \( A \) belongs to \( \mathcal{R} \), and \( 0 = \nu(A) = \sum_n \nu(A_n) \). Otherwise, there is some \( A_n \) which is not in \( \mathcal{R} \), and therefore, \( A \) cannot belongs to \( \mathcal{R} \), i.e., \( \nu(A) = \infty = \sum_n \nu(A_n) \).

If there is a measurable set \( E \) which is not in \( \mathcal{R} \) and \( \mu(E) = \mu_f(E) \) then we can define another set function \( \tilde{\nu}(A) = 0 \) if \( A \) belongs to \( \mathcal{R} \) and \( \tilde{\nu}(A) = \infty \) otherwise, where \( \mathcal{R} \) is the \( \sigma \)-ring generated by \( E \) and \( \mathcal{R} \). Since \( 0 = \tilde{\nu}(E) \neq \nu(E) = \infty \), we have verify that the measure \( \nu \) satisfying \( \mu = \mu_f + \nu \) is not necessarily unique.

(2.3) Inner Approach

**Exercise 2.16.** With the notation of in Theorem 3.19, verify that the class \( \overline{\mathcal{K}} \) of all finite unions of sets in \( \mathcal{K} \) is a lattice (i.e., a \( \pi \)-class stable under the formation of finite unions), and that the expression \( \mu(K \cup K') = \mu(K) + \mu(K') - \mu(K \cap K') \) defines, by induction, a unique extension of \( \mu \) to the class \( \overline{\mathcal{K}} \), which results \( \overline{\mathcal{K}} \)-tight. At this point, instead of (3.8), we could redefine \( \mu_* \) as

\[
\mu_*(A) = \sup \{ \mu(K) : K \subset A, K \in \overline{\mathcal{K}} \},
\]

where \( \mu \) is a \( \overline{\mathcal{K}} \)-tight finite-valued set function with \( \mu(\emptyset) = 0 \) defined on the lattice \( \overline{\mathcal{K}} \), i.e., without any loss of generality we could assume, initially, that the class \( \mathcal{K} \) is a lattice.

**Proof.** By definition, the class \( \overline{\mathcal{K}} \) is stable under finite unions, and the distribute formula, \( \left( \bigcup_{i \leq n} K_i \right) \cap \left( \bigcup_{j \leq m} K_j' \right) = \bigcup_{i \leq n} \bigcup_{j \leq m} (K_i \cap K_j') \), shows that \( \overline{\mathcal{K}} \) is stable under finite intersections.
By Theorem 3.19, \( \mu_* \) an additive measure on \( A \supseteq \mathcal{K} \) and thus \( \mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B) \) for any \( A \) and \( B \) in \( \mathcal{A} \) with \( \mu(A \cap B) < \infty \). In particular, because \( \mu = \mu_* \) (finite valued) on the \( \mathcal{K} \), we deduce that \( \mu \) can be uniquely extended to the lattice \( \bar{\mathcal{K}} \) by means of the expression \( \mu(K \cup K') = \mu(K) + \mu(K') - \mu(K \cap K') \), for every \( K \) and \( K' \) in \( \mathcal{K} \). The \( \bar{\mathcal{K}} \)-tightness of the extension \( \mu \) follows from the additivity property of \( \mu_* \).

Remark that the \( \bar{\mathcal{K}} \)-tightness condition on \( \mu \) means that for every \( \varepsilon > 0 \) and sets \( A \subset B \) in \( \bar{\mathcal{K}} \) there exists a set \( K \subset B \setminus A \) in \( \bar{\mathcal{K}} \) such that \( \mu(B) < \mu(A) + \mu(K) + \varepsilon \).

**Exercise 2.17.** Let \( \mathcal{K} \subset 2^\Omega \) be a lattice (i.e., a class containing the empty set and stable under finite unions and finite intersections) and \( \mu : \mathcal{K} \to [0, \infty) \) be a finite-valued set function with \( \mu(\emptyset) = 0 \). Consider

\[
\mu_*(A) = \sup \{ \mu(K) : K \subset A, K \in \mathcal{K} \}, \quad \forall A \subset \Omega,
\]

and assume that \( \mu \) is \( \mathcal{K} \)-tight. Show that (1) \( \mu \) can be uniquely extended to an (finitely) additive finite-valued set function \( \bar{\mu} \) defined on the ring \( \mathcal{R} \) generated by the lattice \( \mathcal{K} \). Next, prove that (2) if \( \mu \) is \( \sigma \)-smooth on \( \mathcal{K} \) (i.e., \( \lim_n \mu(K_n) = 0 \) whenever \( \{K_n\} \subset \mathcal{K} \) is a decreasing sequence with \( \bigcap_n K_n = \emptyset \)) then the extension \( \bar{\mu} \) is \( \sigma \)-additive on \( \mathcal{R} \).

**Proof.** This is essentially the arguments in Theorem 2.22, \( \mu_* \) is an additive set function on the algebra \( \mathcal{A} \), which is finite-valued on the lattice \( \mathcal{K} \), i.e., the ring \( \mathcal{R} \subset \mathcal{A} \) and \( \bar{\mu} = \mu_* \big|_{\mathcal{R}} \). Because any element in \( \mathcal{R} \) is contained in some set of \( \mathcal{K} \), the set function \( \bar{\mu} \) is finite-valued. Indeed, let \( \mathcal{R}_0 \) the class of all sets \( R \in \mathcal{R} \) such that there exists \( K \in \mathcal{K} \) satisfying \( R \subset K \). Because \( \mathcal{K} \) is stable under the formation of finite unions the class \( \mathcal{R}_0 \) is stable under the formation of differences and finite unions, i.e., \( \mathcal{R}_0 \) is a ring and therefore \( \mathcal{R}_0 \cap \mathcal{R} = \mathcal{R} \).

For any decreasing sequence \( \{A_n\} \subset \mathcal{R} \) with \( \bigcap_n A_n = \emptyset \), and any \( \varepsilon > 0 \) there exist sets \( K'_n \subset A_n \) in \( \mathcal{K} \) such that \( \bar{\mu}(A_n) = \mu_*(A_n) < \mu(K'_n) + \varepsilon 2^{-n} \). Consider the sequence \( \{K_n\} \) with \( K_n = \bigcap_{i \leq n} K'_i \) to obtain

\[
A_n \setminus K_n = A_n \setminus \bigcap_{i \leq n} K'_i = \bigcup_{i \leq n} (A_n \setminus K'_i) \subset \bigcup_{i \leq n} (A_i \setminus K'_i)
\]

Remark that \( A_n \) belongs to \( \mathcal{R} \subset A \), \( \mu_*(A_n) < \infty \) and \( \mu_*(A_n \setminus K'_n) < \varepsilon 2^{-n} \) to have

\[
0 \leq \mu_*(A_n) - \mu_*(K_n) = \mu_*(A_n \setminus K_n) \leq \sum_{i \leq n} \mu_*(A_i \setminus K'_i) < \varepsilon.
\]

This proves that \( \lim_n \bar{\mu}(A_n) = \lim_n \mu(K_n) \), and because \( \bigcap_n K_n \subset \bigcap_n A_n = \emptyset \) and \( \mu \) is \( \sigma \)-smooth on \( \mathcal{K} \), we deduce that \( \bar{\mu} \) is \( \sigma \)-smooth on \( \mathcal{R} \). \( \square \)

### (2.4) Geometric Construction

**Exercise 2.18.** Consider the space \( \mathbb{R}^d \) with the non-Euclidean distance \( d \) derived form norm \( |x| = \max\{|x_i| : i = 1, \ldots, d\} \). Let \( \mathcal{E} \) be the semi-ring of all
Denote by $\mu^*_r, \delta$ and $\mu^*_r$ the Hausdorff measure constructed from these $\mathcal{E}$ and $b_r$. Prove (1) that $\mu^*_r$ with $r = d$ is the Lebesgue outer measure. Next, consider the injection and projection mappings $i_r : x' \mapsto x = (x', 0)$ from $\mathbb{R}^r$ into $\mathbb{R}^d$ and $\pi_r : x \mapsto x'$ from $\mathbb{R}^d$ into $\mathbb{R}^r$, when $r = 1, \ldots, d - 1$. Show (2) that $\mu^*_r(A) = \ell_r^*(i_r(A))$, for every $A \subset i_r(\mathbb{R}^r)$, where $\ell_r$ is the Lebesgue outer measure. Also, show (3) that $\mu^*_r(A) = \infty$ if the projection $\pi_{r+1}(A)$ contains an open $(r+1)$-interval in $\mathbb{R}^{r+1}$, and thus, $\mu^*_r$ is not $\sigma$-finite in $\mathbb{R}^d$ for $r = 1, \ldots, d-1$, but only semi-finite. On the other hand, let $\mathcal{R}$ be the ring of all finite unions of semi-open cubes with edges parallel to the axis and with rational endpoints (i.e., $d$-intervals of the form $[a, b] \subset \mathbb{R}^d$ with $b_i, a_i$ rational numbers and $b_i - a_i = h$). Show (4) that

$$
\mu^*_{r, \delta}(A, \mathcal{R}) = \inf \left\{ \sum_{n=1}^{\infty} d^*(E_n) : E_n \in \mathcal{R}, A \subset \bigcup_{n=1}^{\infty} E_n, d(E_n) \leq \delta \right\},
$$

and $\mu^*_r(\cdot, \mathcal{R}) = \lim_{\delta \to 0} \mu^*_{r, \delta}(\cdot, \mathcal{R})$ satisfy also (1), (2) and (3) above.

Proof. (1) The point is that if every $E$ in $\mathcal{E}$ can be written as a countable disjoint union $\bigcup_{n=1}^{\infty} E_n$ of sets in $\mathcal{E}$ with $d(E_n) < \delta$, and if $b_r$ is $\sigma$-additive on $\mathcal{E}$, then $\mu^*_r, \delta = \mu^*_r$. In this case, $b_r$ is the hyper-volume for $r = d$ and the Lebesgue outer measure is constructed in $\mathbb{R}^d$, i.e., $\mu^*_d = \ell_d$.

(2) Suppose $r = 1, \ldots, d-1$, and $A \subset i(\mathbb{R}^r) \subset \mathbb{R}^d$. For any cover $\{E_n\} \subset \mathcal{E}$ of $A$ the projections of $E_n$ on $\mathbb{R}^r$ produce a cover $\{I_k\}$ of $r$-intervals of $i_r(A)$, which proves that $\mu^*_r(A) \geq \ell_r(i_r(A))$, for every $\delta > 0$. Thus, to verify the reverse inequality when $\mu^*_r(A) < \infty$ take $\varepsilon > 0$ to find a cover $\{I_k\}$ of $A$ with $r$-intervals such that $\ell_r^*(i_r(A)) + \varepsilon \geq \sum_k \ell_r(I_k)$, and hence, another sequence of disjoint $r$-intervals with $d(I_k) \leq \delta$ satisfying the same estimate can be obtain. Therefore, the sequence $\{E_k\}$ with $E_k = I_k \times [-\delta, 0]^d - r$ cover $A$ and $\ell_r^*(i_r(A)) + \varepsilon \geq \sum_k b_r(E_k)$, which yields $\ell_r^*(i_r(A)) \geq \mu^*_r, \delta(A)$, for every $\delta > 0$, i.e., $\ell_r^*(i_r(A)) \geq \mu^*_r, \delta(A) = \mu^*_r(A)$.

(3) Remark that if $E$ is a $d$-interval with $d(E) \leq \delta$ then $\mu^*_r(E) \geq \mu^*_r, \delta(E) = \ell_r^*(\pi_r(E))$, and therefore, $\mu^*_r(E) = \sum_n \mu^*_r, \delta(E_n)$ whenever $\pi_r(E) = \sum_n \pi_r(E_n)$. Thus, if $E$ is a set not necessarily in $\mathcal{E}$ such that $\pi_{r+1}(E) = [a', b']$ is a $(r+1)$-interval in $\mathbb{R}^{r+1}$ with $b_{r+1}' - a_{r+1}' = k \delta$ then any cover $\{E_n\} \subset \mathcal{E}$ of $E$ will produce at least $k$ disjoint covers $\{\pi_r(E_n) : n \in N_i\}$ ($i = 1, \ldots, k$) of $\pi_r(E)$, i.e., $\sum_n b_r(E_n) \geq k \ell_r(\pi_r(E))$. Hence,

$$
\mu^*_r(E) \geq (b_{r+1}' - a_{r+1}') \delta^{-1} \ell_r(\pi_r(E)),
$$

which implies $\mu^*_r(E) = \infty$.

Since each $\mu^*_r, \delta$ is a $\sigma$-finite outer measure, the limit $\mu^*_r$ is a semi-finite regular Borel outer measure. Only $\mu^*_d$ is a $\sigma$-finite Borel outer measure. Indeed,
any set containing an open set has necessarily an infinite measure (for any \( r = 1, \ldots, d - 1 \)), it seems reasonable that \( \mu^*_r \) is not \( \sigma \)-finite, but the actual proof requires more details. For instance, a variation of part (2) shows that the uncountable family of disjoint sets

\[
E_a = \{ x \in \mathbb{R}^d : 0 < x_i \leq 1, i = 1, \ldots, r, x_{r+1} = a, x_i = 0, i \geq r + 2 \}
\]
satisfies \( \mu^*_r(E_a) = 1 \). Hence, any set with finite measure can contain only finite many sets of the family \( \{ E_a : a \in \mathbb{R} \} \), which implies that any \( \sigma \)-finite set cannot contain the whole uncountable family of sets \( \{ E_a : a \in \mathbb{R} \} \), i.e., \( \mathbb{R}^d \) is not \( \sigma \)-finite.

(4) Observe (a) that \( b_r(E) = d(E)^r \) for any \( d \)-cube, (b) that any element in \( \mathcal{R} \) (i.e., any finite union of disjoint \( d \)-cubes) belongs also to the ring generated by the \( d \)-intervals in \( \mathcal{E} \), and (c) that any \( d \)-interval with rational endpoints can be written as a union of disjoint \( d \)-cubes to deduce that \( \mu^*_r(\mathcal{A}, \mathcal{R}) = \mu^*_r(\mathcal{A}) \) as in Exercise 2.27 part (2). It is clear then that \( \mu^*_r(\mathcal{A}, \mathcal{R}) \) satisfies all previous parts (1), (2) and (3), and \( \mu_r \leq \mu^*_r(\cdot, \mathcal{R}) \).

Certainly, we could continue this analysis and perhaps even call \( \mu^*_r \) or \( \mu^*_r(\cdot, \mathcal{R}) \) the Lebesgue-Hausdorff measure of dimension \( r \leq d \) on the space \( \mathbb{R}^d \), with the max-distance, but something is missing.

\[ \square \]

(2.5) Lebesgue Measures

**Exercise 2.19.** Consider the outer Lebesgue measure \( \ell^* \) on \( (\mathbb{R}^d, \mathcal{L}) \). First, (1) verify that any Borel set is measurable and that the boundary \( \partial I \) of any semi-open (semi-close) \( d \)-interval \( I \) in the semi-ring \( \mathcal{I}_d \) has Lebesgue measure zero. Second, (2) show that for any subset \( A \) of \( \mathbb{R}^d \) and any \( \varepsilon > 0 \) there is an open set \( O \) containing \( A \) such that \( \ell^*(A) + \varepsilon \geq \ell(O) \). Deduce that also there is a countable intersection of open sets \( G \) containing \( A \) such that \( \ell^*(A) = \ell(G) \).

**Proof.** (1) Since the Lebesgue measure \( \ell \) on \( (\mathbb{R}^d, \mathcal{L}) \) was defined from the hyper-volume set function \( m \) on the semi-ring \( \mathcal{I}_d \) of semi-open (semi-closed) \( d \)-intervals \( I = [a, b] \) with \( a \) and \( b \) in \( \mathbb{R}^d \), we deduce from Caratheodory’s construction (e.g., Proposition 2.11) that \( \sigma \)-algebra \( \mathcal{L} \) of \( \ell^* \)-measurable sets contains the \( \sigma \)-algebra of Borel sets in \( \mathbb{R}^d \), as being generated by the semi-ring \( \mathcal{I}_d \). The arguments in Proposition 2.26 also show that the boundary of any \( d \)-interval has measure zero. For instance, for every the \( \varepsilon > 0 \), set \( A = \{ a_1 \times \cdots \times [a_d, b_d] \} \) can be covered by \( I_{\varepsilon} = [a_1 - \varepsilon, a_1 + \varepsilon] \times \cdots \times [a_d, b_d] \times \cdots \times [a_d, b_d] \) and then \( \ell(I_{\varepsilon}) \leq 2 \varepsilon \prod_{i=2}^d |a_i, b_i| \), i.e., \( \ell(A) = 0 \). Hence, any \( d \)-interval (open, closed, non-open, non-closed) \( J \) has the same measure as a \( d \)-interval \( I \) in the semi-ring \( \mathcal{I}_d \), i.e., \( \ell(J) = \ell(I) = m(I) \), where \( I \) and \( J \) have the same interior points.

(2) Clearly, if \( \ell^*(A) = \infty \) then choose \( G = \mathbb{R}^d \). Thus take \( A \subset \mathbb{R}^d \), \( \ell(A) < \infty \). By the definition of \( \ell^* \), for any \( \varepsilon > 0 \) there exits a sequence \( \{ I_n \} \) of \( d \)-intervals in \( \mathcal{I}_d \) such that

\[
\ell^*(A) + \varepsilon/2 > \sum_n m(I_n) \quad \text{and} \quad O \subset \bigcup_n I_n.
\]
Now, for each $d$-interval $I_n$ and any $\varepsilon > 0$, we can modify the extremes and construct $d$-intervals open $J_n \supset I_n$ such that $\ell(J_n) \leq m(I_n) + \varepsilon/4$. Thus, the open set $O = \bigcup_n J_n$ satisfies $A \subset O$ and

$$\ell^*(A) + \varepsilon > \sum_n \ell(J_n) \geq \ell(O) \geq \ell^*(A).$$

Now, applying this argument for $\varepsilon = 1/k$, there is a sequence $\{O_n\}$ of open sets satisfying $A \subset O_k$ and $\ell(O_k) \leq \ell^*(A) + 1/k$, which yields $G = \bigcap_k O_k \supset A$ and $\ell(G) = \ell^*(A).$ \hfill \Box

**Exercise 2.20.** Consider the Lebesgue measure $\ell$ on $(\mathbb{R}^d, \mathcal{L})$. First, (1) show that for any measurable set $A$ with $\ell(A) < \infty$ and any $\varepsilon > 0$ there exits an open set $O$ with $\ell(O) < \infty$ and a compact set $K$ such that $K \subset A \subset O$ and $\ell(O \setminus K) < \varepsilon$. Next, (2) prove that for every measurable set $A \subset \mathbb{R}^d$ and any $\varepsilon > 0$ there exists a closed set $C$ and an open set $O$ such that $C \subset A \subset O$ and $\ell(O \setminus C) < \varepsilon$. Finally, if $\mathfrak{F}_\sigma$ denotes the class of countable unions of closed sets in $\mathbb{R}^d$ and $\mathfrak{G}_\delta$ denotes the class of countable intersections of open sets in $\mathbb{R}^d$ then (3) prove that for any measurable set $A$ there exists a set $G$ in $\mathfrak{G}_\delta$ and a set $F$ in $\mathfrak{F}_\sigma$ such that $F \subset A \subset G$ and $\ell(G \setminus F) = 0$.

**Proof.** (1) By the definition of $\ell$, for every $A$ in $\mathcal{L}$ with $\ell(A) < \infty$ and any $\varepsilon > 0$ there exits a sequence $\{I_n\}$ of $d$-intervals in $\mathcal{I}_d$ such that

$$\ell(A) + \varepsilon/4 > \sum_n m(I_n) \quad \text{and} \quad O \subset \bigcup_n I_n.$$

Now, for each $d$-interval $I_n$ and any $\varepsilon > 0$, we can modify the extremes and construct $d$-intervals open $J_n \supset I_n$ such that $\ell(J_n) \leq m(I_n) + \varepsilon/4$. Thus, the open set $O = \bigcup_n J_n$ satisfies $A \subset O$ and

$$\ell(A) + \varepsilon/2 > \sum_n \ell(J_n) \geq \ell(O),$$

which proves that $\ell(O) < \infty$ and $\ell(O \setminus A) < \varepsilon/2$.

To obtain a compact set, the monotone continuity from below shows that for every $\varepsilon > 0$ there is a compact $d$-interval $I$ such that $\ell(A \cap I) + \varepsilon/4 > \ell(A)$, i.e., $\ell(A \setminus I) < \varepsilon/4$. Hence, applying the previous argument to the set $I \setminus A$, we obtain an open set $U \supset I \setminus A$ such that $\ell(U \setminus (I \setminus A)) < \varepsilon/4$. Therefore, the compact set $K = I \setminus U$ satisfies $K \subset \bigcap A$, $(I \setminus A) \setminus K \subset U \setminus (I \setminus A)$, and

$$\ell((I \cap A) \setminus K) \leq \ell(U \setminus (I \setminus A)) < \varepsilon/4.$$ 

Finally, the inclusion $O \setminus K \subset (O \setminus A) \cup ((I \cap A) \setminus K) \cup (A \setminus I)$ yields

$$\ell(O \setminus K) \leq \ell(O \setminus A) + \ell((I \cap A) \setminus K) + \ell(A \setminus I) \leq \varepsilon,$$

with $K \subset A \subset O$. 

(2) Let \( A \) be a measurable set in \( \mathbb{R}^d \), and without assuming \( \ell(A) < \infty \), pick any \( \varepsilon > 0 \). Write \( A \) as a disjoint union of sets with finite measure (i.e., \( A = \bigcap_{k \geq 1} A_k \) with \( \ell(A_k) < \infty \)) to invoke part (2) with the set \( A_k \), and to find an open set \( O_k \) such that \( A_k \subset O_k \) and \( \ell(O_k \setminus A_k) \leq 2^{-k}\varepsilon/2 \). Since the set \( O = \bigcup_k O_k \) is also open, \( A \subset O \), \( O \setminus A \subset \bigcap_k (O_k \setminus A_k) \), and
\[
\ell(O \setminus A) \leq \sum_k \ell(O_k \setminus A_k) \leq \varepsilon/2.
\]
This shows that for any measurable set \( A \) in \( \mathbb{R}^d \) and for any \( \varepsilon > 0 \), there exist an open set \( O \) such that \( F \subset A \subset O \) and \( \ell(O \setminus A) < \varepsilon/2 \).

Now, applying this argument to the complement \( A^c = \Omega \setminus A \) there exists an open set \( U \) such that \( A^c \subset U \) and \( \ell(U \setminus A^c) < \varepsilon/2 \). Since the complement \( C = U^c \) is closed, \( A \supseteq U^c \) and \( A \cap C = U \setminus A^c \), we conclude that \( C \subset A \subset O \) and \( \ell(O \setminus C) < \varepsilon \).

(3) Next, for \( \varepsilon = 1/n \) we find a sequence \( \{O_n\} \) of open sets and a sequence \( \{C_n\} \) of closed sets such that \( C_n \subset A \subset O_n \) and \( \ell(O_n \setminus C_n) \leq 1/n \). Hence, the set \( G = \bigcap_n O_n \) is in \( \mathfrak{G}_\delta \) (usually called a \( \mathfrak{G}_\delta \)-set) and the set \( F = \bigcup_n C_n \) is a in \( \mathfrak{G}_\sigma \) (usually called a \( \mathfrak{G}_\sigma \)-set) satisfy \( F \subset A \subset G \) and \( \ell(G \setminus F) = 0 \). \( \square \)

**Exercise 2.21.** Consider the class \( \mathcal{I}_d \) of open bounded \( d \)-intervals in \( \mathbb{R}^d \) and the hyper-volume set function \( m = m_d \), i.e., of the form \( I = (a_1, b_1) \times \cdots \times (a_d, b_d) \), with \( a_i \leq b_i \) in \( \mathbb{R} \), \( i = 1, \ldots, d \), and \( m(I) = (b_1 - a_1) \cdots (b_d - a_d) \). Even if \( \mathcal{I}_d \) is not a semi-ring, we can define the outer measure
\[
m^* (A) = \inf \left\{ \sum_{n=1}^\infty m(I_n) : I_n \in \mathcal{I}_d, A \subset \bigcup_{n=1}^\infty I_n \right\}, \quad \forall A \subset \mathbb{R}^d,
\]
as in Carathéodory’s construction Proposition 2.6. Compare with the construction of the Lebesgue measure given in Proposition 2.26 and show (1) that both definition are equivalent. Similarly, for any other class \( \mathcal{E} \) of \( d \)-intervals (which is not necessarily a semi-ring) generating the Borel \( \sigma \)-algebra in \( \mathbb{R}^d \), consider the outer measure induced by (B.20) with the class \( \mathcal{E} \) replacing the semi-ring \( \mathcal{I}_d \).

Prove (2) that the same Lebesgue measure is obtained.

**Proof.** Let us denote by \( \ell = \ell_d \) and \( \ell^* = \ell_d^* \) the construction given in Proposition 2.26 via the semi-ring \( \mathcal{I}_d \) of bounded \( d \)-intervals that are semi-open (to the left) and semi-closed (to the right), i.e., of the form \( \lbrack a, b \rbrack \).

(1) For any set \( A \), if \( \{I_n\} \) is a sequence of bounded open \( d \)-interval covering \( A \) then the sequence \( \{I_n\} \) of \( d \)-intervals in \( \mathcal{I}_d \) with exactly the same endpoints will also cover \( A \), \( I_n \subset I_n \), \( A \subset \bigcup_n I_n \), and we deduce that \( \ell^*(A) \leq m^*(A) \). Similarly, if \( \{\varepsilon_n\} \) is a sequence of positive numbers and \( \{I_n\} \) is a sequence of \( d \)-interval in \( \mathcal{I}_d \) covering \( A \) then we can modify the endpoints of each \( I_n \) to construct a sequence \( \{\hat{I}_n\} \) of open \( d \)-intervals such that \( I_n \subset \hat{I}_n \) and \( m(\hat{I}_n) \leq \varepsilon_n + \ell(I_n) \). Thus \( A \subset \bigcup_n \hat{I}_n \), and we obtain \( m^*(A) \leq \ell^*(A) + \sum_n \varepsilon_n \), which proves the reverse inequality. Hence both outer measures construction are the same, i.e., \( m^* = \ell^* \).
It may be convenient to define the hyper-volume $m$ for any $d$-intervals (non necessarily open or bounded) $I$ as follows: $m(I) = m(\hat{I})$, with $\hat{I}$ denoting the interior of $I$, and $m(\hat{I}) = \infty$ if $\hat{I}$ is unbounded. In this way, the hyper-volume $m$ is additive over $d$-intervals, i.e.,

$$I = I_1 + \cdots + I_k \implies m(I) = m(I_1) + \cdots + m(I_d).$$

Actually, if $I$ is a non-overlapping finite union of intervals $\{I_1, \ldots, I_k\}$ (i.e., $I = \bigcup_{i=1}^{k} I_i$ with either $\hat{I}_i \cap \hat{I}_j = \emptyset$ or $m(\hat{I}_i \cap \hat{I}_j) = 0$, for any $i \neq j$) then we also have $m(I) = m(I_1) + \cdots + m(I_d)$.

To complete the construction of the Lebesgue measure via open $d$-intervals, we need to show (independently from the properties of $\ell^*$) that any $d$-interval is $m^*$-measurable and that $m = m^*$ on the class $\mathcal{I}_d$ of bounded open $d$-intervals.

Therefore, for a given $I$ in $\mathcal{I}_d$ and any $\varepsilon > 0$, we can modify the endpoints of $I$ to construct a compact $d$ interval $K$ and another open $d$-interval $J$ satisfying $J \subset K \subset I$ and $m(I) < m(J) + \varepsilon$. Now, if $\{I_n\}$ is a sequence of bounded open $d$-intervals covering $I$ then they necessarily cover $K$, and because $K$ is compact, there exist a finite subcover $J \subset K \subset \bigcup_{n \leq k} I_n$. Hence

$$m(J) \leq \sum_{n \leq k} m(I_n) \leq \sum_{n} m(I_n),$$

which implies that $m(I) \leq m^*(I)$ as $\varepsilon \to 0$. Since $I$ is open and cover itself, we also have the reverse inequality and so $m(I) = m^*(I)$ for every open $I$.

Actually, almost without modification, the above argument can be applied to any $d$-interval (non necessarily open or bounded) to get $m(I) = m^*(I)$ for any $d$-interval. This implies that $m^*$ is additive on $d$-interval, i.e., if a $d$-interval (non necessarily open or bounded) $I$ is written as a disjoint finite union of $d$-intervals (non necessarily open or bounded), $I = I_1 + \cdots + I_k$, then $m^*(I) = m^*(I_1) + \cdots + m^*(I_k)$.

Now, to show that a Borel set (in particular any open bounded $d$-interval) is $m^*$-measurable, we need only to have a family $\mathcal{A}$ of $m^*$-measurable sets (and Borel sets) that generates the Borel $\sigma$-algebra. For such a family, we may choose open $d$-interval of the form $A = (a, +\infty)$ with $a$ in $\mathbb{R}^d$. Thus, to check that $A = (a, +\infty)$ is $m^*$-measurable we have to show that $m^*(E) \geq m^*(E \cap A) + m^*(E \cap A^c)$, for every $E \subset \mathbb{R}^d$, with $m^*(E) < \infty$. Therefore, by definition, for a given subset $E$ of $\mathbb{R}^d$ with $m^*(E) < \infty$ and any $\varepsilon > 0$ there exists a sequence $\{I_n\}$ of open $d$-interval covering $E$ satisfying $\sum_n m(I_n) < m^*(E) + \varepsilon$. The set $I_n \cap A$ is an open interval and $I_n \cap A^c$ is a disjoint finite union of intervals, and so $m(I_n) = m^*(I_n \cap A) + m^*(I_n \cap A^c)$. Because any outer measure is $\sigma$-sub-additive, we obtain

$$m^*(E \cap A) + m^*(E \cap A^c) \leq \sum_{n} \left[ m^*(I_n \cap A) + m^*(I_n \cap A^c) \right] = \sum_{n} m(I_n) < m^*(E),$$

i.e., $A$ is $m^*$-measurable.
(2) It is clear that, by modifying the endpoints of the \( d \)-intervals, the outer Lebesgue measures induced by the various classes of \( d \)-intervals agree with each other. If the class of \( d \)-intervals \( E \) is a semi-ring then Caratheodory’s construction Proposition 2.11 yields the suitable extension after proving that the hyper-volume \( m \) is \( \sigma \)-additive on \( E \). This can be done similar to Proposition 2.26. Alternatively, if the class of \( E \) is not a semi-ring then Caratheodory’s construction Proposition 2.6 yields the outer Lebesgue measure, still denoted \( m^* \) and defined by (B.20) with \( E \) replacing the semi-ring \( I_d \). However, in this case we need to show that the hyper-volume \( m \) agrees with \( m^* \) on \( E \) and that each set \( B \) in \( B_0 \) is \( m^* \)-measurable, for some family \( B_0 \) generating the \( \sigma \)-algebra \( B(\mathbb{R}^d) \) of Borel sets. These is accomplished with arguments as in (1).

\[ \square \]

**Exercise 2.22.** Let \( J_d \) be the class of all \( d \)-intervals in \( \mathbb{R}^d \) (which includes any open, non-open, closed, non-closed, bounded or unbounded intervals) with boundary points \( a = (a_i) \) and \( b = (b_i) \), where \( a_i \) and \( b_i \) belong to \([-\infty, +\infty] \). By considering in some detail the case \( d = 1 \) (with comments to the general case \( d \geq 2 \), do as follow:

(1) Prove that \( J_d \) is a semi-algebra of subsets of \( \mathbb{R}^d \).

(2) Define an additive set function \( \bar{m} \) on \( J_d \) such that \( \bar{m} \) restricted to the semi-ring \( I_d \) of (left-open and right-closed) \( d \)-intervals used in Proposition 2.26 agrees with the expression (2.5). Extend the definition of \( \bar{m} \) to an additive set function on the algebra \( \tilde{J}_d \) generated by the semi-algebra \( J_d \).

(3) Show that 

\[ \bar{m}(J) = \sup \{ \bar{m}(K) : J \supseteq K, \text{ compact } K \in \tilde{J}_d, \} \]

for every \( J \) in \( \tilde{J}_d \).

(4) Deduce from (3) that \( \bar{m} \) is \( \sigma \)-additive on \( \tilde{J}_d \) and show that the extension of \( \bar{m} \) is also the Lebesgue measure.

**Proof.** (1) It is clear that \( J_d \) is a \( \pi \)-class, i.e., stable under the formation of finite intersections. In the case of \( \mathbb{R} \), the complement of an interval is another interval or the union of two intervals. In \( \mathbb{R}^d \), the complement of a \( d \)-interval is another \( d \)-interval or the union of several \( d \)-intervals, i.e., \( J_d \) is indeed a semi-algebra.

(2) Certainly, the set function \( \bar{m} \) on \( J_d \) is defined by \( \bar{m}(J) = 0 \) if the interior \( J \) is nonempty, \( \bar{m}(J) = \infty \) if \( J \) is an unbounded \( d \)-interval with nonempty interior and if \( a \) and \( b \) in \( \mathbb{R}^d \) are the extreme points of \( J \) then define \( I_i = [a, b] \) and \( \bar{m}(J) = m(I_i) = (b_1 - a_1) \cdots (b_d - a_d) \).

As in Exercise 2.13 and Proposition 2.11, using the fact the any set \( E \) in the algebra \( \tilde{J}_d \) is a disjoint finite union of elements in \( J_d \), i.e., \( E = \sum_{i=1}^n J_i \), \( J_i \) in \( J_d \), we define \( \bar{m}(E) = \sum_{i=1}^n \bar{m}(J_i) \), which is an additive set function well defined on \( \tilde{J}_d \).

(3) First, let us verify that for every \( J \) in the semi-algebra and any real number \( 0 < c < \bar{m}(J) \) there is a compact \( d \)-interval \( K \subset J \) such that \( c < \bar{m}(K) \). Indeed, if \( a \) and \( b \) are the extreme points of \( J \) then we can find points \( a' \) and \( b' \) in \( \mathbb{R}^d \) such that \( a_i < a'_i < b'_i < b_i \) and \( c < \bar{m}(K) \) with \( K = [a'_i, b'_i] \times \cdots \times [a'_d, b'_d] \).
Denote by \( m_*(J) \) the right-hand of (B.21), we note that the monotony of \( \bar{m} \) implies that \( m(K) \leq m(J) \), for every \( K \subset J \), i.e. \( m_*(J) \leq m(J) \). For the reverse inequality, if \( J = \sum_{i=1}^{n} \bar{m}(J_i) \) with \( \bar{J} > 0 \) then for any \( d \)-interval \( J_i \) with \( \bar{J}_i > 0 \) and every real number \( 0 < c_i < \bar{J}_i \), we can find a compact \( d \)-interval \( K_i \) satisfying \( K_i \subset J_i \) and \( c_i < \bar{m}(K_i) \). Thus, if \( K_i = \emptyset \) when \( \bar{m}(J_i) = 0 \), then the compact \( d \)-interval \( K = \sum_{i=1}^{n} K_i \subset J \) satisfies

\[
\bar{m}(K) = \sum_{i=1}^{n} \bar{m}(K_i) > \sum_{i=1}^{n} c_i = c > 0,
\]

which yields \( m_*(J) = \bar{m}(J) \).

(4) Since \( \bar{m} \) is additive on the algebra \( \hat{J}_d \), the \( \sigma \)-additivity reduces to the monotone continuity from below, i.e., if \( \{J_n\} \) is an increasing sequence in \( \hat{J}_d \) and and \( J = \bigcup_n J_n \) belongs to \( \hat{J}_d \) then \( \bar{m}(J) = \lim_n \bar{m}(J_n) \). Moreover, if one of the sets \( J_n \) is unbounded then so is \( J \), the limit is infinite and the equality holds. Therefore, we need to consider only the case where all set \( J_n \) are bounded and \( \bar{m}(J) > 0 \).

Arguing similar to the beginning of part (3), for any \( \varepsilon > 0 \) and a given increasing sequence \( \{J_n\} \) of bounded sets in the algebra \( \hat{J}_d \), we can modify the extreme points of \( d \)-interval forming \( J_n \) to construct an increasing sequence \( \{J_n'\} \) of open (bounded) sets in \( \hat{J}_d \) such that

\[
\lim \bar{m}(J_n) \leq \lim \bar{m}(J_n') \leq \varepsilon + \lim \bar{m}(J_n) \quad \text{and} \quad J_n \subset J_n', \quad \forall n.
\]

Also, for every real number \( c < \bar{m}(J) \) there exists a compact set \( K \subset J \) in \( \hat{J}_d \) such that \( \bar{m}(K) > c \). Because \( \{J_n'\} \) is an open cover of the compact set \( K \), there is a finite subcover, i.e., \( K \subset \bigcup_{n \leq m} J_n' = J_m' \), for some number \( m \). Hence

\[
c < \bar{m}(K) \leq \bar{m}(J_m') \leq \varepsilon + \lim \bar{m}(J_n),
\]

and as \( \varepsilon \to 0 \) and \( c \to \bar{m}(J) \) we deduce \( \bar{m}(J) = \lim_n \bar{m}(J_n) \), i.e., \( \bar{m} \) is \( \sigma \)-additive on the algebra \( \hat{J}_d \).

Recall that in Exercise 2.20, we proved that the Lebesgue measure of the boundary points of a bounded \( d \)-interval is zero, i.e., \( \ell(\partial J) = 0 \) for every \( J \) in \( \hat{J}_d \), after expressing unbounded \( d \)-intervals a increasing limits of bounded \( d \)-intervals.

Now, to check that the extension of \( \bar{m} \) is indeed the Lebesgue measure as obtained in Proposition 2.26 via the semi-ring \( I_d \) of bounded \( d \)-intervals that are semi-open (to the left) and semi-closed (to the right), we need to use the fact that \( \bar{m}(J) = \ell(J) \), for every \( d \)-interval (i.e., element in \( \hat{J}_d \)), and use the uniqueness of Caratheodory’s extension Theorem 2.9.

\[\square\]

Exercise 2.23. Consider the Lebesgue measures \( \ell_d, \ell_h \) and \( \ell_{d+h} \) on the spaces \( \mathbb{R}^d, \mathbb{R}^h \) and \( \mathbb{R}^{d+h} \). Discuss the additive product measure \( \ell_d \times \ell_h \) (see Remark 2.14), its outer measure extension \( \ell^* \) in \( \mathbb{R}^{d+h} \) and the \( (d+h) \)-dimensional Lebesgue measure \( \ell_{d+h} \). Actually, verify that \( \ell_{d+h} \) is the completion of the product Lebesgue measure \( \ell_d \times \ell_h \).
Proof. We can view the Lebesgue measure \( \ell_d \) constructed as in Proposition 2.26 via the semi-ring \( \mathcal{I}_d \), or as in Exercise 2.22 via the semi-ring \( \mathcal{J}_d \). In any case, we consider a similar construction \( \ell_h \) in \( \mathbb{R}^h \) and \( \ell_{d+h} \) in \( \mathbb{R}^{d+h} \).

By definition, we have \( \bar{m}_d \times \bar{m}_h = \bar{m}_{d+h} \) on the (product) semi-algebra \( \mathcal{J}_d \times \mathcal{J}_h = \mathcal{J}_{d+h} \) or equivalently, \( m_d \times m_h = m_{d+h} \) on the (product) semi-ring \( \mathcal{I}_d \times \mathcal{I}_h = \mathcal{I}_{d+h} \). Since \( \bar{m}_{d+h} \) and \( m_{d+h} \) are \( \sigma \)-additive, the Caratheodory’s extension Theorem 2.9 insures that \( \bar{m}_{d+h} = \ell_{d+h} \) on \( \mathcal{J}_{d+h} \) and \( m_{d+h} = \ell_{d+h} \) on \( \mathcal{I}_{d+h} \). In any case, \( \ell_{d+h} \) is defined on the complete \( \sigma \)-algebra \( \mathcal{L}_{d+h} \), i.e., the \( \ell_{d+h}^* \)-completion of the Borel \( \sigma \)-algebra (as generated by the semi-ring \( \mathcal{I}_{d+h} \)). \( \Box \)

**Exercise 2.24.** Let \( F_i : \mathbb{R} \to \mathbb{R}, i = 1, \ldots, d \), be non-decreasing functions. Verify that the expression

\[
m_F([a, b]) = \prod_{i=1}^d (F_i(b_i) - F_i(a_i)), \quad \forall a, b \in \mathbb{R}^d
\]

defines an additive set function \( m_F \) on the semi-ring \( \mathcal{I}_d \), which is \( \sigma \)-additive if each \( F_i \) is right-continuous. Give some details on how the alternative construction of the Lebesgue measure presented in the previous Exercise 2.22 can be used for the Lebesgue-Stieltjes measure.

**Proof.** Replacing the hyper-volume with the expression

\[
\text{if } [a, b] = [a_1, b_2] \times \cdots \times [a_d, b_d]
\]

then \( m_F([a, b]) = (F_1(b_1) - F_1(a_1)) \cdots (F_d(b_d) - F_1(a_d)) \),

we see that, by construction, \( m_F \) is an additive set function defined on the semi-ring \( \mathcal{I}_d \). To check the \( \sigma \)-additivity, let \( I = [a, b] \) be a disjoint countable union of the \( d \)-intervals \( I_n = [a_n, b_n] \), i.e., \( I = \bigcup_n I_n \), and note that in view of the continuity from the right of \( F = (F_i) \) we have \( m_F([b, b+t]) \to 0 \) as \( t \downarrow 0 \) with \( t = (t_i) \). Since, as in Proposition 2.26, for every \( \varepsilon > 0 \) there exits some \( \delta = (\delta_i), \delta_n = (\delta_{n,i}), \) with \( \delta_i, \delta_{n,i} > 0 \) such that \( m_F([b_n, b_n + \delta_n]) < 2^{-n-1}\varepsilon \) and \( m_F([a, a + \delta]) < \varepsilon/2 \), we can define the open intervals \( J_n = [a_n, b_n] \cap I_n, n \geq 1 \), and the compact interval \( K = [a + \delta, b] \subseteq I \) to obtain

\[
m_F(I) \leq \varepsilon/2 + m_F(K) \quad \text{and} \quad \sum_n m_F(J_n) \leq \varepsilon/2 + \sum_n m_F(I_n).
\]

Because \( \{J_n\} \) is a cover by open sets of the compact set \( K \), there must be a finite cover, i.e., \( K \subseteq \bigcup_{n \leq k} J_n \), for some \( k \geq 1 \), and in view of the additivity of \( m_F \) we obtain

\[
m_F(I) - \varepsilon/2 \leq m_F(K) \leq \sum_{n \leq k} m_F(J_n) \leq \sum_n m_F(J_n) \leq \varepsilon/2 + \sum_n m_F(I_n),
\]

which shows the \( \sigma \)-additivity of \( m_F \) on the semi-ring \( \mathcal{I}_d \).
Thus, by means of Caratheodory’s construction Proposition 2.11, every Borel set is \( m_F^* \)-measurable and \( m_F^*(]a, b[) = m_F([a, b]) \).

To construct the Lebesgue-Stieltjes measure, we could make use the class of open and bounded \( d \)-intervals \( \mathcal{I}_d \) instead of the semi-ring \( \mathcal{I}_d \), as in Exercise 2.22. However, if we want to obtain the same Lebesgue-Stieltjes measure, we replace the hyper-volume with the expression

\[
\text{if } ]a, b[=]a_1, b_2[ \times \cdots \times ]a_d, b_d[ \\
\text{then } \bar{m}_F([a, b]) = (F_1(b_1) - F_1(a_1)) \cdots (F_d(b_d) - F_1(a_d)),
\]

where \( F_i(b_i) \) means the limit from the left at \( b_i \) of \( F_i \). Thus, Caratheodory’s extension Theorem 2.9 can be also applied, but we need to establish that \( \bar{m}_F^*[]a, b[) = m_F([a, b]) \) and that every Borel set is \( \bar{m}_F^* \)-measurable, independently from the previous construction.

Essentially the same argument to established the \( \sigma \)-additivity of \( m_F \) on the semi-ring \( \mathcal{I}_d \) can be applied to obtain \( \bar{m}_F^*[]a, b[) \geq m_F([a, b]) \), and the converse inequality follows from the monotony, \( ]a, b[ \subset ]a', b'[, b' > b, \bar{m}_F^*[]a, b[) \leq \bar{m}_F^*[]a, b'[] \). Also, since \( \bar{m}_F^*[]a, c'[) \to m_F([a, b]) \) as \( c' \uparrow b \), we also deduce that \( \bar{m}_F^*[]a, b[) = \bar{m}_F([a, b]) \).

To check that any Borel set is \( \bar{m}_F^* \)-measurable, we argue similar to Exercise 2.22, but this time, we have to use the additivity of \( \bar{m}^*_F \) for \( d \)-intervals either open \( ]a, b[ \) or left-open of the form \( [a, b[ \).

**Exercise 2.25.** Recall that the Cantor set \( C \) as the set of all real numbers \( x \) in \([0, 1]\) expressed in the ternary system \( x = \sum_n a_n 3^{-n} \) with \( a_n \) in \( \{0, 2\} \); and consider the Cantor function \( f \) initially defined by \( f(x) = \sum_n b_n 2^{-n} \), with \( b_n = a_n/2 \). Give a quick argument justifying that the Cantor set is uncountable and compact, with empty interior and no isolated points. Show that the Cantor set has Lebesgue measure zero, i.e., \( m(C) = 0 \). Extend \( f \) to the function \( f : [0, 1] \to [0, 1] \), which is constant on the complement \([0, 1] \setminus C\) and strictly increasing on \( C \) (except at the two endpoints of each interval removed). Again, verify that \( f \) is a continuous function, e.g., see Folland [45, Proposition 1.22, pp. 38–39].

**Proof.** The Cantor set \( C \) can also be defined by induction, by removing the open middle third (1/3, 2/3) of closed interval \([0, 1]\), then removing the open middle thirds (1/9, 2/9) and (7/9, 8/9) of the 2 remaining closed intervals \([0, 1/3]\) and \([2/3, 1]\), and so forth, i.e., removing the open middle thirds of the \( 2^n \) remaining closed intervals, in the \( n \)-iteration. Thus, \( C \) is a closed set of \([0, 1]\) (therefore \( C \) is compact), and if \( x < y \) belong to \( C \) then there exists some \( z \) in \((x, y) \setminus C\), i.e., \( C \) is totally disconnected (i.e., the only connected subsets of \( C \) are single points), \( C \) has no isolated points, and \([0, 1] \setminus C\) is dense in \([0, 1]\).

The Lebesgue measure of \( C \) is calculated as \( \ell(C) = 1 - \ell([0, 1] \setminus C) \), and the series of the length of all the open middle thirds removed

\[
\ell([0, 1] \setminus C) = \sum_{n=0}^{\infty} \frac{2^n}{3^{n+1}} = \frac{1}{1 - (2/3)} = 1
\]
yields $\ell(C) = 0$.

Since the Cantor function $f$ is a bijection between the Cantor set $C$ and the closed interval $[0, 1]$, the Cantor set $C$ has the cardinality of the continuum, $\aleph_1$. This implies that the $\sigma$-algebra of all Lebesgue measurable sets has the cardinality of the parts of a continuum, $\aleph_2$.

Because $f$ is non-decreasing and $C$ is dense, $f$ cannot have any jumps as a function defined from $C$ onto $[0, 1]$. Moreover, if $x < y$ are in $C$ then $f(x) < f(y)$, unless $x$ and $y$ are the two endpoints of one of the intervals removed from $[0, 1]$ to construct $C$. Therefore, we can extend $f$ to the whole interval $[0, 1]$ by keeping $f$ constant on each interval that has been removed from $[0, 1]$ to build $C$. This function is continuous from $[0, 1]$ onto $[0, 1]$, strictly increasing on $C$ and constant on the open set $[0, 1) \setminus C$.

Looking at the Cantor by induction, we are removing $2^k - 1$ disjoint open (middle thirds) intervals in step $k$, ordered as $(a_{k,i}, b_{k,i})$, $b_{k,i} < a_{k,i+1}$, $1 \leq i < 2^k$. Thus, its closed complement $C_k = [0, 1] \setminus \bigcup_i (a_{k,i}, b_{k,i})$, so that the Cantor set can be written as

$$C = \bigcap_k C_k, \quad C_k = \bigcup_{1 \leq i \leq 2^k} [b_{k,i-1}, a_{k,i}], \quad \text{with } b_{k,0} = 0, \quad a_{k,2^k} = 1.$$  

Thus, define (a) $f_k(0) = 1$, $f_k(1) = 1$, (b) $f_k$ constant on each closed interval $[a_{k,i}, b_{k,i}]$, $f_k(a_{k,i}) = 2^{-k}i$, $1 \leq i \leq 2^k$, (c) $f_k$ linear $C_k$, and (d) continuous on $[0, 1]$. The function $f_k$ is nondecreasing from $[0, 1]$ into itself, and $|f_{k+1}(x) - f_k(x)| \leq 2^{-k}$, for every $x$. Hence, $f_k \to f$ uniformly on $[0, 1]$.  

**Exercise 2.26.** The translation invariance of the Lebesgue measure can be used to show the existence of a non-measurable set in $\mathbb{R}$. Indeed, consider the addition modulo 1 acting on $[0, 1) \times [0, 1)$, i.e., for any $a$ and $b$ in $[0, 1)$ we set $a + b = c$ mod 1 with $c = a + b$ if $a + b < 1$ or $c = a + b - 1$ if $a + b \geq 1$. Verify that for any Lebesgue measurable set $E$ and any $b$ in $[0, 1)$ the set $A + b$ mod 1 $= \{a + b \mod 1 : a \in A\}$ is Lebesgue measurable and $m(A + b \mod 1) = m(A)$. Next, define the equivalence relation $x \sim y$ if and only if $x - y$ is rational, and consider the family $\bar{x}$ of equivalence classes in $[0, 1)$, i.e., $\bar{x} = \{y \in [0, 1) : y \sim x\}$. Certainly, all rational numbers belong to the same equivalence class, and by means of the axiom of choice, we can select one (and only one) element of each equivalence class to form a subset $E$ of $[0, 1)$ such that (1) any two distinct element $x$ and $y$ in $E$ does not belong to the same equivalence class, and (2) $\bar{x}$ with $x$ in $E$ yield all possible equivalence classes. Let $\{r_n\}$ be an enumeration of the rational numbers in $[0, 1)$ with $r_0 = 0$. Prove that $E_n = E + r_n$ mod 1 defines a sequence of disjoint sets in $[0, 1)$, with $m(E_n) = m(E)$. Finally, show that $[0, 1) = \bigcup_n E_n$ and deduce that $E$ cannot be a Lebesgue measurable set. For instance, the reader may check the book Burk [23, Appendix B, C] or Kharazishvili [69] to find a comprehensive discussion on non-measurable sets.

**Proof.** Define $F_n = E + r_n \subset [0, 2)$, $E'_n = F_n \cap [0, 1)$ and $E''_n = F_n \cap [1, 2) - 1 \subset [0, 1)$ to check that

$$E_n = E'_n \cup E''_n \quad \text{and} \quad E'_n \cap E''_n = \emptyset.$$
Indeed, if some $x$ belongs to $E'_n \cap E''_n$ then $x = x'_n + r_n = x''_n + r_n - 1$ with $x'_n$ and $x''_n$ in $F_n$, so that $x_n - x''_n \in (-1, 1)$ with $x'_n - x''_n = -1$ produce a contradiction.

Now, if $y$ belongs to $F_n \cap F_m$ then $x = x_n + r_n = x_m + r_m$ with $x_n$ in $F_n$ and $x_m$ in $F_m$. In view of the property (1) of $E$, the equality $x_n - x_m = r_m - r_n$ implies that $x_n = x_m$ and $r_n = r_m$, i.e., $\{F_n\}$ is a sequence of disjoint subsets, and so is $\{E_n\}$.

Any $x$ in $[0, 1)$ belongs to the same equivalence class, and in view of the property (2) of $E$, we deduce that $x$ belongs to some $E_n$, i.e., $[0, 1) = \bigcup_n E_n$. Hence, if $E$ is Lebesgue measurable then $\ell(E_n) = \ell(E'_n) + \ell(E''_n) = \ell(F_n) = \ell(E)$ and

$$1 = \ell([0, 1)) = \sum_n \ell(E_n) = \sum_n \ell(E),$$

which is a contradiction. \qed

**Exercise 2.27.** Verify that if $I$ is a bounded $d$-intervals in $\mathbb{R}^d$ (which includes any open, non-open, closed, non-closed) with endpoints $a$ and $b$ then the Lebesgue measure $m(I)$ is equal the product $\prod_{i=1}^d (b_i - a_i)$.

(1) Let $E$ be the class of all open bounded intervals with rational endpoints, i.e., $(a, b)$ with $a$ and $b$ in $\mathbb{Q}^d$. Denote by $m_1^*$ and $m_1$ the outer measure and measure induced by the Caratheodory construction relative to the Lebesgue measure restricted to the (countable) class $E$. Check that a subset $A$ of $\mathbb{R}^d$ is $m_1^*$-measurable if and only if $A$ is Lebesgue measurable, and prove that $m_1 = m$.

Can we show that $m_1^* = m^*$?

(2) Similarly, let $C$ be the class of all open cubes with edges parallel to the axis with rational endpoints (i.e., $(a, b)$ with $a$, $b$ in $\mathbb{Q}^d$ and $b_i - a_i = r$, for every $i$), and let $(m_2^*)$ $m_2$ be the corresponding (outer) measure generated as above, with $C$ replacing $E$. Again, prove results similar to item (1). What if $E$ is the class of semi-open dyadic cubes $[(i - 1)2^{-n}, i2^{-n}]^d$ for $i = 0, \pm 1, \ldots, \pm 4^n$?

(3) How can we extend all these arguments to the Lebesgue-Stieltjes measure. State precise assertions with some details on their proof.

(4) Consider the class $D$ of all open balls with rational centers and radii. Repeat the above arguments and let $(m_3^*)$ $m_3$ (outer) measure associated with the class $D$. How can we easily verify the validity of the previous results for this setup (see later Corollary 2.35).

**Proof.** Let $E$ be a class of $d$-intervals and $\bar{m}$ be the hyper-volume set function, i.e., if $E$ is a $d$-interval with endpoints $a$ and $b$ then $\bar{m}(E) = \prod_{i=1}^d (b_i - a_i)$ (with the convention that $0 \times \infty = 0$, for unbounded intervals in the product formula). Besides the additive property for non-overlapping intervals, i.e., if $E = \bigcup_{i=1}^n E_i$ with disjoint interiors ($\hat{E}_i \cap \hat{E}_j = \emptyset$ for $i \neq j$) then $\bar{m}(E) = \sum_{i=1}^n \bar{m}(E_n)$, an essential property of the inf expression

$$m^*(A) = \inf \left\{ \sum_{n=1}^\infty \bar{m}(E_n) : E_n \in \mathcal{E}, A \subset \bigcup_{n=1}^\infty E_n \right\}, \quad \forall A \subset \mathbb{R}^d$$
(in the Carathéodory construction), used very often is the following approximation: for any given $E$ bounded $d$-interval and for every $\varepsilon > 0$ there exists an open $d$-interval $I$ and a compact $d$-interval $K$ such that $K \subset E \subset I$ and $\bar{m}(I) - \varepsilon < \bar{m}(E) < \bar{m}(K) + \varepsilon$. Certainly, $\bar{m}$ is monotone, i.e., $E \subset I$ implies $\bar{E} \leq \bar{E}$, and if $E$ is be unbounded then for every $c < \bar{m}(E)$ there exists a compact $d$-interval $K$ such that $K \subset E$ and $c < \bar{m}(K)$.

Indeed, the main argument to verify that either the hyper-volume $\bar{m}$ is $\sigma$-additive on $\mathcal{E}$ or the outer Lebesgue measure $m^*(I)$ of a bounded $d$-interval $I$ is equal to $\bar{m}(I)$ goes as follows: (a) for any $\varepsilon > 0$ and for any sequence $\{I_n\}$ of $d$-interval covering $I$ we can modify the endpoints of each interval to obtain a sequence $\{I_n'\}$ of open bounded $d$-intervals and a compact $d$-interval $K$ such that

$$\sum_n \bar{m}(I_n') \leq \varepsilon + \sum_n \bar{m}(I_n) \quad \text{and} \quad \bar{m}(I) \leq \varepsilon + \bar{m}(K),$$

with $K \subset I$ and $I_n \subset I_n'$, for every $n$; and (b) we use the compactness to obtain a finite cover and the additivity of $\bar{m}$ to conclude.

Since the hyper-volume $\bar{m}$ is an additive set function, the outer Lebesgue measure $m^*$ is also additive on the class of $d$-intervals. This is a key point used to deduce that any Borel set is $m^*$-measurable, see the previous Exercise 2.22.

(1) If $\mathcal{E}$ the class of all open bounded $d$-intervals with rational endpoints then the arguments are similar to Exercise 2.21. Indeed, for any $E$ bounded $d$-interval and any $\varepsilon > 0$ there exists an open $d$-interval $I$ and a compact $d$-interval $K$ (both with rational endpoints if desired) such that $K \subset E \subset I$ and $\bar{m}(I) - \varepsilon < \bar{m}(E) < \bar{m}(K) + \varepsilon$. As above, this allows us to show (independently of the initial definition of $m^*$) that $m^*_1$ agrees with the hyper-volume $\bar{m}$, which yields the additivity of $m^*_1$ for $d$-intervals. Hence an open unbounded $d$-interval of the form $(a, +\infty)$ is $m^*_1$-measurable, and therefore any Borel set is also $m^*_1$-measurable. Also, because $m_1$ and $m$ agree on a $d$-intervals, they agree everywhere. Alternatively, we can argue that from any cover by $d$-intervals in $\mathcal{E}$ we can find a cover by $d$-intervals in the semi-ring $\mathcal{I}_d$ with almost the same sum of hyper-volume (and conversely) to deduce that the corresponding Lebesgue outer measures agree, i.e., $m^*_1 = m^*$.

(2) If $\mathcal{C}$ the class of all open (bounded) cubes (or $d$-cubes if we prefer) with edges parallel to the axis with rational endpoints, then the arguments are similar to part (1) above, replacing $\mathcal{E}$ with $\mathcal{C}$. Note that now, for any given bounded $d$-interval $E$ and for every $\varepsilon > 0$ there exists a finite number of open $d$-cubes $\{I_i\}$ and a finite number of compact $d$-cubes $\{K_j\}$ such that $\bigcup_j K_j \subset E \subset \bigcup_i I_i$ and $\sum_i \bar{m}(I_i) - \varepsilon < \bar{m}(E) < \sum_j \bar{m}(K_j) + \varepsilon$. This property suffices to make an argument, similar to part (1), to show (independently) that $m^*_2$ agrees with the hyper-volume on $\bar{m}$ on any $d$-interval and that indeed, $m^*_2$ is the Lebesgue outer measure $m^*$.

We have to point-out that we could use the class $\mathcal{C}_s$ of all $d$-cubes, semi-open from the left, semi-closed from the right, with edges parallel to the axis and with rational endpoints, i.e., $(a, b) \subset \mathbb{Q}^d$ and $b_i - a_i = r$, for every $i$. This class $\mathcal{C}_s$ is not stable under the formation of finite intersections,
i.e., \( C_s \) is not a semi-ring. However, if \( \mathcal{I}_s \) denotes the class of all \( d \)-interval in the semi-ring \( \mathcal{I}_d \) with rational endpoints (having rational endpoints is crucial here) then \( \mathcal{I}_{ds} \) is a semi-ring and any element there can be written exactly as a finite disjoint union of \( d \)-cubes. Hence, both classes (\( \mathcal{I}_{ds} \) and \( C_s \)) generated the same ring \( \mathcal{R} \), composed by all disjoint finite unions of \( d \)-cubes (semi-open from the left, semi-closed from the right, with edges parallel to the axis, and with rational endpoints). Therefore, by means of Caratheodory’s construction Proposition 2.11, we obtain a new (outer) measure denoted by \( (\tilde{m}_2)\) \( \tilde{m}_2 \). Again, because this measure agrees with the hyper-volume on \( C_s \), this is indeed the Lebesgue (outer) measure.

If \( E \) is the class of semi-open dyadic cubes \( Q_{i,n} = [(i-1)2^{-n},i2^{-n}]^d \) for \( i = 0, \pm1, \ldots \pm 4^n \) then \( E \) is a \( \pi \)-class, but is not a semi-ring. However, as mentioned early, what really count is the fact that the class \( \mathcal{R} \) of all finite disjoint unions of sets in \( E \) is a ring. Indeed, this follows from the facts that (a) the intersection of two semi-open dyadic cubes \( Q_{i,n} \cap Q_{j,m} \) is either empty or one of the two cubes and (b) the difference of two semi-open dyadic cubes \( Q_{i,n} \setminus Q_{j,m} \) is a finite disjoint union of semi-open dyadic cubes. The argument is the same as in the case of cubes with rational endpoints, we need to use the fact that the dyadic numbers \( \{i2^{-n} : n \geq 1, i = 0, \pm1, \ldots \} \) is a dense set in \( \mathbb{R} \).

Choosing \( i = 0, \pm1, \ldots \pm 4^n \) is not necessary, the class with \( i = 0, \pm1, \ldots \) plays the same role. However, using the finite ring \( \mathcal{R}_n \) of all semi-open dyadic cubes \( Q_{i,n} \) for \( i = 0, \pm1, \ldots \pm 4^n \) (or even letting \( i = 0, \pm1, \ldots \)) to produce an outer measure \( m^*_n \) and then take the limit (or infimum) \( \tilde{m}^* = \lim_n m^*_n = \inf_n m^*_n \) does not produce the Lebesgue measure. Indeed, any cube \( Q_{i,n} \) in the ring \( \mathcal{R}_n \) has a hyper-volume \( m(Q_{i,n}) \geq 2^{-dn} \), which means that only finite series can be used (since infinite series produces infinite values) when taking the infimum to obtain \( m^*_n \).

(3) For the Lebesgue-Stieltjes measure we follow the arguments in Exercise 2.24. Note that using the semi-ring of \( d \)-cubes \( C_s \), as mentioned in part (2), via the Caratheodory’s construction Proposition 2.11 is very handy. However if we insist in using either open \( d \)-interval as in (1) (i.e., the class \( E \)) or open \( d \)-cubes as in (2) (i.e., the class \( C \)) then we need to adjust the definition of the initial set function \( m_F \) to obtain the same Lebesgue-Stieltjes measure. The crucial point that the outer measure \( m^*_F \) is indeed an extension of the initial set function \( m_F \) is more delicate.

(4) If we decide to use the class \( D \) of all open balls (with rational centers and radii) then more difficulties arise. First, we need a formula for the hyper-volume \( \tilde{m} \) valid for balls, which are known for dimension \( d \leq 3 \), but in general, this requires some calculation involving the integral. A related second point is to show the details on obtaining the following approximation property: for every \( \varepsilon > 0 \) and for any given bounded \( d \)-interval (or \( d \)-ball) \( E \) there exists a finite number of open \( d \)-intervals (or \( d \)-balls) \( \{I_i\} \) and a finite number of compact \( d \)-intervals (or \( d \)-balls) \( \{K_j\} \) such that \( \bigcup_j K_j \subset E \subset \bigcup_i I_i \) and \( \sum_i \tilde{m}(I_i) - \varepsilon < \tilde{m}(E) < \sum_j \tilde{m}(K_j) + \varepsilon \). Another complication is that a \( d \)-ball cannot be written as a finite union of smaller \( d \)-balls (for \( d \geq 2 \)), so that additivity on \( b \)-balls makes nonsense. Moreover, we cannot build a semi-ring of \( d \)-balls to
simplify. Therefore, if we assume a suitable definition of the hyper-volume $\bar{m}$ for $d$-balls then the above arguments let us show that the outer measures agree, i.e., $m_1^* = m^*$, but using Caratheodory’s extension Theorem 2.9 for the actual construction (or definition) of the Lebesgue measure is hard (or impossible) without the support of the $d$-intervals. Some light on this is given by Vitali’s covering as discussed in Section 2.5 If we insist in using $d$-balls then Hausdorff construction in Section 2.4 is more adequate.

Exercise 2.28. Let $F: \mathbb{R} \rightarrow \mathbb{R}$ be a nondecreasing right-continuous functions and $m_F$ be its Lebesgue-Stieltjes measure associated. Verify that (1) $m_F([a, b]) = F(b) - F(a)$. Prove that (2) there is a bijection between the Lebesgue-Stieltjes measures in $\mathbb{R}$ and the semi-space of all nondecreasing right-continuous functions $F$ from $\mathbb{R}$ into itself satisfying $F(0) = 0$. Finally, (3) can we do the same for $\mathbb{R}^d$?

Proof. If $\mathcal{I}$ denote the semi-ring of all intervals for the form $I = [a, b]$ with $a$ and $b$ in $\mathbb{R}$ then the initial set function $\bar{m}_F([a, b]) = F(b) - F(a)$ defines the Lebesgue-Stieltjes (outer) measure $m_F^*$ associated with the nondecreasing right-continuous function $F$.

(1) Given $I = [a, b]$ in $\mathcal{I}$, the inf definition of the outer measure $m_F^*$ ensures that $m_F^*([a, b]) \leq \bar{m}_F([a, b])$. To prove the reverse inequality, if $I_i = [a_i, b_i]$ is a cover of $I$ then for any $\varepsilon > 0$ we can use the right-continuity of $F$ to find $a' > a$ and $b'_i > b_i$ satisfying $F(a') < F(a) + \varepsilon$ and $F(b'_i) < F(b_i) + 2^{-i}\varepsilon$, for $i \geq 1$. This yields

$$
\varepsilon + \sum_{i=1}^{\infty} \bar{m}_F(I_i) \geq \sum_{i=1}^{\infty} \bar{m}_F([a_i, b'_i]) \quad \text{and} \quad m_F([a, b]) \leq \varepsilon + \bar{m}_F([a', b]).
$$

Since $\{(a_i, b'_i)\}$ is an open cover of the compact $[a', b]$, there exists a finite subcover, and re-indexing, we obtain

$$
\sum_{i=1}^{n} (F(b'_i) - F(a_i)) \geq (F(b) - F(a')).
$$

Collecting all the pieces, we get

$$
\varepsilon + \sum_{i=1}^{\infty} \bar{m}_F(I_i) \geq m_F([a, b]) - \varepsilon,
$$

and because $\varepsilon > 0$ and $\{I_i\}$ are arbitrary, we deduce $m_F^*([a, b]) \geq \bar{m}_F([a, b])$.

(2) A Lebesgue-Stieltjes measure in $\mathbb{R}$ satisfies two key properties: (a) any Borel set is Lebesgue-Stieltjes measurable and (b) any compact set has a finite Lebesgue-Stieltjes measure. As seen later, these two properties define the so-called Radon measures. If $\mu$ is a Lebesgue-Stieltjes measure (in $\mathbb{R}$) we can define the real-valued function $F(x) = \mu([0, x])$ if $x > 0$ and $F(x) = -\mu([x, 0])$ if $x \leq 0$. Since $0 < x \leq x'$ implies

$$
F(x') - F(x) = \mu([0, x']) - \mu([0, x]) = \mu([x, x']) \geq 0,
$$
and for any nonincreasing sequence \( \{x_n\}, x_n \downarrow x \), we have

\[
\lim_n F(x_n) = \lim_n \mu([0, x_n]) = [0, x] = F(x),
\]

we deduce that \( F \) is nondecreasing and right-continuous for \( x > 0 \), and analogously for the other cases.

Thus for each Lebesgue-Stieltjes measure \( \mu \) in \( \mathbb{R} \), we construct a nondecreasing right-continuous function \( F \) with \( F(0) = 0 \) and such that \( \mu([a, b]) = F(b) - F(a) \). Conversely, from a nondecreasing right-continuous function \( F \) with \( F(0) = 0 \) we can use the expression \( m_F([a, b]) = F(b) - F(a) \) to construct the Lebesgue-Stieltjes measure associated with \( F \), which in turn, reproduces the function \( F \).

(3) If by a Lebesgue-Stieltjes measure we mean a product form like the expression (2.7), then we do have a bijective between the Lebesgue-Stieltjes measures in \( \mathbb{R}^d \) and the set of all nondecreasing right-continuous functions \( F = (f_1, \ldots, f_d) \) with \( f_i \) from \( \mathbb{R} \) into itself satisfying \( f_i(0) = 0 \), for \( i = 1, \ldots, d \). However, we may consider Radon measures \( \mu \) on \( \mathbb{R}^d \) and \( \mu([a, b]) = F(b) - F(a) \), where \( F : \mathbb{R}^d \to \mathbb{R} \), is a nondecreasing right-continuous in the sense of the partial order \( x \leq y \) if and only \( x_i \leq y_i \) for every \( i \). Since \([a, b] = \bigcup [a', b] \) with \( a_j \leq a_j' = b_j' \leq b_j \) for some \( j \) and \( a_i = a_i', b_i = b_i' \) for any \( i \neq j \) implies that

\[
F(b) - F(a) = \mu([a, b]) = \mu([a, b']) + \mu([a', b]) = [F(b') - F(a)] + [F(b) - F(a')],
\]

we deduce that (c) \( F(b') = F(a') \), if \( a' \leq b' \) and \( a_j' = b_j' \) for some \( j \). Certainly, from a nondecreasing right-continuous \( F : \mathbb{R}^d \to \mathbb{R} \) with \( F(0) = 0 \) and satisfying the property (c), we can define an additive set function \( m_F([a, b]) = F(b) - F(a) \), which is indeed \( \sigma \)-additive. Therefore, \( m_F \) induces a Lebesgue-Stieltjes (Radon measure) on \( \mathbb{R}^d \), which in turn, reproduces the function \( F \).

For instance, besides the product form \( F(b) - F(a) = \prod_i [f_i(b_i) - f_i(a_i)] \), the Riemann \( d \)-dimensional integral of a continuous function on an hyper-rectangle \([a, b] \), i.e.,

\[
F(b) - F(a) = \int_{[a, b]} f(x) \, dx, \quad \forall a \leq b
\]

with \( F(0) = 0 \), satisfies the property (c). \( \square \)

**Exercise 2.29.** Let \( F : \mathbb{R} \to \mathbb{R} \) be a nondecreasing right-continuous functions and \( m_F \) be its corresponding Lebesgue-Stieltjes measure, i.e., \( m_F([a, b]) = F(b) - F(a) \). Now consider the measure \( m_F \) restricted to the interval \([0, 1]\), where \( F = f \) is now the Cantor function as in Exercise 2.25. Verify that \( m_F(C) = 1 \) and \( m_F([0, 1] \setminus C) = 0 \), where \( C \) is Cantor set.

**Proof.** With the Lebesgue-Cantor function \( f \) of Exercise 2.25 we may define the
nondecreasing continuous function

\[
F(x) = \begin{cases} 
  f(0) = 0 & \text{if } x < 0, \\
  f(x) & \text{if } 0 \leq x \leq 1, \\
  f(1) = 1 & \text{if } x > 1.
\end{cases}
\]

Note that \(m_F((0, \infty)) = 0\) and \(m_F((-\infty, -a)) = 0\). Since Cantor set can be written as \(C = [0, 1] \setminus \bigcup_n U_n\), where \(\{U_n\}\) is the sequence of disjoint deleted open intervals, we obtain

\[
m_F(C) = m_F([0, 1]) - \sum_n m_F(U_n) = 1,
\]

because \(F\) is constant on the interval \(U_n\) and therefore \(m_F(U_n) = 0\). \qed

**Exercise 2.30.** Let \(F : \mathbb{R} \to \mathbb{R}\) be a nondecreasing function, define \(G(x) = F(x+) = \lim_{y \to x^+} F(y)\) and consider the (Lebesgue-Stieltjes type) measures \(m_F\) and \(m_G\) induced by \(F\) and \(G\), via the expressions \(\bar{m}_F([a, b]) = F(b) - F(a)\) and \(\bar{m}_G([a, b]) = G(b) - G(a)\), respectively, and the semi-ring \(I\) of intervals \(I = (a, b]\), with \(a\) and \(b\) in \(\mathbb{R}\).

(1) Show that \(G\) is right-continuous, and that

\[
G(x-) = \lim_{y \to x^-} G(x) = \lim_{y \to x^-} F(x) = F(x-).
\]

Also give some details on the construction of the measures \(m_F\) and \(m_G\). Verify that the expressions of \(m_F\) and \(m_G\) does not change if we assume that \(F(0+) = G(0) = 0\).

(2) Prove that any Borel set is measurable relative to either \(m^*_F\) or \(m^*_G\). Check that \(m_G((a, b]) = G(b) - G(a)\) and \(m_F((a, b]) \leq F(b) - F(a)\).

(3) Verify that for every singleton (set of only one point) \(\{x\}\) we have

\[
m_F(\{x\}) = F(x) - F(x-) \leq G(x) - G(x-) = m_G(\{x\}),
\]

and deduce (a) if \(F\) is left-continuous then \(m_F\) has no atoms, (b) any atom of \(m_F\) is also an atom of \(m_G\) and (c) a point \(\alpha\) is an atom for \(m_G\) if and only if \(\alpha\) is a point of discontinuity for \(G\) (or equivalently, if and only if \(\alpha\) is a point of discontinuity for \(F\)). Moreover, calculate \(m_F(I)\) and \(m_G(I)\), for any interval \(I\) (non necessarily of the form \([a, b]\)) with endpoints \(a\) and \(b\), none of them being atoms. Can you find \(m_F(I)\)?.

(4) Deduce that for every bounded interval \([a, b]\) and \(c > 0\), there are only finite many atoms \(\{a_n\} \subset [a, b]\) (possible none) with \(m_F(\{a_n\}) \geq c > 0\). Thus, conclude that \(F\) and \(G\) can have only countable many points of discontinuities, and \(\sum_n b_n \mathbb{1}_{b_n \leq \varepsilon} \to 0\) as \(\varepsilon \to 0\), for either \(b_n = m_F(\{\alpha_n\})\) or \(b_n = m_G(\{\alpha_n\})\), with \(\{\alpha_k\}\) the sequence of all atoms in \([a, b]\).
(5) Assume that $F$ is a nondecreasing purely jump function, i.e., for some sequence $\{\alpha_i\}$ of points and some sequence $\{f_i\}$ of positive numbers we have $H_l \leq F \leq H_r$, where

$$H_l(x) = \sum_{0 < \alpha_i < x} f_i, \quad \forall x > 0 \quad \text{and} \quad H_l(x) = \sum_{x \leq \alpha_i \leq 0} f_i, \quad \forall x \leq 0,$$

$$H_r(x) = \sum_{0 < \alpha_i \leq x} f_i, \quad \forall x > 0 \quad \text{and} \quad H_l(x) = \sum_{x < \alpha_i \leq 0} f_i, \quad \forall x \leq 0.$$

Verify that $H_l$ is a left-continuous and $H_r$ is right-continuous, and if $H_l^\varepsilon$ and $H_r^\varepsilon$ are as above with $\{f_i\}$ replaced with $\{f_i^\varepsilon\}$, $f_i = f_i 1_{f_i \geq \varepsilon}$, then $H_l^\varepsilon$ and $H_r^\varepsilon$ have only a finite number of jumps, and $H_l^\varepsilon \to H_l$ and $H_r^\varepsilon \to H_r$, uniformly on any bounded interval $]a, b].$

(6) Prove that if $F = H_r$ then $m_F^*$ is a purely atomic measure, i.e., $\{\alpha_i\}$ are the atoms of $m_F^*$ and $m_F^*(A) = \sum_{\alpha_i \in A} f_i$, for every $m_F^*$-measurable set $A \subset \mathbb{R}$. Similarly, if $F = H_l$ then $m_F^* = 0.$

(7) If $F(x-)$ denotes the left-hand limit of $F$ at the point $x$ then define the function

$$\bar{F}_r(x) = \sum_{0 < y \leq x} [F(y) - F(y-) ] \quad \text{or} \quad \bar{F}_r(x) = \sum_{x < y \leq 0} [F(y-) - F(y)],$$

depending on the sign of $x$. Prove that $\bar{F}_r$ is a nondecreasing purely jump right-continuous function and that $\bar{F}_l = F - \bar{F}_r$ is a nondecreasing left-continuous function. Mimic the above argument to construct a nondecreasing purely jump left-continuous function $\bar{F}_l$ such that $F_r = F - \bar{F}_l$ is a nondecreasing right-continuous function. Next, show that the measures $m_F$ is actually equal to $m_F^* + m_{F_l}$, and in view of part (5), deduce that actually $m_F = m_{F^l}$. Calculate $m_F(I)$, for every $I$ in $\mathcal{I}$ and check that $m_F^* \leq m_G^*$.

\textbf{Proof.} (1) To check that $G$ is right-continuous, pick a sequence $x_n > x$ with $x_n \to x$. The definition of limit from the right of $G$ implies that for each $x_n$ there exists $y_n$ in $(x_n, x_n + 1/n)$ such that $|F(y_n) - G(x_n)| \leq 1/n$. Since $y_n \to x$, we have $F(y_n) \to G(x)$ and thus $G(x_n) \to G(x)$.

Similarly, if $x_n \to x$ and $x_n < x$ then we can find $y_n$ in $(x_n, x)$ such that $|F(y_n) - G(x_n)| \leq 1/n$. Again, since $y_n \to x$, we have $F(y_n) \to F(x-)$ and $G(x_n) \to G(x)$, i.e., $F(x) = G(x)$. Analogously, if $H(x) = F(x-) = \lim_{y \to x^-} F(y)$ then $H$ is left-continuous and $F(x+) = H(x+)$.

Because $F$ and $G$ are nondecreasing, the expressions $\bar{m}_F$ and $\bar{m}_G$ define two additive measure on $\mathcal{I}$. Actually, $\bar{m}_G$ is $\sigma$-additive but $\bar{m}_F$ is not necessarily $\sigma$-additive. Indeed, if $I = (a, b]$ is decomposed into a disjoint sequence of intervals $\{I_n\}$, $I_n = (a_n, b_n]$, $I = \bigcup_n I_n$, then the right-continuity of $G$ ensures that for every $\varepsilon > 0$ there exits $b'_n > b_n$ and $a' > a$ such that $G(b'_n) < G(b_n) + 2^{-n} \varepsilon$ and $G(a') < G(a) + \varepsilon$. Hence

$$\varepsilon + \sum_n \bar{m}_G(I_n) \geq \sum_n \bar{m}_G((a_i, b'_i]) \quad \text{and} \quad \bar{m}_G((a, b]) \leq \varepsilon + \bar{m}_G((a', b]],$$
and again, there exists a finite subcover of the compact \([a', b]\) and the additivity of \(m_G\) implies
\[
\varepsilon + \sum_n \bar{m}_G(I_n) \geq \bar{m}_G(I) - \varepsilon,
\]
which proves the \(\sigma\)-additivity of \(\bar{m}_G\). If \(F\) is right-continuous at each point in \(\{b_n\}\) and in \(a\), then we also have \(\sum_n \bar{m}_F((a_n, b_n]) = \bar{m}_F((a, b])\).

The expression \(\bar{m}_F\) and \(\bar{m}_G\) are invariant if we add a constant to the function \(F\), so that we can assume \(F(0+) = G(0) = 0\) without any loss of generality.

(2) Since both, \(\bar{m}_F\) and \(\bar{m}_G\) are additive measures on the semi-ring \(\mathcal{I}\), Caratheodory’s construction Proposition 2.11 shows that any Borel set is \(m_F^*\)-measurable and \(m_G^*\)-measurable, for the outer (induced) measures. Since \(\bar{m}_G\) is also \(\sigma\)-additive, we have \(m_G((a, b]) = G(b) - G(a)\), but a priori, we get only \(m_F((a, b]) \leq F(b) - F(a)\).

(3) Given a singleton \(\{x\}\) and \(\varepsilon > 0\) there exists a sequence \(\{I_n\}\) in \(\mathcal{I}\) covering \(\{x\}\) such that
\[
\varepsilon + \sum_n \bar{m}_F(I_n) \geq \sum_n \bar{m}_F(I_n)
\]
and because \(x\) belongs to some \(I_n = (a_n, b_n]\), i.e., \(a_n < x \leq b_n\), we have \(F(b_n) - F(a_n) \geq F(x) - F(x)\), with \(F(x -) = \lim_{y \to x^-} F(y)\). Hence \(m_F^*(\{x\}) = F(x) - F(x)\). The same argument applied to \(G\) yields \(m_G^*(\{x\}) = G(x) - G(x^-)\).

Since \(G(x) \geq F(x)\) and \(F(x^-) = G(x^-)\) we get
\[
m_F(\{x\}) = F(x) - F(x^-) \leq G(x) - G(x^-) = m_G(\{x\}).
\]
It is clear that the only singletons can be atoms, and if \(F\) is left-continuous then \(m_F(\{x\}) = F(x) - F(x^-) = 0\), i.e., \(m_F\) has no atoms. The inequality \(m_F(\{x\}) \leq m_G(\{x\})\) shows that any atom of \(m_F\) is also an atom of \(m_G\). By definition, a point \(\alpha\) is an atom for \(m_G\) if and only if \(\alpha\) is a point of discontinuity for \(G\) (or equivalently, if and only if \(\alpha\) is a point of discontinuity for \(F\)).

For \(m_G\) the situation is easier, we know that \(m_G((a, b]) = G(b) - G(a)\). Thus, if \(I\) is an interval, say \((a, b]\) or \([a, b)\) or \([a, b]\) or \([a, b]\), and \(m_G(\{a\}) = m_G([b]) = 0\) then \(m_G(I) = m_G((a, b]) = m_G([a, b]) = G(b) - G(a)\).

(4) For every bounded interval \([a, b]\) and \(c > 0\), there are only finite many atoms \(\{a_i\} \subset [a, b]\) (possible none) with \(m_F(\{a_i\}) \geq c > 0\), otherwise, the inequality
\[
m_F([a, b]) \geq \sum_{i=1}^n m_F(\{a_i\}) \geq nc
\]
yields a contradiction. Thus, we conclude that \(F\) and \(G\) can have only countable many points of discontinuities. Now, if \(\{\alpha_k\}\) the sequence of all atoms in \([a, b]\) then
\[
\sum_k m_F(\{\alpha_k\}) \leq \sum_k m_G(\{\alpha_k\}) \leq m_G([a, b]) = G(b) - G(a),
\]
i.e., the series is convergence and therefore the remainder vanishes. Hence
\[ \sum_n b_n \mathbb{1}_{b_n < \varepsilon} \to 0 \text{ as } \varepsilon \to 0, \]
for either \( b_n = m_F(\{\alpha_n\}) \) or \( b_n = m_G(\{\alpha_n\}) \).

(5) For any sequences of points \( \{x_n\}, \{y_n\} \) with \( x_n \leq x, \ y_n \geq y \)
y and \( y_n \to y \), we have \( \bigcup_n (0,x_n) = (0,x), \bigcup_n [x_n,0] = [x,0], \bigcup_n (0,y_n) = (0,y) \) and \( \bigcup_n (y_n,0] = (y,0] \). This implies that \( H_l \) is left-continuous and \( H_r \) is right-continuous.

Since \( F \) takes finite values, the same applies for \( H_l \) and \( H_r \), and the series
defining either \( H_l \) or \( H_r \) are (absolutely) convergent. For any \( x \) in \([a,b]\) with
\( a > 0 \), we have
\[ 0 \leq H_l(x) - H_l^+(x) = \sum_{a \leq \alpha_i \leq b} f_i \mathbb{1}_{f_i < \varepsilon} \]
and we deduce that \( H_l^+(x) \to H_l(x) \), uniformly for \( x \) in \([a,b] \), as \( \varepsilon \to 0 \). In a
similar manner we treat all the others cases.

(6) If \( F = H_r \) is a purely jump function as above then \( \bar{m}_F(I) = \sum_{\alpha_i \in I} f_i \), for
every \( I \in \mathcal{I} \) and therefore, for \( E = \bigcup_n I_n \) (i.e., \( E \) in \( E_\sigma \)), we have
\[ \sum_n \bar{m}_F(I_n) = \sum_n \sum_{\alpha_i \in I_n} f_i = \sum_{\alpha_i \in E} f_i \geq \sum_{\alpha_i \in A} f_i, \]
for \( A \subset E \). Hence, from the definition of \( m_F^* \), we obtain \( m_F^*(A) \geq \sum_{\alpha_i \in A} f_i \), and in particular, \( m_F^*(E) = \sum_{\alpha_i \in E} f_i \), for every \( E \) in \( E_\sigma \).

Extending the definition of \( m_F(A) = \sum_{\alpha_i \in A} f_i \) to any subset \( A \) of \( \Omega \), we
have a measure defined on \( 2^\Omega \). Since \( m_F^* = \bar{m}_F \) on the \( \pi \)-class \( E \), by means
of Proposition 2.15 we deduce that \( m_F^* = \bar{m}_F \) on the \( \sigma \)-algebra of all \( m_F^* \)-measurable
sets.

On the other hand, if \( F = H_l \) is a purely jump function as above then \( \sum_{0 < \alpha_i < x} f_i \to 0 \) as \( x \downarrow 0 \), i.e., \( F(0+) = 0 \). Since \( f_i > 0 \) the function \( F \)
is nondecreasing, and the summation expression for \( \alpha_i \) in \((0,x)\) if \( x > 0 \) or in
\([x,0)\) if \( x \leq 0 \) ensures the left-continuity. It is clear that, without any lost of
generality, we may suppose that all points \( \alpha_i \) are distinct. Moreover, because we
implicitly assume that \( F \) takes finite values, the series \( \sum_{0 \leq \alpha_i \leq b} f_i \) is convergent
for every real values \( a \) and \( b \).

Thus, to show that \( m_F^*([a,b]) = 0 \), pick \( \varepsilon > 0 \) and find a finite number \( d \)
of terms such that \( \sum_{a < \alpha_i \leq b, i \geq d} f_i < \varepsilon \). Now, denote by \( a < \alpha'_1 < \alpha'_2 < \cdots < \alpha'_d \leq b \) those \( d \) points with their associated values \( \{f'_i, i = 1, \ldots, d\} \). For
any integer \( k \geq 1 \) sufficiently large, define the finite sequence of disjoint intervals
\( J^k_i = (\alpha'_i, \alpha'_i + 1/k] \), for \( i = 1, \ldots, d \) and its union \( J^k = \bigcup_{i=1}^d J^k_i \). Since \( \mathcal{I} \)
is a semi-ring, for every \( k \), the difference \( \bigcup_n [a,b] \smallsetminus J^k \) can be written as a disjoint
(finite) union of intervals in \( \mathcal{I} \), and as \( k \to \infty \), we find a sequence \( \{[a_n, b_n]\} \) of
disjoint intervals satisfying \( \bigcup_n [a_n, b_n] = \bigcup_{n \leq N_k} [a_n, b_n] \), for some finite index
\( N_k \). Because \( d \) is finite (fixed, independent of \( k \)), we have \( \bigcap_k J_k = \emptyset \), and
therefore \( [a,b] = \bigcup_{n \geq N_k} [a_n, b_n] \). The expression of the (left-continuous) function \( F \)
yields \( m_F([a_n, b_n]) = F(b_n) - F(a_n) = \sum_{n \leq \alpha_i < b_n} f_i \), for every \( n \geq 1 \), and by
construction, each \( \alpha_i \) with \( i = 1, \ldots, d \) does not belong to the (disjoint) union

[Preliminary]  
Menaldi  
November 11, 2016
\[ \bigcup_{n \leq N_k} [a_n, b_n[ \text{, for every } k. \text{ Thus,} \]

\[
\sum_n \bar{m}_F([a_n, b_n]) = \sum_n \sum_{a_n \leq \alpha_i < b_n} f_i \leq \sum_{a < \alpha_i \leq b, i \geq d} f_i < \varepsilon,
\]

which implies that \( m_F^*(|a, b]) < \varepsilon \), i.e., \( m_F^*(|a, b]) = 0 \). Hence, Remark 2.7 shows that \( m_F^* = 0 \).

(7) First, as in part (5), the expression of \( \bar{F}_r \) defines a nondecreasing purely jump right-continuous function. Similarly to what follows, we can show that \( \bar{F}_l = F - \bar{F}_r \) is a nondecreasing left-continuous function.

Now, based on part (5), the expression

\[
F_l(x) = \sum_{0 < y < x} [G(y) - F(y)] \quad \text{or} \quad F_l(x) = \sum_{x \leq y \leq 0} [F(y) - G(y)],
\]

depending on the sign of \( x \), also defines a nondecreasing purely jump left-continuous function. By construction this nondecreasing function \( F_l \) the jump \( F_l(x+) - F_l(x) \) equals to \( G(x) - F(x) = F(x+) - F(x) \). Moreover, the equality

\[
\lim_{y \to x^+} F_l(y) = F_l(x) + F(x+) - F(x)
\]

implies that the function \( F_r = F - F_l \) is right-continuous, and if \( x < y < x' \) then \( F_l(x) - F_l(x') \geq G(y) - F(y) \geq G(x-) - F(x'+) = F(x+) - F(x'+) \), i.e., \( F_r \) is also nondecreasing.

Now, we may consider the outer measures \( m_{F_r}^* \) and \( m_{F_l}^* \) (and the measures \( m_{F_r} \) and \( m_{F_l} \)) induced by \( F_r \) and \( F_l \), respectively. By means of Proposition 2.15, to prove that \( m_F = m_{F_r} + m_{F_l} \), we need only to show that they agree on the class \( \mathcal{E} = \mathcal{I} \), i.e., that \( m_F^* = m_{F_r}^* + m_{F_l}^* \) on the semi-ring \( \mathcal{I} \). The inf definition implies that \( m_F^* \geq m_{F_r}^* + m_{F_l}^* \), and for the converse inequality, pick an interval \( I \) in \( \mathcal{I} \) and \( \varepsilon > 0 \), then there exists a sequence \( \{I_n\} \subset \mathcal{I} \) such that \( I \subset \bigcup_n I_n \) and

\[
\varepsilon + m_{F_l}^*(I) \leq \sum_n \bar{m}_{F_l}(I_n) \leq \sum_n \bar{m}_{F_l}(I \cap I_n)
\]

In view of Remark 2.8, we may assume (without any loss of generality) that the sequence is disjoint, i.e., \( I = \bigcup_n I_n \) and therefore \( m_{F_r}^*(I) = \sum_n \bar{m}_{F_r}(I \cap I_n) \).

Hence

\[
\varepsilon + m_{F_l}^*(I) + m_{F_r}^*(I) \geq \sum_n \bar{m}_{F_l}(I \cap I_n) + \sum_n \bar{m}_{F_r}(I \cap I_n) = \sum_n \bar{m}_F(I \cap I_n) \geq m_F^*(I),
\]

i.e, \( m_{F_l}^*(I) + m_{F_r}^*(I) \geq m_F^*(I) \), for every \( I \) in \( \mathcal{I} \).

At this point we have \( m_F = m_{F_r} + m_{F_l} \) (but not necessarily \( m_F^* = m_{F_r}^* + m_{F_l}^* \)), and because \( F_l \) is a nondecreasing purely jump left-continuous function.
as in part (5), we obtain $m_{F_i}^* = 0$, i.e., $m_F = m_{F_r}$. In particular, we deduce

$$m_F([a,b]) = (F(b) - F(a)) - \sum_{a \leq y < b} [G(y) - F(y)],$$

for any $b > a > 0$. Hence, $m_F([a,b]) \leq m_G([a,b])$ and then Remark 2.7 implies the same for the outer measures, i.e., $m_F^* \leq m_G^*$.

The reader may want to check the book Taylor [114, Section 4.10, pp. 218–224].

### Invariant under Translations

**Exercise 2.31.** Let $D$ be a closed set in $\mathbb{R}^d$ and $f : D \to \mathbb{R}^n$ be a continuous function. Prove that if $A \subset \mathbb{R}^d$ is a $\mathcal{F}_\sigma$-set (i.e., a countable union of closed sets) so is $f(D \cap A)$. Also show that if $f$ maps sets of (Lebesgue) measure zero into sets of measure zero, then $f$ also maps measurable sets into measurable sets.

**Proof.** Since any function preserves unions (but not intersections) and the class $\mathcal{F}_\sigma$ is closed under the formation of unions, we deduce that the class $\mathcal{E}$ of all sets $E \subset \mathbb{R}^d$ for which $f(D \cap E) \in \mathcal{F}_\sigma$ is closed under the formation countable unions.

The continuity of $f$ implies that $f(K \cap D)$ is a compact set, for any compact set $K$ of $\mathbb{R}^d$, i.e., any compact set belongs to $\mathcal{E}$. Moreover, since any closed set $A$ in $\mathbb{R}^d$ is a countable union of compact sets, we deduce that $f(D \cap A)$ is a $\mathcal{F}_\sigma$-set.

Next, given a measurable set $A \subset D$, by means of Exercise 2.20 part (4), there exists a $\mathcal{F}_\sigma$-set $F$ such that $F \subset A$ and $\ell(A \setminus F) = 0$. Since $f$ maps sets of (Lebesgue) measure zero into sets of measure zero, we obtain $\ell(f(A \setminus F)) = 0$, which implies that the set $f(A \setminus F)$ is measurable. Because $f$ preserves $\mathcal{F}_\sigma$-set, the set $f(F)$ is also an $\mathcal{F}_\sigma$-set, thus, $f(F)$ is measurable. Finally, the equality $f(A) = f(F) \cup f(A \setminus F)$ shows that $f(A)$ is measurable. □

**Exercise 2.32.** First give details on how to show that an hyperplane in $\mathbb{R}^d$ has zero Lebesgue measure. Second, verify that if $B$ and $\overline{B}$ denote the open and closed ball of radius $r$ and center $c$ in $\mathbb{R}^d$, then $m_d(B) = m_d(\overline{B}) = c_d r^d$, where $c_d$ is the Lebesgue measure of the unit ball.

**Proof.** First, if the hyperplane $A$ in $\mathbb{R}^d$ then we can rotate and translate $A$ to become an hyperplane $B$ coincident with one of the main planes, i.e., with equation $x_i = 0$ for some $i$. Next, in view of Theorem 2.27, if $m(B) = 0$ then $m(A) = 0$, so we need to consider only the case of an hyperplane perpendicular to one of the axis, e.g., the first axis with equation $A = \{x \in \mathbb{R}^d : x_1 = 0\}$. Moreover, if

$$A_n = \{x \in \mathbb{R}^d : x_1 = 0, -n < x_i \leq n, \forall i = 2, \ldots, d\}$$

then $A_n \subset A_{n+1}$, $A = \bigcup_n A_n$, and the monotone continuity from below of the Lebesgue measure $m$ implies $m(A) = \lim_n m(A_n)$, i.e., we are reduce to show
that \( m(A_n) = 0 \), for every \( n \). Therefore, for every \( \varepsilon > 0 \), we construct an interval

\[
I_{\varepsilon,n} = \{ x \in \mathbb{R}^d : -\varepsilon < x_1 \leq \varepsilon, -n < x_i \leq n, \forall i = 2, \ldots, d \} \]

satisfying \( m(I_{\varepsilon,n}) = (2\varepsilon)(2n)^{d-1} \). Hence, \( m(A_n) \leq (2\varepsilon)(2n)^{d-1} \), for every \( \varepsilon > 0 \), which implies that \( m(A_n) = 0 \), for every \( n \).

Certainly, by means of a translation, we may consider only balls with center at the origin. If \( \omega_d \) is the Lebesgue measure of the unit open ball \( B_1 \) then for any \( r > 0 \), the linear transformation \( x \mapsto rx \), maps \( B_1 \) onto the open ball \( B_r \) of radius \( r \). Therefore, in view of Theorem 2.27, we obtain \( m_d(B_r) = r^d m_d(B_1) = r^d c_d \).

Now, if \( \bar{B}_r \) denotes the closed ball of radius \( r \) then \( \bar{B}_r = \bigcap_n B_{r+1/n} \) and the monotone continuity from above of the Lebesgue measure \( m_d \) yields

\[
m_d(\bar{B}_r) = \lim_n m_d(B_{r+1/n}) = \lim_n (r + 1/n)^d c_d = r^d c_d, \]
i.e., \( m_d(B_r) = m_d(\bar{B}_r) \).

Note, that we may calculate the value of \( c_d = m_d(B_1) \) by using the multidimensional Riemann integral, provided we first establish its connection with the Lebesgue measure, see later sections. \( \square \)

### Vitali’s Covering

**Exercise 2.33.** Actually, give more details relative to the statements in Remark 2.39, i.e., given a subset \( A \) of \( \mathbb{R}^d \) and a number \( r > m^*(A) \) there exist a sequence \( \{B_n\} \) of balls and a sequence \( \{Q_n\} \) of cubes (with edges parallel to the axis) such that \( A \subset \bigcup_n B_n, A \subset \bigcup_n Q_n, r > \sum_n m(B_n) \) and \( r > \sum_n m(Q_n) \). Moreover, relating to the above sequence of cubes, (1) can we make a choice of cubes intersecting only on boundary points (i.e., non-overlapping), and (2) can we take a particular type of cubes defining the class \( \mathcal{E} \) so that the cubes can be chosen disjoint? Finally, compare these assertions with the those in Exercises 2.27 and 2.21.

**Proof.** By using the max-norm \( |x|_\infty = \max\{|x_1|, \ldots, |x_d|\} \) instead of the usual Euclidean norm \( |x| = \sqrt{|x_1|^2 + \cdots + |x_d|^2} \) in \( \mathbb{R}^d \), we see that the ball becomes cubes (with edges parallel to the axis). Therefore, Corollary 2.35 (and Remark 2.37) applies with either ball or cubes, i.e., for every open set \( O \) and any \( \varepsilon > 0 \) there exists a sequence \( \{B_i\} \) of disjoint closed balls and a sequence \( \{Q_i\} \) of disjoint closed cubes (with edges parallel to the axis) both with radii \( 0 < r_i < \varepsilon \) and such that \( B_i \subset O, Q_i \subset O, m(O \setminus \bigcup_i B_i) = 0 \) and \( m(O \setminus \bigcup_i Q_i) = 0 \).

If \( \{B_i\} \) is a sequence of closed ball covering \( A, A \subset \bigcup B_i \), then for every \( \varepsilon > 0 \), we can replace \( B_i \) with an open ball \( B_i' \) with a larger radius to have \( B_i \subset B_i' \) and \( \sum_i m(B_i') < \varepsilon + \sum_i m(B_i) \). Similarly for a sequence \( \{Q_i\} \) of closed cubes.

Since the Euclidean norm and the max-norm are equivalent, i.e., \( |x| \leq \sqrt{d}|x|_\infty \) and \( |x|_\infty \leq |x| \), which means that a cube \( (|x|_\infty \leq r) \) of radius \( r \) is covered by a ball \( (|x| \leq \sqrt{dr}) \) of radius \( \sqrt{dr} \) and ball \( (|x| \leq r) \) of radius \( r \) is
covered by a cube \(|x|_\infty \leq r\) of radius \(r\). For a ball \(B_r\) of radius \(r\) and for a cube \(Q_r\) of radius \(r\) (i.e., size \(2r\)) we have \(m(B_r) = r^dm(B_1)\), \(m(B_1) = c_d\) (which will be calculated after developing the integral), and \(m(Q_r) = r^dm(Q_1)\), \(m(Q_1) = 2^d\). Therefore, if \(\mathcal{E}_b\) and \(\mathcal{E}_q\) denote the classes of all closed balls and all closed cubes (with edges parallel to the axis), respectively, the corresponding outer measures induced by the Lebesgue measure \(m\) (as defined in Section 2.5, Proposition 2.26) via Caratheodory’s extension Theorem 2.9 (i.e., \(m_b^*\) and \(m_q^*\)) have the same sets of measure zero. Moreover, we may use balls or cubes that are not necessarily closed, and the sets of measure zero are the same as those with the Lebesgue measure \(m\).

In view of Remark 2.10, we know that \(m_b^* = m\) on \(\mathcal{E}_b\) and \(m_q^* = m\) on \(\mathcal{E}_q\). However, because neither \(\mathcal{E}_b\) nor \(\mathcal{E}_q\) are \(\pi\)-classes, we cannot ensure (a priori) that any set in \(\mathcal{E}_b\) (or \(\mathcal{E}_q\)) is \(m_b^*\)-measurable (or \(m_q^*\)-measurable).

Nevertheless, we can show that \(m_b^* = m_q^* = m^*\). For instance, to prove that \(m_b^* \geq m_q^*\) we have to establish a way of obtaining a covering by cubes from any covering by balls. To this effect, for every \(\varepsilon > 0\) and for any sequence \(\{B_i\}\) of closed balls with \(\sum_i m(B_i) < \infty\), first, we can find a sequence \(\{B_i'\}\) of open balls satisfying \(B_i \subset B_i'\) and \(\sum_i m(B_i') < \varepsilon/2 + \sum_i m(B_i)\); and by Corollary 2.35 applied to the open set \(O = \bigcup_i B_i'\), we can obtain a sequence \(\{Q_i\}\) of disjoint closed cubes such that \(Q = \bigcup_i Q_i \subset \bigcup_i B_i'\) and \(m(O \setminus Q) = 0\). Essentially, by the definition of sets of measure zero, we can find a sequence \(\{Q_i'\}\) of closed cubes such that \(O \setminus Q \subset \bigcup_i Q_i'\) and \(\sum_i m(Q_i') < \varepsilon/2\). Collecting all pieces, the double sequence \(\{Q_i\}\) and \(\{Q_i'\}\) is a covering of \(O = \bigcup_i B_i' \supset \bigcup_i B_i\) satisfying

\[
m_q^*(\bigcup_i B_i) \leq \sum_i m(Q_i) + \sum_i m(Q_i') < \varepsilon + \sum_i m(B_i).
\]

This yields \(m_q^*(\bigcup_i B_i) \leq m_b^*(\bigcup_i B_i)\), for any covering by closed balls, i.e., \(m_q^*(A) \leq m_b^*(A)\), for every \(A \subset \mathbb{R}^d\). Certainly, the other cases are treated with the same technique.

For (1), we realize that we may consider \(d\)-cubes with rational endpoints, so that any intersection of \(d\)-cubes can be expressed as a finite union of non-overlapping \(d\)-cubes. Hence, for any sequence \(\{Q_n\}\) of \(d\)-cubes we can find another sequence \(\{Q_n'\}\) of non-overlapping \(d\)-cubes such that \(\bigcup_i Q_n = \bigcup_i Q_i'\).

Regarding (2), consider the class \(\mathcal{C}_s\) of all \(d\)-cubes, semi-open from the left, semi-closed from the right, with edges parallel to the axis and with rational endpoints, i.e., \((a, b]\) with \(a, b \in \mathbb{Q}^d\) and \(b - a = r\), for every \(i\). This class is not closed under the formation of finite intersections, but the class \(\mathcal{R}\) of all disjoint finite unions of sets in \(\mathcal{C}_s\) is the minimal ring containing \(\mathcal{C}_s\). Moreover, \(\mathcal{R}\) is also the ring generated by the semi-ring \(\mathcal{I}_d\) of \(d\)-intervals, semi-open from the left and semi-closed from the right. Thus, the outer measures induced by the hyper-volume on either the class \(\mathcal{C}_s\) or the ring \(\mathcal{R}\) are the same Lebesgue outer measure, i.e., for every set \(A \subset \mathbb{R}^d\) with \(m(A) < \infty\) and every \(\varepsilon > 0\) there exists a sequence \(\{Q_n\}\) of disjoint \(d\)-cubes in \(\mathcal{C}_s\) such that \(\sum_n m(Q_n) < m^*(A) + \varepsilon\).  \(\square\)
Exercises - Chapter (3)
Measures and Topology

(3.1) Borel Measures

(3.2) On Metric Spaces

Exercise 3.1. Verify Remark 3.11 and give more details on the passage from a finite measure to a $\sigma$-finite measure in the proof of Proposition 3.9.

Proof. Since $\mu$ is $\sigma$-finite, there exists a sequence $\{O_n\}$ of open sets such that $\Omega = \bigcup_n O_n$, $O_n \subset O_{n+1}$ and $\mu(O_n) < \infty$. Now, each open set $O_n$ is a countable intersection of closed sets, i.e., $O_n = \bigcap_n C_{n,i}$, $C_{n,i} \subset C_{n,i+1}$, and $\mu(C_{n,i}) \leq \mu(O_n) < \infty$. Hence, relabeling the double sequence $\{C_{n,i}\}$, we find a sequence $\{C_n\}$ of closed sets that $\Omega = \bigcup_n C_n$, $C_n \subset C_{n+1}$ and $\mu(C_n) < \infty$.

Therefore, for any Borel set $B$ and every $\varepsilon > 0$, the monotone continuity from below of $\mu$ ensures that $\mu(B) = \lim_n \mu(B \cap C_n)$, meaning that we can deduce that inner regular Borel measure, i.e., (3.4), from $\mu(B \cap C_n) = \sup\{\mu(K) : K \subset B \cap C_n, K \text{ compact}\}$, $\forall B \in B(\Omega)$,

for every $n \geq 1$. Since this last equality has been established during the proof of Proposition 3.9, we conclude.

Note that the tightness condition (3.5), namely,

$\forall \varepsilon > 0$ there exists a compact $K_\varepsilon$ such that $\mu(\Omega \setminus K_\varepsilon) \leq \varepsilon$,

has been obtained only when $\mu(\Omega) < \infty$. \hfill $\square$

Exercise 3.2. With the notation of the previous Remark 3.14, under the conditions of Theorem 3.3. and assuming the restriction $\bar{\mu} = \mu^*|_B$ on the Borel $\sigma$-algebra $B$ is inner regular:

(1) Verify that $\mu^*$ is $\sigma$ super-additive, i.e., if $A_i \in 2^\Omega$ and $A = \bigcup_{i=1}^\infty A_i$ then $\mu^*(A) \geq \sum_{i=1}^\infty \mu^*(A_i)$.

(2) Show that if $A \in 2^\Omega$ and $\{C_i\}$ is a disjoint sequence of closed sets with $C = \bigcup_i C_i$ then $\mu^*(A \cap C) = \sum_i \mu^*(A \cap C_i)$.  

303
(3) Prove that if the topological space \( \Omega \) is countable at infinity (i.e., there exists a monotone increasing sequence \( \{K_n\}, K_n \subset K_{n+1} \) of compact sets such that \( \Omega = \bigcup_n K_n \)) then \( \bar{\mu} \) is inner regular.

(4) Assuming that \( \bar{\mu} \) is inner regular, prove that if \( A \) is a \( \mu^* \)-measurable set then \( \mu^*(A) = \mu_*(A) \), and conversely, if \( A \in 2^\Omega \), \( A = \bigcup_k A_k \), \( \mu^*(A_k) = \mu_*(A_k) < \infty \) then \( A \) is \( \mu^* \)-measurable.

(5) Prove that for any two disjoint subsets \( A \) and \( B \) of \( \Omega \) we have \( \mu_*(A \cup B) \leq \mu_*(A) + \mu^*(B) \leq \mu^*(A \cup B) \).

(6) Assume that \( E \) is \( \mu^* \)-measurable. Show \( \mu_*(A) = \mu^*(E) - \mu^*(E \setminus A) \), that for every \( A \subset E \) with \( \mu^*(E \setminus A) < \infty \).

(7) Discuss alternative definitions of \( \mu_* \), for instance, when \( \bar{\mu} \) is not necessarily inner regular or when \( \Omega \) is not a topological space.

**Proof.** Recall that \( \mu \) is a \( \sigma \)-additive and \( \sigma \)-finite set function (pre-measure) on the algebra \( \mathcal{A} \) generated by all open (or compact) sets, and

\[
\mu^*(A) = \inf \{ \mu(O) : O \supset A, \text{O open} \},
\]

\[
\mu_*(A) = \sup \{ \mu(K) : K \subset A, \text{K compact} \},
\]

for every \( A \) in \( 2^\Omega \).

(1) Since \( \mu \) is additive on the algebra \( \mathcal{A} \), it is also monotone and the sup-expression shows that \( \mu_* \) is also monotone. Thus, to check that \( \mu_* \) is \( \sigma \) super-additive, we need to prove only that \( \mu^* \) is super-additive. To this purpose, pick any disjoint finite sequence \( \{A_i : i = 1, \ldots, n\} \) of sets in \( \mathcal{A} \) and \( r_i < \mu_*(A_i) \) to find compact sets \( K_i \subset A_i \) such that \( r_i < \mu(K_i) \). The set \( K = \sum_{i=1}^n K_i \) is compact, \( K \subset A = \sum_{i=1}^n A_i \) and

\[
\mu_*(A) \geq \mu(K) = \sum_{i=1}^n \mu(K_i) > \sum_{i=1}^n r_i,
\]

which shows that \( \mu_*(A) \geq \sum_{i=1}^n \mu_*(A_i) \).

(2) In view of the \( \sigma \) super-additivity, we have to show only that if \( \{C_i : i \geq 1\} \) is a disjoint sequence of closed sets and \( C = \sum_i C_i \) then \( \mu_*(A \cap C) \leq \sum_i \mu_*(A \cap C_i) \), for every \( A \subset \Omega \). Now, for any \( r < \mu_*(A \cap C) \) there exists a compact set \( K \subset A \cap C \) such that \( r < \mu(K) \), and because \( K_i = K \cap C_i \subset A \cap C_i \) is a compact set and \( \mu \) is \( \sigma \)-additive, we deduce

\[
r \leq \mu(K) = \sum_i \mu(K_i) \leq \sum_i \mu_*(A \cap C_i),
\]

which implies \( \mu_*(A \cap C) \leq \sum_i \mu_*(A \cap C_i) \).

(3) Under the assumptions of Theorem 3.3, we have (3.2), i.e.,

\[
\bar{\mu}(B) = \sup \{ \mu(C) : C \subset B, C^c \in \mathcal{T} \}, \quad \forall B \in \mathcal{B}(\Omega).
\]

Since for every \( \mu^* \)-measurable set \( A \) there exist Borel sets \( B_1 \) and \( B_2 \) such that \( B_1 \supset A \supset B_2 \) and \( \mu^*(B_1 \setminus B_2) = 0 \) (see Remark 3.2), for every \( r < \bar{\mu}(A) = \ldots \)
\( \tilde{\mu}(B_2) \) there exists a closed set \( C \subset B_2 \subset A \) such that \( r < \tilde{\mu}(C) \), i.e., the representation of \( \tilde{\mu} \) in term of the sup remains valid for any \( \mu^* \)-measurable set \( B = A \).

Assuming that there exists a monotone sequence \( \{K_n\} \) of compact sets, and using the fact that any closed set of a compact set is compact, we have

\[
\tilde{\mu}(B \cap K_n) = \sup\{\tilde{\mu}(C) : C \subset B \cap K_n, C^c \in T\} = \mu_*(B \cap K_n) \leq \mu_*(B).
\]

As \( n \to \infty \) we get \( \tilde{\mu}(B) \leq \mu_*(B) \), i.e., \( \tilde{\mu} \) is inner regular.

(4) Since \( \tilde{\mu} \) is inner regular, \( \tilde{\mu} = \mu_* \) on the \( \sigma \)-algebra of all \( \mu^* \)-measurable sets, i.e., \( \mu^*(A) = \mu_*(A) \) for every \( \mu^* \)-measurable set \( A \). Conversely, if \( A = \bigcup_k A_k, \mu^*(A) = \mu_*(A) < \infty \) then there exist a Borel set \( B \supset A_k \) such that \( \tilde{\mu}(B_k) = \mu^*(A_k) \) and monotone increasing sequences \( \{C_{k,n}\} \) of closed sets such that \( C_{k,n} \subset A_k \) and \( \mu_*(A_k) = \lim_n \tilde{\mu}(C_{k,n}) \). Thus, the Borel set \( C_k = \bigcup_n C_{n,k} \subset A_k \) and \( \tilde{\mu}(C_k) = \mu^*(A_k) \), i.e., \( \mu^*(B_k \setminus C_k) = 0 \) and so, \( A_k \) and \( A \) are \( \mu^* \)-measurable.

(5) Given two disjoint subsets \( A \) and \( B \) of \( \Omega \), for every \( r < \mu_*(A \cup B) \) there exists a compact set \( K \subset A \cup B \) such that \( r < \tilde{\mu}(K) \); and also, for any \( \varepsilon > 0 \) there exists a Borel set \( B' \supset B \) and an open set \( O \supset B' \) such that \( \mu^*(B) = \tilde{\mu}(B') \), and \( \tilde{\mu}(O \setminus B') < \varepsilon \). Since \( K \setminus O \) is a compact subset of \( A \), we have

\[
r < \tilde{\mu}(K) \leq \tilde{\mu}(K \setminus O) + \tilde{\mu}(O) \leq \mu_*(A) + \mu^*(B) + \varepsilon,
\]

which yields \( \mu_*(A \cup B) \leq \mu_*(A) + \mu^*(B) \).

Similarly, for any \( \varepsilon > 0 \) there exists a Borel set \( E \supset A \cup B \) and an open set \( O \supset E \) such that \( \mu^*(A \cup B) = \tilde{\mu}(E) \) and \( \tilde{\mu}(O \setminus E) < \varepsilon \); and also, for every \( r < \mu_*(A) \) there exists a compact set \( K \subset A \) such that \( r < \tilde{\mu}(K) \). Since \( O \setminus K \) is an open set containing \( B \) and \( K \subset O \), we have

\[
\mu^*(A \cup B) - \varepsilon \geq \tilde{\mu}(O) = \tilde{\mu}(K) + \tilde{\mu}(O \setminus K) > r + \mu^*(B),
\]

which implies that \( \mu^*(A \cup B) \geq \mu_*(A) + \mu^*(B) \).

(6) Apply part (5) to the sets \( A \) and \( B = E \setminus A \) to get

\[
\mu_*(A \cup E) \leq \mu_*(A) + \mu^*(E \setminus A) \leq \mu^*(A \cup E),
\]

Since \( E \) is \( \mu^* \)-measurable and \( A \subset E \) we obtain \( \mu_*(E) = \mu_*(A \cup E) \) and \( \mu^*(A \cup E) = \mu^*(E) \), which implies \( \mu_*(A) = \mu^*(E) - \mu^*(E \setminus A) \) if \( \mu^*(E \setminus A) < \infty \).

(7) Only part (4) requires \( \tilde{\mu} \) to be inner regular. An alternative definitions could be

\[
\mu_*(A) = \sup\{\mu(C) : C \subset A, K \text{ closed}\},
\]

or even, in the non-topological case,

\[
\mu_*(A) = \sup\{\mu(B) : B \subset A, B \in \mathcal{A}\},
\]

where \( \mathcal{A} \) is an initial algebra or \( \sigma \)-algebra. Under this setting, we can check that parts (1), (2), (4) (without the assumption of inner regular), (5) and (6) remain valid, e.g., see Halmos [57, Section III.14, pp. 58–62]. \( \square \)
(3.3) On Locally Compact Spaces

Exercise 3.3. Elaborate the previous Remark 3.18, i.e., by means of Theorem 3.3 and Proposition 3.9 state and prove under which precise conditions a set function \( \mu \) satisfying the assumptions of Proposition 3.15 can be extended to a (regular and) inner regular Borel measure.

Proof. The initial set function \( \mu \) is only additive, but in view of Proposition 3.15, \( \mu \) results \( \sigma \)-additive on the semi-ring \( S \) (and supposing to satisfy the assumptions of the Caratheodory’s construction in Proposition 2.11, see also Remark 2.12), we can extend \( \mu \) to a measure \( \bar{\mu} \) defined on the \( \sigma \)-ring (which is a \( \sigma \)-algebra) generated by \( S \). Now, assuming that \( S \) generated the Borel \( \sigma \)-algebra \( B \), the extension \( \bar{\mu} \) is a regular Borel measure.

To check that \( \bar{\mu} \) is also inner regular, we have to establish the sup representation (3.6) for any Borel set, i.e., if

\[
\nu(B) = \sup\{\bar{\mu}(K) : K \subset B, K \text{ is compact}, \mu(K) < \infty\}, \quad \forall B \in B,
\]

then we should prove that \( \nu = \bar{\mu} \).

From Proposition 3.15, we know that \( \nu(R) = \mu(R) \) for every \( R \) in the ring \( R \) generated by \( S \). Moreover, we can show that if \( A \) belongs to the class \( R' = \{A \in 2^\Omega : A = \sum_{n=1}^{\infty} A_n, A_n \in R\} \) of countable disjoint unions of sets in \( S \) then \( \nu(A) = \bar{\mu}(A) \). Indeed, as in the proof of Proposition 3.15, for every \( \varepsilon > 0 \) there exists compact sets \( K_n \subset A_n \) in \( R \) with \( \mu(K_n) < \infty \) and \( \mu(A_n) = \nu(A_n) < \mu(K_n) + \varepsilon 2^{-n} \). Since \( \bigcup_{n \leq N} K \subset A \) is a compact set with finite measure belonging to the ring \( R \), we deduce

\[
\sum_{n \leq N} \mu(A_n) < \sum_{n \leq N} \mu(K_n) + \varepsilon \leq \nu(A) + \varepsilon,
\]

i.e., \( \bar{\mu}(A) \leq \nu(A) \), for every \( A \) in \( R' \).

Actually, for a Borel set \( B \) we have \( \nu(B) = \bar{\mu}(B) \) if and only if the Borel set \( B \) is semi-finite (relative to \( \mu \)), i.e., if and only if there exists a sequence \( \{R_k\} \) in the ring generated by \( S \) such that \( R_k \subset B \), \( \mu(R_k) < \infty \) and \( \bar{\mu}(B) = \lim_k \mu(R_k) \).

For instance, if we know that any closed set is semi-finite then under the conditions of Theorem 3.3, we can approximate any Borel set from below with a sequence of closed sets to deduce that any Borel set is semi-finite, i.e., \( \bar{\mu} \) is inner regular. Alternatively, if topological space \( \Omega \) is countable at infinity (i.e., there exists a monotone increasing sequence \( \{K_n\}, K_n \subset K_{n+1} \) of compact sets such that \( \Omega = \bigcup_n K_n \)) then \( \bar{\mu} \) is also inner regular, see Exercise 3.2, part (3).

Certainly, if the \( \Omega \) is a Polish space then Proposition 3.9 implies that \( \bar{\mu} \) is an inner regular Borel measure.

Exercise 3.4. Regarding the sup-formula (3.6), prove that if \( \tilde{S} \) is the ring generated by the semi-ring \( S \) and \( \mu : S \to [0, \infty] \) is an additive set function (which is uniquely extended by additivity to the ring \( \tilde{S} \)) and such that there exists a compact class \( K \subset \tilde{S} \) satisfying

\[
\mu(S) = \sup\{\mu(K) : K \subset S, \mu(K) < \infty, K \in K\}, \quad \forall S \in \tilde{S}.
\]

(C.23)
then $\mu$ is necessarily $\sigma$-additive on $\mathcal{S}$.

**Proof.** Recall that $\mathcal{K} \subset 2^\Omega$ is called a compact class if (a) $\mathcal{K}$ is stable under finite intersections and unions, and (b) for any sequence $\{K_i : i \geq 1\} \subset \mathcal{K}$ with $\bigcap_i K_i = \emptyset$ there exists an index $n$ such that $\bigcap_{i=1}^n K_i = \emptyset$.

The arguments are the same as in Proposition 3.15. Indeed, if $S = \sum_{k \geq 1} S_k$ with $S_k$ and $S$ in the ring $\mathcal{S}$ and $\mu(S) < \infty$ then $\mu(S) = \sum_k \mu(S_k)$ is equivalent to $\lim_n \mu(R_n) = 0$ with $R_n = S \setminus \sum_{k < n} S_k$, $R_n \subset \mathcal{S}$ and $R_n \supset R_{n+1} \downarrow \emptyset$.

Now, for each $n$ and for every $\varepsilon > 0$ there exists a set $K_n$ in the compact class $\mathcal{K}$ such that $K_n \subset A_n$ and $\mu(R_n) < \mu(K_n) + 2^{-n}\varepsilon$. Since $\bigcap_{n=1}^\infty K_n = \emptyset$, there exists an index $N$ such that $\bigcap_{n<N} K_n = \emptyset$, which yields, for any $k \geq N$,

$$R_k = R_k \setminus \bigcap_{k<n} K_n = \bigcup_{n<k} (R_k \setminus K_n) \subset \bigcup_{n<k} (R_n \setminus K_n).$$

Since $\mu$ is additive on the ring $\mathcal{S}$, we have

$$\mu(R_k) \leq \sum_{n<k} \mu(R_n \setminus K_n) \leq \sum_{n<k} 2^{-n}\varepsilon < \varepsilon,$$

which implies that $\lim_k \mu(R_k) = 0$.

If $\mu(S) = \infty$, $S = \sum_{k \geq 1} S_k$ with $S_k$ and $S$ in the ring $\mathcal{S}$ then for every $r > 0$ there exists a set $K$ in the compact class $\mathcal{K}$ such that $K \subset S$ and $r < \mu(K) < \infty$. Since $K = \sum_n S_k \cap K$ and the $\sigma$-additivity holds for sets of finite measure, we have $r < \mu(K) = \sum_n \mu(S_k \cap K) \leq \sum_n \mu(S_k)$, i.e., $\mu(S) = \infty = \sum_n \mu(S_k)$. □

**Exercise 3.5.** In a topological space, (1) Verify that any family of compact sets is indeed a compact class of sets in the above sense. Now, let $\mu$ a finite countable additive (non-necessarily $\sigma$-additive, just additive) and $\sigma$-finite measure defined on a algebra $\mathcal{A} \subset 2^\Omega$. Suppose that there exists a compact class $\mathcal{K}$ such that for every $\varepsilon > 0$ and any set $A$ in $\mathcal{A}$ with $\mu(A) < \infty$ there exists $A_\varepsilon \subset A$ in $\mathcal{A}$ and $K_\varepsilon$ in $\mathcal{K}$ such that $A_\varepsilon \subset K_\varepsilon \subset A$ and $\mu(A \setminus A_\varepsilon) < \varepsilon$. (2) Using a technique similar to Proposition 3.15, show that $\mu$ is necessarily $\sigma$-additive. (3) Also, prove the representation

$$\mu(A) = \sup\{\mu(K) : K \subset A, K \in \mathcal{K}\}, \quad \forall A \in \mathcal{A},$$

provided $\mathcal{K} \subset \mathcal{A}$, see Bogachev [16, Section 1.4, pp. 13-16].

**Proof.** Because compact sets have the finite-intersection property, any class of compact sets (stable under the formation of finite unions and intersections) is indeed a compact class as defined early.

We can make arguments similar to Proposition 3.15 to prove the assertion (2). Indeed, only the monotone continuity from above should be proved, and for a monotone sequence $\{A_n : n \geq 1\} \subset \mathcal{A}$ with $\mu(A_n) < \infty$ and $A_n \supset A_{n+1} \downarrow \emptyset$, and for every $\varepsilon > 0$, there exist sequences $\{B_n\} \subset \mathcal{A}$ and $\{K_n\} \subset \mathcal{K}$ such that $B_n \subset K_n \subset A_n$ and $\mu(A_n \setminus B_n) < 2^{-n}\varepsilon$. Since $\bigcap_n K_n \subset \bigcap_n A_n = \emptyset$ and $\mathcal{K}$
is a compact class, exists an index $N$ such that $\bigcap_{n<N} K_n = \emptyset$, which yields $\bigcap_{n<N} B_n = \emptyset$, and for any $k \geq N$,

$$A_k = A_k \setminus \bigcap_{k<n} B_n = \bigcup_{n<k} (A_k \setminus B_n) \subset \bigcup_{n<k} (A_n \setminus B_n).$$

Since $\mu$ is additive on the algebra $\mathcal{A}$, we have

$$\mu(A_k) \leq \sum_{n<k} \mu(A_n \setminus B_n) \leq \sum_{n<k} 2^{-n} \epsilon < \epsilon,$$

which implies that $\lim_k \mu(A_k) = 0$, i.e., $\mu$ is also $\sigma$-additive on sets with finite measure.

Since $\mu$ is $\sigma$-finite, there exists a sequence $\{\Omega_n\}$ in $\mathcal{A}$ such that $\Omega = \bigcup_n \Omega_n$ and $\mu(\Omega_n) < \infty$. If $\mu(A) = \infty$, $A = \sum_k A_k$ with $A$ and $A_k$ in $\mathcal{A}$ then the $\sigma$-additivity for sets of finite measure implies

$$\mu(A \cap \Omega_n) = \sum_k \mu(A_k \cap \Omega_n) \leq \sum_k \mu(A_k), \quad \forall n,$$

and the monotone continuity from below yields $\infty = \mu(A) = \lim_n \mu(A \cap \Omega_n)$. Hence $\mu$ is $\sigma$-additivity on $\mathcal{A}$.

To check (3), pick $A$ in $\mathcal{A}$ with finite measure, $\mu(A) < \infty$, and $\epsilon > 0$ to find $A_\epsilon$ in $\mathcal{A}$ and $K_\epsilon$ in $\mathcal{K}$ such that $A_\epsilon \subset K_\epsilon \subset A$ and $\mu(A \setminus A_\epsilon) < \epsilon$. Therefore $\mu(A) \leq \mu(K_\epsilon) + \epsilon$, i.e., the sup representation holds for any set $A$ in $\mathcal{A}$ with finite measure.

If $\mu(A) = \infty$ then use a sequence with finite measure to find that the sup representation holds for $A \cap \Omega_n$, and conclude as $n \to \infty$.

**(3.4) Product Measures**

**Exercise 3.6.** Recall that a compact metrizable space is necessarily separable and so it satisfies the second axiom of countability (i.e., there exists a countable basis). A topological space $\Omega$ is called a Lusin space if it is homeomorphic (i.e., there exists a bi-continuous bijection function between them) to a Borel subset of a compact metrizable space. Certainly, any Borel set in $\mathbb{R}^d$ is a Lusin space and actually, any Polish space (complete separable metrizable space) is also a Lusin space. Similarly to Proposition 3.24, if $\{\Omega_i : i = 1, 2, \ldots, n, \ldots\}$ is a sequence of Lusin spaces then verify (1) that the product space $\Omega = \prod_i \Omega_i$ is also a Lusin space. Let $\{\nu_n\}$ be a sequence of $\sigma$-additive set functions defined on the algebra $\mathcal{A}^n$, generated by all cylinder sets of dimension at most $n$. Verify (2) that $\nu^n$ can be (uniquely) extended to a measure on the $\sigma$-algebra $\sigma(\mathcal{A}^n)$ generated by $\mathcal{A}^n$, and that the particular case of a finite product measures $\prod_{i \leq n} \mu_i$ can be taken as $\nu^n$. Assume that $\{\nu_n\}$ is a compatible sequence of probabilities, i.e., $\nu_{n+1}(A) = \nu_n(A)$ for every $A$ in $\mathcal{A}^n$ and $\nu_n(\Omega) = 1$, for every $n \geq 1$. Now, prove (3) that if each $\Omega_i$ is a compact metrizable space then there exists a unique probability measure $\nu$ defined on the product $\sigma$-algebra $\prod_i \mathcal{B}(\Omega_i)$ such that
\( \nu(A) = \nu_n(A) \), for every \( A \) in the algebra \( \mathcal{A}^n \). Finally, show (4) the same result, except for the uniqueness, when each \( \Omega_i \) is only a Lusin space. In probability theory, this construction is referred to as Daniell-Kolmogorov Theorem, e.g., see Rogers and Williams [97, Vol 1, Sections II.3.30-31, pp. 124–127].

Proof. (1)-(2) This is a direct consequence of Tychonoff’s Theorem (the product of compact spaces is a compact space). The extension to \( \sigma(\mathcal{A}^n) \) is given by means of Caratheodory’s extension Proposition 2.11. It is also clear that Proposition 3.24 proves that we may take \( \nu_n = \prod_{i=1}^n \mu_i \), and assuming that \( \mu_i(\Omega_i) = 1 \). Moreover, these product measures make a compatible sequence of probabilities.

(3) Assuming that each \( \Omega_i \) is a compact metrizable space, Tychonoff’s Theorem shows that the product \( \Omega \) is also a compact metrizable space. Therefore, each \( \nu_n \) can be identified to a probability measure on the finite product space \( \prod_{i=1}^n \Omega_i \), and because each compact metrizable space has a countable basis, the \( \sigma \)-algebra \( \sigma(\mathcal{A}^n) \) generated by \( \mathcal{A}^n \) is identified with the Borel \( \sigma \)-algebra \( \mathcal{B}(\prod_{i=1}^n \Omega_i) \). Hence, Theorem 3.3 and Corollary 3.4 ensure that any \( \nu_n \) is tight, i.e. for every \( \varepsilon > 0 \) there exists a compact set \( K_{\varepsilon,n} \) in \( \prod_{i=1}^n \Omega_i \) such that \( \nu_n(K_{\varepsilon,n}) > 1 - \varepsilon \), or equivalently, for every set \( B \) in \( \sigma(\mathcal{A}^n) \) there exists a compact set \( K \) in \( \sigma(\mathcal{A}^n) \) such that \( K \subset B \) and \( \nu_n(B \setminus K) < \varepsilon \).

Consider the algebra \( \mathcal{A}^\infty \) generated by of all cylinder sets, i.e., \( A \) belongs to \( \mathcal{A}^\infty \) if and only if \( A \) belongs to \( \mathcal{A}^n \) for some dimension \( n \). Thus, the compatibility condition allows us to define an additive set function \( \nu \) on \( \mathcal{A}^\infty \), namely, \( \nu(A) = \nu_n(A) \), for every \( A \) in \( \mathcal{A}^n \). In view of Caratheodory’s extension Proposition 2.11, we need to show only that \( \nu \) is \( \sigma \)-additive on \( \mathcal{A}^\infty \) to be able to extend \( \nu \) to the product \( \sigma \)-algebra \( \prod_i \mathcal{B}(\Omega_i) \). Moreover, because \( \mathcal{A}^\infty \) is an algebra, the \( \sigma \)-additivity reduces to the monotone continuity from above at \( \emptyset \), i.e., if \( \{A_k\} \) is a decreasing sequence in \( \mathcal{A}^\infty \) such that \( \bigcap_k A_k = \emptyset \) then \( \lim_k \nu(A_k) = 0 \).

To this purpose, for each \( \varepsilon > 0 \) there exist compact sets \( K_k \subset A_k \) in \( \mathcal{A}^\infty \) such that \( \nu(A_k \setminus K_k) < \varepsilon 2^{-k} \). The finite-intersection property applied to the sequence of compact sets \( \{K_k\} \) implies that \( \bigcap_{k \leq n} K_k = \emptyset \), for some finite index \( n \). Hence, the inclusion

\[
A_k = \bigcap_{i \leq k} A_i \setminus \bigcap_{i \leq k} K_i \subset \bigcup_{i \leq k} (A_i \setminus K_i), \quad \forall k \geq n,
\]

implies \( \nu(A_k) \leq \sum_{i \leq k} \nu(A_i \setminus K_i) \leq \varepsilon \), i.e., \( \lim_k \nu(A_k) = 0 \).

The existence of a countable basis for the topology of each space \( \Omega_i \) ensures that the product \( \sigma \)-algebra \( \prod_i \mathcal{B}(\Omega_i) \) is the Borel \( \sigma \)-algebra \( \mathcal{B}(\Omega) \), thus the uniqueness follows from Caratheodory’s construction.

(4) For each Lusin space \( \Omega_i \) we can find an homeomorphism \( h_i : \Omega_i \to Y_i \) with \( Y_i \) being a Borel subset of a compact metrizable space \( X_i \). Thus, we can construct the Borel measurable (injective) function \( \varphi : \Omega \to \prod_i X_i = X \) with the coordinates functions \( h_i \), namely, \( \varphi(\omega) = (h_1(\omega_1), h_2(\omega_2), \ldots) \). If \( \mathcal{A}^\infty (\mathcal{A}^n) \) denotes the algebra generated by all cylinder sets (of dimension at most \( n \)) in the space \( X = \prod_i X_i \) then, by construction, any set \( A \) in \( \mathcal{A}^n \) has the form \( A = A^n \times \prod_{i> n} X_i \), with \( A^n \) in \( \mathcal{B}(\prod_{i \leq n} X_i) \), so that the pre-image \( \varphi^{-1}(A) \)
is an element of $\sigma$-algebra $\sigma(A^n)$ generated by $A^n$. Thus, the sequence of probabilities on $X$ defined by

$$\tilde{\nu}_n(A) = \nu_n(\varphi^{-1}(A)), \quad \forall A \in \tilde{A}^n,$$

satisfies the conditions of the previous part (3), and so, there exists a probability measure $\tilde{\nu}$ on $B(X)$ such that $\tilde{\nu}(A) = \tilde{\nu}_n(A)$, for every $A$ in $\tilde{A}^n$, for every $n \geq 1$. Hence,

$$\nu(A) = \tilde{\nu}(\varphi(A)), \quad \forall A \in B(\Omega),$$

yields the desired extension.

The uniqueness of the probability $\nu$ on the product $\sigma$-algebra $\prod_i B(\Omega_i)$ follows from Caratheodory’s construction, but without the second axiom of countability, the Borel $\sigma$-algebra $B(\Omega)$ could be strictly larger that the product $\sigma$-algebra $\prod_i B(\Omega_i)$.

**Exercise 3.7.** In a product space $\Omega = \prod_{i \in I} \Omega_i$ the projection mappings $\pi_J : \Omega \to \Omega_J = \prod_{i \in J} \Omega_i$ are defined as $\pi_J((\omega_i : i \in I)) = (\omega_i : i \in J)$, for any subindex $J$ of $I$. Assume that the index $I$ is uncountable, first (1) prove that a set $A$ belongs to the product $\sigma$-algebra of the uncountable product space $\Omega$ if and only if for some countable subset $J$ of indexes the projection $\pi_J(A)$ belongs to the product $\sigma$-algebra of the countable product space $\Omega_J$ and $\pi_{I \setminus J}(A) = \Omega_{I \setminus J}$. Next (2) show that Proposition 3.24 can be extended to the case of an uncountable product of probability spaces. Finally, (3) discuss how to extend Remark 3.25 and Exercise 3.6 to the uncountable case.

**Proof.** (1) For a family of measurable spaces $\{(\Omega_i, F_i) : i \in I\}$, the product $\sigma$-algebra $\prod_{i \in I} F_i$ is generated by the cylinder sets $Q = \prod_{i \in I} Q_i$, where $Q_i$ belongs to $F_i$ and $Q_i = \Omega_i$ for every $i$ in $I \setminus J$ with a finite subset of indexes $J \subset I$. Therefore, any countable operation involving cylinder sets will results in a set having all components $Q_i = \Omega_i$, expect for a countable set of indexes $J$. Thus, denote by $F$ the class of all subsets $F$ of $\Omega = \prod_{i \in I} \Omega_i$ such that for some countable subset of indexes $J$ of $I$ the the projection $\pi_J(F)$ belongs to the product $\sigma$-algebra of the countable product space $\Omega_J$ and $\pi_{I \setminus J}(F) = \Omega_{I \setminus J}$, or equivalently, $F = \pi_J^{-1}(E)$ for some countable subset of indexes $J$ of $I$ and $E$ in the countable product $\sigma$-algebra $\prod_{i \in J} F_i$. It is clear that $F$ is a sub-$\sigma$-algebra of the product $\sigma$-algebra $\prod_{i \in I} F_i$, i.e., $F = \prod_{i \in I} F_i$.

(2) Really, it is a matter of notation. Given a family $\{\mu_i : i \in I\}$ of tight Borel regular measures on $\Omega_i$ with $\mu_i(\Omega_i) = 1$. For any finite subset of indexes $J$ of $I$, define the product set function $\mu_J(Q) = \prod_{i \in J} \mu_i(Q_i)$ on the semi-ring $S_J$ of cylinder sets of the form $Q = \prod_{i \in I} Q_i$, where $Q_i$ belongs to $F_i$ and $Q_i = \Omega_i$ for every $i$ in $I \setminus J$. From the finite-product construction of Proposition 3.22, the product set function $\mu_J$ can be extended to a (unique) probability measure on the sub-$\sigma$-algebra $F_J = \pi_J^{-1}(\prod_{i \in J} F_i)$ of the product $\sigma$-algebra $F = \prod_{i \in I} F_i$. Moreover, in view of Proposition 3.24 and with the same notation, the finite set of indexes $J$ can be taken countable, i.e., for any countable index $K$ there
exists a (unique) probability measure \( \mu_K \) on \( \mathcal{F}_K = \pi_K^{-1}(\prod_{i \in K} \mathcal{F}_i) \) such that \( \mu_K = \mu_J \) on \( \mathcal{F}_J \) for every finite subset of indexes \( J \) of \( K \). In view of part (2), a set \( F \) belongs to uncountable product \( \sigma \)-algebra \( \mathcal{F} \) if and only if \( F = \pi_J^{-1}(E) \) for some \( E \in \prod_{i \in J} \mathcal{F}_i \), i.e., \( F \) belongs to \( \mathcal{F}_J \). Hence, we can define a (unique) the uncountable product probability measure \( \mu \) on \( \mathcal{F} \) by means of countable product probability measures \( \mu_K \) on \( \mathcal{F}_K \), via a countable product \( \prod_{i \in K} \mathcal{F}_i \). Furthermore, the probability measure \( \mu \) is uniquely determinate by the equality \( \mu = \mu_J \) on the semi-ring \( \mathcal{S}_J \), for any finite set of indexes \( J \) of \( I \).

(3) Indeed, following the arguments in part (2), both, Remark 3.25 and Exercise 3.6 are extended to the uncountable case. The only point to stress is the fact that the uncountable product of Borel \( \sigma \)-algebras is not necessarily equals to the Borel \( \sigma \)-algebra of the uncountable product space, even if each space \( \Omega_i \) has a countable basis.

From part (1), we see that the uncountable product of Borel \( \sigma \)-algebras is a rather small \( \sigma \)-algebra for the relatively large uncountable product space \( \Omega \). In probability theory, this is usually solved by endowing the space \( \Omega \) with a stronger (than the product) topology on a reduced subspace \( \Omega_0 \subset \Omega \) with full outer measure, i.e., \( \mu^*(\Omega_0) = 1 \), such that the cylinder sets generate the (new) Borel \( \sigma \)-algebra on \( \Omega_0 \).

Exercise 3.8. Formalize the previous comments, e.g., (1) prove that for any measurable set \( F \) such that \( 0 < c < \mu(F) < \infty \) there is at most a finite family of atoms \( A \subset F \) satisfying \( c \leq \mu(A) \leq \mu(F) \); and, assuming that \( \mu \) is \( \sigma \)-finite, (2) deduce that there are no atoms of infinite measure and there is a countable family (possible finite or empty) containing all atoms, moreover (3) discuss in some details the existence of the partition \( \{F_i : i = 1, \ldots, k\} \).

Proof. (1) If \( A_i \), for \( i = 1, \ldots, k \) are atoms inside \( F \) with \( \mu(A_i) \geq c \) then the additivity of \( \mu \) implies that \( \sum_{i=1}^n \mu(A_i) \leq \mu(F) \). Hence \( k \leq \mu(F)/c \), which implies that any set with finite measure may contain at most countable many atoms.

(2) Next, if \( \mu \) is \( \sigma \)-finite then \( \Omega = \bigcup_n \Omega_n \) with \( \mu(\Omega_n) < \infty \). Each \( \Omega_n \) may contain at most countable many atoms, so there are only countable many atoms (possible a finite number or none) in \( \Omega \). On the other hand, if there is an atom \( A \) with \( \mu(A) = \infty \) then \( \mu \) could not be a \( \sigma \)-finite measure.

(3) From part (1), for every \( \varepsilon > 0 \) there exists at most a finite number of sets \( F_i, i = 1, \ldots, r \) such that \( \mu(F_i) > \varepsilon \) and \( F \setminus \bigcup_{i=1}^r F_i \) does not contain any atom \( A \) with measure greater than \( \varepsilon \).

The next point is to show that any measurable set \( E \subset F \) contains a measurable subset \( B \) with \( 0 < \mu(B) \leq \varepsilon \). Indeed, if such set \( B \) does not exist then \( B \) is not an atom and therefore, there is a measurable set \( B_1 \subset B \) such that \( 0 < \mu(B_1) < \mu(B) \), and the same argument applies to \( B \setminus B_1 \), and by induction, we construct a sequence of disjoint measurable sets \( \{B_k\} \) with \( \sum_k \mu(B_k) < \infty \), which yields the contradiction \( \mu(B_k) < \varepsilon \) for \( k \) sufficiently large.

Now, if

\[
\beta(E) = \sup\{\mu(B) : B \in \mathcal{F}, \mu(B) \leq \varepsilon\}
\]
then the previous assertion implies, by induction, there is a sequence of disjoint measurable sets \( \{E_n\} \), \( E_n \subset F \) such that \( \frac{1}{2} \beta(F \setminus \bigcap_{i=1}^{n} E_i) \leq \mu(E_{n+1}) \leq \varepsilon \).

Remark that \( \sum_n \mu(E_n) \leq \mu(F) < \infty \) and \( \mu(E_0) \leq 2\mu(E_{n+1}) \) with \( E_0 = F \setminus \bigcap_{i=1}^{\infty} E_i \) to obtain \( \mu(E_0) = 0 \), which yields that there exists a finite index \( m \) such that \( \sum_{i>m} \mu(E_i) < \varepsilon \). Hence if \( F_{r+i} = E_i \) with \( i = 1, \ldots, m \) and \( F_k = \bigcup_{i>m} E_i \) with \( k = r + m + 1 \) then \( \{F_i : i = 1, \ldots, k\} \) is the desired partition, see Dunford and Schwartz [40, Vol I, Lemma IV.9.7, pp. 308-309].
Exercises - Chapter (4) 
Integration Theory

(4.1) Definition and Properties

Exercise 4.1. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space.

(1) Show that if $f$ is an integrable function and $N$ is a set of measure zero then
\[ \int_N f \, d\mu = 0. \]

(2) Prove that if $f$ is a strictly positive integrable function and $E$ is a measurable set such that
\[ \int_E f \, d\mu = 0 \]
then $E$ is a set of measure zero.

(3) Suppose that an integrable function $f$ satisfies
\[ \int_E f \, d\mu = 0, \]
for every measurable set $E$, deduce that $f = 0$ a.e.

(4) If $f$ is a measurable function and $g$ is an integrable function such that $|f| \leq g$ a.e., then $f$ is also an integrable function.

Proof. Recall that essentially from Fatou lemma (Theorem 4.6), see Remark 4.8, we deduce that if $f$ is an integrable (measurable, respectively) function and $g$ is a measurable (the $\sigma$-algebra of measurable set contains all sets of measure zero, respectively) function, and $f = g$ a.e. then $g$ is integrable (measurable, respectively) and they have the same integral. Note that if the measure space $(\Omega, \mathcal{F}, \mu)$ is not complete then $f = g$ a.e. (almost everywhere) means that there exists a (measurable) set $N$ of measure zero such that $f = g$ outside $N$; which is not necessarily equivalent to the fact that the set $\{ x \in \Omega : f(x) \neq g(x) \}$ is has measure zero (implying that it has to be measurable).
(1) If \( f \) is integrable and \( N \) a set of measure zero (implicitly assumed to be measurable!) then \( f1_N = 0 \) a.e. and because the integral of the null function is zero, we conclude.

(2) Similarly, if \( f \) is a strictly positive integrable function and \( E \) is a measurable set then the set \( E_n = \{ x \in E : |f(x)| > 1/n \} \) is measurable, \( E_n \subset E_{n+1} \), \( E = \bigcup_n E_n \) and

\[
0 = \int_E f \, d\mu \geq \int_{E_n} \frac{1}{n} \, d\mu \geq 0,
\]

which proves that \( \mu(E_n) = 0 \) and therefore \( \mu(E) = 0 \).

(3) If \( f \) is as indicated then take \( E_n^+ = \{ x \in \Omega : f(x) > 1/n \} \) to deduce that

\[
0 = \int_{E_n^+} f \, d\mu \geq \int_{E_n} \frac{1}{n} \, d\mu \geq 0,
\]

which implies that \( \mu(E_n^+) = 0 \) and hence the set \( \{ x \in \Omega : f(x) > 0 \} \) has zero measure. Similarly, for the set \( \{ x \in \Omega : f(x) < 0 \} \), i.e., \( f = 0 \) a.e.

(4) We have established that changing the values of a function in a set of measure zero does not change its character (measurability, integrability or the value of its integral). Since

\[
\int |f| \, d\mu \leq \int g \, d\mu < \infty
\]

we deduce that \( f \) is integrable.

\[\square\]

Exercise 4.2. Give examples of sequences \( \{f_n\} \) of real-valued measurable functions satisfying one following conditions (a) \( \{f_n\} \) is pointwise decreasing to 0, but the integral of \( f_n \) does not converge to 0; (b) \( f_n \geq 0 \) is integrable for every \( n \) and the inequality in Fatou's Lemma holds strictly; (c) the integral of \( f_n \leq 0 \) is a bounded numerical sequence, \( f_n \) converges pointwise to an integrable function \( f \), but the integral of \( f_n \) does not converge to the integral of \( f \).

Proof. (a) On \( \mathbb{R} \) with the Lebesgue measure, the functions defined by \( f_n(x) = 1_{[|x|>n]} \) satisfy \( f_n \geq f_{n+1}, \quad f_n(x) \to 0 \) as \( n \to \infty \), for each \( x \) in \( \mathbb{R} \), but the integrals of \( f_n \) are all infinite.

(b) On \( \mathbb{R} \) with the Lebesgue measure, the functions defined by \( f_n(x) = [1 + (-1)^n]1_{\{ |x|<n \}} \) satisfy \( f_n \geq 0, \quad \lim \inf f_n(x) = 0, \) for every \( x \), but the integral of \( f_n \) is equal to \( 2n \).

(c) On \( \mathbb{R} \) with the Lebesgue measure, the functions defined by \( f_n(x) = -n1_{\{0<|x|<1/n \}} \) satisfy \( f_n(x) \to 0 \) as \( n \to \infty \), for each \( x \) in \( \mathbb{R} \), the functions \( f_n \) have all integrals equal to 2, so that integral of \( f_n \) does not converge to the integral of \( f \).

\[\square\]

Exercise 4.3. Given a nonnegative measurable function \( h \) on a measure space \((\Omega, \mathcal{F}, \mu)\), define the set function

\[
\lambda(A) = \int_A h \, d\mu, \quad \forall A \in \mathcal{F}.
\]
Prove that $\lambda$ is a measure and that

$$\int_{\Omega} f \, d\lambda = \int_{\Omega} f h \, d\mu,$$

for every nonnegative measurable function $f$.

**Proof.** The monotone convergence implies the $\sigma$-additivity of $\lambda$, i.e., given a sequence of measurable sets $\{A_n\}$ with $A = \sum_n A_n$, we have $1_A = \sum_n 1_{A_n}$, and therefore

$$\lambda(A) = \int_{\Omega} \sum_n 1_{A_n} \, d\mu = \sum_n \int_{\Omega} 1_{A_n} \, d\mu = \sum_n \lambda(A_n).$$

To check the equality, for every measurable set $A$ we have

$$\int_{\Omega} 1_A \, d\lambda = \lambda(A) = \int_A h \, d\mu \int_{\Omega} 1_A h \, d\mu,$$

and by linearity, the same inequality holds for simple functions. Next, if $f$ is a nonnegative measurable function then there exists an increasing sequence $\{\varphi_n\}$ of simple functions such that $\varphi_n \to f$; hence, the equality holds for $f$. \hfill $\square$

**Exercise 4.4.** Let $E$ be a semi-ring of measurable space $(\Omega, \mathcal{F})$ and $E$ be the space of $\mathcal{E}$-step functions, i.e., functions of the form $\varphi = \sum_{i=1}^n r_i 1_{E_i}$, with $E_i$ in $\mathcal{E}$ and $r_i$ in $\mathbb{R}$. (1) Verify that $E$ is a vector lattice and any $\mathcal{R}$-step function belongs to $E$, where $\mathcal{R}$ is the ring generated by $\mathcal{E}$. Given an additive and finite measure $\mu$ on $\mathcal{E}$ we define the integral for functions in $E$ by the formula

$$I(\varphi) = \sum_{i=1}^n r_i \mu(E_i), \quad \text{if} \quad \varphi = \sum_{i=1}^n r_i 1_{E_i}.$$ 

(2) Prove that $\mu$ is $\sigma$-additive on $\mathcal{E}$ if and only if for any decreasing sequence $\{\varphi_k\}$ in $E$ such that $\varphi_k(x) \downarrow 0$ for every $x$, we have $I(\varphi_k) \downarrow 0$.

**Proof.** (1) Since any $\mathcal{R}$-step function $f$ has the form $\sum_{i=1}^n r_i 1_{R_i}$ with $R_i$ in $\mathcal{R}$ and any set $R$ in $\mathcal{R}$ can be expressed as $R = \sum_{i=1}^k E_i$ with $E_i$ in $\mathcal{E}$, we deduce that any $\mathcal{R}$-step function belongs to $E$. On the other hand, a ring is stable under the formation of unions and differences, so that $E$ is a vector space. Therefore, a function $\varphi$ belongs to $E$ if and only if $\varphi$ assumes only a finite number of values and the pre-image $\varphi^{-1}(r)$ belongs to $\mathcal{R}$ for every $r \neq 0$. Now, to realize that $E$ is also a lattice, we note that any two functions $\varphi$ and $\psi$ in $E$ can be written as $\varphi = \sum_{i=1}^n r_i 1_{E_i}$ and $\psi = \sum_{i=1}^n s_i 1_{E_i}$, with $\{E_i\}$ a finite sequence of disjoint sets in $\mathcal{E}$, i.e., where $r_i$ and $s_i$ could be zero or repeated. Thus $\max\{\varphi, \psi\} = \sum_{i=1}^n (r_i \lor s_i) 1_{E_i}$ and therefore, $E$ is a lattice.

(2) Note that finite sequence $\{E_i\}$ used to define $I$ may or may not be disjoint sets, the additivity of $\lambda$ produces a definition of $I$ independent of the way how $\varphi$ is represented. Now, let $\{\varphi_k\}$ a sequence in $E$ such that $\varphi_k(x) \downarrow 0$ for every $x$. If $\mu$ is $\sigma$-additive on the semi-ring $\mathcal{E}$ then $\mu$ can be extended to a
measure on the σ-ring generated by \( E \) and the monotone convergence can be used to deduce that \( I(\varphi_k) \downarrow 0 \). Conversely, for any sequence \( \{E_k\} \subset E \) such that \( E = \sum_k E_k \) belongs to \( E \), the functions \( \varphi_k = 1_E - \sum_{i=1}^k 1_{E_i}, \ k \geq 1 \), are in \( E \) and \( \varphi_k(x) \downarrow 0 \) for every \( x \). Hence, \( \mu(E) - \sum_{i=1}^k \mu(E_k) = I(\varphi_k \downarrow 0) \), proving that \( \mu \) is σ-additive on the semi-ring \( E \).

Exercise 4.5. Let \( E \) be a semi-ring of measurable sets in a measure space \((X, \mathcal{F}, \mu)\), and \( E \) be the class of \( E \)-step functions, see Exercise 4.4. Suppose that any function \( \varphi \in E \) is integrable and define

\[
I(\varphi) = \int_\Omega \varphi \, d\mu,
\]

A subset \( N \) of \( X \) is called a \( I \)-null or \( I \)-negligible set if there exists a increasing sequence \( \{\varphi_k\} \subset E \) and a constant \( C \) such that (a) \( \varphi_k(x) \uparrow +\infty \) for every \( x \) in \( N \) and (b) \( I(\varphi_k) \leq C \), for every \( k \geq 1 \).

(1) Prove that (a) if \( \varphi \in E \) with \( \varphi \geq 0 \) outside of a \( I \)-null set then \( I(\varphi) \geq 0 \), and (b) if \( \varphi \in E \) with \( \varphi \geq 0 \) and \( I(\varphi) = 0 \) then \( \varphi = 0 \) outside of a \( I \)-null set.

(2) Show that a set \( N \) is \( I \)-null if and only if for every \( \varepsilon > 0 \) there exists a sequence \( \{E_k\} \) in \( E \) such that (a) \( N \subset \bigcup_k E_k \) and (b) \( \sum_k I(1_{E_k}) < \varepsilon \).

Proof. (1) (a) By assumption, there exists a sequence \( \{\varphi_k\} \) of functions in the lattice \( E \) such that \( \varphi_k(x) \uparrow +\infty \) for every \( x \) in \( N \) and \( I(\varphi_k) \leq C < \infty \), for every \( k \geq 1 \), and \( \varphi(x) \geq 0 \), for every \( x \) outside of \( N \). Consider the increasing sequence of functions \( \psi_k = \varphi + \varepsilon \max\{\varphi_k, 0\}, \ k \geq 1 \), with limit \( \psi = \lim_k \psi_k \geq 0 \), for every \( \varepsilon > 0 \). Thus, the monotone convergence yields \( \lim_k I(\psi_k) = I(\lim_k \psi_k) \geq 0 \), i.e., \( I(\varphi) + \varepsilon C \geq I(\varphi) + \varepsilon \lim_k I(\varphi_k^+) \geq 0 \). Hence, as \( \varepsilon \to 0 \) we deduce \( I(\varphi) \geq 0 \). (b) Similarly, the sequence \( \{\varphi_k = k\varphi : k \geq 1\} \) satisfies \( \lim_k \varphi_k(x) = +\infty \) for every \( x \) such that \( \varphi(x) > 0 \) and \( I(\varphi_k) = kI(\varphi) = 0 \).

(2) If \( N \) is a \( I \)-null set then there exists a increasing sequence \( \{\varphi_k\} \) satisfying (a) and (b) as above. Since \( 1_N \leq \varepsilon \lim_n \varphi_k^+ \) for every \( \varepsilon > 0 \), and the pre-image \( \{x : \varphi_k(x) \geq 1\} \) forms an increasing sequence of sets belonging to the ring generated by \( E \), we can write \( \varepsilon \varphi_k \geq \sum_{i=1}^k 1_{E_i} \) with \( E_i \in E \), and thus \( N \subset \bigcup_i E_i \). Hence, \( \mu(N) \leq \sum_i \mu(E_i) \leq \varepsilon C \), which implies that \( N \) satisfies (a) and (b) of (2), after replacing \( \varepsilon \) by \( \varepsilon/C \). Conversely, take \( \varepsilon = 2^{-n} \), for every \( n = 1, 2, \ldots \), in (a) and (b) of (2) for the set \( N \) to find a double sequence \( \{E_{k,n}\} \subset E \) such that \( N \subset \bigcup_k E_{k,n} \) and \( \sum_k I(E_{k,n}) = 2^{-n} \). Thus, the functions \( \psi_n = n \sum_{k=1}^n 1_{E_{k,n}} \) belongs to \( E \) and \( \psi_n(x) \geq n \) for every \( x \) in \( \bigcup_{k=1}^n E_{k,n} \). Hence, \( \varphi_k = \sum_{n=1}^k \psi_n \) forms an increasing sequence of functions in \( E \) such that \( \varphi(x) \to +\infty \) for every \( x \) in \( \bigcap_n \bigcup_k E_{k,n} \supset N \) and \( I(\varphi_k) \leq \sum_n n2^{-n} < \infty \).

Exercise 4.6. Let \( E \) be a vector lattice of real-valued function defined on \( X \) and \( I : E \to \mathbb{R} \) be a linear and monotone functional satisfying the condition: if \( \{\varphi_n\} \) is a decreasing sequence in \( E \) such that \( \varphi_n(x) \downarrow 0 \) for every \( x \) in \( X \), then \( I(\varphi) \downarrow 0 \). A functional \( I \) as above is called a pre-integral or a Daniell integral (functional) on \( E \). Now, let \( \overline{E} \) be the semi-space of extended real-valued
functions which can be expressed as the pointwise limit of a monotone increasing sequence of functions in \(E\). Verify that repeating this procedure does not add more functions, i.e., any pointwise limit of a monotone increasing sequence of functions in \(E\) is a function in \(E\). Moreover, also verify that if \(\varphi\) and \(\psi\) belong to \(E\) and \(c\) is a positive real number then \(\varphi + c\psi\), \(\varphi \wedge \psi\) and \(\varphi \vee \psi\) belong to \(E\).

(1) Define \(I\)-null sets and prove assertion (1) as in Exercise 4.5, with \(E\) replaced with \(E\), i.e., (a) if \(\{\varphi_n\}\) is an increasing sequence in \(E\) such that \(\lim_n \varphi_n(x) < 0\) on a \(I\)-null set then \(\lim_n I(\varphi_n) \geq 0\), and (b) if \(\{\varphi_n\}\) is an increasing sequence in \(E\) such that \(\lim_n \varphi_n \geq 0\) and \(\lim_n I(\varphi_n) = 0\) then \(\lim_n \varphi_n = 0\) except in a \(I\)-null set. Similarly, show that a countable union of \(I\)-null sets is a \(I\)-null set.

(2) Prove that the limit \(I(\varphi) = \lim_n I(\varphi_n)\) if \(\varphi = \lim_n \varphi_n\) provides a unique extension of \(I\) as a semi-linear mapping from \(E\) into \((-\infty, \infty]\), i.e., (a) for any two increasing sequences \(\{\varphi_n\}\) and \(\{\psi_n\}\) in \(E\) pointwise convergent to \(\varphi\) and \(\psi\) with \(\varphi \leq \psi\) we have \(\lim_n I(\varphi_n) \leq \lim_n I(\psi_n)\), and (b) for every \(\varphi, \psi\) in \(E\) and \(c\) in \([0, \infty)\) we have \(I(\varphi + c\psi) = I(\varphi) + cI(\varphi)\), under the convention that \(0\infty = 0\). Also show the monotone convergence, i.e., if \(\{\varphi_n\}\) is an increasing sequence of functions in \(E\) then \(I(\lim_n \varphi_n) = \lim_n I(\varphi_n)\). How about if the sequence \(\{\varphi_n\}\) is increasing only outside of a \(I\)-null set?

(3) Check that if \(\varphi\) belongs to \(E\) and \(I(\varphi) < \infty\) then \(\varphi\) assumes (finite) real values except in a \(I\)-null set. A function \(f\) belongs to the class \(L^+\) if \(f\) is equal, except in a \(I\)-null set, to some nonnegative function \(\varphi\) in \(E\). Next, verify that \(I\) can be uniquely extended to \(L^+\), by setting \(I(f) = I(\varphi)\).

(4) The class \(L\) of \(I\)-integrable functions is defined as all functions \(f\) that can be written in the form \(f = g - h\) (except in a \(I\)-null set) with \(g\) and \(h\) in \(E\) satisfying \(I(g) + I(h) < \infty\); and by linearity, we set \(I(f) = I(g) - I(h)\). Verify that (a) \(I\) is well-defined on \(L\), (b) \(L\) is vector lattice space, and (c) \(I\) is a linear and monotone. Moreover, show that any function in \(L\) is (except in a \(I\)-null set) a pointwise decreasing limit of a sequence \(\{f_n\}\) of functions in \(E\) such that \(|I(f_n)| \leq C < \infty\), for every \(n\). Furthermore, prove that \(f\) belongs to \(L\) if and only if \(f^+\) and \(f^-\) belong to \(L^+\) with \(I(f^+) + I(f^-) < \infty\). A posteriori, we write \(f = f^+ - f^-\) with \(f^\pm\) in \(L^+\) and if either \(I(f^+)\) or \(I(f^-)\) is finite then \(I(f) = I(f^+) + I(f^-)\), which may be infinite.

(5) Show that if \(f = f^+ - f^-\) with \(f^+\) and \(f^-\) belonging to \(L^+\) and \(I(f^-) < \infty\) then for every \(\epsilon > 0\) there exist two functions \(g\) and \(h\) in \(E\) satisfying \(f = g - h\), except in a \(I\)-null set, \(h \geq 0\) and \(I(h) < \epsilon\). Moreover, deduce a Beppo Levi monotone convergence result, i.e., if \(\{f_k\}\) is a sequence of nonnegative (except in a \(I\)-null set) functions in \(L\) then \(\sum_k I(f_k) = I(\sum_k f_k)\), in particular, if \(\sum_k f_k < \infty\) then \(\sum_k f_k\) belongs to \(L\).

Hint: the identities \(g \wedge h = (g + h - |h - g|)/2\), \(g \vee h = (g + h + |h - g|)/2\) and \(|h - g| = g \vee h - g \wedge h\) may be of some help.

Proof. Clearly, the properties of limit prove that if \(\varphi\) and \(\psi\) belong to \(E\) and \(c\) is a positive real number then \(\varphi + c\psi\), \(\varphi \wedge \psi\) and \(\varphi \vee \psi\) belong to \(E\).
To verify that repeating the increasing limit does not add more functions, let \( \{ \varphi_n \} \) be a increasing sequence of functions in \( E \). Hence, there exist increasing sequences \( \{ \varphi_{n,k} : k \geq 1 \} \) in \( E \) such that \( \lim_k \varphi_{n,k} = \varphi_n \), and by monotonicity, also \( \lim_k \phi_{n,k} = \varphi_n \) with \( \phi_{n,k} = \max\{\varphi_1, \ldots, \varphi_{n,k}\} \). From \( \phi_{n,k} \leq \phi_{n+p,k+q} \leq \varphi_{n+p} \) we deduce \( \phi_{n,k} \leq \lim_q \lim_p \phi_{p,q} \leq \lim_p \varphi_p \) and \( \phi_{n,k} \leq \lim_{p,q} \phi_{p,q} \leq \lim_p \varphi_p \), i.e.,

\[
\lim_{n,k} \varphi_n = \lim_{n,k} \lim_k \phi_{n,k} = \lim_{n,k} \lim_k \phi_{n,k} = \lim \phi_{n,k},
\]
e.g., the diagonal sequence \( \{ \varphi_{n,n} \} \) satisfies \( \lim_n \varphi_{n,n} = \lim_n \varphi_n \) and \( \lim_n \varphi_n \) belongs to \( E \).

(1) Let \( \{ \varphi_n \} \) be an increasing sequence in \( E \) such that \( \lim_n \varphi_n(x) < 0 \) on a \( I \)-null set \( N \). Thus there exists an increasing sequence \( \{ \phi_k \} \) in \( E \) such that \( \lim_k \phi_k(x) = +\infty \) for every \( x \) in \( N \) and \( I(\phi_k) \leq C < \infty \). Hence, for every \( \varepsilon > 0 \), the functions \( \psi_n = (\varphi_n - \varepsilon \phi_n)^+ \) forms a decreasing sequence in \( E \) pointwise convergence to \( 0 \), so that \( \lim_n I(\psi_n) = 0 \). Since \( \psi_n \geq \varphi_n - \varepsilon \phi_n \), we have \( \varphi_n \geq \varphi_n^+ - \psi_n - \varepsilon \phi_n \), which yields \( I(\varphi_n) \geq I(\varphi_n^+) - I(\psi_n) - \varepsilon I(\phi_n) \), i.e., \( \lim_n I(\varphi_n) \geq 0 \), as \( \varepsilon \to 0 \).

To see that a countable union of \( I \)-null sets is again a null set, for each \( i = 1, 2, \ldots, \), let \( N_i \) be \( I \)-null a set with nonnegative increasing sequence \( \{ \varphi_{i,k} : k \geq 0 \} \) in \( E \) satisfying \( \lim_k \varphi_{i,k}(x) = +\infty \), for every \( x \) in \( N_i \) and \( I(\varphi_{i,k}) \leq 2^{-i} \).

Thus, the sequence \( \{ \psi_k = \sum_{i=1}^k \varphi_{i,k} \} \) forms an increasing sequence of functions in \( E \) such that \( \lim_k \psi_k(x) = +\infty \), for every \( x \) in \( N = \bigcup_i N_i \) and \( \sum_k I(\psi_k) \leq \sum_k 2^{-i} = 1 \), i.e., \( N \) is also a \( I \)-null set.

(2) To check (a), if \( \{ \varphi_n \} \) and \( \{ \psi_n \} \) are two increasing sequences in \( E \) with limits \( \varphi \leq \psi \) then consider the double sequence forms with \( \phi_{i,j} = \min\{\varphi_i, \psi_j\} \).

Since \( \phi_{i,j} - \psi_j \) decreases to \( 0 \) as \( i \to \infty \) the continuity, monotonicity and linearity of the functional \( I \) yields \( I(\phi_{i,j}) = \lim_j I(\phi_{i,j}) \leq \lim_k \varphi_k \).

Now, as \( j \to \infty \) we get \( I(\psi) = \lim_j I(\psi_j) \leq I(\varphi) \) as desired. Similarly, remark that \( \varphi + c \psi = \lim_k (\varphi_k + c \psi_k) \) implies \( I(\varphi + c \psi) = I(\varphi) + c I(\psi) \), which shows (b).

To check the monotone convergence, let \( \{ \varphi_n \} \) be a sequence of nonnegative functions in \( E \). Thus, for each \( n \) there exists a non nondecreasing sequences \( \{ \varphi_{n,k} : k \geq 1 \} \) of functions in \( E \) such that \( \lim_k \varphi_{n,k} = \varphi_n \). Now, for \( \psi_k = \max_{n \leq k} \{ \varphi_n \} \) we have \( \psi_k \geq \varphi_k \geq \varphi_{n,k}, \) any \( k \geq n \).

Thus, as \( k \to \infty \) we get

\[
\varphi = \lim_k \varphi_k \geq \lim_k \psi_k \geq \lim_k \varphi_{n,k} = \varphi_n,
\]

\[
\lim_k I(\varphi_k) \geq \lim_k I(\psi_k) \geq \lim_k I(\varphi_{n,k}) = I(\varphi_n),
\]

and later, as \( n \to \infty \), we deduce

\[
\varphi = \lim_k \psi_k \quad \text{and} \quad I(\varphi) = \lim_k I(\psi_k) = \lim_k I(\varphi_k),
\]

which prove that \( I(\varphi) = \lim_n I(\varphi_n) \).

Now, if the sequence \( \{ \varphi_n \} \) is increasing only outside of a \( I \)-null set, i.e., \( \varphi_n \leq \varphi_{n+1} \) outside of a \( I \)-null set \( N_n \), then (because a countable union of \( I \)-null is again
a $I$-null set) the same argument as above applies, but now, $\bar{\varphi}_k \geq \psi_k \geq \varphi_{n,k}$, any $k \geq n$, holds except in a $I$-null set $N$. Thus, to consider this case, we assume the validity of (3) below, in particular, the fact that if two functions $\phi$ and $\psi$ in $E$ satisfy $\phi = \psi$ except in a $I$-null then $I(\phi) = I(\psi)$. Therefore, the same argument as above goes through, and we get $I(\lim_k \psi_k) = \lim_k I(\psi_k) = \lim_k I(\bar{\varphi}_k)$ and $\bar{\varphi} = \lim_k \psi_k$ except in a $I$-null set, which show that $I(\bar{\varphi}) = I(\psi_k)$.

(3) The statement that if $\varphi$ belongs to $\bar{E}$ and $I(\varphi) < \infty$ then $\varphi$ assumes (finite) real values except in a $I$-null set, is essentially the definition of $I$-null sets. Now, to be able to properly extend the functional $I$ to the class $L^+$, we need to check that if $f$ and $g$ belongs to $\bar{E}$ and $f \leq g$ except in a $I$-null set then $I(f) \leq I(g)$. Indeed, if $f \leq g$ except in a $I$-null set $N$ then for any $\varepsilon > 0$ there exists an increasing nonnegative sequence $\{\phi^\varepsilon_n\}$ in $E$ such that $\lim_n \phi^\varepsilon_n(x) = +\infty$, for every $x$ in $N$ and $I(\phi^\varepsilon_n) \leq \varepsilon$. Thus, for $\varphi_n \uparrow f$ and $\psi_n \uparrow g$ we get $\lim_n \varphi_n \leq \lim_n (\psi_n + \phi^\varepsilon_n)$ everywhere, and applying the previous point (1) (a), we deduce $I(f) + \varepsilon \leq I(g)$, which yields $I(f) \leq I(g)$ as desired.

(4) The properties (a), (b) and (c) are inhered from the initial functional $I$ defined on the lattice $E$, the identities $g \wedge h = (g + h - |h - g|)/2$, $g \vee h = (g + h + |h - g|)/2$ and $|h - g| = g \vee h - g \wedge h$ are handy here. What need to be noted that because $|I(f)| < \infty$ for any $f$ in the class $L$, the function $f$ assume finite values except in a $I$-null set.

From the definition, for every function $f$ in the class $L$, there exist two increasing sequences $\{\varphi_n\}$ and $\{\psi_n\}$ such that $f = \lim_n \varphi_n - \lim_n \psi_n$ except in a $I$-null set. Thus, the double sequence formed by $\phi_{k,n} = \varphi_k - \psi_n$ satisfies $\phi_{k,n} \uparrow \varphi - \psi_n$ with $\varphi - \psi_n = f_n$ in $\bar{E}$ and $f_n \downarrow f$ except in a $I$-null set, and $\lim_n I(f_n) = I(f)$, i.e., $|I(f_n)| \leq C < +\infty$.

Now, if $f = g - h$ (except in a $I$-null set) with $g$ and $h$ in the class $L^+$ then $f^- + f^+ \leq g + h$. Thus, if $f$ belongs to $L$ then $I(g) + I(h) < \infty$ and so $I(f^-) + I(f^+) < \infty$. For the converse, we may take $g = f^+$ and $h = f^-$.

(5) Since $f^+$ and $f^-$ belong to $L^+$ there exist two increasing sequences $\{\varphi_n\}$ and $\{\psi_n\}$ with pointwise limits equal to $f^+$ and $f^-$, except in a $I$-null set. Since $I(f^-) < \infty$, for every $\varepsilon > 0$ there exists a $k = k_\varepsilon$ such that $I(f^- - \psi_k) < \varepsilon$. Thus, take $g = \lim_n (\varphi_n - \psi_k)$ and $h = \lim_n (\psi_n - \psi_k) \geq 0$ to obtain $f^- - f^+ = g - h$ with $I(h) < \varepsilon$.

To check Beppo Levi monotone convergence theorem, given a sequence $\{f_k\}$ of nonnegative (except in a $I$-null set) functions in $L$ then for every $\varepsilon > 0$ there exists a sequence $\{\bar{\varphi}_k\}$ of functions in $\bar{E}$ such that $f_k \leq \bar{\varphi}_k$, except in a $I$-null set, and $I(\bar{\varphi}_k - f_k) < \varepsilon 2^{-k}$. Since $f_k \geq 0$ we have $f_k \leq \bar{\varphi}_k^+$, except in a $I$-null set, and $I(\bar{\varphi}_k^+ - f_k) < \varepsilon 2^{-k}$. Hence, the sequence of partial sum $\sum_{k=1}^n \bar{\varphi}_k$ is increasing, except in a $I$-null set, and (2) above applies to produce $\sum_k I(\bar{\varphi}_k) = I(\sum_k \bar{\varphi}_k)$. Therefore $I(\sum_k f_k) \leq I(\sum_k \bar{\varphi}_k) = \sum_k I(f_k) + \sum_k \varepsilon 2^{-k}$, and as $\varepsilon \to 0$, we deduce $\sum_k I(f_k) = I(\sum_k f_k)$.

\[\square\]

Exercise 4.7. On a measurable space $(\Omega, \mathcal{A})$, let $f$ be a nonnegative measurable, $\mu$ be a measure and $\{\mu_n\}$ be a sequence of measures. Prove that (1) if
\[
\liminf_n \mu_n(A) \geq \mu(A), \text{ for every } A \in \mathcal{A} \text{ then }
\int f \, d\mu \leq \liminf_n \int f \, d\mu_n.
\]

(2) if \( \lim_n \mu_n(A) = \mu(A) \), for every \( A \) in \( \mathcal{A} \) and \( \mu_n(A) \leq \mu_{n+1}(A) \), for every \( A \) in \( \mathcal{A} \) and \( n \), then
\[
\int f \, d\mu = \lim_n \int f \, d\mu_n.
\]

Finally, along the lines of the dominate convergence, what conditions we need to impose on the sequence of measure to ensure the validity of the above limit? Hint: use the monotone approximation by simple functions given in Proposition 1.9, e.g., see Dshalalow [36, Section 6.2, pp. 312–326] for more comments and details.

**Proof.** (1) To check (a) we take a nonnegative measurable simple function \( \varphi = \sum_{i=1}^k \alpha_i 1_{A_i} \) to see that
\[
\liminf_n \int \varphi \, d\mu_n = \liminf_n \sum_{i=1}^k \alpha_i \mu_n(A_i) \geq \sum_{i=1}^k \alpha_i \liminf_n \mu_n(A_i) \geq \sum_{i=1}^k \alpha_i \mu(A_i) = \int \varphi \, d\mu.
\]

Hence, using an increasing sequence \( \{ f_k \} \) of simple functions to approximate any nonnegative measurable function \( f \), we get
\[
\int f \, d\mu = \lim_k \int f_k \, d\mu \leq \liminf_n \int f_k \, d\mu_n \leq \liminf_n \int f \, d\mu_n
\]
as desired.

For (2), essentially the same argument shows that for any nonnegative measurable simple function \( \varphi \) the equality holds. Thus, because \( f_k \leq f \) we get
\[
\int f \, d\mu = \lim_k \int f_k \, d\mu = \lim_k \lim_n \int f_k \, d\mu_n \leq \lim_n \int f \, d\mu_n.
\]
But \( \mu_n \leq \mu \) yields
\[
\int f \, d\mu = \lim_k \int f_k \, d\mu \geq \lim_k \int f_k \, d\mu_n = \int f \, d\mu_n,
\]
and the equality follows.

If the convergence is not necessarily monotone, but dominated, i.e., there exists a measure \( \nu \) such that \( \mu_n(A) \leq \nu(A) \) for every \( A \) in \( \mathcal{A} \), then we can use part (1) with \( \nu \pm \mu_n \) to have
\[
\int f \, d(\nu \pm \mu) \leq \liminf_n \int f \, d(\nu \pm \mu_n),
\]
which implies

$$\int_{\Omega} f \, d\mu = \lim_{n} \int_{\Omega} f \, d\mu_n,$$

for every nonnegative $\nu$-integrable function $f$. \hfill \Box

**Exercise 4.8.** Let $(X, \mathcal{X}, \mu)$ be a measure space, $(Y, \mathcal{Y})$ be a measurable space and $\psi : X \to Y$ be a measurable function. Verify that the set function $\mu_{\psi} : B \mapsto \mu(\psi^{-1}(B))$ is a measure on $\mathcal{Y}$, called image measure. Prove that

$$\int_{X} g(\psi(x)) \mu(dx) = \int_{Y} g(y) \mu_{\psi}(dy),$$

for every nonnegative measurable function $g$ on $(Y, \mathcal{Y})$. In particular, if $\psi$ is also bijective with measurable inverse then

$$\int_{X} f(x) \mu(dx) = \int_{Y} f(\psi^{-1}(y)) \mu_{\psi}(dy),$$

for every nonnegative measurable function $f$ on $(X, \mathcal{X})$.

**Proof.** Since the pre-image of a function preserves unions and intersections, it is clear that $\mu_{\psi}$ is a measure on $\mathcal{Y}$. Now, if $B$ belongs to $\mathcal{Y}$ then

$$\int_{X} 1_B(\psi(x)) \mu(dx) = \mu(\psi^{-1}(B)) = \int_{Y} 1_B(y) \mu_{\psi}(dy),$$

and, by linearity and monotone approximation, the equality holds for any non-negative measurable function $g$ on $(Y, \mathcal{Y})$ as requested. On the other hand, if $\psi$ is bijective, then we can take $g(y) = f(\psi^{-1}(y))$ to deduce the second equality. \hfill \Box

**4.2) Cartesian Products**

**Exercise 4.9.** Let $(\Omega, \mathcal{F}, \lambda)$ be a probability measure space and $\mathcal{G} \subset \mathcal{F}$ be a sub $\sigma$-algebra. Suppose that $\Omega = X \times Y$, where $(X, \mathcal{X}, \mu)$ is another probability measure space and $\mathcal{G}$ is the $\sigma$-algebra generated by the projection $p$ from $\Omega$ into $X$, i.e., $\mathcal{G} = p^{-1}(\mathcal{X})$, and that $\lambda$ restricted to $\mathcal{G}$ coincides with $p^{-1}(\mu)$, i.e., $\mu(A) = \lambda(p^{-1}(A))$, for every $A$ in $\mathcal{X}$. First, show that a real-valued $\mathcal{F}$-measurable function $f = f(\omega)$ is $\mathcal{G}$-measurable if and only if $f$ is independent of the variable $y$, i.e., $f(\omega) = g(x)$, for any $\omega = (x, y)$. Next, suppose that $\nu : X \times \mathcal{Y} \to [0, 1]$ is a probability transition measure (i.e., $\nu(x,Y) = 1$ for every $x$) such that $\lambda = \mu \times \nu$ as in Proposition 4.10, and for any nonnegative $\mathcal{F}$-measurable function $f$ define

$$\nu(f)(\omega) = \int_{Y} f(x, y) \nu(x, dy), \quad \text{with} \quad \omega = (x, y) \in X \times Y = \Omega.$$
Prove that
\[ \int_{\Omega} f g \, d\lambda = \int_{\Omega} \nu(f) g \, d\lambda, \]
for every nonnegative \( \mathcal{G} \)-measurable function \( g \). In probability terms, \( \nu(f) \) is called the conditional expectation of \( f \) given \( \mathcal{G} \), and the transition measure \( \nu \) is a regular conditional probability measure given \( \mathcal{G} \).

\[ \square \]

Proof. If \( f \) is independent of the variable \( y \) then \( f(x, y) = f \circ p(x) \), where \( p \) is the projection from \( X \times Y \) onto \( X \). Thus, for every \( \alpha \) in \( \mathbb{R} \), \( f^{-1}([\alpha, \infty]) = p^{-1} \circ f^{-1}([\alpha, \infty]) \), which is a set in \( \mathcal{G} \) by definition. On the other hand, by Proposition 1.13, if \( f \) is \( \mathcal{G} \)-measurable then there exists a measurable function \( k \) such that \( f = k \circ p \), which means that \( f \) is independent of the variable \( y \).

So check the second equality, note that for any function \( h \) independent of the variable \( y \) we have
\[ \int_{\Omega} h \, d\lambda = \int_X h \, d\mu. \]

Thus, take a \( \mathcal{G} \)-measurable function \( g \) (i.e., independent of the variable \( y \)) then
\[ \int_{\Omega} \nu(f) g \, d\lambda = \int_X \nu(f) g \, d\mu = \int_{\Omega} f(x, y) g(y) \, d\lambda \]
after using Fubini-Tonelli Theorem 4.14. \[ \square \]

**Exercise 4.10.** Let \((X, \mathcal{X}, \mu)\) be a complete measure space and \((Y, \mathcal{Y})\) be a measurable space. Consider a function \( \nu : X \times Y \to [0, \infty] \) satisfying the conditions of a \( \sigma \)-finite transition measures, except in a set of \( \mu \)-measure zero, e.g., (b) means that the mapping \( B \to \nu(x, B) \) is a measure on \( \mathcal{Y} \) for almost every \( x \in X \). Give details to show that for any set \( E \) in \( X \times Y \) the function \( x \mapsto \nu(x, E_x) \) is \( \mathcal{X} \)-measurable. Next, verify that the product measure \( \mu \times \nu \) is constructed even if (b) is weaken as follows: besides the condition \( \nu(x, \emptyset) = 0 \), for every sequence of disjoint sets \( \{B_k\} \subset \mathcal{Y} \) there is a \( \mu \)-null set \( A \) such that \( \nu(x, \sum_k B_k) = \sum_k \nu(x, B_k) \) for every \( x \) in \( X \setminus A \). If necessary, make adequate modifications to Fubini-Tonelli Theorem 4.14 to include this new situation.

\[ \text{Proof.} \] Let \( E = A \times B \) with \( A \) in \( \mathcal{X} \) and \( B \) in \( \mathcal{Y} \). Thus \( E_x = B \) if \( x \) belongs to \( A \) and \( E_x = \emptyset \) if \( x \) does not belong to \( A \), i.e., for \( x \) outside of a set \( N \) with \( \mu(N) = 0 \) we have \( \nu(x, E_x) = \nu(x, B) \) if \( x \) belongs to \( A \) and \( \nu(x, E_x) = 0 \) if \( x \) does not belong to \( A \), which shows that \( x \mapsto \nu(x, E_x) \) is \( \mathcal{X} \)-measurable for every \( E = A \times B \) with \( A \) in \( \mathcal{X} \) and \( B \) in \( \mathcal{Y} \). Now, as in Proposition 4.10, we may assume that \( \nu(x, Y) < \infty \) for every \( x \) outside of another set \( N \) with \( \mu(N) = 0 \), and therefore, the family \( \mathcal{E} \subset \mathcal{X} \times \mathcal{Y} \) of sets of \( E \) such that \( x \mapsto \nu(x, E_x) \) is \( \mathcal{X} \)-measurable still forms a \( \lambda \)-class, i.e., \( \mathcal{E} = \mathcal{X} \times \mathcal{Y} \). Moreover, the construction the product measure \( \mu \times \nu \) uses only the \( \sigma \)-additivity of \( \nu(x, \cdot) \) for almost every \( x \) as stated above.
Note that the expression \( \nu(x, \cdot) \) is a measure for almost every \( x \) implies that there exists a null set \( N \) such that for every sequence \( \{B_k\} \) of disjoint sets in \( \mathcal{Y} \) we have \( \nu(x, \sum_k B_k) = \sum_k \nu(x, B_k) \) for every \( x \) in \( X \setminus N \), i.e., in this case, the null set \( N \) is independent of the sequence of sets. This hypothesis is not necessary for the construction of the product measure \( \mu \times \nu \). Naturally, the change in Fubini-Tonelli Theorem 4.14 is that the conclusion (a) holds only for almost every \( x \) in \( X \).

\( \square \)

### (4.3) Convergence in Measure

**Exercise 4.11.** (a) Verify that if the sequence \( \{f_n\}, f_n : \Omega \to E \), of measurable functions is convergent (or Cauchy) in measure, \( (Z, d_Z) \) is a metric space and \( \psi : E \to Z \) is a uniformly continuous function then the sequence \( \{g_n\}, g_n(x) = \psi(f_n(x)) \) is also convergent (or Cauchy) in measure. (b) In particular, if \( (E, \|\cdot\|_E) \) is a normed space then for any sequences \( \{f_n\} \) and \( \{g_n\} \) of \( E \)-valued measurable functions and any constants \( a \) and \( b \) we have \( af_n + bg_n \to af + bg \) in measure, whenever \( f_n \to f \) and \( g_n \to g \) in measure. Moreover, assuming that the sequence \( \{g_n\} \) takes real (or complex) values, (c) if the sequences are also quasi-uniformly bounded, i.e., for any \( \varepsilon > 0 \) there exists a measurable set \( F \) with \( \mu(F) < \varepsilon \) such that the numerical series \( \{|f_n(x)|_E\} \) and \( \{|g_n(x)|\} \) are uniformly bounded for \( x \) in \( F^c \), then deduce that \( f_n g_n \to fg \) in measure. Furthermore, (d) if \( g_n(x)(g(x) \neq 0 \text{ a.e. } x \) and the sequences \( \{f_n\} \) and \( \{1/g_n\} \) are also quasi-uniformly bounded then show that \( f_n/g_n \to f/g \) in measure. Finally, (e) verify that if the measure space \( \Omega \) has finite measure then the conditions on quasi-uniformly bounded are automatically satisfied.

**Proof. (a)** Since \( \psi \) is uniformly continuous function, for every \( \varepsilon > 0 \) there exists \( \delta > 0 \) such that \( d_E(y, y') < \delta \) implies \( d_Z(y, y') < \varepsilon \). Thus \( d_Z(g_n(x), g_m(x)) \geq \varepsilon \) implies \( d_E(f_n(x), f_m(x)) \geq \delta \), which yields

\[
\mu(\{x \in \Omega : d_Z(g_n(x), g_m(x)) \geq \varepsilon \}) \leq \mu(\{x \in \Omega : d_E(f_n(x), f_m(x)) \geq \delta \}).
\]

Hence, \( \{g_n\} \) is a Cauchy sequence in measure if \( \{f_n\} \) is so. Similarly for a convergent sequence.

(b) The function \( \psi(x, y) = ax + by \) is uniformly continuous from \( E^2 \) into \( E \) and the sequence \( \{(f_n, g_n)\} \) with values in \( E^2 \) converges to \( (f, g) \) in measure. Thus, applying part (a) we conclude.

(c) By assumption, for every \( \varepsilon > 0 \) there exist a set \( F \) with \( \mu(F) < \varepsilon \) and \( r > 0 \) such that \( |f_n(x)|_E \leq r \) and \( |g_n(x)| \leq r \), for every \( n \geq 1 \) and for every \( x \) not in \( F \). Thus, if \( x \) belongs to \( F^c \), \( |f_n(x) - f(x)|_E \leq \varepsilon/r \) and \( |g_n(x) - g(x)| \leq \varepsilon/r \) then \( |f_n(x)g_n(x) - f(x)g(x)|_E < \varepsilon/r |g_n(x) - g(x)| + r |f_n(x) - f(x)|_E < \varepsilon \), which yields

\[
\mu(|f_n g_n - fg|_E \geq \varepsilon) \leq \mu(F) + \mu(|f_n - f|_E \geq \varepsilon/r) + \mu(|g_n - g| \geq \varepsilon/r).
\]

Hence, \( f_n g_n \to fg \) in measure.
(d) Remarking that \( g_n(x)g(x) \neq 0 \) a.e. means that the functions \( 1/g_n \) and \( 1/g \) are defined almost everywhere, we proceed as in part (c).

(e) Because \( \mu \) is a finite measure and \( |f(x)|_E < \infty \) for almost every \( x \), we deduce that \( \mu(|f(x)|_E \geq r) \to 0 \) as \( r \to \infty \), i.e., for every \( \varepsilon > 0 \) there exits \( r > 0 \) such that \( \mu(F) < \varepsilon \) and \( |f(x)|_E \leq r \), for every \( x \) in \( F^c = \{|f(x)|_E \geq r\} \). Now, for every \( \delta > 0 \), if \( |f(x)|_E \leq r \) and \( |f_n(x) - f(x)|_E < \delta \) then \( |f_n(x)|_E \leq \delta + r \), thus

\[
\mu(|f_n|_E \geq \delta + r) \leq \mu(|f|_E \geq r) + \mu(|f_n - f|_E \geq \delta).
\]

Hence, for every \( n \geq n(\delta) \) and \( x \) in \( F_\delta \) we have \( |f_n(x)|_E \leq \delta + r \), with \( F_\delta = \{|f_n(x)|_E \geq \delta + r\} \) and \( \mu(F_\delta) \leq \varepsilon + \delta \). This proves that \( \{f_n\} \) is quasi-uniformly bounded.

Note that we are proving that if a sequence converges in measure to a quasi-bounded function, then the sequence is quasi-bounded. \( \square \)

**Exercise 4.12.** Let \((\Omega, \mathcal{F}, \mu)\) be a measure space, \((E, d)\) a metric space and \( \{f_n\} \) a sequence of measurable functions \( f_n : \Omega \to E \). Show that if \( \{f_n\} \) converges to some function \( f \) pointwise quasi-uniform then \( f_n \to f \) in measure. Prove or disprove the converse.

**Proof.** If \( \{f_n\} \) converges to some function \( f \) pointwise quasi-uniform then for every \( \varepsilon > 0 \) there exists a set \( \Omega_\varepsilon \in \mathcal{F} \) with \( \mu(\Omega \setminus \Omega_\varepsilon) \leq \varepsilon \) such that \( f_n(x) \to f(x) \) uniformly in \( \Omega_\varepsilon \). Thus, there exists an index \( n_\varepsilon \) such that \( |f_n(x) - f(x)| < \varepsilon \), for every \( x \) in \( \Omega_\varepsilon \) and \( n \geq n_\varepsilon \). Hence

\[
\mu(|f_n - f| \geq \varepsilon) \leq \mu(\Omega \setminus \Omega_\varepsilon) \leq \varepsilon,
\]

i.e., \( f_n \to f \) in measure.

Let us show that pointwise quasi-uniform convergence implies pointwise almost everywhere convergence. Indeed, let \( \Omega_0 = \bigcap_k \Omega_{1/k} \), where \( \Omega_{1/k} \) the \( \Omega_\varepsilon \) with \( \varepsilon = 1/k \) obtained from the pointwise quasi-uniform convergence. Since \( \mu(\Omega \setminus \Omega_0) \leq \mu(\Omega \setminus \Omega_{1/k}) \leq 1/k \), and \( f_n(x) \to f(x) \) uniformly, for every \( x \) in \( \Omega_{1/k} \), we deduce \( \mu(\Omega \setminus \Omega_0) = 0 \) and \( f_n(x) \to f(x) \) for every \( x \) in \( \Omega_0 \).

If the converse were true, then convergence in measure would imply pointwise almost everywhere convergence, which contradicts the example given in Remark 4.18 of a sequence convergence in measure but non convergent pointwise almost everywhere. \( \square \)

**Exercise 4.13.** Assume that \( \mu(\Omega) < \infty \) and by means of arguments similar to those of Egorov’s Theorem 4.19, prove (a) if \( (E, |\cdot|_E) \) is a normed space and the numerical sequence \( \{|f_n(x)|_E\} \) is bounded for almost every \( x \) in \( \Omega \) then for every \( \varepsilon > 0 \) there exists a measurable subset \( F \) of \( \Omega \) such that \( \mu(F) < \varepsilon \) and the sequence \( \{|f_n(x)|_E\} \) is uniformly bounded for any \( x \) in \( F^c \). Moreover, (b) show that if \( E = [-\infty, +\infty], \bar{f}(x) = \text{lim sup}_n f_n(x) \) (or \( \bar{f}(x) = \text{lim inf}_n f_n(x) \)) and \( \bar{f} \) (or \( \bar{f} \)) is a real valued (finite) a.e. then for every \( \varepsilon > 0 \) there exists a measurable subset \( F \) of \( \Omega \) such that \( \mu(F) < \varepsilon \) and the \( \text{lim sup} \) (or \( \text{lim inf} \)) is uniformly for \( x \) in \( F^c \). Moreover, if \( \mu \) is a Borel measure then \( F = O \) is an open set of \( \Omega \).
Proof. (a) Define the set \( X(n, m) = \{ x \in \Omega : |f_n(x)|_E \geq m \} \), \( m = 1, 2, \ldots \), to check that the numerical sequence \( \{|f_n(x)|_E \} \) is bounded if and only if \( x \) does not belong to \( \bigcap_m \bigcup_n X(n, m) \). Since \( \mu(\Omega) < \infty \) and \( \bigcap_m \bigcup_n X(n, m) \) has zero measure assumption, if \( F_m = \bigcup_n X(n, m) \) then \( \mu(F_m) \to 0 \) as \( m \to \infty \). Hence for every \( \varepsilon > 0 \) there exits a set \( F = F_m \) such that \( \mu(F) < \varepsilon \) and \( |f_n(x)|_E \leq m \) for every \( x \) in \( F^c \) and for every \( n \).

(b) Essentially, we repeat the arguments proving Egorov’s Theorem 4.19. Consider the \( \bar{f}(x) = \lim \sup_n f(x) \), the \( \lim \inf_n f_n \) is treated similarly. Define the set \( X(\varepsilon, f_n, \bar{f}) = \{ x \in \Omega : \bar{f}(x) < \infty \text{ and } |\sup_{k \geq n} f_k(x) - \bar{f}(x)| \geq \varepsilon \} \) to check that \( \lim \sup_n f(x) = \bar{f}(x) \) and \( \bar{f}(x) < \infty \) if and only if \( x \notin F_\varepsilon = \bigcap_{n=1}^\infty \bigcup_{k=n}^\infty X(\varepsilon, f_k, \bar{f}) \), for every \( \varepsilon > 0 \). Since \( X(\varepsilon, f_n, \bar{f}) \subset F_{\varepsilon,n} = \bigcap_{k=1}^n \bigcup_{i=k}^\infty X(\varepsilon, f_i, \bar{f}) \), we have \( \mu(X(\varepsilon, f_n, \bar{f})) \leq \mu(F_{\varepsilon,n}) \), and therefore

\[
\limsup_n \mu(X(\varepsilon, f_n, \bar{f})) \leq \lim \mu(F_{\varepsilon,n}), \quad \forall \varepsilon > 0.
\]

If \( \sup_{k \geq n} f_k \to \bar{f} \) pointwise almost everywhere and \( \bar{f} < \infty \) almost everywhere, then \( \mu(F_\varepsilon) = 0 \) for every \( \varepsilon > 0 \), and if also \( \mu \) is a finite measure then \( \mu(F_{\varepsilon,n}) \to \mu(F_\varepsilon) = 0 \).

To show the quasi-uniform convergence, let \( k, n \) be positive integers and set

\[
A_k(n) = \bigcup_{m=n}^\infty \big\{ x : |\sup_{i \geq m} f_i(x) - \bar{f}(x)| \geq 1/k \big\} = \bigcup_{m=n}^\infty X(1/k, f_m, f).
\]

It is clear that \( A_k(n) \supset A_k(n + 1) \) for any \( k, n \), and the almost everywhere convergence implies that \( \mu(B_k) = 0 \) with \( \bigcap_{n=1}^\infty A_k(n) = B_k \). Since \( \mu(\Omega) < \infty \) we deduce \( \mu(A_k(n)) \to 0 \) as \( n \to \infty \). Hence, given \( \varepsilon > 0 \) and \( k \), choose \( n_k \) such that \( \mu(A_k(n_k)) < \varepsilon 2^{-k} \) and define \( F = \bigcup_{k=1}^\infty A_k(n_k) \). Thus \( \mu(F) < \varepsilon \), and \( |f_n(x) - \bar{f}(x)| < 1/k \) for any \( n > n_k \) and \( x \notin F \). This yields \( \limsup_n f_n = \bar{f} \) uniformly on \( F^c \).

Finally, if \( \mu \) is a Borel measure then we conclude by choosing (see Theorem 3.3) an open set \( O \supset F \) with \( \mu(O) < 2\varepsilon \).

Exercise 4.14. Prove that if a sequence \( \{f_n\} \) of (extended) real-valued integrable functions on measure space \( (\Omega, \mathcal{F}, \mu) \) converges in mean to \( f \) then \( f_n \to f \) in measure. Moreover, give an example of a sequence of functions defined on the Lebesgue measure space \( ([0,1], \ell) \) which converges in measure but it does not converge in mean.

Proof. From the inequality \( 1_{|f_n - f| \geq \varepsilon} \leq |f_n - f| \) we obtain the estimate

\[
\mu(|f_n - f|) \leq \frac{1}{\varepsilon} \int_\Omega |f_n - f| \, d\mu,
\]

which shows that convergence in mean implies convergence in measure.

Now, on the Lebesgue measure space \( ([0,1], \ell) \) consider the sequence \( \{f_n\} \) with \( f_n = n 1_{[0,1/n]} \), for \( n \geq 1 \). This sequence converges pointwise to the function

\[
\frac{1}{2n} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \ldots + \frac{1}{2} + \frac{1}{2} + \ldots
\]
\[ f(x) = 0 \text{ and } \ell(\lvert f_n - f \rvert \geq \varepsilon) \leq 1/n. \] However,
\[ \int_0^1 f_n(x) \, dx = 1 \quad \text{and} \quad \int_0^1 f(x) \, dx = 0, \]
i.e., the integral of \( f_n \) does not converges to the integral of the limiting function \( f \), so that \( f_n \) cannot converge in mean to \( f \).

\[ \square \]

**Exercise 4.15.** The assumption that \( \mu(\Omega) < \infty \) is essential in Egorov's Theorem 4.19. Indeed, let \( \{A_k\} \) be a sequence of disjoint measurable sets such that \( A = \sum_k A_k, \mu(A) = \infty \) and \( \mu(A_k) \to 0 \). Consider the sequence of measurable functions \( f_k = 1_{A_k} \). Prove that (1) \( f_k \to 0 \) in mean (and in measure, in view of Exercise 4.14), (2) \( f_k(x) \to 0 \) for every \( x \), and (3) \( f_k(x) \to 0 \) uniformly for \( x \) in \( B \) if and only if there exists an index \( n_B \) such that \( B \cap A_k = \emptyset \), for every \( k \geq n_B \). Finally, (4) deduce that \( \{f_k\} \) does not converge pointwise quasi-uniform and (5) make an explicit construction of the sequence \( \{A_k\} \).

**Proof.** (1) It is clear that the integral of \( f_k \) is equal to \( \mu(A_k) \), so \( f_k \to 0 \) in mean and in measure.

(2) For every \( x \) in \( A \) there exists exactly one index \( k(x) \) such that \( x \) belongs to \( A_{k(x)} \), which implies that \( f_k(x) = 0 \), for every \( k > k(x) \), i.e., \( f_k(x) \to 0 \) for every \( x \) in \( A \) and obviously, for every \( x \) in \( \Omega \).

(3) It is clear that \( f_k(x) \to 0 \) uniformly for \( x \) in \( B \) if and only if the index \( k(x) \) found in part (2) satisfies \( \sup_{x \in B} k(x) < \infty \), which is equivalent to say that there exists an index \( n_B \) such that \( B \cap A_k = \emptyset \), for every \( k \geq n_B \).

(4) If \( F \) is a set with finite measure then \( F \) contains only a finite number of the sets \( A_k \). Therefore, the complement \( F^c \) must contain all, but a finite number, of the sets \( A_1, A_2, \ldots \). This means that the convergence cannot be uniform on the complement \( F^c \). Hence \( \{f_k\} \) does not converge pointwise quasi-uniform.

(5) For instance, on the Lebesgue measure space \((\mathbb{R}, \ell)\) consider the sequence \( \{A_k\} \) with \( A_k = (k, k + 1/k) \). Since \( \ell(A_k) = 1/k \) and \( A_k \cap A_n = \emptyset \) for every \( k \neq n \), we have \( A = \sum_k A_k, \ell(A_k) = 1/k \to 0 \) and \( \ell(A) = \sum_k 1/k = \infty \).

\[ \square \]

**Exercise 4.16.** Let \( \Omega \) and \( E \) be two metric spaces. Suppose that \( \mu \) is a regular Borel measure on \( \Omega \) and \( f: \Omega \to E \) is a \( \mu \)-measurable function such that for some \( e_0 \) in \( E \) the set \( \{x \in \Omega : f(x) \neq e_0\} \) is a \( \sigma \)-finite and the range \( \{f(x) \in E : x \in \Omega\} \) is contained in a separable subspace of \( E \). With arguments similar to those of Lusin Theorem 4.22, (a) show that for any \( \varepsilon > 0 \) there exists a close set \( C \) such that \( f \) is continuous on \( C \) and \( \mu(\Omega \setminus C) < \varepsilon \); and (b) establish that there exists a \( \mathcal{F}_\sigma \) set (i.e., a countable union of closed sets) \( F \) such that \( f \) is continuous and \( \mu(F^c) = 0 \). Next (c) deduce that any \( \mu \)-measurable function as above is almost everywhere equal to a Borel function. Is there any other way (without using Lusin Theorem) of proving part (c)?

**Proof.** (a) We repeat the arguments of Lusin Theorem 4.22 replacing \( E \) with the separable subspace \( E_0 \) containing the range \( \{f(x) \in E : x \in \Omega\} \), i.e., for every integer \( i \) there exists a sequence \( \{E_{i,j}\} \) of disjoint Borel sets of diameters
not larger that 1/i such that $E_0 = \sum_j E_{i,j}$. Thus, setting $E'_{i,j} = E_{i,j} \setminus \{e_0\}$, for $j \geq 1$ and adding $E'_{i,0} = \{e_0\}$, we obtain a closed set $C_i = \sum_{j=0}^{n(i)} C_{i,j}$, $C_{i,j} \subset f^{-1}(E'_{i,j})$ such that $\mu(\Omega \setminus C_i) < \varepsilon 2^{-i}$. Note that $C_{i,0} = C_0$ can be taken independent of i. We follow essentially the same argument by choosing $e_{i,j}$ in $E'_{i,j}$ and $e_{i,0} = e_0$, we can define a function $g_i : C_i \rightarrow E$ as $g_i(x) = e_{i,j}$ whenever $x$ belongs to some $C_{i,j}$ with $j = 0, 1, \ldots, n$. This function $g_i$ is continuous because $\{C_{i,j} : j \geq 1\}$ are closed and disjoint, and the distance from $g_i(x)$ to $f(x)$ is not larger then $1/i$ (note that $g_i(x) = f(x) = e_0$ for every $x$ in $C_{i,0}$). Actually, we could define $g_i(x) = e_0$ for every $x$ in $f^{-1}(e_0)$. Thus, we have

$$C = \bigcap_{i=1}^{\infty} C_i = \bigcap_{i=1}^{\infty} \bigcup_{j=0}^{n(i)} C_{i,j}, \quad \mu(\Omega \setminus C) < \varepsilon,$$

and $g_i(x) \rightarrow f(x)$ uniformly for every $x$ in $C$, i.e., the restriction of f to the closed set $C$ is continuous.

Certainly, we could simply define $f_0$ and $\mu_0$ as the restriction to $\Omega_0 = \{x \in \Omega : f(x) \neq e_0\}$ and directly use Lusin Theorem. Moreover, the same argument apply if only $\{x \in \Omega : f(x) \neq e_i, i = 0, 1, \ldots, m\}$ is separable, for a finite number of points $e_i$ in $E$.

Remark that if $\mu$ is also inner regular (see Proposition 3.15) and the set $\Omega_0 = \{x \in \Omega : f(x) \neq e_0\}$ has finite measure then all, but $C_{i,0}$, can be taken compact sets. Usually, $E$ is a normed space and $e_0 = 0$.

(b) For a given measurable function f and any $k \geq 1$ we can use part (a) with $\varepsilon = 1/k$ to find a closed set $C_k$ such that $f$ is continuous on $C_k$ and $\mu(C_k^c) \leq 1/k$. Since $F = \bigcup_k C_k$ is a $\mathcal{F}_\sigma$ set and $\mu(F^c) \leq 1/k$ for every $k$, we deduce that $\mu(F^c) = 0$ and $f$ is continuous on $F$.

(c) Define $g(x) = f(x)$ for every $x$ in $F$ and $g(x) = e_0$ for every $x$ in the negligible set $F^c$. Thus, for any open set $O$ in $E$, $g^{-1}(O) = F \cap f^{-1}(O)$ if $e_0$ is not in $O$ and $g^{-1}(O) = F^c \cup (F \cap f^{-1}(O \setminus \{e_0\}))$ if $e_0$ is in $O$. In both cases the pre-images is a Borel set, i.e., $g$ is Borel function such that $f = g$ almost everywhere.

Note that because $F$ is not closed, in general, the restriction of $f$ to $F$ cannot be extended to a continuous function in the whole $\Omega$.

Remark that we can prove part (b) without using Lusin Theorem. Indeed, assume $E = \mathbb{R}$ to simplify and first express $f = f^+ - f^-$ to reduce the question to a nonnegative function. Next, for a $\mu$-measurable function as above, we can find an increasing sequence of $\mu$-measurable simple functions $\{f_n\}$ such that $f_n(x) \rightarrow f(x)$ for every $x$ in $\Omega$, see Proposition 1.9 and Corollary 1.10. Since for each $\mu$-measurable set $A$ there exists a Borel set $B \supset A$ such that $\mu(B \setminus A) = 0$, we can modify each simple function $f_n$ to get a Borel measurable simple function $g_n$ satisfying $g_n = f_n$ almost everywhere. Hence the sequence $\{g_n\}$ is almost everywhere increasing to the function $f$, i.e., there exists a Borel set $B$ such that $g_n(x)$ increases to $f(x)$ for every $x$ in $B$ and $\mu(B^c) = 0$. Hence the limiting function $g = 1_B \lim_n g_n$ is a Borel function satisfying $g = f$ almost everywhere. 

\[\square\]
Exercise 4.17. Given a closed subset $F$ of $\mathbb{R}^d$ and a real-valued function $f$ defined on $\mathbb{R}^d$, what does the assertion “the restriction of $f$ to $F$ is continuous” actually means, in term of convergent sequences? Let $(\mathbb{R}^d, \mathcal{L}, \ell)$ be the Lebesgue measure space and $\Omega$ a measurable subset of $\mathbb{R}^d$. Prove that a function $f : \Omega \to \mathbb{R}$ is measurable if and only if for every $\varepsilon > 0$ there exits a closed set $F \subset \Omega$ such that $f$ restricted to $F$ is continuous and $\ell(F^c) < \varepsilon$.

Proof. If the restriction of $f$ to $F \subset \mathbb{R}^d$ is continuous then for any $x$ in $F$ and any sequence $\{x_n\} \subset F$ convergent to $x$ we have $\lim_{n \to \infty} f(x_n) = f(x)$. The converse is also true. The restriction of $f$ to $F$ is continuous if for every open interval $]a, b[\}$ the pre-image $f^{-1}(]a, b[\})$ is open relative in $F$, i.e., there exists an open set $O$ in $\mathbb{R}^d$ such that $F \cap O = F \cap f^{-1}(]a, b[\})$.

Take $\varepsilon = 1/k$ and get a closed set $F_k$ such that $f$ restricted to $F_k$ is continuous and $\ell(F_k^c) < 1/k$. Define the set $B = \bigcap_{k} F_k$ to check that $f$ restricted to the Borel set $B$ is continuous and $\ell(B^c) = 0$. Now, the function $g = 1_B f$ is Borel measurable because $g^{-1}(]a, b[\}) = B \cap f^{-1}(]a, b[\})$ if $a \geq 0$ or $b \leq 0$, and $g^{-1}(]a, b[\}) = B^c \cup (B \cap f^{-1}(]a, b[\}))$ if $a < 0 < b$. Since $g = f$ almost everywhere, the function $f$ is $\mu$-measurable.

The converse is part (c) of Exercise 4.16 with $E = \mathbb{R}$, $\Omega = \mathbb{R}^d$ and $\mu = \ell$. \hfill $\Box$

Exercise 4.18. Let $\mu$ be a Radon measure on a locally compact space $X$ and $f : X \to \mathbb{R}$ be a measurable function that vanishes outside a set of finite measure. Prove that for any $\varepsilon > 0$ there exist a closed set $F$ with $\mu(F^c) < \varepsilon$ and a continuous function with compact support $g$ such that $f = g$ on $F$ and if $|f(x)| \leq C$ a.e. then $|g(x)| \leq C$ for every $x$, see Folland [45, Section 7.2, pp. 216–221].

Proof. Repeating the construction of Lusin Theorem 4.22 as in Exercise 4.16 of the closed sets $C_{i,j}$, $i \geq 1$ and $j \geq 0$ with $e_0 = 0$. Because $f$ vanishes outside a set of finite measure, all the sets $C_{i,j}$, with $j \geq 1$, can be taken compact sets, so that the restriction of $f$ to $F = \bigcap_{i=1}^{\infty} \bigcup_{j=0}^{\infty} C_{i,j}$ is continuous, $\mu(F^c) < \varepsilon$, and the set $\{x \in F : f(x) \neq 0\}$ is contained in the compact set $K = \bigcap_{i=1}^{\infty} \bigcup_{j=1}^{\infty} C_{i,j}$. Since $F$ a closed set, we can find a continuous function $g$ defined on the whole space (see Tietze's extension Proposition 0.2) such that $g = f$ on $F$ and $|g(x)| \leq \sup\{|f|\}$. \hfill $\Box$

(4.4) Almost Measurable Functions

Exercise 4.19. A real-valued function $f$ on a measure space $(\Omega, \mathcal{F}, \mu)$ belongs to weak $L^1$ if

$$\mu(\{\omega \in \Omega : r |f(\omega)| > 1\}) \leq C_r, \quad \forall r > 0,$$

for some finite constant $C = C_f$. Prove (1) any integrable function belongs to weak $L^1$ and verify (2) that the function $f(x) = |x|^{-d}$ is not integrable in
\( \Omega = \mathbb{R}^d \) with the Lebesgue measure \( \mu = \ell \), but it does belong to weak \( L^1 \). Now, consider the map

\[
\|f\|_1 = \sup_{r>0} \{ r \mu (\{ \omega \in \Omega : |f(\omega)| > r \}) \},
\]

for every \( f \) in \( L^1(\Omega, \mu; \mathbb{R}) \). Prove that (3) \( \|cf\|_1 = |c| \|f\|_1 \), for any constant \( c \); (4) \( \|f + g\|_1 \leq \|f\|_1 + 2\|g\|_1 \); and (5) if \( \|f_n - f\|_1 \to 0 \) then \( f_n \to f \) in measure. Finally, if \( L^1_w(\Omega, \mu; \mathbb{R}) \) is the subspace of all almost measurable functions \( f \) satisfying \( \|f\|_1 < \infty \) (i.e., the weak \( L^1 \) space) then prove that (6) \( (L^1_w, \|\cdot\|_1) \) is a topological vector space, i.e., besides being a vector space with the topology given by the balls (even if \( \|f\|_1 \) is not a norm and therefore does not induce a metric) makes the scalar multiplication and the addition continuous operations. See Folland [45, Section 6.4, pp. 197–199] and Grafakos [55, Section 1.1].

**Proof.** (1) By means of the inequality \( \mathbb{1}_{|f| \geq 1/r} \leq r |f| \) we obtain

\[
\mu (\{ \omega \in \Omega : r |f(\omega)| > 1 \}) \leq r \|f\|_1,
\]

which proves that any integrable function belongs to weak \( L^1 \).

(2) By using spherical coordinates, it is clear that the function \( f(x) = |x|^{-d} \) is not integrable in \( \Omega = \mathbb{R}^d \) with the Lebesgue measure \( \mu = \ell \). However,

\[
\ell (\{ x \in \mathbb{R}^d : r |f(x)| > 1 \}) = \ell (\{ x \in \mathbb{R}^d : r > |x|^d \}) = r \ell (\{ x \in \mathbb{R}^d : |x| < 1 \}),
\]

i.e., \( x \mapsto |x|^{-d} \) belongs to weak \( L^1(\mathbb{R}^d) \).

(3) From the definition it is clear that the homogeneity property \( \|cf\|_1 = |c| \|f\|_1 \), for any constant \( c \).

(4) The triangular inequality for \( |\cdot| \) yields the inclusion

\[
\{ \omega : |h(\omega) - h(\omega')| \geq r \} \subset \{ \omega : |f(\omega) - f(\omega')| \geq r/2 \} \cup \{ \omega : |g(\omega) - g(\omega')| \geq r/2 \},
\]

for \( h = f + g \), which implies

\[
\mu (\{ \omega : |h(\omega) - h(\omega')| \geq r \}) \leq \mu (\{ \omega : |f(\omega) - f(\omega')| \geq r/2 \}) + \mu (\{ \omega : |g(\omega) - g(\omega')| \geq r/2 \}),
\]

and then \( \|f + g\|_1 \leq 2 \|f\|_1 + 2 \|g\|_1 \).

(5) Since

\[
\mu (\{ \omega \in \Omega : |f(\omega)| > \varepsilon \}) \leq \frac{\|f\|_1}{\varepsilon}, \quad \forall \varepsilon > 0,
\]

we deduce that if \( \|f_n - f\|_1 \to 0 \) then \( f_n \to f \) in measure.

(6) The point here is that \( f \mapsto \|f\|_1 \) is not a norm and \( d(f, g) = \|f - g\|_1 \) is not really a distance. Nevertheless, the topology induced by the balls of center \( g \) and radius \( r > 0 \) \( B(f, r) = \{ f \in L^1_w : \|f - g\|_1 \leq r \} \) (which are not balls in the proper geometric sense) is compatible with the vector structure, i.e., because of the properties proves in (3) and (4), the scalar multiplication and the vector addition are continuous operations. \( \square \)
Exercise 4.20. Paying special attention to the almost everywhere concept, show that if a sequence of functions \( \{f_n\} \) in \( L^1(\Omega, \mu; \mathbb{R}) \) satisfies \( \sum_n \|f_n\|_1 < \infty \) then \( \sum_n f_n \) converges to a function \( f \) in \( L^1(\Omega, \mu; \mathbb{R}) \) and

\[
\int_{\Omega} f \, d\mu = \sum_n \int_{\Omega} f_n \, d\mu.
\]

Again, deduce that \( L^1(\Omega, \mu; \mathbb{R}) \) is a complete space, i.e., a Banach space.

Proof. Taking a representant of each equivalence class, we have a sequence \( \{g_n\} \) of integrable functions such that \( f_n = g_n \) for every \( x \) outside of a negligible set \( N_n \) and \( \sum_n \|g_n\|_1 = \sum_n \|f_n\|_1 < \infty \). Applying Beppo Levi or monotone convergence Theorem 4.4 to the increasing sequence \( \{\sum_{k \leq n} |g_k|\} \) of nonnegative measurable functions the limiting function \( G = \sum_k |g_k| \) is integrable and \( \|G\|_1 = \sum_n \|f_n\|_1 \). Now, redefining \( g_n = 0 \) outside of a negligible set where \( G \) is infinite, we can say that the series \( \sum_k |g_k(x)| < \infty \) for every \( x \) as well as \( f_n(x) = g_n(x) \) outside of a negligible set \( N = \bigcup_{k \geq 0} N_k \), where \( N_0 \) contains the set where \( G \) is infinite. Since \( \sum_{k \leq n} f_k(x) = \sum_{k \leq n} g_k(x) \), for every \( x \) outside of the negligible set \( N \) and \( \{\sum_{k \leq n} g_k(x)\} \leq \sum_k |g_k(x)| < \infty \), for every \( x \), the Lebesgue or dominate convergence Theorem 4.7 implies that the limiting functions \( g = \sum_n g_n \) is integrable, \( f(x) = \sum_n g_n(x) \) for every \( x \) outside of the negligible set \( N \), and

\[
\sum_n \int_{\Omega} f_n \, d\mu = \sum_n \int_{\Omega} g_n \, d\mu = \int_{\Omega} \sum_n g_n \, d\mu = \int_{\Omega} f \, d\mu,
\]

i.e., \( f \) is also integrable. Finally, because \( \|f - \sum_{k \leq n} f_k\|_1 \leq \sum_{k > n} \|f_k\|_1 \to 0 \) as \( n \to \infty \), we deduce that \( \sum_n f_n \) converges to \( f \) in \( L^1(\Omega, \mu; \mathbb{R}) \).

To check that \( L^1(\Omega, \mu; \mathbb{R}) \) is a complete space, for any given a sequence \( \{f_n\} \) we can construct the telescoping series \( g_1 = f_1, g_n = f_n - f_{n-1} \) for \( n \geq 2 \), i.e. \( f_n = \sum_{k \leq n} g_n \). Because \( f_n - f_m = \sum_{m < k \leq n} g_n \), if \( \sum_n \|f_n - f_{n-1}\|_1 < \infty \) then we can use the previous argument to deduce that the limiting function \( f = \lim_n f_n \) exists almost everywhere, \( f \) is integrable and \( \|f_n - f\|_1 \to 0 \). Thus, taking a subsequence \( \{f_{n_k}\} \) of a given Cauchy sequence \( \{f_n\} \) satisfying \( \sum_k \|f_{n_k} - f_{n_k-1}\|_1 < \infty \), we find an integrable function \( f \) such that \( \|f_{n_k} - f\|_1 \to 0 \). However, given \( \varepsilon > 0 \) there exits an index \( N_\varepsilon \) such that \( \|f_n - f_m\|_1 < \varepsilon/2 \) for every \( n, m \geq N_\varepsilon \). Since

\[
\|f_n - f\|_1 \leq \|f_n - f_{n_k}\|_1 + \|f_{n_k} - f\|_1 \leq \varepsilon/2 + \varepsilon/2
\]

if \( k \geq K_\varepsilon, n_k, n \geq N_\varepsilon \), we deduce that the whole sequence converges. \( \square \)

Exercise 4.21. Let \( \mu \) be a Borel measure on a Polish space \( \Omega \). Give some details on most of the statements related to the spaces \( L^1(\Omega, \mu; \mathbb{R}), L^0(\Omega, \mu; \mathbb{R}), L^0(\Omega, \mu; \mathbb{R}), L^0(\Omega, \mu; \mathbb{R}), L^\infty(\Omega, \mu; \mathbb{R}), L^\infty(\Omega, \mu; \mathbb{R}), S^0(\Omega, \mu; \mathbb{R}) \) and \( S^1(\Omega, \mu; \mathbb{R}) \), recall that the \( \sigma \) refers to the \( \sigma \)-finite support. In particular, define a metric (or norm), specify when the space is separable and/or complete, and state any topological...
inclusion. Moreover, for $\Omega = \mathbb{R}^d$ and $\mu = \ell$ the Lebesgue measure, if possible, give examples of functions in each of the above spaces not belonging to any of the others.

**Proof.** First, the element of all these (quotient) spaces are equivalence classes of measurable functions or measurable functions defined almost everywhere or almost measurable functions, i.e., a real-valued measurable function $f$ on the measure space $(\Omega, \mathcal{F}, \mu)$ and all other functions $g$ such that $g = f$ outside of a $\mu$-null set are considered just one function, where the function $g$ is measurable with respect to the completion of $\mathcal{F}$ relative to the measure $\mu$.

The space $L^0(\Omega, \mu) = L^0(\Omega, \mu; \mathbb{R})$ of all $\mu$-almost measurable functions is the largest one, it is a complete metric space with the metric

$$d_\mu(f, g) = \inf \{r > 0 : \mu(\{\omega \in \Omega : |f(\omega) - g(\omega)| > r\}) \leq r\}.$$ 

Moreover, $L^0(\Omega, \mu)$ is also a topological vector space, i.e., a vector space where the addition and scalar multiplication are continuous operations.

The space $L^\infty(\Omega, \mu) = L^\infty(\Omega, \mu; \mathbb{R})$ of all $\mu$-almost measurable ($\mu$-essentially) bounded functions is a Banach space, i.e., a vector space with a norm, namely,

$$\|f\|_\infty = \inf \{b \in [0, \infty] : \mu(|f| > b) = 0\} = \inf \{\sup\{|f(\omega)| : \omega \in \Omega \setminus N\} : N, \mu(N) = 0\},$$

which is complete. It is clear that $L^\infty(\Omega, \mu) \subset L^0(\Omega, \mu)$, and the inclusion is continuous. The subspace $S^0(\Omega, \mu) = S^0(\Omega, \mu; \mathbb{R})$ of almost measurable simple functions (i.e., of all almost measurable functions assuming a finite number of values) is dense in $L^0(\Omega, \mu)$ and in $L^\infty(\Omega, \mu)$ (see Proposition 1.9 and Theorem 4.19). In general, $L^\infty(\Omega, \mu)$ is not separable and the separability of $L^0(\Omega, \mu)$ depends on the separability of $S^0(\Omega, \mu)$, which is related to the structure of the measure space $(\Omega, \mathcal{F}, \mu)$. In $\mathbb{R}^d$ with the Lebesgue measure, the subspace of step functions (i.e., simple functions relative to the semi-ring of $d$-intervals) could be used to show separability.

Most of the attention is given to the space $L^1(\Omega, \mu) = L^1(\Omega, \mu; \mathbb{R})$ of all $\mu$-integrable functions, which is a Banach space with the norm

$$\|f\|_1 = \int_\Omega |f(\omega)| \mu(\mathrm{d}\omega).$$

Certainly, $L^1(\Omega, \mu) \subset L^0(\Omega, \mu)$ with continuous inclusion, and if $\mu$ is a finite measure then $L^\infty(\Omega, \mu) \subset L^1(\Omega, \mu)$ with continuous inclusion. If $\Omega$ is a metric space and $\mu$ is measure on the Borel $\sigma$-algebra constructed via Caratheodory’s arguments then the space $L^1(\Omega, \mu)$ is separable. In particular this applies to the Lebesgue measure.

The support of an integrable function is always a $\sigma$-finite set, so that if $\mu$ is not a $\sigma$-finite measure then the spaces $L^0(\Omega, \mu)$ and $L^\infty(\Omega, \mu)$ are made smaller, i.e., the subspaces $L^0_\sigma(\Omega, \mu) = L^0(\Omega, \mu; \mathbb{R})$ and $L^\infty_\sigma(\Omega, \mu) = L^\infty(\Omega, \mu; \mathbb{R})$ become useful. They have the same character, i.e., $L^0_\sigma(\Omega, \mu)$ is complete metric space and a topological vector space, and $L^\infty_\sigma(\Omega, \mu)$ is a Banach space.
The subspaces $S^0_\sigma(\Omega, \mu) = S^0(\Omega, \mu; \mathbb{R})$ and $S^1(\Omega, \mu) = S^1(\Omega, \mu; \mathbb{R})$ of simple functions are used for approximations, $S^0_\sigma(\Omega, \mu) \subset L^0_\sigma(\Omega, \mu)$ and $S^1_\sigma(\Omega, \mu) \subset L^1_\sigma(\Omega, \mu)$. Elements in $S^0_\sigma(\Omega, \mu)$ are almost measurable assuming a finite number of values and with $\sigma$-finite support, while $S^1_\sigma(\Omega, \mu) \subset S^0(\Omega, \mu)$ are integrable functions, i.e., almost measurable assuming a finite number of values and with $\mu$-finite support. The subspace $S^k(\Omega, \mu)$ is dense in $L^k(\Omega, \mu)$, for $k = 0, 1$, and $S^0_\sigma(\Omega, \mu) \subset L^\infty(\Omega, \mu)$.

If $\Omega = \mathbb{R}$ and $\mu = \ell$ the Lebesgue measure then the subindex of $\sigma$-finite support is not necessary. The function $1_{[0, +\infty)}$ is an example of a function in $S^0(\mathbb{R}, \ell) \setminus S^1(\mathbb{R}, \ell)$, and the function $1_{[0, 1]}$ belongs to $S^1(\mathbb{R}, \ell)$. The function $(\ln x)1_{(0, 1)}$ is in $L^1(\mathbb{R}, \ell) \setminus L^\infty(\mathbb{R}, \ell)$, the functions $x1_{(0, 1)}$ and $e^{-x}1_{[0, +\infty)}$ are in $L^\infty(\mathbb{R}, \ell) \cap L^1(\mathbb{R}, \ell) \setminus S^0(\mathbb{R}, \ell)$. □

**Exercise 4.22.** Let $E$ be a vector lattice of real-valued function defined on $X$ and $I : E \to \mathbb{R}$ be a linear and monotone functional satisfying the condition: if $\{\varphi_n\}$ is a decreasing sequence in $E$ such that $\varphi_n(x) \downarrow 0$ for every $x$ in $X$, then $I(\varphi) \downarrow 0$. Consider the vector lattice $L$ as defined in Exercise 4.6, which are $I$-integrable function defined outside of an $I$-null set, see Exercise 4.5. Let $M^+$ be the semi-space of functions $f : X \to [0, +\infty]$ such that $f \wedge \varphi$ belongs to $L$, for every nonnegative $\varphi$ in $L$. Functions $f$ such that $f^+$ and $f^-$ belong to $M^+$ are called $I$-measurable.

(1) Prove that (a) $M^+$ is stable under the pointwise convergence of sequences, and that (b) $M^+$ is a semi-vector lattice, i.e., for every positive constant $c$ and any functions $f_1$ and $f_2$ in $M^+$, the functions $f_1 + cf_2$, $f_1 \wedge f_2$ and $f_1 \vee f_2$ are also in $M^+$. Moreover, show that for any $f$ and $g$ in $M^+$ with $f - g \geq 0$ we have $f - g$ in $M^+$. For a $f$ in $M^+$ we define

$$I(f) = \sup \{I(f \wedge \varphi) : \varphi \in L, \varphi \geq 0\}.$$ 

Verify that Beppo Levi’s Theorem holds true within the semi-space $M^+$ and $I$ is semi-linear and monotone on $M^+$.

(2) Consider the class $A$ of subsets of $\Omega$ such that $1_A$ belongs to $M^+$. Prove that $A$ is a $\sigma$-ring and the set function $A \mapsto \mu(A) = I(1_A)$ is a measure on $A$.

(3) Assume that the function $1 = 1_X$ is $I$-measurable and verify that $A$ is a $\sigma$-algebra. Prove that a function $f$ is almost everywhere $(A, \mu)$-measurable if and only if $f$ is $I$-measurable.

(4) Suppose that the initial vector lattice $E$ is the vector space generated by all functions of the form $1_A$ with $A$ in $E$, where $E$ is a semi-ring in a measure space $(X, \mathcal{F}, \mu)$, and $I$ is defined as in Exercise 4.5. Assume that $\mathcal{F} = \sigma(\mathcal{E})$ and that there is a sequence $\{E_n\}$ of sets in $\mathcal{E}$ such that $X = \bigcup_n E_n$ with $\mu(E_n) < \infty$. Can we affirm that any almost everywhere measurable function $f$ is an almost everywhere pointwise limit of a sequence of functions in $E$ or perhaps in $E$?

Hint: the identities $(x - y) \wedge z = x \wedge (z + y \wedge z) - y \wedge z$, for real numbers $x, y, z$ with $x \geq 0$, $(a - 1) + b = (a \wedge (b + a \wedge 1) - 1) +$, for any real numbers $a, b$ with $b \geq 0$, and the monotone limit $1_A = \lim_n (n[f(\cdot)/a - 1]) + 1$, with $A = \{x \in \Omega : f(x) > a > 0\}$ may be of some help. □
Exercise 4.23. Let $f$ be a function defined on $\mathbb{R}^d$ and let $B(x, r)$ denote the open ball $\{y \in \mathbb{R}^d : |y - x| < r\}$. (1) Prove that the functions $\underline{f}(x, r) = \inf \{f(y) : y \in B(x, r)\}$ and $\overline{f}(x, r) = \sup \{f(y) : y \in B(x, r)\}$ are lower semi-continuous (lsc) and upper semi-continuous (usc) with respect to $x$. (2) Can we replace the open ball with the closed ball $\overline{B}(x, r) = \{y \in \mathbb{R}^d : |y - x| \leq r\}$? (3) Why the functions $\underline{f}(x) = \inf_{r > 0} \underline{f}(x, r)$ and $\overline{f}(x) = \sup_{r > 0} \overline{f}(x, r)$ are also usc and lsc, respectively? (4) Discuss what could be the meaning of $\overline{f}(x, r)$ and $\underline{f}(x, r)$ if the function $f$ is only almost everywhere defined, see Exercises 1.22 and 5.1. In this case, can we deduce that $\overline{f}(x, r)$ and $\underline{f}(x, r)$ are usc and lsc, almost everywhere?

Proof. (1) To show that $x \mapsto \overline{f}(x, r)$ is upper semi-continuous (usc) we must show $\overline{f}(x, r) \geq \limsup_{y \to x} \overline{f}(x, r)$. Therefore, let $\{y_n\}$ be a sequence convergent to $x$. For any $r > 0$ there exists $n_r$ such that $y_n$ belongs to $B(x, r)$, for every $n \geq n_r$, which implies that $\overline{f}(x, r) \geq f(y_n) \geq \overline{f}(y_n, r)$, for every $n \geq n_r$. Hence $\overline{f}(x, r) \geq \limsup_{y \to x} \overline{f}(x, r)$. Similarly, we deduce that $x \mapsto \underline{f}(x, r)$ lower semi-continuous (lsc).

(2) In the above proof we see that replacing the open ball $B(x, r)$ with the closed ball $\overline{B}(x, r) = \{y \in \mathbb{R}^d : |y - x| \leq r\}$ to define $\underline{f}(x, r)$ and $\overline{f}(x, r)$ does not change their lsc or usc properties.

(3) We may use the fact that the supremum (infimum) of a family of lower (upper) semi-continuous is also lower (upper) semi-continuous. Alternatively, a direct proof goes as follows: if $x_n \to x$ then $f(x_n) \geq f(x_n, r)$, for every $n$ and $r > 0$, which implies that $\liminf_n f(x_n) \geq f(x, r)$ in view of the usc of $x \mapsto f(x, r)$. Taking the supremum in $r > 0$ the proof is completed.

(4) Certainly, if $f = g$ a.e. then $\overline{f}(x, r) = \overline{g}(x, r)$, a.e., or $\underline{f}(x, r) = \underline{g}(x, r)$, a.e., does not follows. To work with almost everywhere defined functions, we need to use the essential supremum and infimum, i.e., redefine $\underline{f}(x, r) = \text{ess-inf}\{f(y) : y \in B(x, r)\}$ and $\overline{f}(x, r) = \text{ess-sup}\{f(y) : y \in B(x, r)\}$, where this means that (a) $f(x, r) \leq f(y)$ a.e. $y$ in $B(x, r)$, and (b) for every $s > f(x, r)$ there exist a set of positive measure $S \subset B(x, r)$ such that $s > f(y)$ for every $y$ in $S$. In particular, we may have the undesirable situation where $f(x) > \overline{f}(x, r)$ or $f(x) < \underline{f}(x, r)$, but we could expect that for each $r > 0$, $\underline{f}(x, r) \leq f(x) \leq \overline{f}(x, r)$ a.e. holds true. This could means that $\overline{f}(x) = \inf_{r > 0} \overline{f}(x, r)$ and $\underline{f}(x) = \sup_{r > 0} \underline{f}(x, r)$ are usc and lsc almost everywhere. Following the idea of inferior and superior limits, we may define the essential inferior and superior limits as $\liminf_{y \to x} f(y) = \lim_{r \to 0} \underline{f}(x, r) = \underline{f}(x)$ and $\limsup_{y \to x} f(y) = \lim_{r \to 0} \overline{f}(x, r) = \overline{f}(x)$. A partial answer to these questions is developed later, see Exercise 7.9. \[\square\]
**Exercise 4.24.** Let \((X, \mathcal{X}, \mu)\) and \((Y, \mathcal{Y}, \nu)\) be two \(\sigma\)-finite measure spaces. Prove Minkowski’s integral inequality, i.e.,

\[
\left[ \int_X \left| \int_Y f(x,y) \nu(dy) \right|^p \mu(dx) \right]^{1/p} \leq \int_Y \left( \int_X |f(x,y)|^p \mu(dx) \right)^{1/p} \nu(dy)
\]

for any real-valued \((\mu \times \nu)\)-measurable function \(f\).

**Proof.** Remark that by means of Hölder inequality and Fubini Theorem 4.14, a relative weaker estimate follows, namely,

\[
\left[ \int_X \left| \int_Y f(x,y) \nu(dy) \right|^p \mu(dx) \right]^{1/p} \leq C \left( \int_Y \int_X |f(x,y)|^p \mu(dx) \nu(dy) \right)^{1/p}
\]

with \(C = (\nu(Y))^{1/q}, 1/p + 1/q = 1\). On the other hand, if the function \(f(x,y)\) is interpreted as a function \(f(x, \cdot)\) taking values in \(L^1(Y, \mathcal{Y}, \nu)\) and the integral in \(\mu\) is regarded as the \(p\)-norm in the \(X\), then Minkowski’s integral inequality can be written as

\[
\left\| \int_Y f(\cdot, y) \nu(dy) \right\|_{L^p(\mu)} \leq \int_Y \left\| f(\cdot, y) \right\|_{L^p(\mu)} \nu(dy).
\]

Since the integral is a limit of sums, this estimate is expected to hold and to be deduced from the triangular inequality (Minkowski’s inequality) of the \(p\)-norm.

To prove this estimate, let us proceed as follow. First, if \(f = \sum_{i=1}^n c_i 1_{C_i}\), with constants \(c_i\) and disjoint measurable sets \(C_i = A_i \times B_i\) in \(X \times Y\) with \(\mu(A_i) < \infty\) and \(\nu(B_i) < \infty\) (note that \(A_1, \ldots, A_n\) or \(B_1, \ldots, B_n\) are not necessarily disjoint in \(X\) or \(Y\), respectively) then the equality

\[
\int_Y \left( \int_X |f(x,y)|^p \mu(dx) \right)^{1/p} \nu(dy) = \sum_{i=1}^n |c_i| \int_Y \left( \int_X \left| 1_{C_i}(x,y) \right|^p \mu(dx) \right)^{1/p} \nu(dy)
\]

and Minkowski’s inequality for \(\mu(dx)\)

\[
\left[ \int_X \left| \int_Y f(x,y) \nu(dy) \right|^p \mu(dx) \right]^{1/p} \leq \sum_{i=1}^n |c_i| \left[ \int_X \left| 1_{A_i}(x) \right|^p \mu(dx) \right]^{1/p} \left| \int_Y \left| 1_{B_i}(y) \right|^p \nu(dy) \right|,
\]

show the validity of Minkowski’s integral inequality for any functions of the above form, i.e., if \(\mathcal{C}\) denotes the ring generated by measurable sets of the
product form $A \times B$ with $\mu(A_i) < \infty$ and $\nu(B_i) < \infty$, and if $\mathcal{S}(\mathcal{C})$ denotes
the class of all simple function with respect to $\mathcal{C}$, then Minkowski’s integral
inequality is valid for any function $f$ in the class $\mathcal{S}(\mathcal{C})$.

Therefore, if $\mathcal{M}$ is the class of all measurable functions in $X \times Y$ such that
Minkowski’s integral inequality holds true, then $\mathcal{S}(\mathcal{C}) \subset \mathcal{M}$.

Now, for any simple integrable function $f$ there exists a sequence $\{f_n\}$ of
functions in $\mathcal{S}(\mathcal{C})$ such that $f_n \to f$ almost everywhere and $|f_n| \leq g$, for some
simple integrable function $g$. Indeed, working with each term of the given simple
function, it suffices to recall that for any measurable set $F$ with finite product
measure $\mu \times \nu$ and any $\varepsilon > 0$, there exists a set $C$ in $\mathcal{C}$ such that $F \subset C$ and
$(\mu \times \nu)(F \setminus C) < \varepsilon$. Therefore, this approximation and the dominate convergence
Theorem prove that $\mathcal{M}$ contains all simple integrable functions.

Since any nonnegative integrable function is a pointwise limit of an increasing
sequence of simple integrable functions, the monotone convergence Theorem
ensures that $\mathcal{M}$ contains all nonnegative integrable functions, and the inequality

$$\left[ \int_X \left( \int_Y \left| f(x,y) \nu(dy) \right|^p \mu(dx) \right]^{1/p} \leq \left[ \int_X \left( \int_Y |f(x,y)| \nu(dy) \right)^p \mu(dx) \right]^{1/p},$$

shows that Minkowski’s integral inequality holds for every integrable function $f$, non necessarily nonnegative.

Note that there are several possible arguments to establish this estimate, e.g., the dual representation, with $1/p + 1/q = 1$,

$$\left[ \int_X \left( \int_Y \left| f(x,y) \nu(dy) \right|^p \mu(dx) \right]^{1/p} =$$

$$= \sup \left\{ \int_X g(x) \int_Y f(x,y) \nu(dy) \mu(dx) : \int_X |g(x)|^q \mu(dx) = 1 \right\},$$

Fubini Theorem 4.14 and Hölder inequality yield

$$\left| \int_X g(x) \int_Y f(x,y) \nu(dy) \mu(dx) \right| \leq \int_Y \nu(dy) \int_X |g(x)f(x,y)| \mu(dx) \leq$$

$$\leq \left( \int_X |g(x)|^q \mu(dx) \right)^{1/q} \int_Y \left( \int_X |f(x,y)|^p \mu(dx) \right)^{1/p} \nu(dy)$$

and Minkowski’s integral inequality follows, e.g., see Hardy et al. [60, Theorem
202, p. 148–150].

**Exercise 4.25.** Actually, Hölder inequality (4.7) can be generalized as follows

$$\|fg\|_r \leq \|f\|_p \|g\|_q, \quad \forall f \in L^p, \ g \in L^q,$$

for any $p, q, r > 0$ and $1/p + 1/q = 1/r$. Indeed, (1) make an argument to reduce
the inequality to the case $r = 1$ and $\|f\|_p = \|g\|_q = 1$; and (2) use calculus to
get first $|fg| \leq |f|^p/p + |g|^q/q$ and next the conclusion.
Proof. (1) If \( \|f\|_p = 0 \) or \( \|g\|_q = 0 \) then \( fg = 0 \) and the equality holds. Thus, define \( p' = p/r, \ q' = q/r, \ f = |f|^r \) and \( g = |g|^r \) to have \( 1/p' + 1/q' = 1 \)
\[ \|fg\|_r^r = \|f\|_p^r, \|f\|_p^r = \|f\|_{p'}^p \] and \( \|g\|_q^q = \|g\|_{q'}^q \), i.e., to effectively reduce the problem to the case \( r = 1 \).
Next, by setting \( \bar{f} = f\|f\|_p^{-p} \) and \( \bar{g} = g\|g\|_p^{-p} \), the problem is reduced to the case \( \|f\|_p = \|g\|_1 = 1 \).

(2) Since the function \( t \mapsto \ln(t) \) is (strictly) concave, if \( a, b, \alpha, \beta \) are any positive numbers satisfying \( \alpha + \beta = 1 \) then \( \ln(\alpha a + \beta b) \geq \alpha \ln(a) + \beta \ln(b) \), which yields the inequality \( \alpha a + \beta b \geq a^\alpha + b^\beta \), i.e., the arithmetic mean is bounded below by the geometric mean. Now, take \( a^\alpha = |f|, \ b^\beta = |g|, \ \alpha = 1/p \) and \( \beta = 1/q \), to deduce
\[
|f|g = a^\alpha b^\beta \leq \alpha a + \beta b = \frac{|f|^p}{p} + \frac{|g|^q}{q},
\]
which implies \( \|fg\|_1 \leq 1 \), after integration, i.e., the desired estimate. \( \square \)

Exercise 4.26. Let \((\Omega, F, \mu)\) be a measure space with \( \mu(\Omega) < \infty \). Prove (a) that \( \|f\|_p \to \|f\|_\infty \) as \( p \to \infty \), for every \( f \in L_\infty \). On the other hand, on a \( \mathbb{R}^d \) with the Lebesgue measure and a function \( f \) in \( L^p \), (b) verify that the function \( x \mapsto \|f(\cdot)\mathbb{1}_{\cdot-x}r\|_p \) is continuous, for any \( r > 0 \). Next, show that (c) the function \( x \mapsto f^\sharp(x, r) = \text{ess-sup}_{|y-x|<r} |f(y)| \) is lower semi-continuous (lsc) for every \( r > 0 \), and therefore (d) the function \( x \mapsto f^\sharp(x) = \lim_{r \to 0} f^\sharp(x, r) \) is Borel measurable, see related Exercises 7.9, 4.23 and 1.22.

Proof. (a) If \( \mu(\Omega) < \infty \) then
\[
\left( \int_\Omega |f(x)|^p \mu(dx) \right)^{1/p} \leq \|f\|_\infty \left( \mu(\Omega) \right)^{1/p},
\]
which yields \( \limsup_{p \to \infty} \|f\|_p \leq \|f\|_\infty \). Similarly, if \( 0 < r < \|f\|_\infty \) then there exists a measurable set \( A \) with \( \mu(A) > 0 \) such that \( |f(x)| > r \) for every \( x \) in \( A \). Hence \( r(\mu(A))^{1/p} \leq \|f\|_p \), which implies \( r \leq \liminf_{p \to \infty} \|f\|_p \), and therefore \( \|f\|_\infty \leq \liminf_{p \to \infty} \|f\|_p \). This proves that \( \|f\|_p \to \|f\|_\infty \) as \( p \to \infty \).

(b) For any fixed \( y \) and \( r \), the function \( x \mapsto \mathbb{1}_{|y-x|<r} \) is continuous whenever \( |y - x| \neq r \), which implies that if \( x \to x' \) then \( \mathbb{1}_{|y-x|<r} \to \mathbb{1}_{|y-x'|<r} \) for almost every \( y \) in \( \mathbb{R}^d \). Because \( |f(y)|\mathbb{1}_{|y-x|<r} \leq |f(x)| \) the Lebesgue dominate convergence implies that the function \( x \mapsto \|f(\cdot)\mathbb{1}_{\cdot-x}r\|_p \) is continuous.

(c) Since \( f^\sharp(x, r) = \|f\mathbb{1}_{\cdot-x}r\|_\infty \), if \( \Omega_n = \{x \in \mathbb{R}^d : |x| \leq n\} \) then, in view of part (a), the essential supremum can be expressed as
\[
f^\sharp(x, r) = \sup_{p,n \geq 1} \left( \mu(\Omega_n) \right)^{-1/p} \|f\mathbb{1}_{\cdot-x}r\mathbb{1}_{\Omega_n}\|_p,
\]
which proves, after invoking part (b), that \( x \mapsto f^\sharp(x, r) \) is lower semi-continuous.

(d) Because \( f^\sharp(x) = \inf_{r>0} f^\sharp(x, r) \) and \( r \mapsto f^\sharp(x, r) \) is decreasing, the infimum can be taken only for \( r \) in a countable subset (e.g., positive rational)
and each function $x \mapsto f^y(x, r)$ is Borel measurable (actually, lsc), the infimum yields a Borel measurable function.

If the expression $1_{|y-x|<r}$ is replaced by a open ball $B(x, r)$ of radius $r$ centered at $x$ then all these arguments can be adapted with any $\sigma$-finite measure on a metric space $\Omega$. Indeed, the function $x \rightarrow 1_{y \in B(x, r)}$ is lsc if the ball is open and usc if the ball is closed. Hence, even if the boundary $\partial B(x, r)$ has zero $\mu$-measure, the function $x \rightarrow f^y(x)$ results Borel measurable.

**Exercise 4.27.** Let $(X, \mathcal{X}, \mu)$ and $(Y, \mathcal{Y}, \nu)$ be two $\sigma$-finite measure spaces. Suppose $k(x, y)$ is a real-valued ($\mu \times \nu$)-measurable function such that

$$
\int_X |k(x, y)| \mu(dx) \leq C, \text{ a.e. } y \text{ and } \int_Y |k(x, y)| \nu(dy) \leq C, \text{ a.e. } x,
$$

for some constant $C > 0$. Prove (1) that the expression

$$(Kf)(x) = \int_Y k(x, y)f(y)\nu(dy), \quad \forall f \in L^p(\nu),$$

defines a linear continuous operator from $L^p(\nu)$ into $L^p(\mu)$, for any $1 \leq p \leq \infty$, i.e., the functions (a) $y \mapsto k(x, y)f(y)$ is $\nu$-integrable for almost every $x$, (b) $x \mapsto (Kf)(x)$ belongs to $L^p(\mu)$, and (c) the mapping $f \mapsto Kf$ is linear and there exists a constant $C > 0$ such that $\|Kf\|_p \leq C\|f\|_p$, for every $f$ in $L^p(\nu)$. Now, same setting, but under the assumption that the kernel $k(x, y)$ belongs to the product space $L^p(\mu \times \nu)$, prove (2) that $K$ is again a linear continuous operator from $L^q(\nu)$ into $L^p(\mu)$ and $\|Kf\|_p \leq \|k\|_p\|f\|_q$, with $1/p + 1/q = 1$.

**Proof.** (1) Write $|k(x, y)| = |k(x, y)|^{1/p} |k(x, y)|^{1/q}$ with $1/p + 1/q = 1$ and use Hölder inequality to obtain

$$
\left| \int_Y k(x, y)f(y)\nu(dy) \right|^p \leq \left( \int_Y |k(x, y)|\nu(dy) \right)^{p/q} \int_Y |f(y)|^p \nu(dy)
$$

Now, exchange the order of the integrals (Fubini Theorem) to get

$$
\|Kf\|_{L^p(\mu)}^p = \int_X \left| \int_Y k(x, y)f(y)\nu(dy) \right|^p \mu(dx) \leq C^{p/q} \int_Y |f(y)|^p \nu(dy) \int_X |k(x, y)|\mu(dx) \leq C^p \|f\|_{L^p(\nu)}^p,
$$

which prove that $K$ is a linear continuous operator from $L^p(\nu)$ into $L^p(\mu)$, for any $1 \leq p \leq \infty$, where the case $p = \infty$ is treated separately.

(2) Use Hölder inequality to obtain

$$
\left| \int_Y k(x, y)f(y)\nu(dy) \right|^p \leq \left( \int_Y |k(x, y)|^p \nu(dy) \right) \left( \int_Y |f(y)|^q \nu(dy) \right)^{p/q},
$$

and integrate in $\mu(dx)$ to deduce

$$
\|Kf\|_{L^p(\mu)}^p \leq \|k\|_{L^p(\mu \times \nu)}^p \|f\|_{L^q(\nu)}^p
$$

as desired. \qed
Exercise 4.28. Let \( \{\psi_n\} \) be a sequence of functions belonging to \( L^1(\mathbb{R}^d) \) such that \( \sup_n \|\psi_n\|_1 < \infty \) and

\[
\lim_n \int_{\mathbb{R}^d} \psi_n(x) \, dx = 1 \quad \text{and} \quad \lim_n \int_{|x| > \delta} |\psi_n(x)| \, dx = 0, \quad \forall \delta > 0.
\]

Prove that for every \( f \) in \( L^p(\mathbb{R}^d) \), \( 1 \leq p < \infty \), we have \( \|f \ast \psi_n - f\|_p \to 0 \).

**Proof.** Let us use the typical truncation and density arguments. First for a given \( f \) in \( L^p(\mathbb{R}^d) \) define \( f_r(x) = f(x)1_{|x| \leq r} \), with \( r > 0 \). The dominate convergence Theorem implies that \( f_r \to f \) in \( L^p(\mathbb{R}^d) \).

Next, any \( \varepsilon > 0 \) and for any function \( f \) in \( L^p(\mathbb{R}^d) \) find \( r \) sufficiently large so that \( \|f - f_r\|_p < \varepsilon/2 \), and because \( f_r \) vanishing outside of the ball \( |x| \leq r \), there exists a (uniformly) continuous function \( g_r \) with compact support in \( \mathbb{R}^d \) such that \( \|f_r - g_r\|_p < \varepsilon/2 \), i.e., \( \|f - g_r\|_p < \varepsilon \).

Now, the estimate

\[
\|\psi_n \ast (f - f)\|_p \leq \left( \sup_n \|\psi_n\|_1 \right) \|f - g_r\|_p + \|\psi_n \ast (g_r - g_{r \varepsilon})\|_p
\]

shows that

\[
\lim_n \|\psi_n \ast (f - f)\|_p \leq \varepsilon + \lim_n \|\psi_n \ast (g_r - g_{r \varepsilon})\|_p,
\]

which effectively reduces the problem to the case when the given function \( f \) is a bounded and uniformly continuous in \( \mathbb{R}^d \).

For a given \( x \) in \( \mathbb{R}^d \), the equality

\[
(\psi_n \ast f)(x) = f(x) = \int_{|y| \leq \delta} \psi_n(y) \left[ f(x - y) - f(x) \right] \, dy +
\]

\[
+ \int_{|y| > \delta} \psi_n(y) \left[ f(x - y) - f(x) \right] \, dy,
\]

yields the estimate

\[
| (\psi_n \ast f)(x) - f(x) | \leq \|\psi_n\|_1 \left( \sup_{|y| \leq \delta} | f(x - y) - f(x) | \right) +
\]

\[
+ 2 \| f \|_\infty \int_{|y| > \delta} | \psi_n(y) | \, dy.
\]

Because \( f \) is a bounded and uniformly continuous in \( \mathbb{R}^d \), the assumptions on the kernels \( \psi_n \) implies that \( | (\psi_n \ast f)(x) - f(x) | \to 0 \), uniformly for \( x \) in \( \mathbb{R}^d \).

Even if \( f \) has a compact support, the functions \( \psi_n \ast f \) do not necessarily have compact supports. However, writing \( |\psi_n| = |\psi_n|^{1/q} |\psi_n|^{1/p} \) with \( 1/p + 1/q = 1 \), Hölder inequality yields

\[
\int_{\mathbb{R}^d} |(\psi_n \ast f)(x) - f(x)|^p \, dx \leq \|\psi_n\|_1^p \sup_{|y| \leq \delta} \left( \int_{\mathbb{R}^d} |f(x - y) - f(x)|^p \, dx \right) +
\]

\[
+ 2 \| f \|_p^p \left( \int_{|y| > \delta} |\psi_n(y)| \, dy \right)^p.
\]
Hence, taking first limit as $n \to \infty$ and then as $\delta \to 0$, we obtain $\| f \star \psi_n - f \|_p \to 0$, for any $f$ in $L^p(\mathbb{R}^d)$.

Remark that if the continuity in $L^p$ of the translation is used, i.e., $\| f(\cdot - y) - f \|_p \to 0$ as $|y| \to 0$, then the last estimate yields the whole argument. $\square$
Exercises - Chapter (5)
Integrals on Euclidean Spaces

(5.1) Multidimensional Riemann Integral

Exercise 5.1. Prove the following list of properties:

1. If $f(x) = -\infty$ (or $f(x) = +\infty$) then $f$ is lsc (or usc) at $x$;
2. If $f(x) = +\infty$ then $f$ is lsc at $x$ if and only if $\lim_{y \to x} f(y) = +\infty$;
3. $\underline{f}$ (or $\overline{f}$) is the largest lsc (or smallest usc) function above (below) $f$;
4. $f$ is continuous at $x$ if and only if $f$ is lsc and usc at $x$;
5. $f$ is lsc (or usc) at $x$ if and only if $f(x) \leq f(y)$ (or $f(x) \geq f(y)$);
6. If $f_i$ is lsc (or usc) for all $i$ then $\sup_i f_i$ (or $\inf_i f_i$) is also lsc (or usc), moreover, if the family $I$ is finite then $\inf_i f_i$ (or $\sup_i f_i$) results lsc (or usc);
7. $f$ is lsc (or usc) if and only if $f^{-1}(\ ]a, +\infty\])$ (or $f^{-1}(\ ]-\infty, a\[)$) is an open set for every real number $a$, as a consequence, any lsc (or usc) function $f$ is Borel measurable;
8. If $f = g$ except in a set $D$ of isolated points then $\underline{f}(x) = g(x)$ and $\overline{f}(x) = g(x)$, for every $x$ not in $D$;
9. If $K$ is a compact set and $f$ is lsc (usc) then the infimum (supremum) of $f$ in $K$ is attained at a point in $K$.

Proof. It is clear that (1) though (6) follow essentially from the definition, so let us check (7) when $f$ is lsc. To this purpose, if $x$ belongs to the pre-image $f^{-1}(\ ]a, +\infty\])$ then $f(x) > a$, and by definition of inferior limit, there exist $\delta > 0$ such that $f(y) > a$ for every $|y - x| < \delta$, i.e., the pre-image $f^{-1}(\ ]a, +\infty\])$ is an open set, and the converse follows in similar manner.

Certainly, isolate points play not role for the inferior limits, so that (8) holds true. Now, to check (9) take a minimizing sequence $\{x_n\}$ for $f$ in $K$, i.e.,
$f(x_n) \to \inf_K f$ with $x_n$ in $K$. Extracting a convergent subsequence $\{x_{n_k}\}$, $x_{n_k} \to x_0$, we have $f(x_0) \leq \lim_{k} f(x_k)$, i.e., the infimum is achieved at $x_0$. □
Exercise 5.2. Let \( \{r_n\} \) be the sequence of all rational numbers and define the function \( f(x) = \sum_{n \geq 1} 2^{-n}(x - r_n)^{-1/2}\mathbb{1}_{x-r_n \in (0,1)} \) for every \( x \) in \( \mathbb{R} \). Prove (a) \( f \) is Lebesgue integrable in \( \mathbb{R} \), (b) \( f \) is unbounded in any interval of \( \mathbb{R} \), and (c) \( f^2 \) is not Lebesgue integrable in \( \mathbb{R} \).

Proof. Consider the sequence \( \{f_k\} \) of functions given by \( f_k(x) = \sum_{n=1}^{k} 2^{-n}(x - r_n)^{-1/2}\mathbb{1}_{x-r_n \in (0,1)} \), for every \( x \) in \( \mathbb{R} \), where the expression make sense. Certainly, the limit function \( f \) is defined for every irrational number. These functions are defined, bounded and continuous in \( \mathbb{R} \), except in a finite number of points, and therefore, they are integrable.

(a) Since the sequence \( \{f_k\} \) is increasing, the monotone convergence implies

\[
\int_{\mathbb{R}} f(x)\,dx = \lim_{k \to \infty} \int_{\mathbb{R}} f_k(x)\,dx \leq \sum_{n} 2^{-n} \int_{r_n}^{1+r_n} (x - r_n)^{-1/2}\,dx \leq 1,
\]
i.e., \( f \) is Lebesgue integrable.

(b) If \((a, b)\) is a nonempty interval then we can find a sequence \( \{s_i\} \subset (a, b) \) of irrational numbers and another sequence \( \{t_i\} \) of rational numbers such that \( 0 < s_i - t_i < 2^{-2i} \). Because the sequence \( \{r_n\} \) contains all rational numbers, for every \( i \) there exists \( k_i \) such that \( \{t_1, \ldots, t_i\} \) are listed in \( \{r_1, \ldots, r_{k_i}\} \). Thus, the functions \( f_k(s_i) \geq 2^{-i}(s_i - t_i)^{-1/2} \geq 2^i \) for any \( k \geq k_i \), which shows that \( \{f(s_i)\} \) is an unbounded sequence, i.e., \( f \) is unbounded in any interval of \( \mathbb{R} \).

(c) Since \( f^2(x) \geq [2^{-n}(x - r_n)^{-1/2}\mathbb{1}_{x-r_n \in (0,1)}]^2 \),
\[
\int_{\mathbb{R}} f^2(x)\,dx \geq 2^{-2n} \int_{r_n}^{1+r_n} (x - r_n)^{-1}\,dx = +\infty,
\]
the function \( f^2 \) is not integrable.

\( \Box \)

Exercise 5.3. Calculate the limit

\[
A = \lim_{n \to \infty} \int_{a}^{\infty} f_n(x)\,dx,
\]
for \( f_n(x) = n/(1 + n^2x^2) \), \( a > 0 \), \( a = 0 \) and \( a < 0 \).

Proof. Using the sequence \( f_{n,k} = f_n\mathbb{1}_{[a,k]} \) for \( k = 1, 2, \ldots \), we deduce
\[
\int_{a}^{\infty} f_n(x) = \lim_{k \to \infty} [\arctan(nk) - \arctan(na)] = \frac{\pi}{2} - \arctan(na).
\]
Thus, \( A = 0 \) if \( a > 0 \), \( A = \pi/2 \) if \( a = 0 \) and \( A = \pi \) if \( a < 0 \).

\( \Box \)

Exercise 5.4. Prove that the function \( x \mapsto |x|^{-\alpha} \) is Lebesgue integrable (a) on the unit ball \( B = \{x \in \mathbb{R}^d : |x| < 1\} \) if an only if \( \alpha < d \) and (b) outside the unit ball \( \mathbb{R}^d \setminus B \) if and only if \( \alpha > d \).
Proof. Using spherical coordinates in \( \mathbb{R}^d \), if \( \omega_n = 2\pi^{d/2}/\Gamma(d/2) \) is the surface area of the unit ball then
\[
\int_{\mathbb{R}^d} |x|^{-\alpha} \mathbbm{1}_{|x|<1} = \omega_d \int_0^1 r^{-\alpha} r^{d-1} dr
\]
and
\[
\int_{\mathbb{R}^d} |x|^{-\alpha} \mathbbm{1}_{|x|>1} = \omega_d \int_1^\infty r^{-\alpha} r^{d-1} dr,
\]
which show (a) and (b). \( \square \)

Exercise 5.5. Prove the first part of the above Theorem 5.3, namely, on the Lebesgue measure space \((\mathbb{R}^d, \mathcal{L}, \ell)\), verify that if \( T : \mathbb{R}^d \to \mathbb{R}^d \) is a homeomorphism of class \( C^1 \) then a set \( N \) is negligible if and only if \( T(N) \) is negligible. Hence, deduce that a set \( E \) (or a function \( f \)) is a (Lebesgue) measurable if and only if \( T(E) \) (or \( f \circ T \)) is (Lebesgue) measurable.

Proof. There are several ways of establishing the first part. First, because \( T \) is invertible, we only need to show that if \( N \) is negligible then so is its image \( T(N) \). Actually, writing \( N = \bigcup_n N \cap B(0, n) \), where \( B(0, n) \) is the ball centered at the origin with radius \( n \), we may assume that \( N \) is also bounded. At this point, it suffices to establish an estimate of the type: for any bounded set \( K \) there exists a constant \( C_K \) such that \( \ell^*(T(B(x, r))) \leq C_K r^d \), for any \( x \) and \( r > 0 \) as long as \( B(x, r) \subset K \). However, since \( T \) satisfies \( |T(y) - T(y')| \leq M_K |y - y'| \), for every \( y, y' \) in \( K \), the image \( T(B(x, r)) \) must be contained in a ball of radius at most \( 2M_K r \), which yields \( C_K = (2M_K)^d \). Actually, we are proving that if \( T \) is Lipschitz continuous in a set \( K \) then the Lebesgue exterior measure \( \ell^*(T(B)) \leq C_K \ell(B) \) for every ball \( B \) inside \( K \).

For the second part, we may use that fact that \( T \) maps Borel sets into Borel sets and that any Lebesgue measurable set is the union of a Borel set and a negligible set.

The reader may want to check Jones [65, Chapter 15, pp. 494-510] or Stroock [112, 2.2.1 Lemma, pp. 30-31], among others. \( \square \)

Exercise 5.6. Let \( \ell \) be the Lebesgue measure in \( \mathbb{R}^d \), \( E \) a measurable set with \( \ell(E) < \infty \), and \( f : \mathbb{R}^d \to [0, \infty] \) be a measurable function. Define \( F_{f,E}(r) = \ell(\{x \in E : f(x) \leq r\}) \) and prove (a) \( r \mapsto F_{f,E}(r) \) is a right-continuous increasing function and (b) that the Lebesgue-Stieltjes measure \( \ell_F \) induced by \( F_{f,E} \) is equal to the \( f \)-image measure of the restriction of \( \ell \) to \( E \), i.e., \( \ell_F = m_{f,E} \), with \( m_{f,E}(B) = \ell(f^{-1}(B) \cap E) \), for every Borel set \( B \) in \( \mathbb{R} \). Now, (c) prove that for any Borel measurable function \( g : [0, \infty) \to [0, \infty] \) we have the equality
\[
\int_E g(f(x)) \ell(dx) = \int_0^\infty g(r) m_{f,E}(dr) = \int_0^\infty g(r) \ell_F(dr).
\]
In particular, if \( \lambda_f(r) = \ell(\{x \in \mathbb{R}^d : |f(x)| > r\}) \), \( m_f(B) = \ell(f^{-1}(B)) \), for every Borel set \( B \) in \( \mathbb{R} \), and \( f \) is any measurable then deduce that

\[
\int_{\mathbb{R}^d} |f(x)|^p \ell(dx) = \int_0^\infty r^p m_f(dr) = \int_0^\infty pr^{p-1} \lambda_f(r)dr, \quad \forall p > 0.
\]

Actually, verify this assertion holds for any measure space \((\Omega, \mathcal{F}, \mu)\), any measurable function \( f : \Omega \to \mathbb{R} \), and any \( \sigma \)-finite measurable set \( E \), e.g., see Wheeden and Zygmund [119, Section 5.4, pp. 77–83] for more details. Hint: consider first the case \( g = 1_B \), for a Borel measurable set \( B \), and regard both sides of the equal sign as measures.

**Proof.** Essentially, by construction \( r \mapsto F(r) = F_{f,E}(r) \) is a right-continuous increasing function. Next, since \( \ell_f([a,b]) = F_{f,E}(b) - F_{f,E}(a) \), it is clear that \( \ell_f([a,b]) = \ell(f^{-1}([a,b]) \cap E) \), which yields (b).

To show (c), we note that for \( g = 1_B \) we have

\[
m_{f,E}(B) = \int_0^\infty \mathbb{1}_B(r)m_{f,E}(dr) = \int_0^\infty \mathbb{1}_B(r)\ell_F(dr) = \int_E \mathbb{1}_B(f(x))\ell(dx),
\]

trivially for any interval \( B = [a,b] \), and because all expressions are measures in \( B \), the equality holds true for any Borel measurable set \( B \). Next, approximating any Borel measurable function \( g : [0, \infty] \to [0, \infty] \) by an increasing sequence of simple functions, we obtain (c).

Finally, we remark that the role of the Lebesgue measure \( \ell \) in \( \mathbb{R}^d \) is unimportant in the previous analysis, and the same arguments remain true for any measure \((\Omega, \mathcal{F}, \mu)\), any measurable function \( f : \Omega \to \mathbb{R} \), and any \( \sigma \)-finite measurable set \( E \) in \( \Omega \).

**Exercise 5.7.** Consider a measurable function \( f : \mathbb{R}^d \times \mathbb{R}^k \to \mathbb{R} \) such that the expression

\[
F(x) = \int_{\mathbb{R}^d} f(x,y)dy, \quad \forall x \in \mathbb{R}^d.
\]

produces a continuous function. Give precise assumptions on the function \( f \) to be able to use the dominate convergence and to show that \( F \) is differentiable at a point \( x_0 \).

**Proof.** To show that \( F \) is differentiable at a point \( x_0 \) we may assume that \( f(x,y) \) is locally bounded at \( x_0 \) by a integrable function, i.e., there exists \( \delta = \delta(x_0) > 0 \) such that \( |\partial_x f(x,y)| \leq g(y) \) for every \( x \) in the interval \((x_0 - \delta, x_0 + \delta)\) for some integrable function \( g \). In this case, for some \( x \) between \( x_0 + h \) and \( x_0 \) we have

\[
|f(x_0 + h, y) - f(x_0, y)|/h = |\partial_x f(x, y)| \leq g(y).
\]

Thus

\[
\lim_{h \to 0} F(x_0 + h) - F(x_0) = \lim_{h \to 0} \int_{\mathbb{R}^d} \frac{f(x_0 + h, y) - f(x_0, y)}{h} \, dy =
\]

\[
= \int_{\mathbb{R}^d} \partial_x f(x_0, y) \, dy.
\]
in view of the dominate convergence. Remark that if \( f \) is a continuously differentiable function in a \( x \)-neighborhood of \((x_0, y)\) for every \( y \), and \( \partial_x f(x_0, y) \) is assumed integrable in \( \mathbb{R}^d \), then the previous assumption (about locally bounded at \( x_0 \) by a integrable function) is not necessarily satisfied.

**Exercise 5.8.** If \( E \) is a non empty subset of \( \mathbb{R}^d \) then (1) prove that the distance function \( x \mapsto d(x, E) = \inf \{|y - x| : y \in E\} \) is a Lipschitz continuous function, moreover, \(|d(x, E) - d(y, E)| \leq |x - y|\), for every \( x, y \) in \( \mathbb{R}^d \). Now, consider Marcinkiewicz integral

\[
M_{E,\alpha}(x) = \int_Q d^\alpha(y, E) |x - y|^{-d-\alpha} dy,
\]

where \( E \) is a compact subset of \( \mathbb{R}^d \), \( \alpha \) is a positive constant and \( Q \) is a cube containing \( E \). Using Fubini-Tonelli Theorem 4.14 and spherical coordinates prove that (2) \( M_{E,\alpha}(x) \) is integrable on \( E \), and that (3)

\[
\alpha \int_E M_{E,\alpha}(x) dx \leq \omega_d \ell(Q \setminus E),
\]

where \( \ell \) is the Lebesgue measure in \( \mathbb{R}^d \) and \( \omega_d \) is the measure of the unit sphere in \( \mathbb{R}^d \), see DiBenedetto [32, Section III.15, pp. 148–151].

**Proof.** To check (1), note that \( d(x, E) \leq |x - e| \leq |x - y| + |y - e| \), for every \( e \) in \( E \) implies \( d(x, E) \leq |x - y| + d(y, E) \), probing that \(|d(x, E) - d(y, E)| \leq |x - y|\).

Since \( d(y, E) = 0 \), for every \( y \) in \( E \), by Fubini-Tonelli Theorem 4.14 we obtain

\[
\int_E M_{E,\alpha}(x) dx = \int_{Q \setminus E} d^\alpha(y, E) \left( \int_E \frac{dx}{|x - y|^{d+\alpha}} \right) dy.
\]

Because \( E \) is a closed set, we have \(|x - y| \geq d(y, E)\), for every \( y \) in \( Q \setminus E \) and \( x \) in \( E \). Thus, using spherical coordinates

\[
\int_E \frac{dx}{|x - y|^{d+\alpha}} \leq \int_{|x-y|^d \geq d(y, E)} \frac{d(x-y)}{|x-y|^{d+\alpha}} = \omega_d \int_{d(y, E)}^{\infty} \frac{ds}{s^{1+\alpha}} = \frac{\omega_d}{\alpha \omega_d \alpha(y, E)},
\]

which shows (2) and (3).

5.2 Riemann-Stieltjes Integrals

**Exercise 5.9** (cag-lad modulo of continuity). Suggested by the arguments in Proposition 5.4, a modulo of continuity for a cag-lad function \( f : [a, b] \to \mathbb{R} \) (i.e., \( f(t+) = \lim_{s \to t, s > t} f(s) \) exists and is finite for every \( t \) in \([a, b] \), \( f(t-) = \lim_{s \to t, s < t} f(s) \) exists and is finite, and \( f(t) = f(t-) \) for every \( t \) in \([a, b] \)) should be defined as

\[
\rho(r; f, [a, b]) = \inf \left\{ \max_{i=1, \ldots, n} \text{osc}(f, [s_{i-1}, s_i]) : a = s_0 < s_1 < \cdots < s_n = b, s_i - s_{i-1} > r \right\}.
\]
Verify that a function $f$ is cag-lad if and only if $\rho(r) \to 0$ as $r \to 0$. Moreover, a more convenient modulo of continuity could be the following

$$
\rho'(r; f, [a, b]) = \sup \{|f(t') - f(t)| \wedge |f(t'') - f(t)| : t' \leq t \leq t'', t' - t' \leq r, t', t'' \in [a, b]\},
$$

with $\wedge$ and $\vee$ denoting the min and the max. Prove that $\rho'(r; f, [a, b]) \leq \rho(r; f, [a, b])$, but the converse inequality does not hold. Finally, state an analogous for result for cad-lag functions. See Billingsley [14, Chapter 3, p. 109-153].

**Proof.** The beginning of the proof in Proposition 5.4 established that for any cag-lad function $f$ and any $\varepsilon > 0$ there exists a finite $\varepsilon$-decomposition, i.e.,

$$
\max_{1 \leq i \leq n} \osc(f, [s_{i-1}, s_i]) \leq \varepsilon \text{ for some } a = s_0 < s_1 < \cdots < s_n = b, \text{ which means exactly that } \rho(r) \to 0 \text{ as } r \to 0.
$$

For any $\varepsilon, r > 0$ find a decomposition $a = s_0 < s_1 < \cdots < s_n = b, s_i - s_{i-1} > r$ such that $\max_{1 \leq i \leq n} \osc(f, [s_{i-1}, s_i]) < \rho(r; f, [a, b]) + \varepsilon$. If $t' \leq t \leq t''$ with $t'' - t' \leq r$ and $t', t''$ in $[a, b]$ then either $t'$ and $t''$ belong to only on little interval $[s_{i-1}, s_i]$ and so $|f(t') - f(t)| \vee |f(t'') - f(t)| < \rho(r; f, [a, b]) + \varepsilon$, or else $t'$ and $t''$ belong to only on double-little interval $[s_{i-1}, s_{i+1}]$ and so $|f(t') - f(t)| < \rho(r; f, [a, b]) + \varepsilon$ when $t' \leq t < s_i$ and $|f(t'') - f(t)| < \rho(r; f, [a, b]) + \varepsilon$ when $s_i \leq t \leq t''$, i.e., $|f(t') - f(t)| \vee |f(t'') - f(t)| < \rho(r; f, [a, b]) + \varepsilon$. Hence $\rho'(r; f, [a, b]) \leq \rho(r; f, [a, b])$.

Taking $f_n(t) = 1$ if $0 \leq t \leq 1/n$ and $f_n(t) = 0$ if $1/n < t \leq 1$ then $\rho'(r; f_n, [a, b]) = 0$, while $\rho(r; f_n, [a, b]) = 1$ for $r \geq 1/n$.

For cad-lag functions, the modulo of continuity is defined as

$$
\rho(r; f, [a, b]) = \inf \{ \max_{i=1, \ldots, n} \osc(f, [s_{i-1}, s_i]) : a = s_0 < s_1 < \cdots < s_n = b, s_i - s_{i-1} > r \},
$$

but the expression of $\rho'(r; f, [a, b])$ is the same. \qed

**Exercise 5.10** (change-of-time). Let $\nu$ be a $\sigma$-finite measure on the measurable space $(X, \mathcal{F})$, and $c$ be a nonnegative real-valued Borel measurable function on $X \times [0, \infty)$ and, for every fixed $x$ in $X$, define

$$
\tau^{-1}(x, s) = \int_0^s c(x, r)dr \text{ and } \tau(x, t) = \inf\{s > 0 : \tau^{-1}(x, s) > t\},
$$

and $\tau(x, t) = \infty$ if $\tau^{-1}(x, s) \leq t$ for every $s \geq 0$. Assume that $\tau^{-1}(x, s)$ is finite for every $(x, s)$ in $X \times [0, \infty)$, and that $\tau^{-1}(x, s) \to \infty$ as $s \to \infty$. Consider $\tau^{-1}$ and $\tau$ as functions from $X \times [0, \infty)$ into $[0, \infty)$ and verify (1) that both functions are Borel measurable, and also cad-lag increasing functions in the second variable. Now, consider the product measure $\mu = \nu(dx)dt$ on the product space $X \times [0, \infty)$ and apply the transformation $(x, t) \mapsto (x, \tau(t))$ to obtain the following set function

$$
\mu_r(A \times [0, s]) = \mu(\{(x, t) : x \in A, 0 < \tau(x, t) \leq s\}).
$$
Verify (2) that \( \mu_\tau \) extends to a unique measure defined on \( \mathcal{F} \times \mathcal{B} \), where \( \mathcal{B} \) is the Borel \( \sigma \)-algebra on \([0, \infty)\). Prove (3) that \( \mu_\tau = c(x,t)\nu(dx)dt \), i.e.,

\[
\mu_\tau(A \times [0, t]) = \int_{A \times [0, t]} c(x, r)\nu(dx)dr,
\]

for every measurable set \( A \) and any \( t \geq 0 \).

**Proof.** First remark that if \( c \) is strictly positive (e.g., \( c(x, r) + \varepsilon, \varepsilon > 0 \)) then \( s \mapsto \tau^{-1}(x, s) \) is the inverse of \( t \mapsto \tau(x, t) \), and the time-derivatives satisfy \( \tau'(x, t) = 1/c(x, \tau(x, t)) \). Also, the mapping \( c \mapsto \tau \) is decreasing, i.e., if \( c_1(x, r) \leq c_2(x, r) \) for every \( r \geq 0 \) then \( \tau_1(x, t) \geq \tau_2(x, t) \) for every \( t \geq 0 \). Certainly, if \( 0 \leq c_1 \leq 1 \) then \( \tau^{-1}(x, s) \leq s \) for every \( s \geq 0 \) and therefore \( \tau(x, t) \geq t \) for every \( t \geq 0 \).

It is clear that from Proposition 5.5 follows part (1), moreover, the mapping \( \vartheta : (x, t) \mapsto (x, \tau(x, t)) \) is also measurable and \( \{(x, t) : x \in A, 0 < \tau(x, t) \leq s\} = \vartheta^{-1}(A \times [0, s]) \). By linearity, the expression of \( \mu_\tau \) can be extended to any set of the form \( A \times ]a, b[ \) with \( A \) in \( \mathcal{F} \) and \( b > a \geq 0 \). Hence, part (2) is verified.

To prove (3), use equality (5.8) of Proposition 5.5 to obtain

\[
\int_{[0,\infty[} \mathbb{1}_{0 < \tau(x,t) \leq s}dt = \int_0^s d\tau^{-1}(x, r) = \int_0^s c(x, r)dr,
\]

and integrate on \( A \) to deduce

\[
\mu_\tau(A \times [0, s]) = \int_A \nu(dx) \int_{[0,\infty[} \mathbb{1}_{0 < \tau(x,t) \leq s}dt = \int_A \nu(dx) \int_0^s c(x, r)dr,
\]

which shows that \( \mu_\tau = c(x, t)\nu(dx)dt \).

Note that if \( s^*(x) = \sup\{s : \tau^{-1}(x, s) < \infty\} \) and the assumption that \( \tau^{-1}(x, s) \) is finite is dropped then \( t \mapsto \tau(x, t) \) maps \([0, \infty[\) into \([0, s^*(x)\] \), and \( \mu_\tau \) remains a measure on \( X \times [0, \infty) \) satisfying \( \mu_\tau = c(x, t)\nu(dx)dt \).

If \( \lim_{s \to \infty} \tau^{-1}(x, s) = t^*(x) < \infty \) then \( \tau(x, t) \) is finitely defined only for \( 0 \leq t < t^*(x) \), i.e., \( t \mapsto \tau(x, t) \) maps \([0, t^*(x)]\) into \([0, \infty[\). In this case, the definition of \( \mu_\tau \) can be modified as

\[
\mu_\tau(A \times [0, s]) = \mu(\{(x, t) : x \in A, t^*(x) < \infty, 0 < \tau(x, t) \leq s\}),
\]

so that \( \mu_\tau \) remains a measure on \( X \times [0, \infty) \) satisfying \( \mu_\tau = c(x, t)\nu(dx)dt \). \( \square \)

**5.3 Diadic Riemann Integrals**

**5.4 Lebesgue Measure on Manifolds**

**5.5 Hausdorff Measure**

**Exercise 5.11.** Consider the 1-dimensional Hausdorff measure \( h_1 \) on \( \mathbb{R}^1 \) and \( \mathbb{R}^2 \). Verify that \( h_{1,\delta}(A) \) is independent of \( \delta \) and is equal to the Lebesgue measure
\( \ell_1(A) \), for any \( A \subset \mathbb{R}^1 \). Now, show that if \( S_a = \{(x,a) : x \in [0,b]\} \subset \mathbb{R}^2 \) is an isometric copy of the interval \([0,b]\) then \( \ell_1(S_a) = b \), for every \( b > 0 \). Since the \( S_a \) are disjoint, we should deduce that \( \ell_1 \) is not a \( \sigma \)-finite measure in \( \mathbb{R}^2 \). Next, consider \( S_0 \) and \( S_a \) with \( a < \alpha < 1 \) and discuss the role of the parameter \( \delta \) in the above definition. Check that \( h_{1, \delta}(S_0 \cup S_a) = 2b \) if \( \delta < a \), but the diameter of \( S_0 \cup S_a \) is \( \sqrt{b^2 + a^2} \).

**Proof.** Recall that the Lebesgue outer measure \( \ell_1^* \) was define by

\[
\ell_1^*(A) = \inf \left\{ \sum_n \ell_1(I_n) : \bigcup_n I_n \supset A \right\}, \quad \forall A \subset \mathbb{R},
\]

where initially \( \{I_n\} \) is a sequence of intervals of the form \( I_n = ]a_n, b_n[ \). However, the infimum is unchanged if instead we use closed intervals. Similarly, the Hausdorff measure \( h_1 \) on \( \mathbb{R} \) is defined as \( \lim_{\delta \to 0} h_{1, \delta} \), where

\[
h_{1, \delta}(A) = \inf \left\{ \sum_n d(E_n) : \bigcup_n E_n \supset A, \ d(E_n) \leq \delta \right\}, \quad \forall \delta > 0, \ \forall A \subset \mathbb{R}.
\]

Since any subset \( E_n \) of \( \mathbb{R} \) can be cover by a closed interval \( I_n \) with boundaries \( a_n = \inf\{x : x \in E_n\} \) and \( b_n = \sup\{x : x \in E_n\} \) such that \( E_n \subset I_n \) and \( d(E_n) = d(I_n) \), we deduce that also

\[
h_{1, \delta}(A) = \inf \left\{ \sum_n d(I_n) : \bigcup_n I_n \supset A, \ d(I_n) \leq \delta \right\}, \quad \forall \delta > 0, \ \forall A \subset \mathbb{R},
\]

where now, only sequences \( \{I_n\} \) of closed intervals are allowed when taking the infimum. However, for any one-dimensional interval \( I \) we have \( d(I) = \ell_1(I) \), and thus \( \ell_1^* = h_{1, \delta} \) for every \( \delta > 0 \).

Now, we consider the Hausdorff measure \( h_1 \) on \( \mathbb{R}^2 \). Because a translation in \( \mathbb{R}^2 \) is an isometry and \( h_1 \) is invariant under any isometry, we obtain \( h_1(S_a) = h_1(S_0) = b > 0 \). Hence, \( \{S_a : a > 0\} \) is an uncountable family of disjoint sets with the same positive measure. Therefore, if \( \{A_n\} \) is a sequence of subsets of \( \mathbb{R} \) with \( h_1(A_n) < \infty \) then each \( A_n \) may contain at most a finite number of sets in the uncountable family \( \{S_a : a > 0\} \), which implies that \( \bigcup_n A_n \) may contain at most a countable subfamily of \( \{S_a : a > 0\} \). Thus, the Hausdorff measure \( h_1 \) on \( \mathbb{R}^2 \) is not \( \sigma \)-finite.

As mentioned early, the fact that parameter \( \delta \to 0 \) is essential to the construction. Indeed, because the distance from \( S_0 \) to \( S_a \) is equal to \( a \geq \delta > 0 \), any set \( E \) intersecting the set \( S_0 \) cannot intersect the set \( S_a \). Thus, for any sequence \( \{E_n\} \) of sets in \( \mathbb{R}^2 \) covering \( S_0 \cup S_a \), we obtain two sequences, one covering \( S_0 \) and another covering \( S_a \). This implies that \( h_{1, \delta}(S_0 \cup S_a) = 2b \) after projecting over the first coordinate. However, if \( \delta \geq a \) then we can cover \( S_0 \cup S_a \) with the rectangle \([0,b] \times [0,a] \) to find that \( h_{1, \delta}(S_0 \cup S_a) \leq \sqrt{a^2 + b^2} < 2b \), if \( b > a > 0 \). In any case, it is clear that the diameter of the set \( S_0 \cup S_a \) is \( \sqrt{a^2 + b^2} \), and as \( \delta \to 0 \) we obtain \( h_1(S_0 \cup S_a) = b + a = 2b \), for any \( a > 0 \). \( \square \)

**Exercise 5.12.** Regarding the Hausdorff dimension prove for any Borel sets (1) if \( A \subset B \) then \( \dim(A) \leq \dim(B) \) and (2) \( \dim(A \cup B) = \max\{\dim(A), \dim(b)\} \)
Moreover, prove that (3) for a sequence \( \{A_k\} \) of Borel sets we have \( \dim(\bigcup_k A_k) = \sup_k \{\dim(A_k)\} \).

**Proof.** If \( s > \dim(A) \) then \( h_s(A) = 0 \), which implies \( h_s(B) \leq h_s(A) = 0 \), i.e., \( s > \dim(B) \), proving that \( \dim(A) \leq \dim(B) \).

Similarly, if \( s > \max\{\dim(A), \dim(B)\} \) then \( s > \dim(A) \) and \( s > \dim(B) \), which implies \( h_s(A \cup B) \leq h_s(A) + h_s(B) = 0 \), i.e., \( \dim(A \cup B) < s \). Thus \( \dim(A \cup B) \leq \max\{\dim(A), \dim(B)\} \). From part (1) \( \dim(A \cup B) \geq \dim(A) \) and \( \dim(A \cup B) \geq \dim(B) \), i.e., \( \dim(A \cup B) = \max\{\dim(A), \dim(B)\} \).

This same argument applies to a countable union \( A = \bigcup_k A_k \), the only point is the \( \sigma \)-additivity \( h_s(\bigcup_k A_k) \leq \sum_k h_s(A_k) = 0 \).

**Exercise 5.13.** It is simple to establish that if \( T : \mathbb{R}^d \to \mathbb{R}^n \), with \( d \leq n \), is a linear map and \( T^* : \mathbb{R}^n \to \mathbb{R}^d \) is its adjoint then \( T^*T \) is a positive semi-definite linear operator on \( \mathbb{R}^n \), i.e., \( (T^*T x, x) \geq 0 \), for every \( x \in \mathbb{R}^d \) and with \((\cdot, \cdot)\) denoting the scalar product on the Euclidean space \( \mathbb{R}^d \). Now, use Theorem 2.27 and Proposition 5.14 to show that for every subset \( A \) of \( \mathbb{R}^d \) and any linear transformation \( T \) as above we have \( h_d(T(A)) = \sqrt{\det(T^*T)} h_d(A) \), where \( h_d \) is the \( d \)-dimensional Hausdorff measure considered either in \( \mathbb{R}^n \) or in \( \mathbb{R}^d \). See Folland [45, Proposition 11.21, pp. 352].

**Proof.** First, if \( d = n \) then \( \sqrt{\det(T^*T)} = \det(T) \) and Proposition 5.14 yields the desired equality. Next, if \( d < n \) then the range of \( T \) has at most dimension equal to \( k \), so we can find a rotation \( R : \mathbb{R}^n \to \mathbb{R}^n \) such that the range of \( T \) is mapped into a subspace of \( \mathbb{R}^d \times \{0\}^{n-d} \). Thus, if \( S = RT \) then \( S^*S = T^*R^*RT = T^*T \), which yields \( h_d(T(A)) = h_d(S(A)) \), after using the fact that \( h_d \) is rotational invariant. However, identifying the space \( \mathbb{R}^d \times \{0\}^{n-d} \) with \( \mathbb{R}^d \), the expression \( S^*S \) is unchanged and so \( h_d(S(A)) = \sqrt{\det(S^*S)} h_d(A) \), and the desired result follows.

**Exercise 5.14.** Let \( (X, d) \) be a metric space and \( d_1 \) be another metric equivalent to \( d \), i.e., such that \( a d \leq d_1 \leq b d \) for some positive constants \( a \) and \( b \). Prove that if \( h_s \) and \( h_s^1 \) denote the Hausdorff measures corresponding to \( d \) and \( d_1 \) then \( a^s h_s(A) \leq h_s^1(A) \leq b^s h_s(A) \), for every subset \( A \) of \( X \). Briefly discuss the particular cases where \( X = \mathbb{R}^n \) and either \( d_1(x, y) = \max_i |x_i - y_i| \) or \( d_1(x, y) = \sum_i |x_i - y_i| \). Can we easily compute \( h_s^1(B^1) \), where \( B^1 \) is the unit ball in the \( d_1 \) metric?

**Proof.** The inequality \( a^s h_s(A) \leq h_s^1(A) \leq b^s h_s(A) \), follows directly from the definition of the Hausdorff measures corresponding to \( d \) and \( d_1 \).

If \( d_1(x, y) = \max_i |x_i - y_i| \) the “balls” are actually cubes with edges parallel to the axis, i.e., \( \{y : d_1(y, x) < a\} = \prod_{i=1}^n (x_i - a, x_i + a) \). In this case, any ball (actually, cube) can be written as a finite number of non-overlapping smaller balls (actually, cubes). Moreover, for any subset \( E \) of \( \mathbb{R}^n \) we can find (as in the one-dimensional case) a closed ball (actually, a cube in this metric) \( I \) such that \( E \subset I \) and \( d_1(E) = d_1(I) \). Thus, \( h_s^1 = 2^{-n} \ell_s^n \), for \( s = n \). However, for \( s \neq n \) the situation is much harder, since \( h_s^1 \) is not necessarily invariant under rotations.
In any case, we can easily calculate the value of $h_s(B^n)$ as, $= 0$ if $s > n$, $= \infty$ if $s < n$, and $= 1$ if $s = n$.

The case $d_1(x, y) = \sum_i |x_i - x_i|$ is very similar, after using a rotation, essentially the same results are valid.

(5.6) Area and Coarea Formulae
(6.1) Signed Measures

Exercise 6.1. Regarding the above statements, first (a) prove that a series \( \sum_{i=1}^{\infty} a_i \) of real numbers converges absolutely if and only if the series \( \sum_{i=1}^{\infty} |a_{\iota(i)}| \) converges, for any bijective function \( \iota \) between the positive integers. Next, let \( \nu \) be a signed measure on \( (\Omega, \mathcal{F}) \) and \( \{F_k\} \) be a sequence of disjoint sets in \( \mathcal{F} \) such that \( |\nu(\bigcup_k F_k)| < \infty \). Prove that (b) the series \( \sum_k \nu(F_k) \) is absolutely convergence.

Proof. (a) First, let \( \sum_{i=1}^{\infty} a_i \) be series real numbers such that \( \sum_{i=1}^{\infty} |a_i| < \infty \). We need to show that, for any bijective function \( \iota \) between the positive integers, the series \( \sum_{i=1}^{\infty} a_{\iota(i)} \) converges, i.e., for every \( \varepsilon > 0 \) there exists an index \( N \) such that \(|\sum_{i>N} a_{\iota(i)}| < \varepsilon\). Indeed, by assumption there is an index \( M \) such that \( \sum_{i>M} |a_i| < \varepsilon \), so that if \( i > N = \max \{\iota(i) : 1 \leq i < M\} \) then \( i > M \) proving that \( |\sum_{i>N} a_{\iota(i)}| \leq \sum_{i>M} |a_i| < \varepsilon \).

Conversely, let \( \sum_{i=1}^{\infty} a_i \) be series real numbers such that \( \sum_{i=1}^{\infty} a_{\iota(i)} \) converges, for any bijective function \( \iota \) between the positive integers. Since the series \( \sum_{i=1}^{\infty} a_i \) converges, if the series \( \sum_{i=1}^{\infty} |a_i| = \infty \) then the two sub-series of the positive and the negative parts are divergent, i.e., \( \sum_{i=1}^{\infty} (a_i)^+ = \infty \) and \( \sum_{i=1}^{\infty} (a_i)^- = \infty \). Thus, for instance, adding sufficiently many positive terms we have \( \sum_{i=1}^{k_1} (a_i)^+ > 1 \) and then adding sufficiently many negative terms we have \( \sum_{i=1}^{k_1} (a_i)^+ + \sum_{i=1}^{h_2} (a_i)^- < 0 \), so that we can construct a bijective function \( \iota \) between the positive integers satisfying \( \sum_{i=1}^{k_j} a_{\iota(i)} > 1 \) and \( \sum_{i=1}^{k_j+h_j} a_{\iota(i)} < 0 \), for every \( j \geq 1 \). Therefore, this particular ”order of summation” produces a non convergent series \( \sum_{i=1}^{\infty} a_{\iota(i)} \), i.e., a contradiction.

(b) If \( \nu \) is a signed measure on \( (\Omega, \mathcal{F}) \) and \( \{F_k\} \) is a disjoint sequence of sets in \( \mathcal{F} \) such that \( |\nu(\bigcup_k F_k)| < \infty \) then the \( \sigma \)-additivity of \( \nu \) implies that \( \nu(F) = \sum_k \nu(F_k) \), with \( F = \bigcup_k F_k \) and \( |\nu(F)| < \infty \). Since for any bijective function \( \iota \) between the positive integers, we also have \( F = \bigcup_k F_{\iota(k)} \) which yields
\[ \nu(F) = \sum_{k} \nu(F_{i(k)}) \]. In view of part (a), the series \( \sum_{k} \nu(F_{k}) \) is absolutely convergent. \( \square \)

**Exercise 6.2.** Prove that (a) \( \nu \ll \mu \) if and only if (b) \( |\nu| \ll |\mu| \) if and only if (c) \( \nu^+ \ll \mu \) and \( \nu^- \ll \mu \).

**Proof.** From the (abstractly continuous measures) Definition 6.2 \( \nu \ll \mu \) if and only if \( |\nu| \ll \mu \), and according to Proposition 6.1, there exists a measurable positive set \( A \) such that \( \mu(F \cap A) \geq 0 \geq \mu(F \cap A^c) \) for every measurable set \( F \). Also, \( |\nu|(F) = 0 \) if and only if \( \nu^+(F) = \nu^-(F) = 0 \). Hence, if \( |\nu|(F) = 0 \) then \( |\nu|(F \cap A) = |\nu|(F \cap A^c) = 0 \), which implies that \( \mu^+(F) = \mu^-(F) = 0 \), i.e., \( |\mu|(F) = 0 \), proving that (a) implies (b).

Since \( |\nu|(F) = 0 \) if and only if \( \nu^+(F) = \nu^-(F) = 0 \) (and similarly for \( \mu \)), we deduce that (b) implies (c) and (c) implies (a), closing the circle. \( \square \)

**Exercise 6.3.** Give more details on the how to reduce the proof of Theorem 6.5 to the case where \( \mu \) and \( \nu \) are finite measures.

**Proof.** As in Theorem 6.3, since \( \nu \) is \( \sigma \)-finite, the whole space \( \Omega \) can be written as a disjoint countable union \( \bigcup_{n} \Omega^\nu_n \) with \( |\nu(\Omega^\nu_n)| < \infty \). Next, because \( \mu \) is also \( \sigma \)-finite, each \( \Omega^\mu_n \) can be written as a disjoint countable union \( \bigcup_{k} \Omega^\mu_{n,k} \) with \( |\mu(\Omega^\mu_{n,k})| < \infty \). Hence, relabeling the double sequence, we have \( \Omega = \bigcup_{n} \Omega_n \), with \( \Omega_n \cap \Omega_m = \emptyset \) if \( n \neq m \) and \( |\nu_n(\Omega_n)| + |\mu(\Omega_n)| < \infty \), for every \( n \).

Define \( \nu^n(A) = \nu(A \cap \Omega_n) \) and \( \mu^n(A) = \mu(A \cap \Omega_n) \), for each measurable set \( A \) and each \( n \), which are finite signed measures on \( \Omega_n \), supported only on \( \Omega_n \). In view of Theorem 6.5, for each \( n \), there exist two signed measures \( \nu^n_A \) and \( \nu^n_s \), uniquely determinate on \( \Omega_n \), such that \( \nu^n = \nu^n_A + \nu^n_s \), \( \nu^n_A \ll \mu^n \) and \( \nu^n_s \perp \mu^n \).

Finally, if \( \nu(A) = \sum_n \nu^n_A(A \cap \Omega_n) \) and \( \nu_s(A) = \sum_n \nu^n_s(A \cap \Omega_n) \), for any measurable set \( A \), then \( \nu = \nu_A + \nu_s \), \( \nu_A \ll \mu \) and \( \nu_s \perp \mu \). \( \square \)

**Exercise 6.4.** With the above notation, fill in details for the previous assertions on the essential supremum and infimum. Compare with Exercise 1.23.

**Proof.** Recall, given a family \( \{f_i : i \in I\} \) of real-valued (or extended real-valued) measurable functions defined on a \( \sigma \)-finite measure space \((X, \mathcal{X}, \mu)\), the essential supremum (or infimum) is an extended real-valued measurable function \( f^* \) (or \( f_* \)) such that for every \( i \) there exists a null set \( N_i \) satisfying \( f_i(x) \leq f^*(x) \) (or \( f_i(x) \geq f_*(x) \)) for every \( x \) in \( X \setminus N_i \), and \( f^* \) (or \( f_* \)) is the smallest (or largest) with the above property. If the family is countable then the pointwise supremum (or infimum) serves as the essential supremum (or infimum). Therefore, only the case when the index set \( I \) is uncountable requires further discussion. Moreover, replacing \( f_i \) with \(-f_i \), only the case of the essential supremum should be analyzed.

Thus, retaking the argument for the existence of the essential supremum, and assuming \( 0 \leq f_i(x) \leq 1 \), for every \( i \) in \( I \) and any \( x \) in \( X \), define

\[ \nu(A) = \sup \left\{ \sum_{k=1}^{n} \int_{A_k} f_{ik}(x) \mu(dx) \right\}, \]
where the supremum is taken over all finite measurable partitions \( A = \sum_{k=1}^{n} A_k \) and any choice of indexes \( i_k \) in \( I \). Note that since \( A_1, \ldots, A_n \) are measurable, it does not matter whether the choices of the indexes \( i_k \) are repeated or not.

To check that \( \nu \) is an additive set function, consider two finite partitions \( A = \sum_{k=1}^{n} A_k \) and \( B = \sum_{h=1}^{m} B_k \) of two disjoint sets \( A \) and \( B \), and choose indexes \( i_1, \ldots, i_n \) and \( j_1, \ldots, j_m \) belonging to \( I \). Because \( A \cap B = \emptyset \), the sets \( A_1, \ldots, A_n, B_1, \ldots, B_m \) are disjoint, and moreover, the choices of indexes \( i_1, \ldots, i_n, j_1, \ldots, j_m \) could be repeated, in view of the additivity of \( \mu \). Hence \( \mu(A + B) = \nu(A) + \nu(B) \).

Next, to \( \sigma \)-additivity follows from the monotone continuity, i.e., if \( \{A_k\} \) is an increasing sequence of measurable sets then \( \lim_{n} \bigcup_{k \geq n} A_k = \emptyset \) and

\[
\lim_{n} \nu\left( \bigcup_{k \geq n} A_k \right) \leq \lim_{n} \mu\left( \bigcup_{k \geq n} A_k \right) = 0.
\]

Therefore, because \( \nu \) is absolutely continuous with respect to \( \mu \), Radon-Nikodym Theorem 6.3 shows that \( f^* = d\nu/d\mu \) satisfies

\[
\int_A f_i d\mu \leq \nu(A) = \int_A f^* d\mu, \quad \forall i \in I, \forall A \in \mathcal{X},
\]

and in fact that \( f^* \) is the essential supremum of the (possible uncountable) family \( \{f_i : i \in I\} \).

Now, if the family is stable under the pairwise maximization (i.e., if \( i \) and \( j \) belong to \( I \) then there exists \( k \) in \( I \) such that \( \max\{f_i(x), f_j(x)\} = f_k(x) \), for almost every \( x \)) then \( \max\{f_{i_1}(x), \ldots, f_{i_k}(x)\} = f_i(x) \) for some index \( i \), which implies that

\[
\nu(A) = \sup \left\{ \int_A f_i(x) \mu(dx) \right\},
\]

where the supremum is taken over indexes \( i \) in \( I \). If \( \{i_n\} \) is a maximizing sequence for \( \mu(\Omega) \) we have

\[
\mu(\Omega) = \lim_{n} \int_{\Omega} f_{i_n}(x) \mu(dx),
\]

which proves that \( \{f_{i_n}\} \) is an almost everywhere increasing sequence satisfying \( f_{i_n}(x) \to f^*(x) \) almost everywhere \( x \).

Comparing with Exercise 1.23 on the pointwise supremum, we can state that if \( \{f_t : t \in T\} \) is a family of measurable function in a \( \sigma \)-finite measure space then constructing the family of measurable functions \( f_i(x) = \max\{f_{t_1}(x), \ldots, f_{t_n}(x)\} \) indexed by \( I = \{i = (t_1, \ldots, t_n) : t_i \in T, n \geq 1\} \), we can define the essential supremum \( f^* = \text{ess-sup}\{f_i : i \in I\} \), which is the almost pointwise supremum in the sense that for every \( t \) in \( T \) there exists a negligible set \( N \) such that \( f_t(x) \leq f^*(x) \) for every \( x \) outside of \( N \), and that there exist an almost everywhere increasing sequence \( \{f_{i_n}\} \) satisfying \( f_{i_n}(x) \to f^*(x) \) almost everywhere \( x \).  \( \Box \)
Exercise 6.5. State the Radon-Nikodym Theorem 6.3 and the Lebesgue decomposition Theorem 6.5 for the case of complex-valued measures (and give details of the proof, if necessary).

Proof. By definition, complex-valued measures have finite values, i.e., \( \mu = \mu_1 + i\mu_2 \), where \( \mu_1 \) and \( \mu_2 \) are two (real valued) finite signed measures. Moreover, Definition 6.2 on absolutely continuous (\( \ll \)) and singular (\( \perp \)) measures can be clearly applied to complex-valued measures, and if \( \mu = \mu_1 + i\mu_2 \) and \( \nu = \nu_1 + i\nu_2 \) then \( \mu \ll \nu \) if and only if \( \mu_k \ll \nu_k \), for \( k = 1, 2 \) (similarly for \( \perp \) instead of \( \ll \)). Furthermore, Hahn-Jordan decomposition Proposition 6.1 can be applied to the real part and the imaginary part, and the variations \( |\mu_k| = |\mu_k^+ + \mu_k^-| \) can be defined. If the variation \( |\mu| \) defined by means of the modulus of a complex number needs some further consideration, and usually, \( |\mu| = |\mu_1| + |\mu_1| \) can be used instead. As in Exercise 6.2, for complex-valued measures, we can check that \( \mu \ll \nu \) (or \( \mu \perp \nu \)) if and only if \( |\mu| \ll |\nu| \) (or \( |\mu| \perp |\nu| \)).

Radon-Nikodym Theorem 6.3 for complex-valued measures can be stated as follows: If \( \mu \) and \( \nu \) are two complex-valued measures on \((\Omega, \mathcal{F})\) such that \( \nu \ll \mu \) then there exists a complex-valued integrable function \( f \) satisfying

\[
\nu(F) = \int_F f \, d\mu, \quad \forall F \in \mathcal{F},
\]

and the function \( f \) is uniquely defined, except in a set of \( \mu \)-measure zero.

Lebesgue decomposition Theorem 6.5 for complex-valued measures can be stated as follows: If \( \mu \) and \( \nu \) are two complex-valued measures on \((\Omega, \mathcal{F})\) then there exist two uniquely determinate complex-valued measures \( \nu_a \) and \( \nu_s \) such that \( \nu = \nu_a + \nu_s \), \( \nu_a \ll \mu \) and \( \nu_s \perp \mu \).

\( 6.2 \) Essential Supremum

Exercise 6.6. Based on Theorem 6.7, discuss and compare the statements in Exercises 1.22, 1.23, 4.23, 6.4 and 7.9. Consider also Remark 7.14.

Proof. Regarding Exercise 1.22, it should be clear by now that taken infimum or supremum on uncountable families of a just measurable functions is not so good. A complicate alternative is used in the general theory of processes, this is the notion of separability discussed in Exercise 1.23. Indeed, it can be proved (e.g., Doob [34, Theorem 2.4, pp. 60], Billingsley [15, Section 7.38, pp. 551-563], Gikhman and Skorokhod [52, Section IV.2]), or Neveu [87, Proposition III.4.3, pp. 84-85]) that any family of measurable functions \( \{f_t : t \in T\} \) has a version (i.e., there is another family of measurable functions \( \{g_t : t \in T\} \) such that \( f_t = g_t \) a.e., for every \( t \) in \( T \)), which is separable (i.e., in short, there is a countable dense subset of indexes which can be used -instead of \( T \)- to compute the sup or inf on the family \( \{g_t : t \in T\} \)). Clearly, this result can be easily extended from a probability spaces to \( \sigma \)-finite spaces.

In Exercise 4.23 the particular case of local supremum (or infimum) \( \overline{f}(x) = \sup \{f(y) : |y - x| < r\} \) is discussed, but it uses the definition of the function
Theorem 6.7 proposes a good solution, but it is not exactly the pointwise infimum or supremum, something is lost in the process. For instance, if \( f \) is a measurable function and \( F_r \) is the multi-valued function \( x \mapsto \{ f(y) : |y-x| \leq r \} \), and \( G_r \) is the family of all measurable sections of \( F_r \), i.e., all measurable functions \( g \) satisfying \( g(x) \in F_r(x) \) then the infimum of the family \( G_r \) is another alternative to the pointwise expression \( \overline{f} \) discussed early, which is stable if \( f \) is modified in a null set.

Suppose that the interest is on the pointwise essential supremum \( f^*(x) = \text{ess-sup}_{y \in Y} f(x, y) \), where \( f(x, y) \) is a measurable function on \( X \times Y \). Because the function \( f^* \) is not necessarily measurable, consider the family \( G \) indexed by the null sets \( N \) in \( X \times Y \), of all measurable functions \( g(x) \leq \text{sup}_{y \in Y, (x,y) \notin N} f(x, y) \), for any \( x \) in \( X \). In this case, if \( \overline{f} = \text{ess-sup}\{G\} \) then, because almost every sections of a null set is null, the function \( \overline{f} \) is indeed the larger measurable function not greater than \( f^* \) almost everywhere. Moreover, if \( f^* \) is measurable then \( \overline{f} = f^* \) a.e., and if \( h = f \) a.e in \( X \times Y \) then \( \overline{h} = \overline{f} \) a.e. \( \Box \)

**Exercise 6.7.** Let \( H \) be a Hilbert space with inner product \((\cdot, \cdot)\). First use Zorn’s Lemma to show that any orthonormal set can be extended to an orthonormal basis \( \{e_i : i \in I\} \), i.e., a maximal set of orthogonal vectors with unit length. Now, prove (1) that any element \( x \) in \( H \) can be written uniquely as \( x = \sum_{i \in I} (x, e_i) e_i \), where only a countable number of \( i \) have \((x, e_i) \neq 0\). Next, (2) verify that the cardinal of \( I \) is invariant for any orthogonal basis. Finally, prove (3) that \( H \) is isomorphic to the Hilbert space \( \ell^2(I) \) of all functions \( c : I \to \mathbb{R} \) (or complex valued) such that \( \sum_{i \in I} |c_i|^2 < \infty \) (i.e., only a countable number of \( c_i \) are nonzero and the series is convergent).

**Proof.** Recall that an orthonormal set \( A \) in \( H \) is a nonempty subset of elements in \( H \) satisfying (i) \((a, a) = 1\) for every \( a \) in \( A \) and (ii) \((a, a') = 0\) for every \( a \neq a' \) in \( A \). Therefore, as in the existence of a Hamel Basis, Lemma 0.5, for a given orthonormal set \( A \), consider the partially ordered set \( \mathcal{S} \) whose elements are all the orthonormal subsets of \( H \), with the partial order given by the inclusion. If \( \{A_\alpha\} \) is a chain or totally ordered subset of \( \mathcal{S} \) then \( A = \bigcup_\alpha A_\alpha \) is also an orthonormal set, i.e., an upper bound. Hence, Zorn’s Lemma implies the existence of a maximal element, denoted by \( \{x_i : i \in I\} \), Now, \( A \subset \{x_i : i \in I\} \) and \( \{x_i : i \in I\} \) is a orthonormal set. Moreover, if \( x \) is a nonzero orthogonal element to \( \{x_i : i \in I\} \), i.e., \((x, x_i) = 0\), for every \( i \in I \) then \( \{x_i : i \in I\} \cup \{x/(x, x)\} \) would be an orthonormal set strictly large to \( \{x_i : i \in I\} \), which contradict the maximal character, proving that \( \{x_i : i \in I\} \) is indeed a basis in \( H \).

(1) Given a basis \( \{x_i : i \in I\} \), take \( x \) in \( H \) and consider an finite subset \( J \subset I \)
of subindexes. The linearity of the inner product and the expression of norm \( \|y\|^2 = (y, y) \) yields the equality \( \|x - \sum_{i \in J} (x, x_i)x_i\|^2 = \|x\|^2 - \sum_{i \in J}|(x, x_i)|^2 \). This proves the so-called Bessel's inequality, namely, \( \sum_{i \in I}|(x, x_i)|^2 \leq \|x\|^2 \), proving that for every \( x \) in \( H \) there can be only a countable number of indexes such that \( (x, x_i) \neq 0 \). Moreover, given \( \varepsilon > 0 \) there exists a finite set \( J_\varepsilon \subset I \) such that \( \sum_{i \in J_\varepsilon}|(x, x_i)|^2 < \varepsilon^2 \), which implies that \( \|\sum_{i \in J_\varepsilon}(x, x_i)x_i\| < \varepsilon \). Since \( H \) is a complete space, the series \( \sum_{i \in I}(x, x_i)x_i \) convergence to some element \( y \) in \( H \) such that \( x - y \) is orthogonal to any vector in the orthonormal basis \( \{x_i : i \in I\} \), which yields \( x - y = 0 \), i.e.,

\[
x = \sum_{i \in I}(x, x_i)x_i \quad \text{and} \quad \sum_{i \in I}|(x, x_i)|^2 \leq \|x\|^2, \quad \forall x \in H,
\]
as desired. The last identity is referred to as Parseval's equality.

(2) If \( \{x_i : i \in I\} \) and \( \{y_j : j \in J\} \) are two orthonormal bases in \( H \) then for any subset \( I_0 \subset I \) there exists a unique subset \( J_0 \subset J \) defined by the condition \( j \in J_0 \) if and only if \( (x_j, x_i) \neq 0 \) for some \( i \) in \( I_0 \). This proves that the cardinal of \( 2^I \) is not larger that the cardinal of \( 2^J \), which means (after reversing roles between \( I \) and \( J \)) that both indexes have the same cardinal.

(3) Parseval’s equality shows that \( H \) is isomorphic to the Hilbert space \( \ell^2(I) \), reinforcing the fact that any two orthonormal bases must have the same cardinal. \( \square \)

**Exercise 6.8.** If \( (X, \mathcal{X}, \mu) \) is a measure space and \( E \) belongs to \( \mathcal{X} \) then we identify \( L^2(E, \mu) \) with the subspace of \( L^2(X, \mu) \) consisting of functions vanishing outside \( E \), i.e., an element \( f \) in \( L^2(X, \mu) \) is in \( L^2(E, \mu) \) if and only if \( f = 0 \) a.e. on \( E^c \). Let \( \{X_i\} \) be a sequence in \( \mathcal{X} \) such that \( X = \bigcup_i X_i \), and \( \mu(X_i \cap X_j) = 0 \) whenever \( i \neq j \). Prove that (a) \( \{L^2(X_i, \mu)\} \) is a sequence of mutually orthogonal subspaces of \( L^2(X, \mu) \) and (b) every \( f \) in \( L^2(X, \mu) \) can be written uniquely as \( f = \sum_{i=1}^\infty f_i \), where \( f_i \) belongs to \( L^2(X_i, \mu) \) and the series converges in norm. Moreover, show that (c) if \( L^2(X_i, \mu) \) is separable for every \( i \) then so is \( L^2(X, \mu) \).

**Proof.** (a) It is clear that if \( i \neq j \) then \( f_i f_j = 0 \) a.e., for every \( f_i \) in \( L^2(X_i, \mu) \), i.e., the subspaces are mutually orthogonal.

(b) If \( f \) belongs to \( L^2(X, \mu) \) then \( f = \sum_i f_i \) a.e., with \( f_i = f 1_{X_i} \). Each \( f_i \) is in \( L^2(X_i, \mu) \) and because the sequence of spaces \( \{L^2(X_i, \mu)\} \) is mutually orthogonal, the series also converge in \( L^2(X, \mu) \).

(c) If \( \{g_{i,j} : j \geq 1\} \) is a countable dense set in \( L^2(X_i, \mu) \) then, in view of (a) and (b), the countable set \( \{g_{i,j} : i, j \geq 1\} \) results a dense set in \( L^2(X, \mu) \). \( \square \)


**6.4 Uniform Integrability**

**Main Properties**

**Mean Convergence**

**Convergence in Norm**

**Exercise 6.9.** Consider the Lebesgue measure on the interval \((0, \infty)\) and define the functions \(f_i = \frac{1}{i} 1_{(i,2i)}\) and \(g_i = 2^i 1_{(2^{-i-1},2^{-i})}\) for \(i \geq 1\). Prove that (a) the sequence \(\{f_i : i \geq 1\}\) is uniformly integrable of any order \(p > 1\), but not of order \(0 < p \leq 1\). On the contrary, show that (b) the sequence \(\{g_i : i \geq 1\}\) is uniformly integrable of any order \(0 < p < 1\), but the sequence is not equi-integrable of any order \(p \geq 1\).

**Proof.** Since \(0 \leq f_i(x) \leq f(x) = \min\{1, 2/x\}\) and \(f\) belongs to \(L^p([1, \infty])\) for every \(p > 1\), the sequence \(\{f_i : i \geq 1\}\) is uniformly integrable of any order \(p > 1\). Clearly, for \(0 < p \leq 1\), the difficulty is the tightness condition. If \(A\) is a subset of \((1, \infty)\) with finite Lebesgue measure then the Lebesgue dominate convergence implies

\[
\lim_i \int_A |f_i(x)|^p dx = 0, \quad \forall p > 0, 
\]

and because

\[
\int_{(0, \infty)} |f_i(x)|^p dx = i^{1-p} \geq 1^{1-p} > 0, \quad \forall i \geq 1, \quad 0 < p \leq 1, 
\]

we deduce that for every \(\varepsilon > 0\) and any set \(A \subset (1, \infty)\) with finite Lebesgue measure there exists an index \(i \geq 1\) such that

\[
\int_A |f_i(x)|^p dx > \varepsilon,
\]

i.e., the sequence \(\{f_i : i \geq 1\}\) is neither uniformly integrable nor equi-integrable of order \(0 < p \leq 1\). Note that \(f_i(x) \to 0\) as \(i \to \infty\) for every \(x\) in \((0, \infty)\).

If \(0 < p < 1\) then

\[
\int_{(0, \infty)} |g_i|^p dx = 2^i (p-1) \to 0 \quad \text{as} \quad i \to \infty,
\]

which show that the sequence \(\{g_i : i \geq 1\}\) is uniformly integrable of any order \(0 < p < 1\). However, if \(p \geq 1\) then open interval \(I_i = (2^{-i-1}, 2^{-i})\) satisfies

\[
\int_{I_i} |f_i(x)|^p dx = 2^i (p-1) \geq 1 \quad \forall i \geq 1
\]

but the Lebesgue measure of \(I_i\) vanish as \(i \to \infty\), which proves that the sequence \(\{g_i : i \geq 1\}\) is not equi-integrable integrable of order \(p \geq 1\). Note that \(g_i(x) = 0\) for every \(x \geq 1\) and \(g_i(x) \to 0\) as \(i \to \infty\) for every \(x\) in \((0, \infty)\).

(6.5) Representation Theorems
(7.1) Differentiation and Approximation

Approximation by Smooth Functions

Exercise 7.1. Let $Q$ be the class of dyadic cubes in $\mathbb{R}^d$, i.e., for $d = 1$ we have the dyadic intervals $[k2^{-n}, (k+1)2^{-n}]$, with $n = \pm 1, \pm 2, \ldots$ and any integer $k$. Consider the set $D$ of all finite linear combinations of characteristic functions of cubes in $Q$ and rational coefficients. Verify that (1) $D$ is a countable set and prove (2) that $D$ is dense in $L^1$, i.e., for any integrable function $f$ and any $\varepsilon > 0$ there is an element $\varphi$ in $D$ such that $\|f - \varphi\|_1 < \varepsilon$. Moreover, by means of Weierstrauss approximation Theorem 0.3, (3) make an alternative argument to show that there is a countable family of truncated (i.e., multiplied by $1_{\{|x|<r\}}$) polynomial functions which is dense in $L^1$, proving (again) that the space $L^1$ is separable.

Proof. A way to approximate integrable functions goes as follows:

(a) Any integrable function is a limit in the $L^1$-norm $\| \cdot \|_1$ of a sequence of integrable simple functions (i.e., integrable functions assuming only a finite number of values), this implies that for any integrable function $f$ and any $\varepsilon > 0$ there exists a simple function $f_\varepsilon$ such that $\|f - f_\varepsilon\|_1 < \varepsilon$, which reduces the problem to the approximation of a characteristic function $f = 1_E$ for a measurable set $E$ with finite measure.

(b) Next, for each measurable set $E$ with finite measure and any $\varepsilon > 0$ there exists an open set $U \supset E$ and a closed set $F \subset E$ such that the measure of $U \setminus F$ is less than $\varepsilon$. This implies that there exist a finite unions of non-overlapping cubes $C$ and a continuous function $g$ such that $\|1_C - 1_U\|_1 < \varepsilon$ and $\|1_F - g\|_1 < \varepsilon$, as discussed in the proof of Proposition 7.2.

(c) Now, the problem of approximation is reduces to the case of either a integrable bounded continuous function or a characteristic function of bounded cube of finite measure. In any of these two type of function can be approximated (in the $L^1$-norm $\|\cdot\|_1$) with a finite linear combinations of characteristic functions of cubes in $Q$ with rational coefficients.
Going back to the questions, (1) $D$ is a countable set because it can be enumerated with the finite parts of the pair of integer numbers $(n,k)$ and the finite parts of the rational numbers, and (2) follows from the above (a)–(c). This prove that the space $L^1$ is separable.

Alternatively, by means of Weierstrauss approximation Theorem 0.3 we deduce that truncated polynomials approximate (uniformly and therefore, in the $L^1$-norm $\| \cdot \|_1$) any continuous with compact support. Hence, any continuous with compact support can be approximated in the $L^1$-norm $\| \cdot \|_1$ by a polynomial with rational coefficients multiplied by $1\{ |x| < n \}$ with $n$ a natural number. Therefore, because (i) this family of truncated polynomials is countable (enumerated by the finite parts of rational numbers and the natural numbers) and (ii) any integrable function can be approximate by a continuous function with compact support, we proved (again) that the space $L^1$ is separable.

Exercise 7.2. Consider the various cases of Proposition 7.5 and verify the claims by doing more details, e.g., consider the case when $g$ is only locally integrable. What if the $\partial_i f$ exists everywhere, $f$ and $\partial_i f$ are locally integrable function, and $g$ is essentially bounded with compact support?

Proof. If $g$ is only locally integrable then $f$ is locally essentially bounded, and either $f$ or $g$ has a compact support. Considering the bound

$$|(f * g)(x + a) - (f * g)(x)| \leq \int_{\mathbb{R}^d} |g(x + a - y) - g(x - y)||f(y)| \, dy,$$

where the integral is actually limited to either the support of $f$ or the support of $y \mapsto g(z - y)$, for $z$ within a ball centered at $x$ with radius $|a| \leq 1$. Thus, in view of the continuity of the translation in $L^1$ (i.e., Proposition 7.3), we deduce that $\lim_{|a| \to 0} (f * g)(x + a) = (f * g)(x)$.

If both $f$ and $g$ are locally integrable and either $f$ or $g$ has a compact support then $f * g$ belongs to $L^1$ (i.e., a priori, only defined almost everywhere). If the partial derivative $\partial_i f$ exists everywhere and is a locally bounded function then the Mean Value Theorem implies that

$$\left| f(x + he_i - y) - f(x - y) \right| |g(y)| \leq \| \partial_i f \|_{\infty} 1_K(y) |g(y)|,$$

for every $0 < |h| \leq 1$, where $e_i$ is the unit vector in the direction $i$ and $K$ is some suitable compact set in $\mathbb{R}^d$. Hence, the dominate convergence applied to

$$\left| \frac{(f * g)(x + he_i) - (f * g)(x)}{h} \right| \leq \int_{\mathbb{R}^d} \left| \frac{f(x + he_i - y) - f(x - y)}{h} \right| |g(y)| \, dy$$

yields $\partial_i (f * g) = (\partial_i f) * g$.

Note that if the partial derivative $\partial_i f$ exits almost everywhere and is a essentially locally bounded function then the Mean Value Theorem cannot be used, but we may assume independently that the quotient $|f(x + he_i - y) - f(x - y)|/h$ is essentially bounded, to be able to use the dominate convergence and reach the desired conclusion.
Actually, even if partial derivative $\partial_i f$ is only locally integrable, we may deduce the equality $\partial_i(f \ast g) = (\partial_i f) \ast g$ as soon as passing to limit inside the integral is allowed. For instance, if $f(y)$ is a convex (or concave) function in the $i$ coordinate for $y_i$ within a small interval $[x_i - \delta, x_i + \delta]$ then $|f(x + h\epsilon_i - y) - f(x - y)|/|h|$ can be bounded by $|\partial_i f(x - y)| + |\partial_i f(x +\epsilon_i - y)|$ for $0 < |h| < \delta$. \qed

**Exercise 7.3.** Let $f$ and $g$ be two nonnegative Lebesgue locally integrable functions in $\mathbb{R}^d$. Prove (a) if $\liminf_{|x| \to \infty} f(x)/g(x) > 0$ and $g$ is not integrable then $f$ is also non integrable, in particular, $\liminf_{|x| \to \infty} f(x)|x|^d = 0$, for any integrable function $f$, (b) if $\limsup_{|x| \to \infty} f(x)/g(x) < \infty$ and $g$ is integrable then $f$ is also integrable. Finally, show that (c) if $f$ is integrable and uniformly continuous in $\mathbb{R}^d$ then $\limsup_{|x| \to \infty} |f(x)| = 0$. Given an example of a nonnegative integrable and continuous function on $[0, \infty)$ such that $\limsup_{x \to \infty} f(x) = +\infty$.

*Proof.* (a) It is clear that if $\liminf_{|x| \to \infty} f(x)/g(x) > 0$ then there exists $r > 0$ such that $f(x)/g(x) > 1/r$ whenever $|x| > r$. Hence, $f(x) \geq g(x)/r$ if $|x| > r$, which implies that $f$ cannot be integrable on $\mathbb{R}^d$. Actually, it suffices that the essential inferior limit has this property, i.e., $\text{ess-liminf}_{|x| \to \infty} f(x)/g(x) > 0$ or equivalently, there exist a null set $N$ and $r > 0$ such that $f(x)/g(x) > 1/r$ for every $x$ outside of $N$ with $|x| > r$.

Moreover, since the function $x \mapsto |x|^{-d}$ is not integrable in $\mathbb{R}^d$, we deduce that $\liminf_{|x| \to \infty} f(x)|x|^d = 0$, for any integrable function $f$, in particular $\liminf_{|x| \to \infty} f(x) = 0$.

(b) On the other hand, if $\limsup_{|x| \to \infty} f(x)/g(x) < \infty$ then there exists $r > 0$ such that $f(x)/g(x) < r$ whenever $|x| > r$. Hence, $f(x) \leq rg(x)$ if $|x| > r$, which implies that $f$ is integrable on $\mathbb{R}^d$. As in part (a), it suffices that the essential superior limit has this property, i.e., $\text{ess-limsup}_{|x| \to \infty} f(x)/g(x) < \infty$ or equivalently, there exist a null set $N$ and $r > 0$ such that $f(x)/g(x) < r$ for every $x$ outside of $N$ with $|x| > r$.

(c) Finally, if $f$ is an integrable and uniformly continuous function in $\mathbb{R}^d$, and $\limsup_{|x| \to \infty} |f(x)| > 0$, then there exist $\varepsilon > 0$ and a sequence $\{x_k\}$ satisfying $|x_k| \to \infty$ and $|f(x_k)| > 2\varepsilon$. In view of the uniform continuity on $\mathbb{R}^d$, there exist $\delta > 0$ such that $|f(x)| > \varepsilon$ whenever $|x - x_k| < \delta$. Defining the function $h(x) = \varepsilon$ if $|x - x_k| < \delta$ and $h(x) = 0$ otherwise, we deduce $|f| \geq h$ and $h$ is non integrable on $\mathbb{R}^d$, i.e., $f$ is cannot be integrable.

In other words, if the limit of $f(x)$ exits as $|x| \to \infty$ then it must be zero, otherwise $f$ is not integrable in $\mathbb{R}^d$. Moreover, if $f$ is integrable and uniformly continuous in $\mathbb{R}^d$ then, as $|x| \to \infty$, the limit of $f(x)$ exits.

Consider a triangle with base and height of sizes $2b \leq 1 < h$ with area $bh$, and set $g(x, b, h) = (1 - |x|/b|h)$ for $|x - b| \leq 1$ and $g(x) = 0$ otherwise, so that $x \mapsto g(x, b, h)$ is a continuous function with integral $bh$ and satisfying $g(0, b, h) = h$. Define $f(x) = \sum_{k \geq 1} g(x - k, 2^{-k}, k)$ to check that

$$\int_{[0, \infty)} f(x) \, dx = \sum_{k \geq 1} 2^{-k} k < \infty$$

and $f(k) = k$, which implies that $\limsup_{x \to \infty} f(x) = +\infty$. \qed
Exercise 7.4. Prove a local version of Corollary 7.6, i.e., if the kernel \( k \) has a compact support then \( f \ast k_\varepsilon \to f \) in \( L^1_{\text{loc}} \), for every \( f \) in \( L^1_{\text{loc}} \), and locally uniformly if \( f \) belongs to \( L^\infty_{\text{loc}} \).

Proof. If the kernel \( k \) has a compact support then for any compact set \( K \) and any \( x \) in \( K \), the integral in the convolution

\[
(f \ast k_\varepsilon)(x) = \int_{\mathbb{R}^d} f(x - \varepsilon y)k(y) \, dy
\]

uses the values of \( f \) only within the compact set \( K + \varepsilon \text{supp}\{k\} \subset F \), for every \( \varepsilon \) in \((0, 1]\). Hence \((f \ast k_\varepsilon)(x) = ((f \mathbb{1}_F) \ast k_\varepsilon)(x)\), for every \( x \) in \( K \). Thus, conclude by applying Corollary 7.6 to the function \( f \mathbb{1}_F \).

\[\square\]

Exercise 7.5. Let \( \{k_n\} \) be a sequence of infinite-differentiable functions such that \( k_n(x) = 1 \) if \(|x| < n \) and \( k_n(x) = 0 \) if \(|x| > n + 1 \), and let \( Q \) be the family of functions of the form \( qk_n \), where \( q \) is a polynomial with rational coefficients.

By means of Weierstrauss approximation Theorem 0.3, prove that \( Q \) is a dense family of \( L^p(\mathbb{R}^d, \mathcal{B}, \mu) \) under the \( p \)-norm, see last part of Exercise 7.1.

Proof. If \( f \) is in \( L^p(\mathbb{R}^d, \mathcal{B}, \mu) \) then the dominate convergence shows that \( \|f - fk_n\|_p \to 0 \) as \( n \to \infty \). Next, each \( fk_n \) has a compact support and can be approximate by a continuous function \( g \) (with a compact support) in the \( p \)-norm (actually, in the case of the Lebesgue measure, using convolution the function \( fk_n \) can be approximated uniformly by smooth functions with compact support). Alternatively, directly by means of Proposition 7.7, we can find a sequence \( \{\varphi_i\} \) of infinity differentiable functions with compact support which converges in the \( p \)-norm to \( f \), and later, each \( \varphi_i \) satisfies \( \varphi_i = \varphi_i k_n \) for \( n \) sufficiently large (depending on \( i \)).

Next, by means of Weierstrauss approximation Theorem 0.3, for every continuous function \( g \) and for every compact set \( K \) there exist a sequence of polynomials \( \{q_n\} \) satisfying \( \lim_n q_n(x) = g(x) \), uniformly within \( x \) in \( K \). Certainly, we can modify the coefficients of \( q_n \) to be rational numbers and still preserve the uniform convergence to \( g \) within the compact set \( K \). This implies that the family \( Q \) is dense in \( L^p(\mathbb{R}^d, \mathcal{B}, \mu) \).

\[\square\]

(7.2) Partition of the Unity

(7.3) Lebesgue Points

Exercise 7.6. With the notation of Proposition 7.10 prove that for any \( 1 < p < \infty \) and \( f \) in \( L^p(\mathbb{R}^d) \) we have that \( f^* \) is in \( L^p(\mathbb{R}^d) \) and \( \|f^*\|_p \leq C_p\|f\|_p \), i.e.,

\[
\int_{\mathbb{R}^d} |f^*(x)|^p \, dx \leq (C_p)^p \int_{\mathbb{R}^d} |f(x)|^p \, dx,
\]

with \( (C_p)^p = 5^d2^p p/(p - 1) \). We may proceed as follows: first use the distribution of \( f^* \), namely, \( m(f^*, r) = \ell(\{x \in \mathbb{R}^d : f^*(x) > r\}) \), and estimate (7.5) to obtain
the inequality
\[ m(f^*, r) \leq m(g_r^*, r/2) \leq \frac{5d_2}{r} \|g_r\|_1, \]
where \( g_r(x) = f(x) \mathbb{1}_{\{|f(x)| \geq r/2\}} \). Next, based on the formula
\[ \int_{\mathbb{R}^d} |f^*(x)|^p \, dx = p \int_0^\infty r^{p-1} m(f^*, r) \, dr, \]
see Exercise 5.6, deduce the estimate
\[ \int_{\mathbb{R}^d} |f^*(x)|^p \, dx \leq p \int_0^\infty r^{p-1} \left( \frac{5d_2}{r} \int_{\{2|f(x)| \geq r\}} |f(y)| \, dy \right) \, dr, \]
which implies (G.24).

**Proof.** Adapting the details in Wheeden and Zygmund [119, Section 9.3, pp. 155–157], and as mentioned early, the Hardy-Littlewood maximal function \( f^* \) obtained from a (locally) integrable function \( f \) is not integrable in \( \mathbb{R}^d \) unless \( f = 0 \). However, for \( p \)-integrability with \( p > 1 \) follows from the above estimate (G.24).

Using the \( m(f^*, r) = \ell(\{x \in \mathbb{R}^d : f^*(x) > r\}) \), and since \( |f(x)| \leq |g_r(x)| + r/2 \), with \( g_r = f(x) \mathbb{1}_{\{|f(x)| \geq r/2\}} \), we deduce
\[ f^*(x) \leq \sup \left\{ \frac{1}{|B(x, r)|} \int_{B(x, r)} |f(y)| \, dy + \frac{r}{2} \right\} = g_r^*(x) + \frac{r}{2}, \]
which implies
\[ m(f^*, r) \leq m(g_r^*, r/2) \leq \frac{5d_2}{r} \|g_r\|_1, \]
after using estimate (7.5) of Proposition 7.10. Next, by means of
\[ \|g_r\|_1 = \int_{\{2|f(x)| \geq r\}} |f(y)| \, dy \]
and Exercise 5.6, we deduce
\[ \int_{\mathbb{R}^d} |f^*(x)|^p \, dx \leq p \int_0^\infty r^{p-1} \left( \frac{5d_2}{r} \int_{\{2|f(x)| \geq r\}} |f(y)| \, dy \right) \, dr, \]
and interchanging the order of integration,
\[ \int_0^\infty r^{p-1} \left( \frac{5d_2}{r} \int_{\{2|f(x)| \geq r\}} |f(y)| \, dy \right) \, dr = 5d_2 \int_{\mathbb{R}^d} |f(x)| \, dx \int_0^{2|f(x)|} r^{p-2} \, dr. \]
Because \( p > 1 \), the inner integral is finite and equal to \((2|f(x)|)^{p-1}/(p-1)\), as long as \( f(x) \) is finite. Hence
\[
\int_{\mathbb{R}^d} |f^*(x)|^p \, dx \leq \frac{2^p p^d}{p - 1} \int_{\mathbb{R}^d} |f(x)|^p \, dx,
\]
as desired.

The reader may want to take a look at Wheeden and Zygmund [119, Sections 9.3 and 9.4] to find some application of this maximal function, for instance, if \( k \) is a kernel such that \( |k(x)| \leq \overline{k}(x) = \phi(|x|) \) for a monotone decreasing function \( t \mapsto \phi(t) \) with \( \overline{k} \) being integrable on \( \mathbb{R}^d \) then
\[
\sup_{\varepsilon > 0} |(f \ast k_{\varepsilon})(x)| \leq C f^*(x),
\]
where \( C \) is the product of the volume of the unit ball (i.e., \(|B(0,1)|\)) and the integral of \( \overline{k} \) (i.e., \(|\overline{k}|_1\)).

**Exercise 7.7.** Verify that if \( E \) is a Lebesgue measurable set then almost every \( x \) in \( E \) is a point of density of \( E \), i.e., we have \( |E \cap B(x,r)|/|B(x,r)| \to 1 \) as \( r \to 0 \). Similarly, almost every \( x \) in \( E^c \) is a point of dispersion of \( E \), i.e., we have \( |E \cap B(x,r)|/|B(x,r)| \to 0 \) as \( r \to 0 \).

**Proof.** Based on Theorem 7.12
\[
\lim_{r \to 0} \frac{1}{|B(x,r)|} \int_{B(x,r)} |f(y) - f(x)| \, dy = 0, \quad \forall x \in \mathbb{R}^d \setminus N,
\]
for any Lebesgue locally integrable function \( f \), which yields
\[
\lim_{r \to 0} \frac{1}{|B(x,r)|} \int_{B(x,r)} f(y) \, dy = f(x), \quad \forall x \in \mathbb{R}^d \setminus N.
\]
In particular, for \( f = \mathbb{1}_E \) we deduce \( |E \cap B(x,r)|/|B(x,r)| \to 1 \) as \( r \to 0 \), for every \( x \) in \( E \setminus N \), while \( |E \cap B(x,r)|/|B(x,r)| \to 0 \) as \( r \to 0 \), for every \( x \) in \( E^c \setminus N \).

**Exercise 7.8.** Prove that if \( f \) is \( p \)-locally integrable, i.e., \(|f|^p\) is integrable on any compact subset of \( \mathbb{R}^d \), \( 1 \leq p < \infty \), then
\[
\lim_{r \to 0} \frac{1}{|B(x,r)|} \int_{B(x,r)} |f(y) - f(x)|^p \, dy = 0, \quad \forall x \in \mathbb{R}^d \setminus N,
\]
or equivalently
\[
\lim_{r \to 0} \int_{|y| \leq R} |f(x + ry) - f(x)|^p \, dy = 0, \quad \forall x \in \mathbb{R}^d \setminus N,
\]
for some negligible set \( N = N_f, |N| = 0 \) and any radius \( R > 0 \).
Proof. Revise the last argument in Theorem 7.12. Given a p-locally (1 ≤ p < ∞) integrable f, consider the countable family \{g_q : q ∈ \mathbb{Q}\} of locally integrable functions \(g_q(x) = |f(x) - q|^p\), with \(\mathbb{Q}\) being the set of rational numbers. For each \(g_q\) there is a negligible subset \(N_q \subset \mathbb{R}^d\), where

\[
\lim_{r \to 0} G_q(x, r) = g_q(x), \quad G_q(x, r) = \frac{1}{|B(x, r)|} \int_{B(x, r)} g_q(y) \, dy,
\]
does not hold true. Hence, for \(x\) in \(\mathbb{R}^d \setminus N\), with \(N = \bigcup_q N_q\), we have

\[
\lim_{r \to 0} \frac{1}{|B(x, r)|} \int_{B(x, r)} |f(y) - q|^p \, dy = |f(x) - q|^p, \quad \forall q \in \mathbb{Q}.
\]

By taking a sequence of rational \(\{q\}\) convergent to \(f(x)\) and using the estimate

\[
|f(y) - f(x)|^p \leq 2^p|f(y) - q|^p + 2^p|q - f(x)|^p,
\]

we obtain the desired result.

It is clear that the condition

\[
\lim_{r \to 0} \int_{|y| \leq R} |f(x + ry) - f(x)|^p \, dy = 0, \quad \forall x \in \mathbb{R}^d \setminus N,
\]

for some negligible set \(N = N_f, |N| = 0\) and any radius \(R > 0\) follows after the change of variables \(y = x + rz\).

\[\square\]

Exercise 7.9. With the notation of Remark 7.14, consider a locally essentially bounded function \(f\). Prove that (a) \(f(x) \leq \underline{f}(x) \leq \overline{f}(x)\) at every Lebesgue point \(x\) of \(f\). Next, (b) if \(g(x) = \overline{f}(x) = f(x)\) is finite for every \(x\) in an open set \(U\) then show that \(f = g\) a.e. in \(U\) and that \(g\) is almost continuous in \(U\), i.e., for any \(x\) in \(U\) and every \(\varepsilon > 0\) there exist a null set \(N\) and \(\delta > 0\) such that \(|y - x| < \delta\) and \(y\) in \(U \cap N^c\) imply \(|g(y) - g(x)| < \varepsilon\), see Exercise 4.23. Therefore, a measurable function \(f\) is called almost upper (lower) semi-continuous if \(f = \overline{f} (f = \overline{f})\) almost everywhere, and almost continuous if \(f = \overline{f} = \underline{f}\) a.e. Finally, (c) verify that if \(f\) is a continuous (or upper/lower semi-continuous) a.e. then \(f\) is almost continuous (or almost upper/lower semi-continuous) a.e. Is there any function which is not continuous a.e. but nevertheless it is almost continuous a.e.?

Proof. (a) Recalling that \(\underline{f}(x) = \text{ess-liminf}_{y \to x} f(y) = \lim_{r \to 0} \text{ess-inf}\{f(y) : y \in B(x, r)\}\) and \(\overline{f}(x) = \text{ess-limsup}_{y \to x} f(y) = \lim_{r \to 0} \text{ess-sup}\{f(y) : y \in B(x, r)\}\), if \(x\) is a Lebesgue point for \(f\) then the inequality

\[
\underline{f}(x) \leq \lim_{r \to 0} \frac{1}{|B(x, r)|} \int_{B(x, r)} f(y) \, dy \leq \overline{f}(x),
\]

implies \(\text{ess-liminf}_{y \to x} f(y) \leq f(x) \leq \text{ess-limsup}_{y \to x} f(y)\).

(b) Note that if \(\{r_k\}\) is a monotone sequence of positive numbers converging to zero then \(\text{ess-inf}\{f(y) : y \in B(x, r)\} \leq \text{ess-inf}\{f(y) : y \in B(x, r_k)\}\), for
every $r \geq r_k$, and $\text{ess-inf}\{f(y) : y \in B(x,r)\} \geq \text{ess-inf}\{f(y) : y \in B(x,r_k)\}$, for every $r \leq r_k$. Therefore, for every $x$, the essential limit $r \to 0$ is the same as $r_k \to 0$, i.e., we could define $\overline{f}(x) = \lim_k \text{ess-inf}\{f(y) : y \in B(x,r_k)\}$ and similarly, $\underline{f}(x) = \lim_k \text{ess-sup}\{f(y) : y \in B(x,r_k)\}$.

Now, if $\overline{f}(x) < \bar{\ell}$ is finite then there exists an index $K$ such that $\overline{f}(x) \leq \text{ess-sup}\{f(y) : y \in B(x,r_k)\} < \bar{\ell}$, for every $k > K$, and thus a null set $N$ such that $\overline{f}(x) \leq \sup\{f(y) : y \in B(x,r_k) \setminus N\} < \bar{\ell}$, for every $k > K$.

Similarly, if $\overline{f}(x) > \underline{\ell}$ is finite then there exists an index $K$ and a null set $N$ such that $\overline{f}(x) \geq \inf\{f(y) : y \in B(x,r_k) \setminus N\} > \underline{\ell}$, for every $k > K$.

It is clear that the index $K$ and the null set $N$ depend on $\bar{\ell}$ (or $\underline{\ell}$) and the point $x$. Certainly, the null set $N$ may be taken independent of the particular sequence $x_k$ converging to $x$, i.e., given $x$ and $\varepsilon > 0$ there exists $\delta > 0$ and a null set $N$ such that $|y - x| < \delta$ and $y$ not in $N$ implies $\overline{f}(x) > f(y) - \varepsilon$ and $\underline{f}(x) < f(y) + \varepsilon$. In particular, if $\overline{f}(x) = \underline{f}(x) < \infty$ we have the so-called almost continuity at the point $x$.

Therefore $\overline{f}$ (respectively, $\underline{f}$) is the larger almost upper semi-continuous (respectively, smaller almost lower semi-continuous) function below (respectively, above) $f$, and if $\overline{f}(x) = \underline{f}(x) < \infty$ then $f$ is almost continuous at $x$, i.e., the essential limit of $f$ as $y \to x$ exits and $f(x) = \text{ess-lim}_{y \to x} f(x)$ is a finite value. As in Exercise 4.23, we have $\overline{f}(x) = \inf_{r > 0} \overline{f}(x,r)$, with $\overline{f}(x,r) = \text{ess-sup}\{f(y) : y \in B(x,r)\}$, and similarly for $\underline{f}$. Moreover, for each $r > 0$, $x \mapsto \overline{f}(x,r)$ (or $\underline{f}(x,r)$) is almost usc (almost lsc) and the essential supremum (infimum) of a family of almost usc (almost lsc) functions is also almost usc (almost lsc).

(c) If $f$ is continuous almost everywhere, then there exists a null set $N$ such that $f$ is continuous on the complement of $N$, while a function $f$ is almost everywhere continuous when the set of all points of discontinuity is a null set. Naturally, a function $f$ is continuous almost everywhere in a neighborhood $U$ of $x$ if there exists a null set $N$ such that $f$ is continuous in $U \setminus N$. Clearly, in this local continuity almost everywhere the null set $N$ may depend on the point $x$, but if the whole space can be covered by a countable number of opens, then global continuity almost everywhere follows. In any case, all these three concepts are not pointwise, they refer to a neighborhood.

In principle, continuity almost everywhere and almost everywhere continuity are not equivalent, i.e., it is clear that if $f$ is continuous almost everywhere and $f = g$ a.e., then $g$ is also continuous almost everywhere. However, an equivalent property for the almost everywhere continuity is not so clear. For instance, the function $1_{\mathbb{Q}}$, which is equal to 1 on the rational points and to 0 otherwise, is continuous nowhere, but it is equal to the 0 function almost everywhere. Hence, when working with functions defined almost everywhere (i.e., functions are now equivalence classes of functions which are equal almost everywhere) usually a function $f$ (or properly speaking, the class of equivalence of a function) is called continuous (or continuous almost everywhere) if there exists a continuous (or continuous almost everywhere) function $g$ such that $f = g$ outside of a null set (or set of measure zero).
On the other hand, the almost everywhere continuity is clearly a pointwise property, i.e., the almost continuity is a property adapted to functions defined almost everywhere, i.e., and if \( f \) is almost continuous and \( f = g \) a.e., then \( g \) is also almost continuous.

Moreover, at each point \( x \) where \( f \) is continuous (or upper/lower semi-continuous) we have \( f(x) = \lim_{y \to x} f(y) \) (or inequality with the superior/inferior limits), which implies the same equality (inequality) with the essential limits; and the same conclusion holds true if \( f \) is continuous almost everywhere. This means that \( f \) is almost continuous (or almost upper/lower semi-continuous) at \( x \), whenever \( f \) is either continuous at \( x \) or continuous almost everywhere in a neighborhood of \( x \).

Assume that \( f \) is continuous almost everywhere on an interval \( [a, b] \) (i.e, the set of discontinuity of \( f \) is a null set in \( [a, b] \)). Now, take two countable dense sets \( Q_1 \) and \( Q_2 \) in \( [a, b] \) (with empty interception), and define \( g(x) = f(x) + 1 \) if \( x \) belongs to \( Q_1 \), \( g(x) = f(x) - 1 \) if \( x \) belongs to \( Q_2 \), and \( g(x) = f(x) \) otherwise. Then \( f = g \) a.e., but in view of the density of \( Q_1 \) and \( Q_2 \) the limit \( \lim_{y \to x} g(y) \) cannot exists, indeed, for any \( x \), point of continuity of \( f \), if \( y_k \to x \) and \( y_k \) belongs to \( Q_1 \) (respectively, \( Q_2 \)) then \( g(x_k) \to f(x) + 1 \) (respectively, \( g(x_k) \to f(x) - 1 \)), i.e., \( g \) is discontinuous at every point where \( f \) is continuous, so \( g \) is not continuous almost everywhere (it is discontinuous in a set of full measure), but \( f = g \) almost everywhere. However, \( g \) could be called continuous almost everywhere because there exits an equivalent function \( f \) which is continuous almost everywhere.

The difference between almost everywhere continuity (i.e., almost continuity at almost every point) and continuity almost everywhere (i.e., continuous on the complement of a null set) is subtle. For the almost continuity, the null set where the continuity does not hold, may depend on the point where this continuity is considered. As mentioned early, if \( f = g \) almost everywhere and \( f \) is continuous (or continuous except in a null set) then \( g \) may even be discontinuous everywhere Note that an almost continuous function is not necessarily Riemann integrable, as in the case of a function which is discontinuous in a null set. \( \square \)

**Exercise 7.10.** Consider a locally integrable function \( f \) defined on an open set \( \mathcal{O} \) of \( \mathbb{R}^d \) and a bounded kernel \( k \) with compact support, e.g., like (7.3). If \( k_{\varepsilon}(x) = \varepsilon^{-d} k(x/\varepsilon) \) then prove that \( (f \ast k_{\varepsilon})(x) \to f(x) \) almost everywhere, indeed, for any Lebesgue point (Theorem 7.12) of \( f \).

**Proof.** If \( f \) is a locally integrable function defined on an open set \( \mathcal{O} \) of \( \mathbb{R}^d \) then \( f \) is extended by zero outside \( \mathcal{O} \) so that the convolution

\[
(f \ast k_{\varepsilon})(x) = \int_{\mathcal{O}} f(y) k_{\varepsilon}(x-y) \, dy = \int_{\mathbb{R}^d} f(x-\varepsilon y) k(y) \, dy
\]

make sense when the kernel \( k \) has a compact support. Therefore

\[
|(f \ast k_{\varepsilon})(x) - f(x)| \leq \int_{\mathbb{R}^d} |f(x-\varepsilon y) - f(x)||k(y)| \, dy,
\]
and since $|k(y)| \leq R$ and $k(y) = 0$ if $|y| \geq R$ for some $R > 0$,

$$|(f \ast k_\epsilon)(x) - f(x)| \leq R \int_{|y| \leq R} |f(x - \epsilon y) - f(x)| \, dy =$$

$$= R^{d-1} \frac{1}{|B(x, \epsilon R)|} \int_{B(x, \epsilon R)} |f(y) - f(x)| \, dy.$$

In view of Theorem 7.12, as $\epsilon \to 0$ we deduce $|(f \ast k_\epsilon)(x) - f(x)| \to 0$ for every Lebesgue point $x$ of $f$.

Note that if the extension by zero of $f$ is integrable in $\mathbb{R}^d$ and the essentially bounded and integrable kernel $k$ has not necessarily a compact support then the convolution $f \ast k_\epsilon$ is defined as a function in $L^1 \cap L^\infty$, and

$$\int_{\mathbb{R}^d} |(f \ast k_\epsilon)(x) - (f \ast \overline{k}_\epsilon)(x)| \, dx \leq \|f\|_1 \|k - \overline{k}\|_1,$$

$$|(f \ast k_\epsilon)(x)| \leq \|f\|_1 \|k\|_\infty,$$

i.e., the almost everywhere convergence is valid in this case, as the kernel $k$ is approximated by kernels $\overline{k}$ with compact supports. In particular, if $f$ is essentially bounded then

$$\int_{\mathbb{R}^d} |f(x - \epsilon y) - f(x)| \, |k(y)| \, dy \leq$$

$$\leq \int_{|y| < R} |f(x - \epsilon y) - f(x)| \, |k(y)| \, dy + 2\|f\|_{\infty} \int_{|y| \geq R} |k(y)| \, dy,$$

and the last integral goes to 0 as $R \to \infty$. Thus, if $f$ and $k$ are essentially bounded, and $k$ is integrable in $\mathbb{R}^d$ (not necessarily with a compact support) then $|(f \ast k_\epsilon)(x) - f(x)| \to 0$ for every Lebesgue point $x$ of $f$. \qed

**Exercise 7.11.** Let $g_1$ be a real-valued Lebesgue measurable function of two real variables, say $g_1(x, y)$ with $x, y$ in $\mathbb{R}$. Assume that $g_1$ is locally integrable in $\mathbb{R}^2$ and define the function

$$g(x, y) = \int_0^x g_1(t, y) \, dt, \quad \forall x \in \mathbb{R},$$

for almost every $y$ in $\mathbb{R}$. (1) Verify that $g$ is a locally integrable function, which is continuous in the first variable. Now, consider the set $A$ in $\mathbb{R}^2$ of all points $(x, y)$ such that $[g(x + h, y) - g(x, y)]/h \to g_1(x, y)$ as $h \to 0$. (2) Prove that the complement $N = A^c$ is a set of Lebesgue measure zero. Next, if $f$ is a continuously differentiable function with compact support then (3) show that the convolution $f \ast g$ is a continuously differentiable function and $\partial_x(f \ast g) = (\partial_x f) \ast g$, where $\partial_x$ denotes the partial derivative in the variable $x$, and finally, by means of Fubini-Tonelli Theorem 4.14, (4) prove that $\partial_x(f \ast g) = f \ast (\partial_x g)$, if $f$ is a measurable (essentially) locally bounded function with compact support. Hint: regarding (4), first assume that

$$g(x, y) = \int_{-\infty}^x g_1(t, y) \, dt, \quad \forall x, y \in \mathbb{R}.$$
to show that
\[
\int_{-\infty}^b (f \ast g_1)(x, y) \, dx = (f \ast g)(b, y), \quad \forall b, y \in \mathbb{R},
\]
and then, consider the general case.

**Proof.** (1) It is clear that for any compact set \( K \subset \mathbb{R}^d \),
\[
\int_K |g(x, y)| \, dx \, dy \leq \int_K \, dx \, dy \int_0^x |g_1(t, y)| \, dt \leq \left( \sup\{ |x| : (x, y) \in K \} \right) \int_K |g_1(x, y)| \, dx \, dy
\]
and
\[
|g(x, y) - g(x', y)| \leq \left| \int_x^{x'} |g_1(t, y)| \, dt \right|
\]
proving the continuity in \( x \) for almost every \( y \).

(2) Recall that a measurable subset \( N \) of \( \mathbb{R}^2 \) is negligible (or null) if and only if almost each \( y \)-section \( N_y = \{ x : (x, y) \in N \} \) is a negligible set in \( \mathbb{R} \). Now, to show that the complement of \( A, N = A^c \), is a set of zero Lebesgue measure on \( \mathbb{R}^2 \), we verify that almost every \( y \)-sections have zero Lebesgue measure on \( \mathbb{R} \). Indeed, for almost every \( y \), the function \( t \mapsto g_1(t, y) \) is locally integrable, and in view of Theorem 7.12, almost every \( x \) is a Lebesgue point of \( g_1(\cdot, y) \), which implies that
\[
\frac{g(x + h, y) - g(x, y)}{h} = \frac{1}{h} \int_x^{x+h} g_1(t, y) \, dt \to g_1(x, y),
\]
proving that almost every \( y \)-sections of \( N \) is null.

(3) Let \( f \) be a continuously differentiable function and suppose that either \( g \) or \( f \) has a compact support. Then the expression
\[
\left| \frac{f(x + h - t, y) - f(x - t, y)}{h} \cdot g(t, y) \right| = |\partial_x f(x + h' - t, y)| |g(t, y)|
\]
can be bounded by an integrable function, uniformly in \( |h| \leq 1, (t, y) \) in \( \mathbb{R}^2 \) and \( x \) within a bounded interval. Actually, we only need to know that the partial derivative \( \partial_x f \) exists (so that the Mean Value Theorem can be applied) and is bounded within a neighborhood of the support of \( g \). This suffices to take limit inside the integral
\[
\int_{\mathbb{R}^d} \frac{(x + h - t, y) - f(x - t, y)}{h} \cdot g(t, y) \, dt \, dy
\]
to deduce that \([(f \ast g)(x + h, y) - (f \ast g)(x, y)]/h \to (\partial_x f \ast g)(x)\) as \( h \to 0 \).
(4) Now, let \( f \) be a measurable (essentially) locally bounded function and suppose that either \( g \) or \( f \) has a compact support. First, by means of Fubini-Tonelli Theorem 4.14, check that

\[
\int_a^b (f \ast g_1)(x,y) \, dx = \int_{\mathbb{R}^2} f(x',y') g(b-x',y-y') \, dx' \, dy' = (f \ast g)(b,y) - (f \ast g)(a,y),
\]

and thus

\[
\frac{(f \ast g)(x+h,y) - (f \ast g)(x,y)}{h} = \frac{1}{h} \int_x^{x+h} (f \ast g_1)(t,y) \, dt.
\]

For almost every \( y \) the function \( x \mapsto (f \ast g_1)(x,y) \) is locally integrable, which implies that almost every \( x \) is a Lebesgue point, and so, the limit of the previous expression approaches \((f \ast g_1)(x,y)\), as \( h \to 0 \), i.e., \((f \ast g)\) is differentiable in \( x \), for almost every \((x,y)\) in \( \mathbb{R}^2 \). Moreover, because the limit is precisely \((f \ast g_1) = f \ast (\partial_x g)\), we conclude. \( \square \)

(7.4) Functions of one variable

**Exercise 7.12.** With the above notation verify that (1) \( \var \), \( \var^\pm \) are additive functions on intervals and sub-additive on the functions, e.g., if \( a < c < b \) then \( \var(f, [a, c]) + \var(f, [c, b]) = \var(f, [a, b]) \) and \( \var(f + \alpha g, [a, b]) \leq \var(f, [a, b]) + |\alpha| \var(g, [a, b]) \), for any real constant \( \alpha \). Next, assuming that \( f \) is a function with bounded variation, show that (2) \( \var(f, [a, b]) = \var^+(f, [a, b]) + \var^-(f, [a, b]) \); (3) \( f \) can be written as the difference of two monotone functions, namely, \( f(x) = \var(f, [a, x]) - (f(x) - \var(f, [a, x])) \) or (4) \( f(x) - f(a) = \var^+(f, [a, x]) - \var^-(f, [a, x]) \). Moreover, if \( \pi = \{a = x_0 < x_1 < \ldots < x_n = b\} \) is a partition of \([a, b]\) with mesh (or norm) \( |\pi| = \max_i \{(x_i - x_{i-1})\} \) then prove that (5) \( \var(f, \pi) = \sum_{i=1}^n |f(x_i) - f(x_{i-1})| \) converges to \( \var(f, [a, b]) \) as \( |\pi| \to 0 \), provided \( f \) is also continuous. Furthermore, if \( f \) is also right-continuous then prove (6) the variation functions \( x \mapsto \var(f, [a, x]) \) and \( x \mapsto \var^\pm(f, [a, x]) \) are right-continuous and \( \var(f, [a, x]) \to 0 \) as \( x \to 0^+ \).

**Proof.** There several sources, e.g., Gordon [54, Chapter 4, pp. 49–68], Dshalalow [36, Section 9.2, pp. 528–534], Leoni [75, Chapter 2, pp. 39–72], Wheeden and Zygmund [119, Section 2.1, pp. 15–21], among many others.

(1) Let us check that if \( a < c < b \) then \( \var(f, [a, c]) + \var(f, [c, b]) = \var(f, [a, b]) \). Since, form a partition of \([a, c]\) and a partition of \([c, b]\) we can construct a partition of the whole \([a, b]\), it follows that \( \var(f, [a, c]) + \var(f, [c, b]) \leq \var(f, [a, b]) \). On the other hand, the triangular inequality implies that if \( a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b \) then adding an extra point \( c \) (if necessary) to the partition, i.e., \( a = x_0 < x_1 < \ldots < x_{k-1} \leq c \) and \( c \leq x_k < \ldots < x_{n-1} < x_n = b \),
we obtain
\[
\text{var}(f, [a, c]) + \text{var}(f, [c, b]) \geq |f(c) - f(x_{k-1})| + \sum_{i=1}^{k-1} |f(x_i) - f(x_{i-1})| + \\
+ |f(x_k) - f(c)| + \sum_{i=k+1}^{n} |f(x_i) - f(x_{i-1})| \geq \\
\geq \sum_{i=1}^{n} |f(x_i) - f(x_{i-1})|
\]
which implies \(\text{var}(f, [a, c]) + \text{var}(f, [c, b]) \geq \text{var}(f, [a, b])\), after taking the supremum on all partitions.

Similarly, the triangular inequality yields \(\text{var}(f + \alpha g, [a, b]) \leq \text{var}(f, [a, b]) + |\alpha| \text{var}(g, [a, b])\), for any real constant \(\alpha\).

(2) The identity \(|f(x) - f(y)| = [f(x) - f(y)]^+ + [f(x) - f(y)]^-\) yields \(\text{var}(f, [a, b]) = \text{var}^+(f, [a, b]) + \text{var}^-(f, [a, b])\).

(3) Assume that \(\text{var}^\pm(f, [a, b]) < \infty\). In view (1), if \(y > x\) then \(\text{var}(f, [a, y]) - \text{var}(f, [a, x]) = \text{var}(f, [x, y])\), i.e., the function \(x \mapsto \text{var}(f, [a, x])\) is monotone increasing. Also, it is clear that \(|f(a) - f(b)| \leq \text{var}(f, [a, b])\), which implies that if \(x < y\) then
\[
[f(x) - \text{var}(f, [a, x])] - [f(y) - \text{var}(f, [a, y])] = \\
= [f(x) - f(y)] - \text{var}(f, [x, y]) \leq 0,
\]
i.e., the function \(x \mapsto g(x) = f(x) - \text{var}(f, [a, x])\) is also monotone increasing. This show that \(f(x) = \text{var}(f, [a, x]) - g(x)\), both functions being monotone increasing.

(4) Now, assume that \(\text{var}^\pm(f, [a, b]) < \infty\). If \(a < c < b\) then \(\text{var}^\pm(f, [a, c]) + \text{var}^\pm(f, [c, b]) = \text{var}^\pm(f, [a, b])\), and therefore, the functions \(x \mapsto \text{var}^\pm(f, [a, x])\) are monotone increasing. Hence, note that if \(\pi\) is a partition of \([a, b]\) then \(\text{var}^+(f, \pi) = [f(b) - f(a)] + \text{var}^-(f, \pi)\), so that taking the supremum over all partitions, we deduce \(f(b) - f(a) = \text{var}^+(f, [a, b]) - \text{var}^-(f, [a, b])\). This shows that \(f(x) - f(a) = \text{var}^+(f, [a, x]) - \text{var}^-(f, [a, x])\) as desired.

(5) First, in view of the continuity of \(f\), actually, uniform continuity on the compact interval \([a, b]\), (i) for every \(\varepsilon' > 0\) there exists \(\delta' > 0\) such that \(|f(x) - f(y)| < \varepsilon'\) provided \(|x - y| < \delta'\). Next, as mentioned early, adding new points to a partition does not decrease the variation, i.e., (ii) if two partitions satisfy \(\pi' \subset \pi\) (meaning that \(\pi\) is a refinement of \(\pi'\)) then \(\text{var}(f, \pi') \leq \text{var}(f, \pi)\).

We need to show that given \(r < \text{var}(f, [a, b])\) there exists \(\delta > 0\) such that \(r < \text{var}(f, \pi)\) for any partition with mesh \(|\pi| < \delta\). To this effect, if \(r < r' < \text{var}(f, [a, b])\) then there exists a partition \(\pi^*\) such that \(r' < \text{var}(f, \pi^*) \leq \text{var}(f, [a, b])\), and now, if \(k(\pi^*)\) is the number of points in the partition \(\pi^*\) then take \(\varepsilon' = (r' - r)/k(\pi^*)\) and find \(\delta'\) (due to the uniform continuity) as in (i) above. Our claim is that \(r < \text{var}(f, \pi)\) for any partition \(\pi\) with mesh \(|\pi| < \delta'\) and \(|\pi| < |\pi^*|\), i.e., \(\delta = \min\{\delta', |\pi^*|\}\). Indeed, for any partition \(\pi\) with mesh
\(|\pi| < \delta\) we can find \(k(\pi^*)\) points in the partition \(\pi\) within a distance less than \(\delta'\) from some point in the partition \(\pi^*\), so if \(x\) and \(x^*\) are two such a points then \(|f(x) - f(x^*)| < \varepsilon' = (r' - r)/k(\pi^*)\). Denote by \(\pi'\) the partition formed by all those \(k(\pi^*)\) points \(\{x\}\) of \(\pi\), and add all previous \(k(\pi^*)\) inequalities for \(|f(x) - f(x^*)|\) to deduce the estimate \(|\text{var}(f, \pi^*) - \text{var}(f, \pi')| < r' - r\), i.e., \(\text{var}(f, \pi') > r\). Finally, conclude the argument by means of (ii), this is, by invoking the fact that \(\pi\) is a refinement of \(\pi'\).

(6) Let \(f\) be right-continuous and \(\text{var}^\pm(f, [a, b]) < \infty\). Since \(\text{var}^\pm(f, [a, x]) - \text{var}^\pm(f, [a, y]) = \text{var}^\pm(f, [x, y])\), for any \(a < x < y < b\), to check that the variation functions \(x \mapsto \text{var}(f, [a, x])\) and \(x \mapsto \text{var}^\pm(f, [a, x])\) are right-continuous, we need to show only that \(\text{var}(f, [a, x]) \to 0\) as \(x \to a+\). By contradiction, assume that \(\text{var}(f, [a, x]) \to \alpha > 0\) as \(x \to a+\). Thus, choose \(x - a > 0\) so small to have \(\text{var}(f, [a, x]) < 4\alpha/3\) and a partition \(\pi = \{a < x_1 < \cdots < x_n = x\}\) of the interval \([a, x]\) satisfying \(\text{var}(f, \pi) > 2\alpha/3\). In view of the right-continuity of \(f\) at \(a\), take a point \(c\) so close to \(a < c < x_1\) to have \(|f(c) - f(a)| < \alpha/3\). Hence, the partition \(\pi_c = \{c < x_1 < \cdots < x_n = x\}\) satisfies \(\text{var}(f, \pi_c) + \alpha/3 \geq \text{var}(f, \pi) > 2\alpha/3\), which implies that \(\text{var}(f, [c, x]) > \alpha/3\). Therefore

\[4\alpha/3 > \text{var}(f, [a, x]) = \text{var}(f, [a, c]) + \text{var}(f, [c, x]) > \alpha + \alpha/3,\]

which is a contradiction.

Note that if we define

\[\text{var}(f, [a, x]) = \sup_{\varepsilon > 0} \text{var}(f, [a + \varepsilon, x]) = \lim_{\varepsilon \to 0} \text{var}(f, [a + \varepsilon, x])\]

then the previous argument shows that \(\text{var}(f, [a, x]) \to 0\) as \(x \to a+\), for any bounded variation function \(f\), non-necessarily right-continuity. Indeed, it suffices to remark that \(\text{var}(f, [a, b]) = \text{var}(g, [a, b])\), where \(f(x) = g(x)\) if \(a < x \leq b\) and \(g(a) = \lim_{x \to a+} f(x)\).

**Exercise 7.13.** If \(f\) is a function Lebesgue integrable in \((0, a)\), with \(a > 0\) and

\[g(x) = \int_x^a \frac{f(t)}{t} \, dt, \quad \forall x \in (0, a),\]

then \(g\) is Lebesgue integrable in \((0, a)\). When the equality

\[\int_0^a g(x) \, dx = \int_0^a f(x) \, dx\]

is valid?

**Proof.** For any \(\varepsilon\) in \((0, a)\), the function \(g\) is a continuously differentiable on \([\varepsilon, a]\) with \(g(a) = 0\), and \(g'(x) = -f(x)/x\). Thus, the integration by part formula yields

\[\int_\varepsilon^a g(x) \, dx = xg(x) \big|_\varepsilon^a - \int_\varepsilon^a xdg(x) = -\varepsilon g(\varepsilon) + \int_\varepsilon^a f(x) \, dx.\]
If $f \geq 0$ then $g \geq 0$ and

$$0 \leq \int_{\varepsilon}^{a} g(x) \, dx \leq \int_{\varepsilon}^{a} f(x) \, dx \leq \int_{0}^{a} f(x) \, dx,$$

and because $f$ is Lebesque integrable on $[0, a]$, so is $g$. Applying this argument to the positive and negative parts of $f$, we deduce that $g$ is Lebesgue integrable on $[0, a]$, without assuming $f \geq 0$.

Once we know that $g$ is integrable on $[0, a]$, the integration by part formula yields

$$\int_{0}^{a} g(x) \, dx + \lim_{\varepsilon \to 0} \varepsilon g(\varepsilon) = \int_{0}^{a} f(x) \, dx.$$  

If the limit vanishes then both integrals agree, e.g., if $f$ is bounded by a constant $C$ then $|g(x)| \leq C(\ln a - \ln x)$, which implies that $\varepsilon |g(\varepsilon)| \to 0$ as $\varepsilon \to 0$.

**Exercise 7.14.** If $f$ and $g$ are two absolutely continuous functions on $[a, b]$ then prove that the following integration-by-parts formula

$$\int_{a}^{b} f'(x)g(x) \, dx + \int_{a}^{b} f(x)g'(x) \, dx = f(b)g(b) - f(a)g(a)$$

is meaningful and correct.

**Proof.** If $f$ and $g$ are absolutely continuous functions on $[a, b]$ then $f'$ and $g'$ exist almost everywhere and are integrable functions, being $[a, b]$ a bounded interval, $f$ and $g$ are also bounded functions, so that the product $f'g$ and $fg'$ are integrable functions. Moreover, the function

$$H(x) = \int_{a}^{x} f'(t)g(t) \, dt + \int_{a}^{b} f(t)g'(t) \, dt$$

is absolutely continuous and $H' = f'g + fg'$ almost everywhere. On the other hand, the function $G(x) = f(x)g(x) - f(a)g(a)$ is also absolutely continuous and

$$\frac{G(x+h) - G(x)}{h} = \left(\frac{f(x+h) - f(x)}{h}\right)g(x) + f(x)\left(\frac{g(x+h) - g(x)}{h}\right)$$

shows that $G' = f'g + fg'$ almost everywhere. Hence, because an absolutely continuous function is constant if its derivative vanishes almost everywhere, and $H(a) = G(a) = 0$, we deduce that $H = G$, i.e., the integration by part formula holds.

**Exercise 7.15.** Let $f$ be a right-continuous increasing function and $m_f$ be its corresponding Lebesgue-Stieltjes measure on $\mathbb{R}$. First prove that (a) the function $f$ is absolutely continuous if and only if the measure $m_f$ is absolutely continuous with respect to the Lebesgue measure $\ell$. Next, (b) extend these results to functions with bounded variation.
Proof. A function \( f \) is an absolutely continuous if and only if for any \( \varepsilon > 0 \) there exists \( \delta > 0 \) such that
\[
\sum_{i=1}^{n} |f(b_i) - f(a_i)| < \varepsilon \quad \text{if} \quad \sum_{i=1}^{n} (b_i - a_i) < \delta, \quad a_{i+1} \geq b_i > a_i, \quad n \geq 1.
\]

If \( m_f \) denotes the Lebesgue-Stieltjes measure associated with a right-continuous increasing function \( f \) then the above condition take the form \( m_f(I) < \varepsilon \) whenever \( \ell(I) < \delta \), with \( I \) being the finite union of disjoint semi-open intervals \( \sum_{i=1}^{n} [a_i, b_i] \). Also, by definition, the measure \( m_f \) is absolutely continuous with respect to the Lebesgue measure \( \ell \) if and only if for every \( \varepsilon > 0 \) there exists a \( \delta > 0 \) such that \( m_f(A) < \varepsilon \) whenever \( \ell(A) < \delta \).

Thus, if \( m_f \ll \ell \) then \( f \) is absolutely continuous. Conversely, if \( A \) is a measurable set with \( \ell(A) < \delta \) then there exists a countable cover by semi-open intervals \( \{I_i\} \), \( A \subset \bigcup_i I_i \), such that \( \sum_i \ell(I_i) < \delta \). Because a difference of any two semi-open intervals is a disjoint union of semi-open intervals, the (non-necessarily disjoint) union \( \bigcup_i I_i \) can be written as disjoint union \( \sum_i I_i' \) of semi-open intervals, i.e.,
\[
A \subset \sum_i I_i' \quad \text{and} \quad \sum_i \ell(I_i') = \ell(\sum_i I_i) \leq \sum_i \ell(I_i) < \delta.
\]

Hence, using the absolutely continuity of \( f \), for any finite sub-sequence \( \{I_i' \colon i < n\} \) we have \( \sum_{i<n} m_f(I_i') < \varepsilon \), which implies \( m_f(A) \leq \sum_i m_f(I_i') \leq \varepsilon \), i.e., \( m_f \ll \ell \). \( \square \)

**Exercise 7.16.** Let \( f \) be a monotone increasing function on \([0, \infty)\), and denote by \( f(a-) \) and \( f(b+) \), \( a > 0 \), the lateral limits from the left and from the right. Verify (1) that \( f(x+) = f(x-) \) for any \( x \) except for a countable number of points, and that the functions \( x \mapsto f_-(x) = f(x-) \) with \( f_-(0) = 0 \) and \( x \mapsto f_+(x) = f(x+) \) with \( f_+(0) = f(0) \) are monotone increasing functions, \( f_+ \) is continuous from the right and \( f_- \) is continuous from the left. Prove (2) that \( f_- \leq f \leq f_+ \) and if
\[
f_i(0) = 0 \quad \text{and} \quad f_i(x) = \sum_{0 \leq y < x} (f_+(y) - f(y)) \quad \forall x > 0,
\]
\[
\bar{f}_r(0) = f(0) \quad \text{and} \quad \bar{f}_r = \sum_{0 \leq y < x} (f(y) - f_-(y)) \quad \forall x > 0,
\]
then show that \( f_i \) and \( \bar{f}_i = f - \bar{f}_r \) are continuous from the left and \( f_r = f - f_i \) and \( \bar{f}_r \) are continuous from the right, and all four functions are monotone increasing.

**Proof.** (1) First, note that for every \( \varepsilon > 0 \) there can be only a finite number of jumps larger than \( \varepsilon \) within a bounded interval \([0, b]\), i.e.,
\[
f(b) - f(0) \geq \sum_{i=1}^{n} [f(x_i+) - f(x_i-)] \geq \varepsilon n,
\]
which yields a countable number of points satisfying \( f(x+) > f(x-) \).

It is clear that the functions \( f_\pm \) are monotone increasing, and to check that \( f_+ \) is continuous from the right on \([0, \infty)\) we may argue as follows.

If \( \{x_n\} \) is a monotone increasing sequence to \( x \) then for every \( n \) there exits \( x_n - 1/n < x'_n < x_n \) such that \( f(x'_n) > f(x_n +) - 1/n \), which shows that

\[
f(x+) = \lim_{n} f(x'_n) = \lim_{n} f^+(x_n) = f^+(x).
\]

Similarly, we obtain that \( f_- \) is continuous from the left on \((0, \infty)\).

(2) It is clear that \( f_- \leq f \leq f_+ \), as well as the facts that \( f_i \) is continuous from the left, \( f_r \) is continuous from the right, and both are monotone increasing.

Consider

\[
\tilde{f}_i(x) = f(x) - \tilde{f}_r(x) = f(x) - \sum_{0 \leq y \leq x} (f(y) - f_- (y)),
\]

which yields

\[
\tilde{f}_i(x-) = f_-(x) - \sum_{0 \leq y < x} (f(y) - f_- (y)) = f(x) - \sum_{0 \leq y \leq x} (f(y) - f_- (y)),
\]

proving that \( \tilde{f}_i \) is continuous from the left. Similarly,

\[
f_r(x) = f(x) - f_l(x) = f(x) - \sum_{0 \leq y < x} (f_+ (y) - f (y)),
\]

which yields

\[
f_r(x+) = f_+(x) - \sum_{0 \leq y \leq x} (f_+ (y) - f (y)) = f(x) - \sum_{0 \leq y < x} (f_+ (y) - f (y)),
\]

showing that \( f_r \) is continuous from the right.

On the other hand, if \( x' > x \) then the series satisfies

\[
\sum_{x < y \leq x'} (f(y) - f_- (y)) \leq (f(x') - f(x)),
\]

which implies that

\[
\tilde{f}_i(x') - \tilde{f}_i(x) = (f(x') - f(x)) - \sum_{x \leq y \leq x'} (f(y) - f_- (y)) \geq 0,
\]

i.e., \( \tilde{f}_i \) is a monotone increasing function. Similarly, the series also satisfies

\[
\sum_{x \leq y < x'} (f_+ (y) - f (y)) \leq (f(x') - f(x)),
\]

which implies that

\[
f_r(x') - f_r(x) = (f(x') - f(x)) - \sum_{x \leq y < x'} (f_+ (y) - f (y)) \geq 0,
\]
i.e., $f_\tau$ is a monotone increasing function.

Note that this proves that either $f = \tilde{f}_\tau + \tilde{f}_i$ or and $f = f_\tau + f_i$ expresses the monotone increasing function $f$ as a sum of two monotone increasing functions, one continuous from the right and the other continuous from the left. \hfill \square

**Exercise 7.17.** A function $f$ has no discontinuities of the second class at a point $x$ of $\mathbb{R}$ if the limits from the left and from the right exists and are finite whenever they are defined (e.g., if a point $x$ is isolated from the left then the limit from the left can not be defined). Show that a function $f$ defined on a compact subset $D$ of $\mathbb{R}$ has no discontinuities of the second class if and only if there exists a sequence $\{f_k\}$ of piecewise functions defined on $\mathbb{R}$ such that $f_k$ converges to $f$ uniformly on $D$. Question: A uniform limit of a function with no discontinuities of the second class is again a function with no discontinuities of the second class?

**Proof.** ($\Rightarrow$) If $f$ is above then for each $x$ in $D$ and any $\varepsilon > 0$ there exists an open interval $I_x$ containing $x$ such that $|f(x') - f(x'')| < \varepsilon$ for any $x', x''$ in $I_x \cap D$ such that either $x', x'' < x$ or $x', x'' > x$. Because $D$ is compact, there exists a finite subcover and therefore, there exists $-\infty < x_1 < \cdots < x_n < +\infty$ such that $x_1, \ldots, x_n$ belong to $D$, and $|f(x') - f(x'')| < \varepsilon$ for any $x', x''$ in $D$ that belong to any of the $n + 1$ open subintervals $(x_i, x_{i+1})$, for $i = 0, \ldots, n$. Now, take $\varepsilon = 1/k$ and make a choice of points $x'_j$ in $D \cap (x_j, x_{j+1})$, for $j = 0, 1, \ldots, n$ (if $D \cap (x_j, x_{j+1})$ is empty then pick any $x'_j$ in $D$). Thus define $f_k(x_i) = f(x_i)$ for $i = 1, \ldots, n$, and $f_k(x) = f(x'_j)$ for any $x$ in the open subinterval $(x_j, x_{j+1})$, for $j = 0, 1, \ldots, n$. The sequence $\{f_k\}$ of functions defined on $\mathbb{R}$ satisfies $|f_k(x) - f(x)| < 1/k$ for every $x$ in $D$, which prove the uniform convergence.

(\Leftarrow) If $f$ is defined on $D$ and $\{f_k\}$ is a sequence of piecewise functions convergent uniformly to $f$ then for each $\varepsilon > 0$ there exists $k_\varepsilon$ such that $|f(x) - f_k(x)| < \varepsilon/2$ for every $x$ in $D$ and any $k \geq k_\varepsilon$. Take a sequence $\{x_n\} \subset D$ convergent to $x$ with $x_n < x$ for all $n$ (i.e., from the left) or $x_n > x$ for all $n$ (i.e., from the right) to check that the inequality

$$|f(x_n) - f(x_m)| \leq |f(x_n) - f_k(x_n)| + |f_k(x_n) - f_k(x_m)| +$$

$$+ |f_k(x_m) - f(x_m)| \leq \varepsilon + |f_k(x_n) - f_k(x_m)|$$

implies that $\{f(n_n)\}$ is a Cauchy sequence, which proves that the limit from the left exists and it is finite. Hence, the function $f$ has no discontinuities of the second class on $D$.

Note that the argument used on the second part shows that a uniform limit of a function with no discontinuities of the second class is again a function with no discontinuities of the second class. Moreover, if $f$ is nonnegative (or left-hand or right-hand continuous) then the approximation functions $f_k$ can be chosen also nonnegative (or left-hand or right-hand continuous). The reader may want to check Amann and Escher [2, Chapter VI, Theorem 1.2, pp. 6-7] for more details. \hfill \square
Exercise 7.18. Complete the details on the above statements on convex functions, in particular, show that if \( f \) is convex in \((a, b)\), i.e., \( f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y) \), for every \( x, y \) in \((a, b)\) and \( t \) in \([0,1]\), then (1) \( f \) is necessarily continuous on \((a, b)\), (2) the left-hand derivative \( f'_-(t) = \lim_{h \to 0^-} \frac{f(t + h) - f(t)}{h} \) exists and is right-continuous and increasing at any \( t \), and (3) \( f \) is absolutely continuous and the derivative \( f' \) is continuous on the complement of a countable set of points. Moreover, (4) show that any slope \( m = m_x \) satisfying \( f'_-(x) \leq m \leq f'_+(x) \) provides a supporting line \( x \) in \((a, b)\), i.e., \( f(y) \geq f(x) + m(y - x) \) for every \( y \) in \((a, b)\). Finally, consider a function \( f \) defined only on the dyadic points of \((a, b)\), i.e., for \( x = a + k2^{-n}(b - a) \), \( k = 1, \ldots, n - 1, n = 1, 2, \ldots \), such that \( 2f((x + y)/2) \leq f(x) + f(y) \) for any dyadic point \( x \) and \( y \) and prove that (5) \( f \) can be uniquely extended to a convex function defined on the whole interval \((a, b)\).

**Proof.** (1) Take three points \( a < x_1 < x < x_2 < b \) and write \( x = (1 - t)x_1 + tx_2 \), and because \( f \) is convex, \( f(x) \leq (1 - t)f(x_1) + tf(x_2) \). Since \( t = (x - x_1)/(x_2 - x_1) \), we deduce

\[
\frac{f(x_2) - f(x_1)}{x_2 - x_1} \geq \frac{f(x) - f(x_1)}{x - x_1} \quad \forall a < x_1 < x < x_2 < b.
\]

Now, fix \( x_1 \) and let \( x \to x_1^- \) to obtain \( \lim \sup_{x \to x_1^-} f(x) \leq f(x_1) \). Similarly, writing \( x = tx_1 + (1 - t)x_2 \) we deduce

\[
\frac{f(x_2) - f(x_1)}{x_2 - x_1} \geq \frac{f(x_2) - f(x)}{x_2 - x} \quad \forall a < x_1 < x < x_2 < b.
\]

Again, fix \( x_2 \) and let \( x \to x_2^+ \) to obtain \( \lim \inf_{x \to x_2^+} f(x) \geq f(x_2) \). Hence \( f \) is continuous on the open interval \((a, b)\).

(2) These same inequalities show that left-hand and right-hand derivatives exist as monotone limits, and moreover, the inequality

\[
\frac{f(y) - f(a)}{y - a} \leq f'_+(y) \leq \frac{f(x) - f(y)}{x - y} \leq f'_-(x) \leq \frac{f(b) - f(x)}{b - x},
\]

for every \( a < y < x < b \), imply that \( f'_+ \) and \( f'_- \) are monotone increasing functions.

(3) Actually, the previous inequalities prove that

\[
|f(x) - f(y)| \leq M|x - y|, \quad \forall a < a' < x < y < b' < b,
\]

with \( M = \max\{|f'(b')|, |f'(a')|\} \). This means that \( f \) is Lipschitz continuous on the bounded open interval \((a, b)\), which implies that \( f \) is absolutely continuous on \((a, b)\).

Even if \( f \) is convex on the closed interval \([a, b]\), the continuity at the boundary is not ensured, but the lateral limits must exit at \( a \) and \( b \), i.e., a jumps may appear, \( f(a) \leq f(a+) \) or \( f(b) \geq f(b-) \).
(4) Indeed, fix \( x \) in \((a,b)\) and choose a slope \( m = m_x \) such that \( f'_-(x) \leq m \leq f'_+(x) \). Now, in the previous inequalities, take \( x \to x_1+ \) and \( x \to x_2- \) to deduce

\[
\frac{f(x_2) - f(x_1)}{x_2 - x_1} \geq f'_+(x_1) \quad \text{and} \quad \frac{f(x_2) - f(x_1)}{x_2 - x_1} \geq f'_-(x_2),
\]

e.i., for \( x_2 = y \) and \( x_1 = x \) we have

\[
f(y) \geq f(x) + f'_+(x)(y - x) \geq f(x) + m(y - x), \quad \forall y \in (x, b),
\]

and for \( x_2 = x \) and \( x_1 = y \) we have

\[
f(y) \geq f(x) + f'_-(x)(y - x) \geq f(x) + m(y - x), \quad \forall y \in (a, x).
\]

Hence, \( f(y) \geq f(x) + m(y - x) \) for every \( y \) in \((a, b)\).

(5) It is clear that the condition \( 2f((x + y)/2) \leq f(x) + f(y) \) for any dyadic point \( x \) and \( y \) in \((a, b)\) is equivalent to \( f(tx + (1 - t)y)) \leq tf(x) + (1 - t)f(y) \) for any dyadic point \( x \) and \( y \) in \((a, b)\) and any dyadic number \( t \) in \([0,1]\). The previous argument (3) can be used to show that \( f \) satisfies a Lipschitz condition for any dyadic point in \([a', b']\) any dyadic \( a' > a \) and \( b' < b \). Hence, \( f \) can be uniquely extended to a continuous function satisfying \( f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y) \) for any point \( x \) and \( y \) in \((a, b)\) and any dyadic number \( t \) in \([0,1]\), and by continuity, this implies that \( f \) is convex.

**Exercise 7.19.** Prove Jensen’s inequality, i.e., if \( \Phi: \mathbb{R} \to \mathbb{R} \) is a convex function and \( \psi \) is a real-valued integrable function in a probability space \((\Omega, \mathcal{F}, P)\) then

\[
\Phi \left( \int_\Omega \psi(\omega) P(d\omega) \right) \leq \int_\Omega \Phi(\psi(\omega)) P(d\omega).
\]

What can be said when \( \phi \) is not necessarily integrable? In particular, deduce that if \( f \) and \( k \) are real-valued measurable functions on a measure space \((X, \mathcal{X}, \mu)\) such \( fk \) is integrable and \( k \geq 0 \) is a kernel (i.e., an integrable function with integral equals to 1) then

\[
\Phi \left( \int_X fk \, d\mu \right) \leq \int_X \Phi(f)k \, d\mu.
\]

Hint: verify first that because \( \Phi \) is convex then for every \( t_0 \) there exists a slope \( \alpha(t_0) \) such that \( \Phi(t) \geq \Phi(t_0) + \alpha(t_0)(t - t_0) \), for every \( t \) in \( \mathbb{R} \).

**Proof.** The idea is simple, because \( \Phi \) is a convex function, for any convex combination \( x = \sum_{i=1}^n a_i x_i, \quad 0 \leq a_i \leq 1, \quad \sum_{i=1}^n a_i = 1 \), we have \( \Phi(x) \leq \sum_{i=1}^n a_i \Phi(x_i). \) Thus, Jensen’s inequality follows after treating the integral as a limit of sums.

However the details are more involved, for instance, first from (4) of Exercise 7.18, for every \( t_0 \) there exists a slope \( \alpha(t_0) \) such that \( \Phi(t) \geq \Phi(t_0) + \alpha(t_0)(t - t_0) \), for every \( t \) in \( \mathbb{R} \). Now, if the integral \( \psi \) is finite and equal to \( t_0 \)
then $\Phi(t_0) + \alpha_0[\psi(\omega) - t_0] \leq \Phi(\psi(\omega))$, for every $\omega$ in $\Omega$. Integrating in $\omega$ with respect to the probability $P$ and because

$$t_0 = \int_\Omega \psi(\omega) P(d\omega),$$

we deduce Jensen’s inequality.

If $\mathbb{E}\{\cdot\}$ denotes the mathematical expectation (i.e., the integral) with respect to the probability $P$ then Jensen’s inequality takes the form $\Phi(\mathbb{E}\{\psi\}) \leq \mathbb{E}\{\Phi(\psi)\}$, valid for any measurable function $\Phi$ satisfying $\mathbb{E}\{|\psi|\} < \infty$.

When $\mathbb{E}\{\psi\} = \infty$, probably a conflicting case is when $\mathbb{E}\{\psi\} = \infty$, i.e., when $\mathbb{E}\{\psi^-\} < \infty$ and $\mathbb{E}\{\psi^+\} = \infty$. In this case, consider separately two possible situations, the convex function $\Phi$ is either (a) decreasing (i.e., $\Phi'(t) \leq 0$ for every $t$) or (b) there exists a point $t_0$ where $\Phi(t) \geq \Phi(t_0) + \alpha(t_0)(t - t_0)$, holds for every $t$ in $\mathbb{R}$ with $\alpha(t_0) > 0$ (i.e., $\Phi'_+(t_0) > 0$ for some $t_0$).

Assuming (a), consider the sequence $\psi \equiv \min\{\psi, n\}$ of integrable functions, which converges monotonically (increasing) to $\psi$. Since the monotone (decreasing) sequence $\{\Phi(\psi_n)\}$ is uniformly bounded from below by an integrable function, $\Phi(\psi_n) \leq \Phi(0) + \Phi'_-(0)\psi^-$, we can apply Jensen’s inequality to each $\psi_n$ and letting $n \to \infty$ to show that Jensen’s inequality remain valid for $\psi$, even if $\mathbb{E}\{\psi^+\} = \infty$, provided that $\Phi'_+(t) \leq 0$ for every $t$.

Similarly, assuming (b), take $t = \psi(\omega)$ in the inequality $\Phi(t) \geq \Phi(t_0) + \alpha(t_0)(t - t_0)$, and integrate to get

$$\mathbb{E}\{\Phi(\psi)\} \geq \Phi(t_0) + \alpha(t_0)(\mathbb{E}\{\psi\} - t_0).$$

Hence, because $\mathbb{E}\{\psi\} = \infty$ we deduce that $\mathbb{E}\{\Phi(\psi)\} = \infty$ and Jensen’s inequality holds true anyway.

If $\mathbb{E}\{\psi^-\} = \infty$ and $\mathbb{E}\{\psi^+\} < \infty$ then $\mathbb{E}\{\psi\} = -\infty$ and Jensen’s inequality holds true provided $\mathbb{E}\{\Phi(\psi)\}$ is meaningful, i.e., either $\mathbb{E}\{\Phi(\psi)^+\} < \infty$ or $\mathbb{E}\{\Phi(\psi)^-\} < \infty$. For instance, the convex function $\Phi(\psi) = (\psi + 1)^2$ if $\psi \geq 0$ and $\Phi(\psi) = 2(\psi + 1)$ if $\psi < 0$ could give a random variable $\Phi(\psi) = 2(\psi + 1)1_{\psi<0} + (\psi + 1)^21_{\psi\geq0}$ with $\mathbb{E}\{\Phi(\psi^-)\} = \mathbb{E}\{\Phi(\psi)^+\} = \infty$.

Summing-up, if $\phi$ is a measurable function and $\Phi$ a convex function then Jensen’s inequality holds meaningful, whenever either (i) $\mathbb{E}\{\psi^-\} < \infty$ or (ii) both sides are defined with possible infinite values.

**Exercise 7.20.** Use Proposition 7.15 to prove that (a) if $\{f_k\}$ is a sequence of monotone increasing functions defined on the interval $[a, b]$ such that the numerical series $g(x) = \sum_k f_k(x)$ converges for every $x$ in $[a, b]$ then we have $g'(x) = \sum_k f'_k(x)$, for almost every $x$ in $[a, b]$. Next, show that (b) the derivative of the variation satisfies $V'_f(x) = |f'(x)|$, for almost every $x$ in $(a, b)$.

**Proof.** For instance, we can follow the arguments in Wheeden and Zygmund [119, Section 7.4, pp. 113–115] or Jones [65, Section 16.A, pp. 511–521].

To show (a), consider $g_n = \sum_{k=1}^n f_k$ and $r_n = \sum_{k=n+1}^\infty f_k$, and define the set $F_n$ of all $x$ in $[a, b]$ such that $f_1, \ldots, f_n, g_n, r_n$ and $g$ are differentiable and $g'(x) = g'_n(x) + r'_n(x)$ to check that, in view of Proposition 7.15, the complement
$F'_n$ of $F_n$ is a set of zero Lebesgue measure. Because $r'_n \geq 0$, it follows that $g'_n \geq g_n$ except in $F'_n$, which implies that $\sum_k f'_k \leq g'$ except on the null set $\bigcup_n F'_n$.

We will prove that $r'_n \to 0$ a.e. to complete part (a), actually, because the whole sequence is convergent, we need to use only some suitable subsequence $\{n_i\}$. Indeed, since $r_n(x) \to 0$ for every $x$ in $[a,b]$, choose a sub-sequence such that the series $\sum_i r_{n_i}(x)$ is convergent, for $x = a$ and $x = b$. Hence, Proposition 7.15 implies

$$0 \leq \int_a^b \sum_i r'_{n_i}(x)dx = \sum_i \int_a^b r'_{n_i}(x)dx \leq \sum_i \left[ r_{n_i}(b-) - r_{n_i}(a+) \right] \leq \sum_i \left[ r_{n_i}(b) - r_{n_i}(a) \right] < \infty.$$ 

The function $\sum_i r'_{n_i}$ is integrable over $(a,b)$, and therefore it is finite a.e., which yields $r'_{n_i} \to 0$ almost everywhere in $(a,b)$.

Next, to show (b), let $f$ be a function with bounded variation on $[a,b]$ and let $V(x) = \text{var}(f, [a,x])$ be the variation of $f$ on $[a,x]$, for any $x$ in $[a,b]$. Choose a sequence $\{\pi_k\}$ of partitions such that $0 \leq \text{var}(f, [a,b]) - \text{var}(f, \pi_k) < 2^{-k}$. If 

$$\pi_k = \{a = x_{0,k} < x_{1,k} < \cdots < x_{n_k-1,k} < x_{n_k,k} = b\}$$

then define $f_k(a) = 0,$

$$f_k(x) = \begin{cases} f(x) + f_k(x_{i-1,k}) & \text{if } x_{i-1,k} < x \leq x_{i,k}, \quad f(x_{i-1,k}) \geq f(x_{i,k}), \\ -f(x) + f_k(x_{i-1,k}) & \text{if } x_{i-1,k} < x \leq x_{i,k}, \quad f(x_{i-1,k}) > f(x_{i,k}), \end{cases}$$

for $i = 1, \ldots, n_k$ to deduce, for every $i$ and $k$, that $f_k(x_{i,k}) - f_k(x_{i-1,k}) = |f(x_{i,k}) - f(x_{i-1,k})|$ and $f_k(x') - f_k(x) \leq |f(x') - f(x)| \leq V(x') - V(x)$ if $x_{i-1,k} < x < x' \leq x_{i,k}$. This implies that $V(a) = f_k(a) = 0$, $|f(x, \pi_k) - f_k(b)|$, and $0 \leq V(x_{i,k}) - f_k(x_{i,k}) < 2^{-k}$. Moreover, the function $|V(x) - f_k(x)|$ results monotone increasing on $[a,b]$, and because

$$\sum_k [V(x) - f_k(x)] \leq \sum_k e^{-k} = 1, \quad \forall x \in [a,b],$$

Proposition 7.15 implies that the series $\sum_k [V'(x) - f'_k(x)]$ converges a.e. in $(a,b)$, and therefore $V' = \lim_k f'_k$. Next, consider the set $F$ of points $x$ where $f''(x)$ exists and $x$ does not belong to the partitions $\{\pi_k\}$. From the definition of $f_k$ it is clear that $f'_k(x) = 0$ if $f'(x) = 0$ and $x$ belongs to $F$. Moreover, if $x$ belongs to $F$ and $f'(x) \neq 0$ then the convergence

$$\frac{f(x_{i,k}) - f(x_{i-1,k})}{x_{i,k} - x_{i-1,k}} \to f'(x) \quad \text{as } k \to \infty \quad \text{and} \quad x_{i-1,k} < x \leq x_{i,k},$$

implies that, for every $x$ in $F$, the sequence $f'_k(x) \to |f'(x)|$ as $k \to \infty$, even when $f'(x) \neq 0$. Hence, $f'_k = |f'|$ a.e., and thus, $V' = |f'|$ a.e. in $(a,b)$.

Exercise 7.21. Prove, as much as possible, the above claims 1, \ldots, 7, and then check the references below for full details.
Proof. For instance, we can follow the arguments in Wheeden and Zygmund [119, Section 7.4, pp. 113–115] or Jones [65, Section 16.A, pp. 511–521].

1.- An absolutely continuous function is continuous and has bounded variation.
Indeed, first it is also clear that continuity follows directly from the definition. Next, take \( \varepsilon = 1 \) in the definition of the absolutely continuous function \( f \) on \([a, b]\) to obtain

\[
\sum_{i=1}^{n} |f(b_i) - f(a_i)| < 1 \quad \text{if} \quad \sum_{i=1}^{n} (b_i - a_i) < \delta, \quad b \geq a_{i+1} \geq b_i > a_i \geq a, \quad n \geq 1,
\]

for some \( \delta > 0 \). Therefore, the variation \( \text{var}(f, [a', b']) \leq 1 \), on any subinterval \([a', b'] \subset [a, b]\) with \( b' - a' \leq \delta \). Hence, partition the interval \([a, b]\) into small subintervals of length not greater then \( \delta \) to deduce that \( \text{var}(f, [a, b]) \leq |b - a|/\delta \).

2.- If \( f \) is both absolutely continuous and singular on \([a, b]\) then \( f \) is constant in \([a, b]\).
Indeed, because \( f \) is singular there is a measurable subset \( F \) of \((a, b)\) with Lebesgue measure \( |F| = b - a \) and such that the derivative \( f'(x) = 0 \) for every \( x \) in \( F \). This means that given \( \varepsilon > 0 \) and \( x \) in \( F \) there is \( \bar{h} = \bar{h}(x, \varepsilon) > 0 \) such that \([x, x + h] \subset (a, b)\) and \( |f(x + h) - f(x)| < \varepsilon h \), for every \( 0 < h < \bar{h} \). Invoking Vitali’s covering (Theorem 2.29, Remark 2.36), for any \( \delta > 0 \) there exists a finite number of disjoint intervals \([a_i, b_i]\), \( i = 1, \ldots, n \) such that (1) \( b_i = a_i + h_i \) with \( 0 < h_i < \bar{h}(a_i, \varepsilon) \) and (2) \( \sum_{i=1}^{n} (b_i - a_i) > (b - a) - \delta \). This yields

\[
\sum_{i=1}^{n} |f(b_i) - f(a_i)| < \varepsilon \sum_{i=1}^{n} (b_i - a_i) \leq \varepsilon (b - a).
\]

By ignoring the endpoints \( a \) and \( b \) if necessary, express the complement of \( \sum_{i=1}^{n} [a_i, b_i] \) as \( \sum_{i=0}^{n+1} [a'_i, b'_i] \), with \( a'_i = b_i, \quad b'_i = a_{i+1}, \quad b_0 = a \) and \( a_{n+1} = b \) to check that condition (2) implies \( \sum_{i=0}^{n+1} (a'_i - b'_i) < \delta \). Hence, if \( \delta \) is taken as the number corresponding to \( \varepsilon \) in the definition of absolute continuity for the function \( f \), then \( \sum_{i=1}^{n} |f(b'_i) - f(a'_i)| < \varepsilon \). After using the triangular inequality, this means that \( |f(b) - f(a)| < \varepsilon (b - a) + \varepsilon \), which yields that \( f(b) = f(a) \). Since this is also valid on any subinterval of \([a, b]\), the function \( f \) is constant.

3.- A function \( f \) is absolutely continuous on \([a, b]\) if and only if \( f' \) exists almost everywhere, \( f' \) is integrable on \([a, b]\) and

\[
\int_{a}^{x} f'(x)dx = f(x) - f(a), \quad \forall x \in [a, b].
\]

Indeed, if \( f \) is the integral of its derivative then the absolute continuity of the Lebesgue measure implies that \( f \) is absolute continuous. To check the converse, suppose \( f \) is absolute continuous on \([a, b]\) and define

\[
F(x) = \int_{a}^{x} f'(x)dx, \quad \forall x \in [a, b].
\]
Actually, in view of Proposition 7.15, the expression defining $F$ makes sense for any function $f$ of bounded variation, and $F'(x) = f'(x)$, for any Lebesgue point $x$ in $(a, b)$. Thus, $F - f$ is both, absolute continuous and singular on $[a, b]$, which implies that $F - f$ is constant in $[a, b]$, proving that $F(x) = f(x) - f(a)$ as desired.

4.- If $f$ is a function of bounded variation on $[a, b]$ then the function

$$g(x) = f(x) - \int_a^x f'(y) \, dy, \quad \forall x \in [a, b]$$

is singular. Indeed, $g$ is also a function of bounded variation and $g' = f' - f' = 0$ almost everywhere, i.e., $g$ is a singular function. Moreover, this prove that any function of bounded variation on $[a, b]$ can be expressed as the sum of an absolutely continuous function and a singular function, in a unique manner, up to an additive constant.

5.- If $f$ is an absolutely continuous function then

$$\text{var}(f, [a, b]) = \int_a^b |f'(x)| \, dx \quad \text{and} \quad \text{var}^\pm(f, [a, b]) = \int_a^b (f'(x))^\pm \, dx,$$

are the total, positive and negative variations.

Indeed, if $x \mapsto V(x) = \text{var}(f, [a, b])$ then, because $V(x)$ is a function of bounded variation, part (b) of Exercise 7.20 shows that $V'(x) = |f'(x)|$ almost everywhere. Since $f$ is absolutely continuous, if $\pi = \{a' = x_0 < x_1 < \cdots < x_{n-1} < x_n = b'\}$ denotes a partition of the subinterval $[a', b'] \subset [a, b]$, then

$$\text{var}(V, [a', b']) = \sup_\pi \sum_{i=1}^n |f(x_i) - f(x_{i-1})| \leq$$

$$\leq \sup_\pi \int_{x_{i-1}}^{x_i} |f(x)| \, dx = \int_{a'}^{b'} |f(x)| \, dx,$$

which yields

$$\sum_{i=1}^n |V(b_i) - V(a_i)| \leq \sum_{i=1}^n \int_{a_i}^{b_i} |f(x)| \, dx,$$

for any finite sequence $a \leq a_i < b_i \leq a_{i+1} \leq b$, and in view of the absolutely continuity of the Lebesgue measure, the function $V$ results absolutely continuous. This implies that $V$ is the integral of its derivative $V'$, and the desired equality follows. Finally, if $V_\pm(x) = \text{var}^\pm(f, [a, x])$ then, from the relations $V_\pm = \frac{1}{2}[V(x) \pm f(x) \mp f(a)]$ the argument is completed.

6.- If $f$ is (a) continuous on $[a, b]$, (b) $f'$ exists almost everywhere, (c) $f'$ is integrable, and (d) $f$ maps null sets (i.e., set of Lebesgue measure zero) into null sets, then $f$ is an absolutely continuous function on $[a, b]$.

The converse of this statement could be refereed to as Lusin (N) Property, see [Preliminary]

Menaldi

November 11, 2016
Hence proves the estimate such that

\[ \sum_{i=1}^{n} |I_i| < \delta. \]

Because \( f \) is continuous, the image \( f(I_i) \) is again an interval and

\[ |f(I_i)| \leq V(b_i - V(a_i)), \]

where \( V(x) = \text{var}(f, [a, x]) \) and \( I_i = [a_i, b_i] \). Thus,

\[ f(A) \subset \bigcap_{i} f(I_i) \quad \text{and} \quad |f(A)| \leq \sum_{i} (V(b_i) - V(a_i)). \]

Since \( V \) is also an absolutely continuous function on \([a, b]\), for every \( \varepsilon > 0 \) there exists \( \delta > 0 \) such that

\[ \sum_{i} (V(b_i) - V(a_i)) < \varepsilon, \]

proving that \( f(A) \) has Lebesgue measure zero \( |f(A)| = 0 \).

For a given measurable set \( A \) of \((a, b)\) and any given \( r \geq 0 \), consider the set \( A_r \subset (a, b) \) of all points \( x \) such that \( f'(x) \) exists and \( 0 \leq |f'(x)| < r \). Because \( f' \) exists almost everywhere and \( f' \) is (Lebesgue) measurable, the set \( A_r \) is measurable and, by definition of the derivative \( f' \), for every \( x \) in \( A_r \) there exists and open interval \((a_x, b_x)\) containing \( x \) such that

\[ |f(y) - f(z)| \leq r(z - y), \]

for every \( a_x < y < x < z < b_x \). Invoking Vitali’s covering (Theorem 2.29, Corollary 2.35), there exists a sequence \( \{I_i = [a_i, b_i]\} \) of disjoint intervals such that (1) \( |f(y) - f(z)| \leq r(z - y) \), for every \( a_i < y < z < b_i \), and (2) \( \sum_i |I_i| \leq |A_r| + \varepsilon \) and \( |A_r \setminus \bigcup_i I_i| = 0 \). Hence, if \( I_i = [a_i, b_i] \) then

\[ |f(I_i)| = \sup_{z \in I_i} f(z) - \inf_{y \in I_i} f(y) \leq r(b_i - a_i), \]

which implies \( \sum_i |f(I_i)| \leq r \sum_i (b_i - a_i) \leq r(|A_r| + \varepsilon) \). Thus, if \( B = \bigcup_i I_i \) then

\[ f(A) \subset f(B) \cup f(A_r \setminus B), \quad |A_r \setminus B| = 0 \quad \text{and} \quad |f(B)| \leq r(|A_r| + \varepsilon). \]

Because \( f \) maps null sets into null sets, \( |f(A_r \setminus B)| = 0 \), i.e., \( |f(A_r)| \leq r(|A_r| + \varepsilon) \), which proves the estimate \( |f(A_r)| \leq r|A_r| \).

Now, since \( f' \) is an integrable function, the set \( A_{k,n} = \{ x \in A : (k-1)2^{-n} \leq |f'(x)| < k2^{-n} \} \), \( k, n = 1, \ldots \), is measurable and \( A = \sum_k A_{k,n} \) for every \( n \). Hence

\[ |f(A)| \leq \sum_{k} |f(A_{k,n})| \leq \sum_{k} k2^{-n}m(A_{k,n}) = \sum_{k} (k-1)2^{-n}m(A_{k,n}) + \sum_{k} 2^{-n}m(A_{k,n}) \leq \sum_{k} \int_{A_{k,n}} f'(x)dx + 2^{-n}m(A), \]

and as \( n \to \infty \), for every measurable set \( A \) of \((a, b)\), we deduce

\[ |f(A)| \leq \int_{A} |f'(x)|dx \quad \text{if} \quad m(A) < \infty, \]

which effectively show that \( f \) is absolutely continuous function on \([a, b]\).

7.- If \( f \) is Lipschitz continuous on \([a, b]\), there exists a constant \( M > 0 \) such that \( |f(x) - f(y)| \leq M|x - y| \), for every \( x, y \) in \([a, b]\), then \( f \) is absolutely continuous on \([a, b]\).

Indeed, because

\[ \sum_{i=1}^{n} |f(b_i) - f(a_i)| \leq M \sum_{i=1}^{n} (b_i - a_i) \]
for any finite sequence $a \leq a_i < b_i \leq a_{i+1} \leq b$, the condition on absolutely continuous for $f$ is satisfied by taking $\delta = \varepsilon/M$. \hfill \Box

**Exercise 7.22.** Prove the integration by part formula, namely, if $\alpha$ and $\beta$ are two right-continuous function of finite variation then

$$
\int_{[a,b]} \beta(s) \, d\alpha(s) + \int_{[a,b]} \alpha_-(s) \, d\beta(s) = \alpha(b)\beta(b) - \alpha(a)\beta(a),
$$

for every real numbers $b > a$ and where $\alpha_-(s) = \alpha(s-)$ is the left-continuous version obtained from $\alpha$.

**Proof.** The equality

$$
\beta_1[\alpha_1 - \alpha_0] + \beta_2[\alpha_2 - \alpha_1] + \cdots + \beta_n[\alpha_n - \alpha_{n-1}] =
$$

$$
-\beta_1\alpha_0 + [\beta_2 - \beta_1]\alpha_1 + \cdots + [\beta_n - \beta_{n-1}]\alpha_{n-1} + \beta_n\alpha_n,
$$

applied to a partition $\pi = \{a = s_0 < s_1 < \cdots < s_{n-1} < s_n = b\}$ yields

$$
\sum_{i=1}^{n} \beta(s_i)[\alpha(s_i) - \alpha(s_{i-1})] =
$$

$$
= \beta(s_n)\alpha(s_n) - \beta(s_1)\alpha(s_0) + \sum_{i=1}^{n} [\beta(s_{i+1}) - \beta(s_i)]\alpha(s_i).
$$

Thus, if the integral is taken in the sense of Riemann-Stieltjes then the existence of one of the integrals (say $\beta \, d\alpha$) and the continuity from the right of $\beta$ at $a$ implies the existence of the other integral (say $\alpha \, d\beta$) and the validity of the integral by parts. This is neat and clean if one of the function (e.g., $\beta$) is continuous and the other (e.g., $\alpha$) has bounded variation. Note that if $\beta$ is only continuous then $d\beta$ cannot be interpreted as a signed measure (both variations may be infinite).

However, if both, integrand $\beta$ and the integrator $\alpha$ (in the expression $\beta \, d\alpha$) are discontinuous then some problems appear, for instance, it is know that if $\beta$ and $\alpha$ have a jumps at the same point then the Riemann-Stieltjes may not exist. Thus, we take a look at the interpretation as the Lebesgue-Stieltjes integral, i.e., when both $\alpha$ and $\beta$ are function of bounded (or finite) variation.

Combining the decomposition of a function of bounded variation into the difference of two monotone functions, and Exercise 7.16, we deduce that function of bounded variation can be expressed as the sum of two functions of bounded variation, one continuous from the right and the other continuous from the left. Alternatively, the Lebesgue-Stieltjes measure can be considered for a function of bounded variation, as in Exercise 2.30, without any continuity assumption. Nevertheless, the functions $\alpha$ and $\beta$ are supposed to be continuous from the right. In this case, $d\alpha$ and $d\beta$ represent the Lebesgue-Stieltjes measures generated by them, and any Borel function is measurable, in particular, both integrals $\beta \, d\alpha$ and $\alpha_- \, d\beta$ make sense are are finite. The above integration by parts holds, provided it is true for monotone increasing functions instead of functions of bounded variation.
variation. Thus, define the sequence of functions \( \beta_\pi(s) = \beta(s_i) \) if \( s_{i-1} < s \leq s_i \), \( i = 1, 2, \ldots, n \) and \( \alpha_\pi(s) = \alpha(s_i) \) if \( s_i < s \leq s_{i+1} \), \( i = 0, 1, \ldots, n - 1 \). Taking a sequence of dyadic partitions \( \pi_k = \{ s_i = i2^{-k}(b-a), k = 0, 1, 2^k \} \), because \( \alpha \) and \( \beta \) are increasing, the sequences \( \{ \alpha_\pi \} \) and \( \{ \beta_\pi \} \) are monotone, and \( \alpha_\pi \uparrow \alpha \) and \( \beta_\pi \downarrow \beta \) pointwise everywhere on \((a,b]\). Hence, the monotone convergence can be used on the summation equality

\[
\int_{[a,b]} \beta_k(s) \, d\alpha(s) = \beta(b)\alpha(b) - \beta(a + 2^{-k})\alpha(a) + \int_{[a,b]} \alpha_k(s) \, d\beta(s).
\]

to deduce, as \( k \to \infty \), the above integration by parts.

Note that

\[
\int_{[a,b]} \beta(s) \, d\alpha(s) = \int_{[a,b]} \beta_-(s) \, d\alpha(s) + \sum_{a<s\leq b} \delta\beta(s)\delta\alpha(s)
\]

and in a symmetric way,

\[
\int_{[a,b]} \beta_-(s) \, d\alpha_-(s) = \int_{[a,b]} \beta(s) \, d\alpha_-(s) - \sum_{a<s<b} \delta\beta(s)\delta\alpha(s),
\]

where \( \delta \) is the jump operator, i.e., \( \delta\alpha(s) = \alpha(s) - \alpha_-(s) \), which is equal to zero except for a countable number of \( s \), so that the series are absolutely convergent. The integral with \( \beta_- \) do and \( \beta \, d\alpha_- \) can be interpreted in either the Lebesgue-Stieltjes sense or the Riemann-Stieltjes. It is also clear that the first equality (with \( \beta \, d\alpha \) holds true for any function \( \beta \) with only possible jumps discontinuities (neither of bounded variation nor right-continuous), as long as \( d\alpha \) is interpreted as the Lebesgue-Stieltjes measure generated by the right-continuous monotone (or of bounded variation) function \( \alpha \). Thus, the integration by parts becomes

\[
\int_{[a,b]} \beta_-(s) \, d\alpha(s) + \int_{[a,b]} \alpha_-(s) \, d\beta(s) = \\
= \beta(b)\alpha(b) - \beta(a)\alpha(a) - \sum_{a<s\leq b} \delta\beta(s)\delta\alpha(s),
\]

when interpreted as Riemann-Stieltjes integrals. The correction term is expressed as an absolutely convergent series, which vanishes if \( \alpha \) and \( \beta \) do not jump simultaneously.

In general, because \( \alpha \) is right continuous and has bounded variation, \( \alpha = \alpha^c + \alpha^j \), where \( \alpha^c \) is a continuous function of bounded variation and \( \alpha^j(s) = \sum_{a<t\leq s} \delta\alpha(t) \) is right continuous function representing the sum of jumps, which is given as an absolutely convergent series. The (signed) measure associated with the continuous part \( d\alpha^c \) is a diffuse measure (i.e., without atoms) and the jumps part \( d\alpha^j \) yields a series of Dirac measures, i.e.,

\[
\int_{[a,b]} \beta(s) \, d\alpha(s) = \int_{[a,b]} \beta(s) \, d\alpha^c(s) + \sum_{a<s\leq b} \beta(s)\delta\alpha^j(s),
\]
for any nonnegative Borel measurable function $\beta$.

For instance, if $\alpha$ and $\beta$ are jumps functions, i.e.,

$$\alpha(s) = \sum_i \alpha_i \mathbb{1}_{a_i \leq s}, \quad \beta(s) = \sum_i \beta_i \mathbb{1}_{b_i \leq s},$$

with $a_i, b_i$ in $[a, b]$ and $\alpha(a) = \beta(a) = 0$, or if $\alpha$ is a continuous function and $\beta$ a single jump function $\beta(s) = \mathbb{1}_{[a', b]}(s)$, $a < b' \leq b$, then the elementary equalities

$$\sum_j \sum_i \beta_i \mathbb{1}_{b_i \leq s_j} \alpha_j + \sum_j \sum_i \alpha_i \mathbb{1}_{a_i < b_j} \beta_j = \alpha(b)\beta(b),$$

$$\int_{[a,b]} \beta(s) \, d\gamma(s) + \int_{[a,b]} \gamma(s) \, d\beta(s) = [\gamma(b) - \gamma(a')] + \gamma(a') = \gamma(b),$$

represents the integration by part. Actually, this argument gives a constructive prove of the equality

$$\int_{[a,b]} \beta(s) \, d\alpha^c(s) + \sum_{a<s \leq b} \beta(s) \delta \alpha^j(s) + \int_{[a,b]} \alpha_-(s) \, d\beta(s) = \alpha(b)\beta(b) - \alpha(a)\beta(a),$$

for any right-continuous function $\beta$ of bounded variation, any continuous function $\alpha^c$ (not necessarily of bounded variation), and any right-continuous jump function $\alpha^j$ of bounded variation. In this case, if the function $\alpha$ is only continuous then the integral with $\beta \alpha^c$ should be interpreted in the Riemann-Stieltjes sense, while the other integral with $\alpha_-$ $d\beta$ could be interpreted in either sense. Also remark that the series of jumps $\alpha^j(s) = \sum_{0<t \leq s} \delta \alpha^j(t)$ defining the jump function is absolutely convergent.

**Exercise 7.23.** Let $\alpha$ be a right-continuous function of finite variation on $[0, +\infty]$ and let $a$ and $b$ be two non-decreasing right-continuous functions on $[0, +\infty]$. If

$$|\alpha(t) - \alpha(s)| \leq \sqrt{a(t) - a(s)} \sqrt{b(t) - b(s)}, \quad \forall t > s \geq 0,$$

then for any Borel functions $f$ and $g$ on $\mathbb{R}$ we have

$$\int_{[0, +\infty]} |f(s)| \, g(s) \, |d\alpha(s)| \leq \left( \int_{[0, +\infty]} |f(s)|^2 \, da(s) \right)^{1/2} \left( \int_{[0, +\infty]} |g(s)|^2 \, db(s) \right)^{1/2},$$

where $|d\alpha(t)| = \text{var}(\alpha, [0, t])$ is the variation function associated with $\alpha$.

**Proof.** Remark that the assumption on the functions $a$, $b$ and $\alpha$ implies the the following estimate on the derivative $|\alpha'(t)| \leq \sqrt{a'(t)} \sqrt{b'(t)}$, for almost everywhere point $t$. However, because the functions are not necessarily absolutely
continuous, the derivative does not reproduce the Lebesgue-Stieltjes measures \( da, db \) and \( da, \) and the desired estimate cannot be obtained in this way. This estimate is very useful in Probability (when studying martingales), and

\[
t \mapsto r^2a(t) + 2ra(t) + b(t) \text{ is non decreasing } \forall r \in \mathbb{R},
\]

is an equivalent statement for the bounds on the functions \( a, b, \) and \( \alpha. \)

For any rational \( r, \) define the right-continuous function \( v(t) = r^2a(t) + 2ra(t) + b(t) \) and consider the quadratic form \( r \mapsto [v(t) - v(s)], \) \( t > 0. \) In view of the assumption on \( a \) and \( b \) on \( \alpha, \) the discriminant of the quadratic form is non-negative, which implies that the function \( t \mapsto v \) is non-decreasing. Therefore, if \( d\mu \) is the Lebesgue-Stieltjes measure associated with the non-decreasing right-continuous function \( \mu(t) = \text{var}(\alpha, [0, t]) + a(t) + b(t), \) by means of Radon-Nikodym Theorem 6.3, the derivatives \( \alpha' = da/d\mu, a' = da/d\mu \) satisfy \( r^2\alpha' + 2ra' + b' \geq 0, \) except in a \( d\mu \)-null set, which may depend on \( r. \)

Because, \( r \) ranges in the rational set (a countable set), this inequality actually holds everywhere, except in a \( \mu \)-null set independent of \( r. \) Again, looking at the quadratic form \( r \mapsto r^2\alpha' + 2ra' + b', \) we deduce that \( |\alpha'| \leq \sqrt{\alpha^2b'}, \) except in a \( d\mu \)-null set.

Now, the equality

\[
\int_{[0, +\infty]} |f(s)||g(s)||d\alpha(s)| = \int_{[0, +\infty]} |f(s)||g(s)||\alpha'(s)||d\mu(s)
\]

and Hölder inequality (for the case \( p = q = 2, \) i.e., Cauchy inequality) imply

\[
\int_{[0, +\infty]} |f(s)||g(s)||d\alpha(s)| \leq \left( \int_{[0, +\infty]} f(s)^2 \alpha'(s) d\mu(s) \right)^{1/2} \left( \int_{[0, +\infty]} g(s)^2 b'(s) d\mu(s) \right)^{1/2},
\]

and the Kunita-Watanabe inequality follows. \( \square \)

### (7.5) Lebesgue Spaces

**Exercise 7.24.** With the Lebesgue measure on \([0, +\infty[\), consider the Haar-type functions \( f_i(s) = \mathbb{1}_{2i-1 < s \leq 2i} - \mathbb{1}_{2(i-1) < s \leq 2i-1} \) and \( f_{i,n}(s) = 2^{-n/2} f_i(s2^n), \) for \( i = 1, \ldots, 4^n, \) \( n \geq 0. \) First, show that if \( n \geq m \) then \( f_{i,m}f_{j,m} = 0 \) except for \( i \) within \((j-1)2^{m-1} + 1 \) and \( j2^{m-1}. \) Secondly, show that \( \{f_{i,n}\} \) is an orthonormal system in \( L^2([0, \infty[). \) Thirdly, prove that \( \{f_{i,n}\} \) can be completed to be a basis by adding the functions \( \tilde{f}_i(s) = \tilde{f}_{i,0}(s) = \mathbb{1}_{i-1 < s \leq i}, \) for \( i = 1, 2, \ldots, \) for instance, show that \( 1/2 \{\tilde{f}_{i,0}\} \pm 1/2 \{f_{i,0}\} \) yields \( \{\tilde{f}_{i,1}(s) = \mathbb{1}_{i-1 < s < i}\}, \) and \( 1/2 \{\tilde{f}_{i,1}\} \pm 1/2 \{f_{i,1}\} \) yields \( \{\tilde{f}_{i,2}(s) = \mathbb{1}_{i-1 < 4s < i-1}\} \) and so on. Finally, discuss a similar construction on \( L^2([a, b[), \) for \( b > a. \)

**Proof.** First, by construction, the support of the functions \( f_i \) are disjoint if \( i \neq j. \) The function \( f_i(s2^n) = f_{i,n}(s) \neq 0 \) only when \( 2s2^n \) belongs to \((2i-1, 2i+1], \)
i.e., \( s \) in \( I_{i,n} = (i - 1/2)2^{-n}, (i + 1/2)2^{-n} \). Thus, if \( n \geq m \) then \( I_{i,n} \cap I_{j,m} \neq \emptyset \) only when \( i \) is within \((j - 1)2^{n-m} + 1 \) and \( j2^{n-m} \).

Secondly, to check that \( \{f_{i,n}\} \) forms an orthonormal system in \( L^2(0, \infty) \), on the case \( n \geq m \) and \( i \) is within \((j - 1)2^{n-m} + 1 \) and \( j2^{n-m} \) should be considered. The change of variable \( s \mapsto 2^m s \) reduces the problem to the case \( f_j \) and \( f_{i,n} \), and by checking the graph of the functions, it follows the desired conclusion.

Thirdly, it is clear that the functions \( \tilde{f}_i \) are orthogonal to each \( f_{i,n} \), so that \( F = \{f_{i,n} : i = 1, \ldots, 4^n, n \geq 0\} \cup \{\tilde{f}_i\} \) is an orthonormal system. The fact that \( \tilde{f}_{i,0}(s)/2 + f_{i,0}(s)/2 = 1_{2i-1 < 2s \leq 2i} \) and \( \tilde{f}_{i,0}(s)/2 - f_{i,0}(s)/2 = 1_{2(i-1) < 2s \leq 2i-1} \) proves that the function \( \tilde{f}_{i,1}(s) = 1_{1 < 2s \leq i} \) belongs to subspace spanned by the orthonormal system \( F \), and, by induction, so the functions \( \tilde{f}_{i,n}(s) = 1_{i-1 < 2^n s \leq i} \). Since any piece-constant function with compact support can be approximate by a linear combination of the function \( \tilde{f}_{i,n} \), the orthonormal system \( F \) is complete.

Finally, for the space \( L^2([a, b]) \), the functions \( f_{i,n}(s) = f_i(2^n(b - a)s), i = 1, \ldots, 2^{n-1}, n \geq 0 \) yields a orthonormal system, which can be completed by adding only the function \( f(s) = 1 \), for every \( s \) in \( (a, b) \).

\( \square \)

### (7.6) Trigonometric Series

**Exercise 7.25.** The Dirichlet kernel of index \( n \) is the function \( D_n(x) = 1/2 + \cos x + \cdots + \cos nx \) Prove (1) that \( D_n(x) = n + 1/2 \) if \( x = 0, \pm 2\pi, \pm 4\pi, \ldots \), and \( D_n(x) = \left[ \sin \left( (n + 1/2)x \right) \right]/\left[ 2\sin(x/2) \right] \) otherwise, and (2)

\[
\int_{-\pi}^{\pi} D_n(x) dx = \pi, \forall n = 0, 1, 2, \ldots,
\]

and (3) if \( f \) belongs to \( L^1(-\pi, \pi) \) then

\[
S_n(f, x) = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x \pm y)D_n(y)dy, \forall n = 0, 1, 2, \ldots,
\]

where either sign + or − can be chosen, and \( S_n(f, x) \) is the Fourier sum associated with \( f \), i.e., \( S_n(f, x) = a_0[f]/2 + \sum_{k=1}^{n} (a_k[f] \cos kx + b_k[f] \sin kx) \).

**Proof.** For instance, the interested reader may take a look at the book by Kirkwood [70, Chapter 9, pp 238-265].

Assertion (1) is a direct consequence of adding (from \( k = 1 \) to \( k = n \)) on the trigonometric formula

\[
\sin ((k + 1/2)x) - \sin ((k - 1/2)x) = \cos kx \sin(x/2),
\]

and assertion (2) follows from remarking that the integral of \( \cos kx \) on any period is zero.

Next, use the trigonometric formula \( \cos (\alpha - \beta) = \cos \alpha \cos \beta + \sin \alpha \sin \beta \) to write

\[
D_n(x - y) = \frac{1}{2} + \sum_{k=1}^{n} \left[ \cos kx \cos ky + \sin kx \sin ky \right],
\]
which yields
\[
\frac{1}{\pi} \int_{-\pi}^{\pi} f(y) D_n(x-y)dy = S_n(f, x).
\]

The change of variable \( z = x - y \) and the symmetry of the kernel, i.e., \( D_n(x) = D_n(-x) \), complete the argument. \( \square \)

(7.7) Complements on Lebesgue Spaces

(7.7) Weak Convergence

(7.7) Totally Bounded Sets

Exercise 7.26. If \( A \) is a totally bounded set of a normed space \( (X, \| \cdot \|) \) then prove that the convex hull (or convex envelope) \( \text{co}(A) \) of \( A \) (i.e., the smallest convex set containing \( A \)) is also totally bounded. In particular, the closed convex hull of a compact set of a Banach space is also compact. Hint: Use the following argument (1) if \( F \subset X \) is a finite set then the convex hull \( \text{co}(F) \) of \( F \) is a totally bounded set. Next, let \( A \) be a totally bounded subset of \( X \) and let \( B_1 \) be an open balls containing the origin. By using the previous result, (2) find a finite set \( F \) such that \( A \subset F + B_1 \) and deduce that \( \text{co}(A) \) lies inside \( K + B_1 \) for some totally bounded set \( K \). Now, take any two open balls \( B_1 \) and \( B \) containing the origin and satisfying \( B_1 + B_1 \subset B \). Finally, because \( K \) is totally bounded, (3) find another finite \( E \) such that \( \text{co}(A) \subset (E + B_1) + B_1 \subset E + B \), and deduce that \( \text{co}(A) \) is indeed totally bounded.

Proof. (1) First, note that \( x \) belongs to \( \text{co}(F) \) if and only if \( x \) is a convex combination of points in \( F \), i.e., if and only if there exist \( n \geq 2 \), \( a_i \) in \( [0, 1] \), and points \( x_i \), for \( i = 1, \ldots, n \) such that \( \sum_{i=1}^{n} a_i = 1 \) and \( x = \sum_{i=1}^{n} a_i x_i \). Thus, if \( F \) is a finite set, say \( F = \{x_1, \ldots, x_n\} \), then consider the dyadic approximation in \([0,1]\), i.e., \( D_k = \{ j2^{-k} : j = 0, 1, \ldots, 2^k \} \), \( k = 1, 2, \ldots \), with \( D_k(a) = \max \{ d \in D_k : d \leq a \} \), and the finite set \( F_k = \{ y = \sum_{i=1}^{n} d_i x_i : d_i \in D_k, 1 = \sum_{i=1}^{n} d_i \} \subset \text{co}(F) \). Hence, for any point \( x \) in \( \text{co}(F) \) there exist \( a_i \) in \([0,1]\) such that \( \sum_{i=1}^{n} a_i = 1 \) and \( x = \sum_{i=1}^{n} a_i x_i \). For each \( a_i \) define \( d_i = D_k(a_i) \) for \( i = 1, \ldots, n-1 \) and \( d_n = 1 - d_1 - \cdots - d_{n-1} \) to deduce that \( y = \sum_{i=1}^{n} d_i x_i \) belongs to \( F_k \) and

\[
\|x - y\| \leq \sum_{i=1}^{n} |a_i - d_i| \|x_i\| \leq [2(n-1)]2^{-k} \left( \max_i \|x_i\| \right).
\]

Therefore, given any \( \varepsilon > 0 \) there exists \( k \) such that \([2(n-1)]2^{-k} < \varepsilon\), which implies that any point in \( \text{co}(F) \) is within a distance less than \( \varepsilon \) from the finite set \( F_k \), i.e., \( \text{co}(F) \) is totally bounded.

(2) Since the open \( B_1 \) contains the origin, there exists \( \varepsilon > 0 \) such that \( \|x\| \leq \varepsilon \) implies \( x \) is in \( B_1 \), and because \( A \) is totally bounded there exits a finite set \( F \) such that every point in \( A \) lies within a distance less than \( \varepsilon \) from \( F \). This
yields \( A \subset F + B_1 \). The ball \( B_1 \) is convex, and in view of (1), the convex hull \( K = \text{co}(F) \) is totally bounded, therefore \( \text{co}(A) \subset K + B_1 \).

(3) Because \( K \) is totally bounded and \( B_1 \) is a ball containing the origin, invoke the property (1) to find another finite set \( E \) such that \( K \subset E + B_1 \). Hence, the inclusion (2) implies \( \text{co}(A) \subset (E + B_1) + B_1 \subset E + B_1. \) Since the ball \( B \) is also arbitrary, this shows that \( \text{co}(A) \) is totally bounded.

Finally, remark that in a Banach space (i.e., a complete normed space) a set is totally bounded if and only if it is pre-compact. Recall that closure and the interior of a convex set is convex, and that the convex hull of an open set is open. However, the convex hull of a closed set is not necessarily closed. In a finite-dimensional space, the convex hull of a compact set is compact.

**Exercise 7.27.** Banach-Saks Theorem states that if \( \{f_n\} \) is a weakly convergence sequence to \( f \) in \( L^p(\Omega, \mathcal{F}, \mu), 1 \leq p < \infty \) then there exists a subsequence \( \{f_{n_k}\} \) such that the arithmetic means \( g_k = (f_{n_1} + \cdots + f_{n_k})/k \) strongly converges to \( f \), i.e., \( \|g_k - f\|_p \to 0 \). Prove this result for a Hilbert space \( H \) with scalar product \((\cdot, \cdot)\) and norm \( \|\cdot\| \), in particular for \( p = 2 \). Hint: First reduce the problem to the case where \( f = 0 \), and \( \|f_n\| \leq 1 \) for every \( n \geq 1 \). Next, construct a subsequence satisfying \( |(f_{n_i}, f_{n_{k+1}})| \leq 1/k \), for every \( i = 1, \ldots, k \), and deduce that \( \|g_k\|^2 \leq 3/k. \) see Riesz and Nagy [94, Section 38, pp. 80–81].

**Proof.** First, if \( f_n \to f \) weakly then \( \|f_n\|_p \leq C \), for every \( n \), and then \( f_n - f \to 0 \) weakly. Hence, the sequence of functions \( f_n' = (f_n - f)/(2C) \) converges weakly to 0 and \( \|f_n'\|_p \leq 1 \).

Now, let \( \{f_n\} \) be a sequence weakly convergence sequence to 0 in a Hilbert space \( H \) satisfying \( \|f_n\| \leq 1 \). Beginning with \( f_{n_1} = f_1 \), note that \( (f, f_n) \to 0 \), as \( n \to \infty \), for every \( f \) in \( H \), to choose \( f_{n_2} \) such that \( |(f_{n_1}, f_{n_2})| \leq 1/k \), and next, by induction, to choose \( f_{n_k} \) such that \( |(f_{n_i}, f_{n_{k+1}})| \leq 1/k \), for every \( i = 1, \ldots, k \). Define \( g_k = (f_{n_1} + \cdots + f_{n_k})/k \) to check that

\[
\|g_k\|^2 \leq \frac{2}{k^2} \sum_{i<j}^k |(f_{n_i}, f_{n_j})| + \frac{1}{k^2} \sum_{i=1}^k |(f_{n_i}, f_{n_i})|,
\]

and because the first sum has \( k(k - 1) \) terms, all bounded by \( 1/(k - 1) \) and the second sum has \( k \) terms all bounded by \( 1 \), deduce that \( \|g_k\|^2 \leq 3/k. \) This shows that the sequence \( g_k \) strongly converges to 0. \( \square \)
Notation

Some Common Uses:

\( \mathbb{N}, \mathbb{Q}, \mathbb{R}, \mathbb{C} \): natural, rational, real and complex numbers.

\( i, \mathbb{R}(\cdot), I \): imaginary unit, the real part of complex number and the identity (or inclusion) mapping or operator.

\( 1_A \): usually denotes the characteristic function of a set \( A \), i.e., \( 1_A(x) = 1 \) if \( x \) belongs to \( A \) and \( 1_A(x) = 0 \) otherwise. Sometimes the set \( A \) is given as a condition on a function \( \tau \), e.g., \( \tau < t \), in this case \( 1_{\tau < t}(\omega) = 1 \) if \( \tau(\omega) < t \) and \( 1_{\tau < t}(\omega) = 0 \) otherwise.

\( \delta \): most of the times this is the \( \delta \) function or Dirac measure. Sometimes one write \( \delta_x(dy) \) to indicate the integration variable \( y \) and the mass concentrated at the point \( x \).

\( d\mu, \mu(dx), d\mu(x) \): together with the integration sign, usually these expressions denote integration with respect to the measure \( \mu \). Most of the times \( dx \) means integration respect to the Lebesgue measure in the variable \( x \), as understood from the context.

\( E^T, \mathcal{B}(E^T), \mathcal{B}^T(E) \): for \( E \) a Hausdorff topological (usually a separable complete metric, i.e., Polish) space and \( T \) a set of indexes, usually this denotes the product topology, i.e., \( E^T \) is the space of all function from \( T \) into \( E \) and if \( T \) is countable then \( E^T \) is the space of all sequences of elements in \( E \). As expected, \( \mathcal{B}(E^T) \) is the \( \sigma \)-algebra of \( E^T \) generated by the product topology in \( E^T \), but \( \mathcal{B}^T(E) \) is the product \( \sigma \)-algebra of \( \mathcal{B}(E) \) or generated by the so-called cylinder sets. In general \( \mathcal{B}^T(E) \subset \mathcal{B}(E^T) \) and the inclusion may be strict.

Most Commonly Used Function Spaces:

\( C(X) \): for \( X \) a Hausdorff topological (usually a separable complete metric, i.e., Polish) space, this is the space of real-valued (or complex-valued) continuous functions on \( X \). If \( X \) is a compact space then this space endowed with
sup-norm is a separable Banach (complete normed vector) space. Sometimes this space may be denoted by $C^0(X)$, $C(X, \mathbb{R})$ or $C(X, \mathbb{C})$ depending on what is to be emphasized.

$C_b(X)$: for $X$ a Hausdorff topological (usually a complete separable metric, i.e., Polish) space, this is the Banach space of real-valued (or complex-valued) continuous and bounded functions on $X$, with the sup-norm.

$C_0(X)$: for $X$ a locally compact (but not compact) Hausdorff topological (usually a complete separable metric, i.e., Polish) space, this is the separable Banach space of real-valued (or complex-valued) continuous functions vanishing at infinity on $X$, i.e., a continuous function $f$ belongs to $C_0(X)$ if for every $\varepsilon > 0$ there exists a compact subset $K = K_\varepsilon$ of $X$ such that $|f(x)| \leq \varepsilon$ for every $x$ in $X \setminus K$. This is a proper subspace of $C_b(X)$ with the sup-norm.

$C_0(X)$: for $X$ a compact subset of a locally compact Hausdorff topological (usually a Polish) space, this is the separable Banach space of real-valued (or complex-valued) continuous functions vanishing on the boundary of $X$, with the sup-norm. In particular, if $X = X_0 \cup \{\infty\}$ is the one-point compactification of $X_0$ then the boundary of $X$ is only $\{\infty\}$ and $C_0(X) = C_0(X_0)$ via the zero-extension identification.

$C_0(X)$, $C^0_0(X)$: for $X$ a proper open subset of a locally compact Hausdorff topological (usually a Polish) space, this is the separable Fréchet (complete locally convex vector) space of real-valued (or complex-valued) continuous functions with a compact support $X$, with the inductive topology of uniformly convergence on compact subset of $X$. When necessary, this Fréchet space may be denoted by $C^0_0(X)$ to stress the difference with the Banach space $C_0(X)$, when $X$ is also regarded as a locally compact Hausdorff topological. Usually, the context determines whether the symbol represents the Fréchet or the Banach space.

$C^k_b(E)$, $C^k_0(E)$: for $E$ a domain in the Euclidean space $\mathbb{R}^d$ (i.e., the closure of the interior of $E$ is equal to the closure of $E$) and $k$ a nonnegative integer, this is the subspace of either $C_b(E)$ or $C^0_0(E)$ of functions $f$ such that all derivatives up to the order $k$ belong to either $C_b(E)$ or $C^0_0(E)$, with the natural norm or semi-norms. For instance, if $E$ is open then $C^k_0(E)$ is a separable Fréchet space with the inductive topology of uniformly convergence (of the function and all derivatives up to the order $k$ included) on compact subset of $E$. If $E$ is closed then $C^k_b(E)$ is the separable Banach space with the sup-norm for the function and all derivatives up to the order $k$ included. Clearly, this is extended to the case $k = \infty$.

$B(X)$: for $X$ a Hausdorff topological (mainly a Polish) space, this is the Banach space of real-valued (or complex-valued) Borel measurable and bounded functions on $X$, with the sup-norm. Note that $\mathcal{B}(X)$ denotes the $\sigma$-algebra of Borel subsets of $X$, i.e., the smaller $\sigma$-algebra containing all open sets in
$X$, e.g., $B(\mathbb{R}^d), B(\mathbb{R}^d)$, or $B(E), B(E)$ for a Borel subset $E$ of $d$-dimensional Euclidean space $\mathbb{R}^d$.

$L^p(X, m)$: for $(X, \mathcal{X}, m)$ a complete $\sigma$-finite measure space and $1 \leq p < \infty$, this is the separable Banach space of real-valued (or complex-valued) $\mathcal{X}$-measurable (class) functions $f$ on $X$ such that $|f|^p$ is $m$-integrable, with the natural $p$-norm. If $p = 2$ this is also a Hilbert space. Usually, $X$ is also a locally compact Polish space and $m$ is a Radon measure, i.e., finite on compact sets. Moreover $L^\infty(X, m)$ is the space of all (class of) $m$-essentially bounded (i.e., bounded except in a set of zero $m$-measure) with essential-sup norm.
Bibliography


Index

\(\mu\)-equicontinuous, 187
\(\mu\)-uniformly integrable, 187
  of order \(p\), 198
\(\mu^*\)-measurable, 37
\(\sigma\)-additive or countably additive, 33
\(p\)-uniformly integrable, 198

absolutely continuous, 228
additive or finitely additive, 33
almost everywhere, ix
  measurable, 35
  statement, 35, 115
almost surely, ix, 35
approx. by smooth functions, 211
atomic or discrete, 91
atoms, 19, 30
axiom of choice, 4, 66

Banach spaces, 7
  \(L^p\)-duality, 205
  reflexive, 206
Beppo Levi theorem, 96
Besicovitch covering, 70
Bessel inequality, 241
Borel measure
  definition, 54
  inner, 79
  outer, regular, 71
  regular, outer, 54
  regular, Thm, 73
  tight, 87
Borel space, 22
bounded set, 7
bounded variation, 227

canonical sample space, 26
Cantor set, 60, 288

Caratheodory
  construction of measures, 38, 48
  criterium of measurability, 55
  extension of measures, 41
cardinal, count, 1
Cauchy sequence
  in mean, 112
  in measure, 108
  in probability, 108
change of variable
  in RS-integrals, 140
  smooth, 133
complete space, 6
continuity of the translation, 212
continuum, hypothesis, 2
convergence
  \(\lim\inf\) or Fatou, 97
  dominate or Lebesgue, 97
  in mean, 192
  in measure, 109
  in norm, 197
  monotone or Beppo Levi, 96
  strong, 246
  weak, 246
  weak*, 247
converges weakly, 246
convolution in \(\mathbb{R}^d\), 213
countability
  first axiom of, 5
  second axiom of, 6, 79
countable generated, 17
cumulative distribution, 59
cylinder set, 87
cylindrical sets, 20

Daniell-Riesz approach, 99, 101
definition
\(\lambda\)-class, 17
\(\mu\)-equicontinuous, 187
\(\sigma\)-ring, \(\sigma\)-algebra, 15
(quasi-)integrable, 95
additive measure, 34
additive probability, 34
Borel outer measure, 71
classes of sets, 13
compact class, 82
inner or interior measure, 48
integral, 95
lattice class, 19
locally integrable, 213
measure, 34
measure space, 34
monotone class, 16
outer or exterior measure, 37
probability measure, 34
probability space, 34
Radon measure, 74
regular Borel outer meas., 71
separable class, 17
separable measurable space, 19
signed measure, 171
simple function, 27
summable, 95
transition measure, 103
transition probability, 103
uniformly integrable, 187
denumerable, countable, 2
diagonal argument, 248
Dirac measure, 34
discrete measure, viii
dual norm, 125
Dunford-Pettis criterium, 204, 249
Dynkin class, 18, 254

Egorov theorem, 111
essential supremum, 176, 177

Fatou thm or lemma, 97
first axiom of countability, 5
first-countable, 5
Frechet-Kolmogorov thm, 251
Fubini-Tonelli theorem, 106
functionals on \(C_0(X, \mathbb{R}^m)\), 207
functions
almost measurable, 116
equivalence class of, 115
essentially bounded, 119
locally integrable, 213
Hölder inequality, 122
Haar measure, 234
Hahn-Jordan
decomposition, 171
Hamel basis, 10
Hausdorff
dissecting system, 85
maximal principle, 3
measure, 54
spaces, 5
Hausdorff dimension, 156
Hausdorff invariance, 160
Hausdorff measure, 155
Hilbert space, 7, 183
inner product, 7
integrable, viii
Jacobian, 161
Jensen’s inequality, 231
Jordan content, 147
Lebesgue
decomposition, 175
invariant under translations, 63
measure locally compact, 57
non-measurable set, 60, 289
points, 219, 221
primitive, 233
surface measure, 134, 153, 157
Lipschitz function, 162
locally, 213
locally compact, 7
Lusin
space, 90, 308
theorem, 113
maximal element, 3
maximal function, 220
measurable
essential sup or inf, 176
functions, 23
simple-function, 27
space, 19
step-function, 98
measure
$\mu_*$-measurable set, 48
additive, 45
Caratheodory's constr., 37
complete, 36
definition, 34
disintegration, 107
inner regular, 77
product of, 104
semi-finite, 47, 49, 55
support, 77
universal completion, 36
Minkowski inequality, 123, 124
mollification, 212
monotone class, 16
argument of, 17
monotone continuity
at $\emptyset$, 34
from above, 34
from below, 34
monotone convergence thm, 96
negligible, viii
negligible set, 35
norm, 7
normed space, 7
ordinal, 4
orthogonal
basis, 237
complement, 183
elements, 183
projection, 183, 184
set, 237
orthonormal basis, 237
orthonormal set, 237
partition of the unity, 152, 217, 218
point of density, 223, 364
point of dispersion, 223, 364
polar decomposition, 161
Polish space, 6, 22
pre-compact in $L^p$, 249
product $\sigma$-algebra, 20
product decomposition, 131
product measure, 104
quasi-integrable, viii, 95
Rademacher's theorem, 169
Radon measure, 74, 80
Radon outer measure, 80
Radon-Nikodym theorem, 173
relation, RS and LS integrals, 137
Riemann integrable, 129, 130
Riemann integral, 127
Riemann integral definition, 129
Riemann-Lebesgue lemma, 243
Riemann-Stieltjes integral, 135
Riesz representation, 185
scalar product, 7
second axiom of countability, 6, 79
second-countable, 6, 79
sections, 25
semi-continuous Functions, 128
separability, 236
separable, 6, 17, 19, 20
sequentially compact, 6
set of atoms, viii
set-function
content or additive measure, 45
pre-measure, 45
signed measure
$\sigma$-finite, 171
absolutely continuous, 173
definition, 171
finite, 171
Hahn-Jordan thm, 171
Lebesgue thm, 175
negative set, 172
positive set, 172
Radon-Nikodym thm, 173
singular, 173
simple function, 27
singular function, 228
space of measures, 206
span or linear span, 237
spherical coordinates, 131
step function, 127
stochastic convergence, 119
Stone-Weierstrauss thm, 10
strong convergence, 246
surface measure, 134, 153, 157
Tietze’s extension, 8
topology, 20
totally bounded, 7
transfinite induction principle, 5
trigonometric series, 239
truncation, 212
uniform absolutely continuous, 191
uniform integrability, 187, 196
upper bound, 3
Urysohn’s lemma, 8
variation operator, 227
vector topological space, 7
Vitali theorem, 192
Vitali’s covering, 64, 65
Vitali-Hahn-Saks Theorem, 245
weak convergence, 246
weak-$\mathcal{L}^1$, 221
weakest topology, 20
Weierstrauss approximation, 9
Young inequality, 234
Zermelo’s axiom, 3
Zorn’s lemma, 3, 66