STUDENT MATHEMATICAL LIBRARY
Volume 48

# A (Terse) Introduction to Lebesgue Integration

John Franks

AMS
AMERICAN MATHEMATICAL SOCIETY

# A (Terse) Introduction to Lebesgue Integration

STUDENT MATHEMATICAL LIBRARY
Volume 48

# A (Terse) Introduction to Lebesgue Integration

John Franks

American Mathematical Society
Providence, Rhode Island

2000 *Mathematics Subject Classification*. Primary 28A20, 28A25, 42B05.

The images on the cover are representations of the ergodic transformations in Chapter 7. The figure with the implied cardioid traces iterates of the squaring map on the unit circle. The "spirograph" figures trace iterates of an irrational rotation. The arc of + signs consists of iterates of an irrational rotation. I am grateful to Edward Dunne for providing the figures.

For additional information and updates on this book, visit
**www.ams.org/bookpages/stml-48**

**Library of Congress Cataloging-in-Publication Data**

Franks, John M., 1943–
A (terse) introduction to Lebesgue integration / John Franks.
p. cm. – (Student mathematical library ; v. 48)
Includes bibliographical references and index.
ISBN 978-0-8218-4862-3 (alk. paper)
1. Lebesgue integral. I. Title. II. Title: Introduction to Lebesgue integration.

QA312.F698 2009
515′.43–dc22 2009005870



10 9 8 7 6 5 4 3 2 1 14 13 12 11 10 09

To my family: Judy, Josh, Mark and Alex

# Contents

# Preface

This text is intended to provide a student's first encounter with the concepts of measure theory and functional analysis. Its structure and content were greatly influenced by my belief that good pedagogy dictates introducing difficult concepts in their simplest and most concrete forms. For example, the study of abstract metric spaces should come after the study of the metric and topological properties of $\mathbb{R}^n$. Multidimensional calculus should not be introduced in Banach spaces even if the proofs are identical to the proofs for $\mathbb{R}^n$. And a course in linear algebra should precede the study of abstract algebra.

Hence, despite the use of the word "terse" in the title, this text might also have been called "A (Gentle) Introduction to Lebesgue Integration". It is terse in the sense that it treats only a subset of those concepts typically found in a substantive graduate level analysis course. I have emphasized the motivation of these concepts and attempted to treat them in their simplest and most concrete form. In particular, little mention is made of general measures other than Lebesgue until the final chapter. Indeed, we restrict our attention to Lebesgue measure on $\mathbb{R}$ and no treatment of measures on $\mathbb{R}^n$ for $n > 1$ is given. The emphasis is on real-valued functions but complex functions are considered in the chapter on Fourier series and in the final chapter on ergodic transformations. I consider the narrow selection of topics to be an approach at one end of a spectrum whose other end is represented, for example, by the excellent graduate text [**Ru**] by Rudin which introduces Lebesgue measure as a corollary of the Riesz representation theorem. That is a sophisticated and elegant approach, but, in my opinion, not one which is suited to a student's first encounter with Lebesgue integration.

In this text the less elegant, and more technical, classical construction of Lebesgue measure due to Caratheodory is presented, but is relegated to an appendix. The intent is to introduce the Lebesgue integral as a tool. The hope is to present it in a quick and intuitive way, and then go on to investigate the standard convergence theorems and a brief introduction to the Hilbert space of $L^2$ functions on the interval.

This text should provide a good basis for a one semester course at the advanced undergraduate level. It might also be appropriate for the beginning part of a graduate level course if Appendices B and C are covered. It could also serve well as a text for graduate level study in a discipline other than mathematics which has serious mathematical prerequisites.

The text presupposes a background which a student should possess after a standard undergraduate course in real analysis. It is terse in the sense that the density of definition-theorem-proof content is quite high. There is little hand holding and not a great number of examples. Proofs are complete but sometimes tersely written. On the other hand, some effort is made to motivate the definitions and concepts.

Chapter 1 provides a treatment of the "regulated integral" (as found in Dieudonné [**D**]) and of the Riemann integral. These are treated briefly, but with the intent of drawing parallels between their definition and the presentation of the Lebesgue integral in subsequent chapters.

As mentioned above the actual construction of Lebesgue measure and proofs of its key properties are left for an appendix. Instead the text introduces Lebesgue measure as a generalization of the concept of length and motivates its key properties: monotonicity, countable additivity, and translation invariance. This also motivates the concept of $\sigma$-algebra. If a generalization of length has these three key properties, then it needs to be defined on a $\sigma$-algebra for these properties to make sense.

In Chapter 2 the text introduces null sets and shows that any generalization of length satisfying monotonicity and countable additivity must assign zero to them. We then *define* Lebesgue measurable sets to be sets in the $\sigma$-algebra generated by open sets and null sets.

At this point we state a theorem which asserts that Lebesgue measure exists and is unique, i.e., there is a function $\mu$ defined for measurable subsets of a closed interval which satisfies monotonicity, countable additivity, and translation invariance.

The proof of this theorem (Theorem 2.4.2) is included in an appendix where it is also shown that the more common definition of measurable sets (using outer measure) is equivalent to being in the $\sigma$-algebra generated by open sets and null sets.

Chapter 3 discusses bounded Lebesgue measurable functions and their Lebesgue integral. The last section of this chapter, and some of the exercises following it, focus somewhat pedantically on the concept of "almost everywhere." The hope is to develop sufficient facility with the concept that it can be treated more glibly in subsequent chapters.

Chapter 4 considers unbounded functions and some of the standard convergence theorems. In Chapter 5 we introduce the Hilbert space of $L^2$ functions on an interval and show several elementary properties leading up to a definition of Fourier series.

Chapter 6 discusses classical real and complex Fourier series for $L^2$ functions on the interval and shows that the Fourier series of an $L^2$ function converges in $L^2$ to that function. The proof is based on the Stone-Weierstrass theorem which is stated but not proved.

Chapter 7 introduces some concepts from measurable dynamics. The Birkhoff ergodic theorem is stated without proof and results on Fourier series from Chapter 6 are used to prove that an irrational rotation of the circle is ergodic and the squaring map $z \mapsto z^2$ on the complex numbers of modulus 1 is ergodic.

Appendix A summarizes the needed prerequisites providing many proofs and some exercises. There is some emphasis in this section on the concept of countability, to which I would urge students and instructors to devote some time, as countability plays an very crucial role in the study of measure theory.

In Appendix B we construct Lebesgue measure and prove it has the properties cited in Chapter 2. In Appendix C we construct a non-measurable set.

Finally, at the website http://www.ams.org/bookpages/stml-48 we provide solutions to a few of the more challenging exercises. These exercises are marked with a $(\star)$ when they occur in the text.

This text grew out of notes I have used in teaching a one quarter course on integration at the advanced undergraduate level. With some selectivity of topics and well prepared students it should be possible to cover all key concepts in a one semester course.

Chapter 1

# The Regulated and Riemann Integrals

## 1.1. Introduction

This text is devoted to exploring the definition and properties of the definite integral. We will consider several different approaches to defining the definite integral

$$\int_a^b f(x)\ dx$$

of a function $f : [a, b] \to \mathbb{R}$. These definitions will all assign the same value to the definite integral, but they differ in the size of the collection of functions to which they apply. For example, we might try to evaluate the Riemann integral (the ordinary integral of beginning calculus) of the function

$$f(x) = \begin{cases} 0, & \text{if } x \text{ is rational;} \\ 1, & \text{otherwise.} \end{cases}$$

The Riemann integral $\int_0^1 f(x)\ dx$ is, as we will see, undefined; but the Lebesgue integral, which we will develop, has no difficulty with $f$ and indeed $\int_0^1 f(x)\ dx = 1$.

There are several properties which we want an integral to satisfy no matter how we define it. It is worth enumerating them at the beginning. We will need to check them and refine them for our different definitions.

## 1.2. Basic Properties of an Integral

We will consider the value of the integral of functions in various collections. These collections all have a common domain which, for our purposes, is a closed interval. They are also closed under the operations of addition and scalar multiplication. Such a collection is a *vector space of real-valued functions* (see, for example, Definition A.9.1). More formally, recall that a non-empty set of real-valued functions $\mathcal{V}$ defined on a fixed closed interval is a vector space of functions provided:

(1) If $f, g \in \mathcal{V}$, then $f + g \in \mathcal{V}$.

(2) If $f \in \mathcal{V}$ and $c \in \mathbb{R}$, then $cf \in \mathcal{V}$.

Notice that this implies that the constant function 0 is in $\mathcal{V}$. All of the vector spaces we consider will contain all of the constant functions.

Three simple examples of vector spaces of functions defined on some closed interval $I$ are the constant functions, the polynomial functions, and the continuous functions.

An "integral" defined on a vector space of functions $\mathcal{V}$ is a way to assign a real number to each function in $\mathcal{V}$ and each subinterval of $I$. For the function $f \in \mathcal{V}$ and the subinterval $[a, b]$ we denote this value by $\int_a^b f(x)\,dx$ and call it "the integral of $f$ from $a$ to $b$."

All the integrals we consider will satisfy five basic properties which we now enumerate.

**I. Linearity:** For any functions $f, g \in \mathcal{V}$, any $a, b \in I$, and any real numbers $c_1, c_2$,

$$\int_a^b c_1 f(x) + c_2 g(x)\,dx = c_1 \int_a^b f(x)\,dx + c_2 \int_a^b g(x)\,dx.$$

In particular, this implies that $\int_a^b 0\,dx = 0$.

**II. Monotonicity:** If the functions $f, g \in \mathcal{V}$ satisfy $f(x) \geq g(x)$ for all $x$ and $a, b \in I$ satisfy $a \leq b$, then

$$\int_a^b f(x)\ dx \geq \int_a^b g(x)\ dx.$$

In particular, if $f(x) \geq 0$ for all $x$ and $a \leq b$, then

$$\int_a^b f(x)\ dx \geq 0.$$

**III. Additivity:** For any function $f \in \mathcal{V}$, and any $a, b, c \in I$,

$$\int_a^c f(x)\ dx = \int_a^b f(x)\ dx + \int_b^c f(x)\ dx.$$

In particular, we allow $a, b$ and $c$ to occur in any order on the line and we note that two easy consequences of additivity are

$$\int_a^a f(x)\ dx = 0 \quad \text{and} \quad \int_a^b f(x)\ dx = -\int_b^a f(x)\ dx.$$

**IV. Constant functions:** The integral of a constant function $f(x) = C$ should be given by

$$\int_a^b C\ dx = C(b-a).$$

If $C > 0$ and $a < b$, this just says the integral of $f$ is the area of the rectangle under its graph.

**V. Finite sets don't matter:** If $f$ and $g$ are functions in $\mathcal{V}$ with $f(x) = g(x)$ for all $x$ except possibly a finite set, then for all $a, b \in I$,

$$\int_a^b f(x)\ dx = \int_a^b g(x)\ dx.$$

Properties III, IV and V are not valid for all mathematically interesting theories of integration. Nevertheless, they hold for all the integrals we will consider, so we include them in our list of basic properties. It is important to note that these are *assumptions*, however, and there are many mathematically interesting theories where they do not hold.

There is one additional property which we will need. It differs from the earlier ones in that we can *prove* that it holds whenever the properties above are satisfied.

**Proposition 1.2.1. (Absolute value).** *Suppose we have defined the integral $\int_a^b f(x)\ dx$ for all $f$ in some vector space of functions $\mathcal{V}$ and for all $a, b \in I$. Suppose this integral satisfies properties I-III above and both $f$ and $|f|$ are in $\mathcal{V}$. Then for any $a, b \in I$ with $a \leq b$,*

$$\Big| \int_a^b f(x)\ dx \Big| \leq \int_a^b |f(x)|\ dx.$$

*If $a > b$, then*

$$\Big| \int_a^b f(x)\ dx \Big| \leq -\int_a^b |f(x)|\ dx.$$

**Proof.** Suppose first that $a \leq b$. Since $f(x) \leq |f(x)|$ for all $x$ we know that

$$\int_a^b f(x)\ dx \leq \int_a^b |f(x)|\ dx$$

by monotonicity. Likewise, $-f(x) \leq |f(x)|$, so

$$-\int_a^b f(x)\ dx = \int_a^b -f(x)\ dx \leq \int_a^b |f(x)|\ dx.$$

But $| \int_a^b f(x)\ dx|$ is either equal to $\int_a^b f(x)\ dx$ or to $-\int_a^b f(x)\ dx$. In either case $\int_a^b |f(x)|\ dx$ is greater, so

$$\Big| \int_a^b f(x)\ dx \Big| \leq \int_a^b |f(x)|\ dx.$$

If $b < a$, then

$$\Big| \int_a^b f(x)\ dx \Big| = \Big| \int_b^a f(x)\ dx \Big| \leq \int_b^a |f(x)|\ dx = -\int_a^b |f(x)|\ dx.$$

$\square$

## 1.3. Step Functions

The easiest functions to integrate are *step functions*, which we now define.

**Definition 1.3.1. (Step function).** *A function $f : [a, b] \to \mathbb{R}$ is called a* step function *provided there are numbers*

$$x_0 = a < x_1 < x_2 < \cdots < x_{n-1} < x_n = b$$

*such that $f(x)$ is constant on each of the open intervals $(x_{i-1}, x_i)$.*

It is not difficult to see that the collection of all step functions defined on $[a, b]$ is a vector space of real-valued functions (see part (1) of Exercise 1.3.4).

We will say that the points $x_0 = a < x_1 < \cdots < x_{n-1} < x_n = b$ define an *interval partition* for the step function $f$. Note that the definition states that on the open intervals $(x_{i-1}, x_i)$ of the partition $f$ has a constant value, say $c_i$, but it says nothing about the values at the endpoints. The value of $f$ at the points $x_{i-1}$ and $x_i$ may or may not be equal to $c_i$. Of course when we define the integral this won't matter because the endpoints form a finite set.

Since the area under the graph of a positive step function is a finite union of rectangles, it is fairly obvious what the integral should be. The $i^{th}$ of these rectangles has width $(x_i - x_{i-1})$ and height $c_i$ so we should sum up the areas $c_i(x_i - x_{i-1})$. If some of the $c_i$ are negative then the corresponding $c_i(x_i - x_{i-1})$ are also negative, but that is appropriate since the area between the graph and the $x$-axis is below the $x$-axis on the interval $(x_{i-1}, x_i)$.

**Definition 1.3.2. (Integral of a step function).** *Suppose $f(x)$ is a step function with partition $x_0 = a < x_1 < x_2 < \cdots < x_{n-1} < x_n = b$ and suppose $f(x) = c_i$ for $x_{i-1} < x < x_i$. Then we define*

$$\int_a^b f(x)\ dx = \sum_{i=1}^{n} c_i(x_i - x_{i-1}).$$

We made the "obvious" definition for the integral of a step function, but in fact, we had absolutely no other choice if we want the integral to satisfy properties I–V above.

**Theorem 1.3.3.** *The integral as given in Definition 1.3.2 is the unique real-valued function defined on step functions which satisfies properties I–V of §1.2.*

**Proof.** Suppose that there is another "integral" defined on step functions and satisfying I–V. We will denote this alternate integral as

$$\oint_a^b f(x)\ dx.$$

What we must show is that for every step function $f(x)$,

$$\oint_a^b f(x)\ dx = \int_a^b f(x)\ dx.$$

Suppose that $f$ has partition $x_0 = a < x_1 < \cdots < x_{n-1} < x_n = b$ and satisfies $f(x) = c_i$ for $x_{i-1} < x < x_i$.

Then, from the additivity property,

$$\oint_a^b f(x)\ dx = \sum_{i=1}^{n} \oint_{x_{i-1}}^{x_i} f(x)\ dx. \tag{1.3.1}$$

But on the interval $[x_{i-1}, x_i]$ the function $f(x)$ is equal to the constant function with value $c_i$ except at the endpoints. Since functions which are equal except at a finite set of points have the same integral, the integral of $f$ is the same as the integral of $c_i$ on $[x_{i-1}, x_i]$. Combining this with the constant function property we get

$$\oint_{x_{i-1}}^{x_i} f(x)\ dx = \oint_{x_{i-1}}^{x_i} c_i\ dx = c_i(x_i - x_{i-1}).$$

If we plug this value into equation (1.3.1) we obtain

$$\oint_a^b f(x)\ dx = \sum_{i=1}^{n} c_i(x_i - x_{i-1}) = \int_a^b f(x)\ dx.$$

$\square$

---

**Exercise 1.3.4.**

(1) Prove that the collection of all step functions on a closed interval $[a, b]$ is a vector space of functions which contains the constant functions.

(2) Prove that if $x_0 = a < x_1 < x_2 < \cdots < x_{n-1} < x_n = b$ is a partition for a step function $f$ with value $c_i$ on $(x_{i-1}, x_i)$ and

$y_0 = a < y_1 < y_2 < \cdots < y_{n-1} < y_m = b$ is another partition for the same step function with value $d_j$ on $(y_{j-1}, y_i)$, then

$$\sum_{i=1}^{n} c_i(x_i - x_{i-1}) = \sum_{j=1}^{m} d_i(y_j - y_{j-1}).$$

In other words, the value of the integral of a step function depends only on the function, not on the choice of partition. *Hint:* the union of the sets of points defining the two partitions defines a third partition and the integral using this partition is equal to the integral using each of the partitions.

(3) Prove that the integral of step functions as given in Definition 1.3.2 satisfies properties I–V of §1.2.

## 1.4. Uniform and Pointwise Convergence

Throughout the text we will be interested in the following question: If a sequence of functions $\{f_n\}$ "converges" to a limit function $f$ does the sequence of numbers $\{\int_a^b f_n(x)\, dx\}$ converge to a limit equal to the integral of the limit function? Put another way, we are interested in when lim and $\int$ commute, i.e., when is

$$\lim_{n\to\infty} \int_a^b f_n(x)\, dx = \int_a^b \lim_{n\to\infty} f_n(x)\, dx?$$

The answer, as we will see, depends on what we mean by the sequence of functions "converging," i.e., what does $\lim f_n$ mean. It turns out there are many interesting (and very different) choices for what we might mean. Among the types of convergence that we will consider the strongest is called *uniform convergence*. We recall its definition.

**Definition 1.4.1. (Uniform convergence).** *A sequence of functions* $\{f_m\}$ *is said to* converge uniformly *on* $[a, b]$ *to a function* $f$ *if for every* $\varepsilon > 0$ *there is an* $M$ (independent of $x$) *such that for all* $x \in [a, b]$,

$$|f(x) - f_m(x)| < \varepsilon \quad \textit{whenever} \quad m \geq M.$$

We contrast this with the following much weaker notion of convergence.

**Definition 1.4.2. (Pointwise convergence).** *A sequence of functions* $\{f_m\}$ *is said to* converge pointwise *on* $[a,b]$ *to a function* $f$ *if for each* $\varepsilon > 0$ *and each* $x \in [a,b]$ *there is an* $M_x$ (depending on $x$) *such that*

$$|f(x) - f_m(x)| < \varepsilon \quad \textit{whenever} \quad m \geq M_x.$$

On first encountering these two types of convergence of functions it is difficult to appreciate how different they are and how different their consequences can be. The point of Examples 1.4.3 and 1.5.4 below and Exercise 1.5.6 parts (6) and (7) is to illustrate some of the ways these concepts differ and to emphasize the importance of the distinction. It should be immediately clear that a sequence of functions which converges uniformly to $f$ also converges pointwise to $f$. The following example shows that the converse of this statement is not true.

**Example 1.4.3.** For $m \in \mathbb{N}$ define the functions $f_m : [0,1] \to \mathbb{R}$ by

$$f_m(x) = \begin{cases} 0, & \text{if } x = 1; \\ x^m, & \text{otherwise.} \end{cases}$$

Then for any fixed $x_0 \in [0,1]$ it is clear that $\lim_{n\to\infty} f^n(x_0) = 0$. That is, the sequence $\{f_n\}$ converges *pointwise* to the constant function 0.

On the other hand, it does not converge *uniformly* to 0. For example, if $\varepsilon = 1/3$ we can never have $|f_m(x) - 0| < \varepsilon$ for all $x \in [0,1]$ since if $x_m = 1/2^{1/m}$, then $f(x_m) = 1/2$.

## 1.5. Regulated Integral

We now want to define the integral of a more general class of functions than just step functions. Since we know how to integrate step functions it is natural to try to take a sequence of better and better step function approximations to a more general function $f$ and define the integral of $f$ to be the limit of the integrals of the approximating step functions. For this to work we need to know that the limit of the integrals exists and that it does not depend on the choice of approximating step functions. It turns out that all of this works if the more general function $f$ can be uniformly approximated by step functions, i.e., if there is a sequence of step functions which *converges* *uniformly* to $f$. As is typical in mathematics when we have a collection of objects which behave in a way we like we make it into a definition.

**Definition 1.5.1. (Regulated function).** *A function* $f : [a, b] \to \mathbb{R}$ *is called* regulated *provided there is a sequence* $\{f_m\}$ *of step functions which converges* uniformly *to* $f$.

Another way to state this is to say a regulated function is one which can be uniformly approximated as closely as we wish by a step function. We can now prove that the limit of the integrals of the approximating step functions always exists and does not depend on the choice of approximating step functions.

**Theorem 1.5.2.** *Suppose* $\{f_m\}$ *is a sequence of step functions on* $[a, b]$ *converging uniformly to a regulated function* $f$. *Then the sequence of numbers* $\{\int_a^b f_m(x)\ dx\}$ *converges. Moreover, if* $\{g_m\}$ *is another sequence of step functions which also converges uniformly to* $f$ *then,*

$$\lim_{m\to\infty} \int_a^b f_m(x)\ dx = \lim_{m\to\infty} \int_a^b g_m(x)\ dx.$$

**Proof.** Let $z_m = \int_a^b f_m(x)\ dx$. We will show that the sequence $\{z_m\}$ is a Cauchy sequence and hence has a limit. To show this sequence is Cauchy we must show that for any $\varepsilon > 0$ there is an $M$ such that $|z_p - z_q| \leq \varepsilon$ whenever $p, q \geq M$.

Since $\{f_m\}$ is a sequence of step functions on $[a, b]$ converging uniformly to $f$, if we are given $\varepsilon > 0$, there is an $M$ such that for all $x \in [a, b]$,

$$|f(x) - f_m(x)| < \frac{\varepsilon}{2(b-a)} \quad \text{whenever} \quad m \geq M.$$

Hence, whenever $p, q \geq M$,

$$\begin{aligned} |f_p(x) - f_q(x)| &< |f_p(x) - f(x)| + |f(x) - f_q(x)| \\ &< \frac{\varepsilon}{2(b-a)} + \frac{\varepsilon}{2(b-a)} \\ &= \frac{\varepsilon}{b-a}. \end{aligned} \tag{1.5.1}$$

Therefore, whenever $p, q \geq M$,

$$\begin{aligned}|z_p - z_q| &= \left| \int_a^b f_p(x) - f_q(x)\ dx \right| \\ &\leq \int_a^b |f_p(x) - f_q(x)|\ dx \\ &\leq \int_a^b \frac{\varepsilon}{b-a}\ dx = \varepsilon,\end{aligned}$$

where the first inequality comes from the absolute value property of Proposition 1.2.1 and the second follows from the monotonicity property and equation (1.5.1). This shows that the sequence $\{z_m\}$ is Cauchy and hence converges.

Now suppose that $\{g_m\}$ is another sequence of step functions which also converges uniformly to $f$. Then for any $\varepsilon > 0$ there is an $M$ such that for all $x$,

$$|f(x) - f_m(x)| < \varepsilon \ \text{ and } \ |f(x) - g_m(x)| < \varepsilon$$

whenever $m \geq M$. It follows that

$$|f_m(x) - g_m(x)| \leq |f_m(x) - f(x)| + |f(x) - g_m(x)| < 2\varepsilon.$$

Hence, using the absolute value and monotonicity properties, we see

$$\begin{aligned}\left| \int_a^b f_m(x) - g_m(x)\ dx \right| &\leq \int_a^b |f_m(x) - g_m(x)|\ dx \\ &\leq \int_a^b 2\varepsilon\ dx = 2\varepsilon(b-a),\end{aligned}$$

for all $m \geq M$. Since $\varepsilon$ is arbitrarily small we may conclude that

$$\begin{aligned}&\lim_{m\to\infty} \left| \int_a^b f_m(x)\ dx - \int_a^b g_m(x)\ dx \right| \\ &= \lim_{m\to\infty} \left| \int_a^b f_m(x) - g_m(x)\ dx \right| = 0.\end{aligned}$$

This implies

$$\lim_{m\to\infty} \int_a^b f_m(x) = \lim_{m\to\infty} \int_a^b g_m(x)\ dx.$$

$\square$

This result enables us to define the regulated integral.

**Definition 1.5.3. (Regulated integral).** *If $f$ is a regulated function on $[a, b]$, we define the* regulated integral *by*

$$\int_a^b f(x)\ dx = \lim_{n\to\infty} \int_a^b f_n(x)\ dx$$

*where $\{f_n\}$ is any sequence of step functions converging uniformly to $f$.*

One might well ask if we can take the same approach and define an integral for functions which are the limits of pointwise convergent sequences of functions. Unfortunately, this does not work as the following example shows.

**Example 1.5.4.** For each $n \in \mathbb{N}$ define a step function on $[0, 1]$ by

$$f_n(x) = \begin{cases} 2n^2, & \text{if } x \in [\frac{1}{2n}, \frac{1}{n}]; \\ 0, & \text{otherwise.} \end{cases}$$

Notice that if $x_0 \in (0, 1]$ and $1/n < x_0$, then $f_n(x_0) = 0$; so clearly $\lim f_n(x_0) = 0$ for every $x_0$. Also, $f_n(0) = 0$ for all $n$. In other words, the sequence of functions $\{f_n\}$ converges *pointwise* to the constant function $f = 0$.

However, $\int_0^1 f_n(x)\ dx = n$ so the sequence of integrals diverges while the integral of the limit function $f = 0$ has the value 0.

This should be contrasted with part (6) of Exercise 1.5.6 which shows that if a sequence of functions $\{f_n\}$ converges *uniformly* to a function $f$, then under very general hypotheses,

$$\lim_{n\to\infty} \int_a^b f_n(x)dx = \int_a^b f(x)dx.$$

For the definition of regulated integral to be interesting it is important that there are lots of regulated functions which we might want to integrate. This is indeed the case since the regulated functions include all continuous functions on a closed interval $[a, b]$.

**Theorem 1.5.5. (Continuous functions are regulated).** *Every continuous function $f : [a, b] \to \mathbb{R}$ is a regulated function.*

**Proof.** A continuous function $f(x)$ defined on a closed interval $[a, b]$ is *uniformly continuous* (see Theorem A.8.2). That is, given $\varepsilon > 0$ there is a corresponding $\delta > 0$ such that $|f(x) - f(y)| < \varepsilon$ whenever $|x - y| < \delta$. Let $\varepsilon_n = 1/2^n$ and let $\delta_n$ be the corresponding $\delta$ guaranteed by uniform continuity.

Fix a value of $n$ and choose a partition $x_0 = a < x_1 < x_2 < \cdots < x_m = b$ with $x_i - x_{i-1} < \delta_n$. For example, we could choose $m$ so large that if we define $\Delta x = (b - a)/m$, then $\Delta x < \delta_n$ and then we could define $x_i$ to be $a + i\Delta x$. Next we define a step function $f_n$ by

$$f_n(x) = f(x_i) \quad \text{for all } x \in [x_{i-1}, x_i).$$

That is, on each half open interval $[x_{i-1}, x_i)$ we define $f_n$ to be the constant function whose value is the value of $f$ at the left endpoint of the interval. The value of $f_n(b)$ is defined to be $f(b)$.

Clearly, $f_n(x)$ is a step function with the given partition. We must estimate its distance from $f$. Let $x$ be an arbitrary point of $[a, b]$. It must lie in one of the open intervals of the partition or be an endpoint of one of them; say $x \in [x_{i-1}, x_i)$. Then since $f_n(x) = f_n(x_{i-1}) = f(x_{i-1})$ we may conclude that

$$|f(x) - f_n(x)| \leq |f(x) - f(x_{i-1})| < \varepsilon_n$$

because of the uniform continuity of $f$ and the fact that $|x - x_{i-1}| < \delta_n$.

Thus we have constructed a step function $f_n$ with the property that for all $x \in [a, b]$,

$$|f(x) - f_n(x)| < \varepsilon_n.$$

So the sequence $\{f_n\}$ converges uniformly to $f$ and $f$ is a regulated function. □

---

**Exercise 1.5.6.**

(1) Show that the continuous function $f(x) = 1/x$ on the *open* interval $(0, 1)$ is not regulated, i.e., it cannot be uniformly approximated by step functions.

(2) Give an example of a bounded continuous function on the *open* interval $(0, 1)$ which is not regulated.

(3) Give an example of a sequence of step functions which converges uniformly to $f(x) = x$ on $[0, 1]$.

(4) Prove that the collection of all regulated functions on a closed interval $I$ is a vector space which contains the constant functions.

(5) Prove that the regulated integral, as given in (1.5.3), satisfies properties I–V of §1.2.

(6) Suppose an integral $\int$ satisfying properties I–V of §1.2 has been defined for all functions $f : [a, b] \to \mathbb{R}$ in some vector space of functions $\mathcal{V}$. Prove that if $\{f_n\}$ is a sequence of functions in $\mathcal{V}$ which converges *uniformly* to $f \in \mathcal{V}$, then

$$\lim_{n\to\infty} \int_a^b f_n(x)\, dx = \int_a^b f(x)\, dx.$$

(7) Suppose $f : [0, 1] \to \mathbb{R}$ is continuous on $(0, 1)$. Prove there is a sequence of step functions $\{f_n\}$ which converge pointwise to $f$ on $[0, 1]$.

(8) ($\star$) Prove that $f$ is a regulated function on $I = [a, b]$ if and only if both of the limits

$$\lim_{x\to c+} f(x) \quad \text{and} \quad \lim_{x\to c-} f(x)$$

exist for every $c \in (a, b)$. (See section VII.6 of Dieudonné [**D**].)

## 1.6. The Fundamental Theorem of Calculus

The most important theorem of elementary calculus asserts that if $f$ is a continuous function on $[a, b]$ then its integral $\int_a^b f(x)\, dx$ can be evaluated by finding an anti-derivative. More precisely, if $F(x)$ is an anti-derivative of $f$ then

$$\int_a^b f(x)\, dx = F(b) - F(a).$$

We now can present a rigorous proof of this result. We will actually formulate the result in a slightly different way and show that the result above follows easily from that formulation.

**Theorem 1.6.1.** *If $f$ is a continuous function and we define*

$$F(x) = \int_a^x f(t)\ dt,$$

*then $F$ is a differentiable function and $F'(x) = f(x)$.*

**Proof.** By definition

$$F'(x_0) = \lim_{h\to 0} \frac{F(x_0+h) - F(x_0)}{h};$$

so we need to show that

$$\lim_{h\to 0} \frac{F(x_0+h) - F(x_0)}{h} = f(x_0),$$

or, equivalently,

$$\lim_{h\to 0} \left| \frac{F(x_0+h) - F(x_0)}{h} - f(x_0) \right| = 0.$$

To do this we note that

$$\begin{aligned} (1.6.1) \quad \left| \frac{F(x_0+h) - F(x_0)}{h} - f(x_0) \right| &= \left| \frac{\int_{x_0}^{x_0+h} f(t)\ dt}{h} - f(x_0) \right| \\ &= \left| \frac{\int_{x_0}^{x_0+h} f(t)\ dt - f(x_0)h}{h} \right| \\ &= \frac{\left| \int_{x_0}^{x_0+h} (f(t) - f(x_0))\ dt \right|}{|h|}. \end{aligned}$$

Proposition 1.2.1 tells us that

$$\left| \int_{x_0}^{x_0+h} (f(t) - f(x_0))\ dt \right| \leq \left| \int_{x_0}^{x_0+h} |f(t) - f(x_0)|\ dt \right|$$

Combining this with equation (1.6.1) above we obtain

$$(1.6.2) \quad \left| \frac{F(x_0+h) - F(x_0)}{h} - f(x_0) \right| \leq \frac{\left| \int_{x_0}^{x_0+h} |f(t) - f(x_0)|\ dt \right|}{|h|}.$$

But the continuity of $f$ implies that given $x_0$ and any $\varepsilon > 0$ there exists $\delta > 0$ such that whenever $|t - x_0| < \delta$ we have $|f(t) - f(x_0)| < \varepsilon$. Thus, if $|h| < \delta$, then $|f(t) - f(x_0)| < \varepsilon$ for all $t$ between $x_0$ and $x_0 + h$. It follows that

$$\left| \int_{x_0}^{x_0+h} |f(t) - f(x_0)|\ dt \right| < \varepsilon |h|$$

and hence that

$$\frac{\left|\int_{x_0}^{x_0+h} |f(t) - f(x_0)|\ dt\right|}{|h|} < \varepsilon.$$

Putting this together with the inequality (1.6.2) above we have that

$$\left|\frac{F(x_0+h) - F(x_0)}{h} - f(x_0)\right| < \varepsilon$$

whenever $|h| < \delta$, which is exactly what we needed to show. $\square$

**Corollary 1.6.2. (Fundamental theorem of calculus).** *If $f$ is a continuous function on $[a, b]$ and $F$ is any anti-derivative of $f$, then*

$$\int_a^b f(x)\ dx = F(b) - F(a).$$

**Proof.** Define the function $G(x) = \int_a^x f(t)\ dt$. By Theorem 1.6.1 the derivative of $G(x)$ is $f(x)$ which is also the derivative of $F$. Hence $F$ and $G$ differ by a constant, say $F(x) = G(x) + C$ (see Corollary A.8.4).

Then

$$\begin{aligned} F(b) - F(a) &= (G(b) + C) - (G(a) + C) \\ &= G(b) - G(a) \\ &= \int_a^b f(x)\ dx - \int_a^a f(x)\ dx \\ &= \int_a^b f(x)\ dx. \end{aligned}$$

$\square$

---

**Exercise 1.6.3.**

(1) Prove that if $f : [a, b] \to \mathbb{R}$ is a regulated function and $F : [a, b] \to \mathbb{R}$ is defined by $F(x) = \int_a^x f(t)\ dt$, then $F$ is continuous.

(2) Let $\mathcal{S}$ denote the set of all functions $F : [a, b] \to \mathbb{R}$ which can be expressed as $F(x) = \int_a^x f(t)\ dt$ for some *step function* $f$. Prove that $\mathcal{S}$ is a vector space of functions, each of which

has a derivative except at a finite set. Suppose that $f, g \in \mathcal{S}$ and $g([a,b]) \subset [a,b]$. Prove that $h(x) = f(g(x))$ is in $\mathcal{S}$.

(3) Let $f : [a,b] \to \mathbb{C}$ be a complex-valued function and suppose its real and imaginary parts, $u(x) = \Re(f(x))$ and $v(x) = \Im(f(x))$, are both continuous. We can then define the derivative (if it exists) by $df/dx = du/dx + idv/x$ and the integral by

$$\int_a^b f(x)dx = \int_a^b u(x)\ dx + i\int_a^b v(x)\ dx.$$

(a) Prove that if $F : [a,b] \to \mathbb{C}$ has a continuous derivative $f(x)$ then

$$\int_a^b f(x)\ dx = F(b) - F(a),$$

i.e., the fundamental theorem of calculus holds.

(b) Prove that, if $c \in \mathbb{C}$ and $F(x) = e^{cx}$ for $x \in [a,b]$, then $dF/dx = ce^{cx}$. *Hint:* Use *Euler's formula*:

$$e^{i\theta} = \cos\theta + i\sin\theta$$

for all $\theta \in \mathbb{R}$.

(c) Prove that, if $c \in \mathbb{C}$ is not 0, then

$$\int_a^b e^{cx}\ dx = \frac{e^{cb} - e^{ca}}{c}.$$

## 1.7. The Riemann Integral

We can obtain a larger class of functions for which a good integral can be defined by using a different method of comparison with step functions.

Suppose that $f(x)$ is a bounded function on the interval $I = [a,b]$ and that it is an element of a vector space of functions which contains the step functions and for which there is an integral defined satisfying properties I–V of §1.2. If $u(x)$ is a step function satisfying $f(x) \leq u(x)$ for all $x \in I$, then monotonicity implies that if we can define $\int_a^b f(x)\ dx$ it must satisfy $\int_a^b f(x)\ dx \leq \int_a^b u(x)\ dx$.

This is true for every step function $u$ satisfying $f(x) \leq u(x)$ for all $x \in I$. Let $\mathcal{U}(f)$ denote the set of all step functions with this property.

Then if we can define $\int_a^b f(x)\,dx$ in a way that satisfies monotonicity it must also satisfy

$$\int_a^b f(x)\,dx \le \inf\left\{ \int_a^b u(x)\,dx \;\middle|\; u \in \mathcal{U}(f)\right\}. \tag{1.7.1}$$

The *infimum* exists because all of the step functions in $\mathcal{U}(f)$ are bounded below by a lower bound for the function $f$.

Similarly, we define $\mathcal{L}(f)$ to be the set of all step functions $v(x)$ such that $v(x) \le f(x)$ for all $x \in I$. Again, if we can define $\int_a^b f(x)\,dx$ in such a way that it satisfies monotonicity it must also satisfy

$$\sup\left\{ \int_a^b v(x)\,dx \;\middle|\; v \in \mathcal{L}(f) \right\} \le \int_a^b f(x)\,dx. \tag{1.7.2}$$

The *supremum* exists because an upper bound for the function $f$ is an upper bound for all of the step functions in $\mathcal{U}(f)$.

Putting inequalities (1.7.1) and (1.7.2) together, we see that if $\mathcal{V}$ is any vector space of bounded functions which contains the step functions and we manage to define the integral of functions in $\mathcal{V}$ in a way that satisfies monotonicity, then this integral must satisfy

$$\begin{aligned}\sup\left\{ \int_a^b v(x)\,dx \;\middle|\; v \in \mathcal{L}(f)\right\} &\le \int_a^b f(x)\,dx \\ &\le \inf\left\{ \int_a^b u(x)\,dx \;\middle|\; u \in \mathcal{U}(f)\right\}\end{aligned}$$

for every $f \in \mathcal{V}$. We next observe that even if we cannot define an integral for $f$ we still have the inequality relating the expressions at the ends.

**Proposition 1.7.1.** *Let $f$ be any bounded function on the interval $I = [a.b]$. Let $\mathcal{U}(f)$ denote the set of all step functions $u(x)$ on $I$ such that $f(x) \le u(x)$ for all $x$ and let $\mathcal{L}(f)$ denote the set of all step functions $v(x)$ such that $v(x) \le f(x)$ for all $x$. Then*

$$\sup\left\{ \int_a^b v(x)\,dx \;\middle|\; v \in \mathcal{L}(f)\right\} \le \inf\left\{ \int_a^b u(x)\,dx \;\middle|\; u \in \mathcal{U}(f)\right\}.$$

**Proof.** If $v \in \mathcal{L}(f)$ and $u \in \mathcal{U}(f)$, then $v(x) \le f(x) \le u(x)$ for all $x \in I$, so monotonicity implies that $\int_a^b v(x)\,dx \le \int_a^b u(x)\,dx$. Hence, if

$$V = \left\{ \int_a^b v(x)\ dx \;\middle|\; v \in \mathcal{L}(f) \right\} \text{ and } U = \left\{ \int_a^b u(x)\ dx \;\middle|\; u \in \mathcal{U}(f) \right\},$$

then every number in the set $V$ is less than or equal to every number in the set $U$. Thus, $\sup V \leq \inf U$ as claimed. □

It is not difficult to see that sometimes the two sides of this inequality are not equal (see Exercise 1.7.7 below); but if it should happen that

$$\sup \left\{ \int_a^b v(x)\ dx \;\middle|\; v \in \mathcal{L}(f) \right\} = \inf \left\{ \int_a^b u(x)\ dx \;\middle|\; u \in \mathcal{U}(f) \right\},$$

then we have only one choice for $\int_a^b f(x)\ dx$; it must be this common value.

This motivates the definition of the next vector space of functions that can be integrated. Henceforth, we will use the more compact notation

$$\sup_{v \in \mathcal{L}(f)} \left\{ \int_a^b v(x)\ dx \right\} \text{ instead of } \sup \left\{ \int_a^b v(x)\ dx \;\middle|\; v \in \mathcal{L}(f) \right\}$$

and

$$\inf_{u \in \mathcal{U}(f)} \left\{ \int_a^b u(x)\ dx \right\} \text{ instead of } \inf \left\{ \int_a^b u(x)\ dx \;\middle|\; u \in \mathcal{U}(f) \right\}.$$

**Definition 1.7.2. (Riemann integral).** *Suppose $f$ is a bounded function on the interval $I = [a, b]$. Let $\mathcal{U}(f)$ denote the set of all step functions $u(x)$ on $I$ such that $f(x) \leq u(x)$ for all $x$ and let $\mathcal{L}(f)$ denote the set of all step functions $v(x)$ such that $v(x) \leq f(x)$ for all $x$. The function $f$ is said to be* Riemann integrable *provided*

$$\sup_{v \in \mathcal{L}(f)} \left\{ \int_a^b v(x)\ dx \right\} = \inf_{u \in \mathcal{U}(f)} \left\{ \int_a^b u(x)\ dx \right\}.$$

*In this case its* Riemann integral $\int_a^b f(x)\ dx$ *is defined to be this common value.*

There is a simple test for when a function $f$ is Riemann integrable. For any $\varepsilon > 0$ we need only find a step function $u$ greater than $f$ and a step function $v$ less than $f$ such that the difference of the integrals of $u$ and $v$ is less than $\varepsilon$.

**Theorem 1.7.3.** *A bounded function $f : [a,b] \to \mathbb{R}$ is Riemann integrable if and only if, for every $\varepsilon > 0$ there are step functions $v_0$ and $u_0$ such that $v_0(x) \le f(x) \le u_0(x)$ for all $x \in [a,b]$ and*

$$\int_a^b u_0(x)\ dx - \int_a^b v_0(x)\ dx \le \varepsilon.$$

**Proof.** Suppose the functions $v_0 \in \mathcal{L}(f)$ and $u_0 \in \mathcal{U}(f)$ have integrals within $\varepsilon$ of each other. Then

$$\begin{aligned}\int_a^b v_0(x)\ dx &\le \sup_{v\in\mathcal{L}(f)} \left\{ \int_a^b v(x)\ dx \right\} \\ &\le \inf_{u\in\mathcal{U}(f)} \left\{ \int_a^b u(x)\ dx \right\} \\ &\le \int_a^b u_0(x)\ dx,\end{aligned}$$

where the second inequality follows from Proposition 1.7.1.

This implies

$$\inf_{u\in\mathcal{U}(f)} \left\{ \int_a^b u(x)\ dx \right\} - \sup_{v\in\mathcal{L}(f)} \left\{ \int_a^b v(x)\ dx \right\} \le \varepsilon.$$

Since this is true for all $\varepsilon > 0$, we conclude that $f$ is Riemann integrable.

Conversely, if $f$ is Riemann integrable, then from the properties of the infimum there exists a step function $u_0 \in \mathcal{U}(f)$ such that

$$\int_a^b u_0(x)\ dx \ < \inf_{u\in\mathcal{U}(f)} \left\{ \int_a^b u(x)\ dx \right\} + \frac{\varepsilon}{2} = \int_a^b f(x)\ dx + \frac{\varepsilon}{2}.$$

Thus,

$$\int_a^b u_0(x)\ dx \ - \int_a^b f(x)\ dx < \frac{\varepsilon}{2}.$$

Similarly, there exists a step function $v_0 \in \mathcal{L}(f)$ such that

$$\int_a^b f(x)\ dx \ - \int_a^b v_0(x)\ dx < \frac{\varepsilon}{2}.$$

Hence,

$$\int_a^b u_0(x)\ dx\ - \int_a^b v_0(x)\ dx < \varepsilon/2 + \varepsilon/2 = \varepsilon,$$

and $u_0$ and $v_0$ are the desired functions. □

There are several facts about the relation with the regulated integral that must be established. Every regulated function is Riemann integrable, but there are Riemann integrable functions which have no regulated integral. Whenever a function has both types of integral the values agree. We start by giving an example of a function which is Riemann integrable, but not regulated.

**Example 1.7.4.** Define the function $f : [0, 1] \to \mathbb{R}$ by

$$f(x) = \begin{cases} 1, & \text{if } x = \frac{1}{n} \text{ for } n \in \mathbb{N}; \\ 0, & \text{otherwise.} \end{cases}$$

Then $f(x)$ is Riemann integrable and $\int_0^1 f(x)\ dx = 0$, but it is not regulated.

**Proof.** We define a step function $u_m(x)$ by

$$u_m(x) = \begin{cases} 1, & \text{if } 0 \leq x \leq \frac{1}{m}; \\ f(x), & \text{otherwise.} \end{cases}$$

A partition for this step function is given by

$$x_0 = 0 < x_1 = \frac{1}{m} < x_2 = \frac{1}{m-1} < \cdots < x_{m-1} = \frac{1}{2} < x_m = 1.$$

Note that $u_m(x) \geq f(x)$. Also, $\int_0^1 u_m(x)\ dx = \frac{1}{m}$. This is because it is constant and equal to 1 on the interval $[0, \frac{1}{m}]$ and except for a finite number of points it is constant and equal to 0 on the interval $[\frac{1}{m}, 1]$. Hence,

$$\inf_{u \in \mathcal{U}(f)} \left\{ \int_0^1 u(x)\ dx \right\} \leq \inf_{m \in \mathbb{N}} \left\{ \int_0^1 u_m(x)\ dx \right\} = \inf_{m \in \mathbb{N}} \left\{ \frac{1}{m} \right\} = 0.$$

Also, the constant function 0 is $\leq f(x)$ and its integral is 0, so

$$0 \leq \sup_{v \in \mathcal{L}(f)} \left\{ \int_0^1 v(x)\ dx \right\}.$$

Putting together the last two inequalities with Proposition 1.7.1 we obtain

$$0 \leq \sup_{v \in \mathcal{L}(f)} \left\{ \int_0^1 v(x)\ dx \right\} \leq \inf_{u \in \mathcal{U}(f)} \left\{ \int_0^1 u(x)\ dx \right\} \leq 0.$$

So all of these inequalities are equalities and by definition, $f$ is Riemann integrable with integral 0.

To see that $f$ is not regulated suppose that $g$ is an approximating step function with partition $x_0 = 0 < x_1 < \cdots < x_m = 1$ and satisfying $|f(x) - g(x)| \leq \varepsilon$ for some $\varepsilon > 0$. Then $g$ is constant, say with value $c_1$ on the open interval $(0, x_1)$.

There exist points $a_1, a_2 \in (0, x_1)$ with $f(a_1) = 0$ and $f(a_2) = 1$. So

$$|c_1| = |c_1 - 0| = |g(a_1) - f(a_1)| \leq \varepsilon$$

and

$$|1 - c_1| = |f(a_2) - g(a_2)| \leq \varepsilon.$$

But

$$|c_1| + |1 - c_1| \geq |c_1 + 1 - c_1| = 1,$$

so at least one of $|c_1|$ and $|1 - c_1|$ must be $\geq 1/2$. This implies that $\varepsilon \geq 1/2$. That is, $f$ cannot be uniformly approximated by any step function to within $\varepsilon$ if $\varepsilon < 1/2$. So $f$ is not regulated. □

**Theorem 1.7.5. (Regulated implies Riemann integrable).** *Every regulated function $f$ is Riemann integrable and the regulated integral of $f$ is equal to its Riemann integral.*

**Proof.** If $f$ is a regulated function on the interval $I = [a, b]$, then, for any $\varepsilon > 0$, it can be uniformly approximated within $\varepsilon$ by a step function. In particular, if $\varepsilon_n = 1/2^n$, there is a step function $g_n(x)$ such that $|f(x) - g_n(x)| < \varepsilon_n$ for all $x \in I$. The regulated integral $\int_a^b f(x)\ dx$ was defined to be $\lim \int_a^b g_n(x)\ dx$.

We define two other approximating sequences of step functions for $f$. Let $u_n(x) = g_n(x) + 1/2^n$ and $v_n(x) = g_n(x) - 1/2^n$. Then $u_n(x) \geq f(x)$ for all $x \in I$ because $u_n(x) - f(x) = 1/2^n + g_n(x) - f(x) \geq 0$ since $|g_n(x) - f(x)| < 1/2^n$. Similarly, $v_n(x) \leq f(x)$ for all $x \in I$ because $f(x) - v_n(x) = 1/2^n + f(x) - g_n(x) \geq 0$ since $|f(x) - g_n(x)| < 1/2^n$.

Since

$$u_n(x) - v_n(x) = g_n(x) + 1/2^n - (g_n(x) - 1/2^n) = 1/2^{n-1},$$

we have

$$\begin{aligned}\int_a^b u_n(x)\ dx - \int_a^b v_n(x)\ dx &= \int_a^b u_n(x) - v_n(x)\ dx\\ &= \int_a^b \frac{1}{2^{n-1}}\ dx\\ &= \frac{b-a}{2^{n-1}}.\end{aligned}$$

Hence, we may apply Theorem 1.7.3 to conclude that $f$ is Riemann integrable.

Also,

$$\lim_{n\to\infty}\int_a^b g_n(x)\ dx = \lim_{n\to\infty}\int_a^b v_n(x) + \frac{1}{2^n}\ dx = \lim_{n\to\infty}\int_a^b v_n(x)\ dx,$$

and

$$\lim_{n\to\infty}\int_a^b g_n(x)\ dx = \lim_{n\to\infty}\int_a^b u_n(x) - \frac{1}{2^n}\ dx = \lim_{n\to\infty}\int_a^b u_n(x)\ dx.$$

Since for all $n$,

$$\int_a^b v_n(x)\ dx \le \int_a^b f(x)\ dx \le \int_a^b u_n(x)\ dx,$$

we conclude that

$$\lim_{n\to\infty}\int_a^b g_n(x)\ dx = \int_a^b f(x)\ dx.$$

That is, the regulated integral equals the Riemann integral. □

**Theorem 1.7.6.** *The set $\mathcal{R}$ of bounded Riemann integrable functions on an interval $I = [a, b]$ is a vector space containing the vector space of regulated functions.*

**Proof.** We have already shown that every regulated function is Riemann integrable. Hence, we need only show that whenever $f, g \in \mathcal{R}$ and $r \in \mathbb{R}$ we also have $(f + g) \in \mathcal{R}$ and $rf \in \mathcal{R}$. We will do only the sum and leave the product as an exercise.

Suppose $\varepsilon > 0$ is given. Since $f$ is Riemann integrable there are step functions $u_f$ and $v_f$ such that $v_f(x) \le f(x) \le u_f(x)$ for $x \in I$ (i.e., $u_f \in \mathcal{U}(f)$ and $v_f \in \mathcal{L}(f)$) and with the property that

$$\int_a^b u_f(x)\ dx - \int_a^b v_f(x)\ dx < \varepsilon.$$

Similarly, there are $u_g \in \mathcal{U}(g)$ and $v_g \in \mathcal{L}(g)$ with the property that

$$\int_a^b u_g(x)\ dx - \int_a^b v_g(x)\ dx < \varepsilon.$$

This implies that

$$\int_a^b (u_f + u_g)(x)\ dx - \int_a^b (v_f + v_g)(x)\ dx < 2\varepsilon.$$

Since $(u_f + u_g) \in \mathcal{U}(f+g)$ and $(v_f + v_g) \in \mathcal{L}(f+g)$, we may conclude that

$$\inf_{u \in \mathcal{U}(f+g)} \left\{ \int_a^b u(x)\ dx \right\} - \sup_{v \in \mathcal{L}(f+g)} \left\{ \int_a^b v(x)\ dx \right\} < 2\varepsilon.$$

As $\varepsilon > 0$ is arbitrary, we conclude that

$$\inf_{u \in \mathcal{U}(f+g)} \left\{ \int_a^b u(x)\ dx \right\} = \sup_{v \in \mathcal{L}(f+g)} \left\{ \int_a^b v(x)\ dx \right\}$$

and hence $(f+g) \in \mathcal{R}$. □

---

**Exercise 1.7.7.**

(1) At the beginning of this chapter we mentioned the function $f : [0,1] \to \mathbb{R}$ which has the value $f(x) = 0$ if $x$ is rational and 1 otherwise. Prove that for this function

$$\sup_{v \in \mathcal{L}(f)} \left\{ \int_0^1 v(x)\ dx \right\} = 0$$

and

$$\inf_{u \in \mathcal{U}(f)} \left\{ \int_0^1 u(x)\ dx \right\} = 1.$$

Hence, $f$ is not Riemann integrable.

(2) Prove that the absolute value of a Riemann integrable function is Riemann integrable.

(3) Suppose $f$ and $g$ are Riemann integrable functions defined on $[a, b]$. Prove that if $h(x) = \max\{f(x), g(x)\}$, then $h$ is Riemann integrable. This generalizes to the max of a finite set of functions, but not of infinitely many. Show there exists a family $\{f_n\}_{n\in\mathbb{N}}$ of step functions such that for each $n$ and each $x \in [a, b]$ the value of $f_n(x)$ is either 0 or 1 and yet the function defined by $g(x) = \max\{f_n(x)\}_{n\in\mathbb{N}}$ is not Riemann integrable.

(4) Prove that if $f$ and $g$ are bounded Riemann integrable functions on an interval $[a, b]$, then so is $fg$. In particular, if $r \in \mathbb{R}$, then $rf$ is a bounded Riemann integrable function on $[a, b]$.

Chapter 2

# Lebesgue Measure

## 2.1. Introduction

In the previous section we studied two definitions of integration that were based on two important facts: (1) There is only one obvious way to define the integral of step functions assuming we want it to satisfy certain basic properties, and (2) these properties force the definition for the integral for more general functions which are uniformly approximated by step functions (regulated integral) or squeezed between step functions whose integrals are arbitrarily close (Riemann integral).

To move to a more general class of functions we first find a more general notion to replace step functions. For a step function $f$ there is a partition of $I = [0, 1]$ into intervals on each of which $f$ is constant. We now would like to allow functions for which there is a finite partition of $I$ into sets on each of which $f$ is constant, but with the sets *not necessarily intervals.* For example, we will consider functions such as

$$f(x) = \begin{cases} 3, & \text{if } x \text{ is rational;} \\ 2, & \text{otherwise.} \end{cases} \tag{2.1.1}$$

The interval $I$ is partitioned into two sets, $A = I \cap \mathbb{Q}$ and $B = I \cap \mathbb{Q}^c$, i.e., the rational points of $I$ and the irrational points. Clearly, the integral of this function should be $3\,\mathrm{len}(A) + 2\,\mathrm{len}(B)$, but only if we can make sense of $\mathrm{len}(A)$ and $\mathrm{len}(B)$. That is the problem to which this chapter is devoted. We want to generalize the concept of length to include as many subsets of $\mathbb{R}$ as we can. We proceed in much the same way as in previous chapters. We first decide what are the "obvious" properties this generalized length must satisfy to be of any use, and then try to define it by approximating with simpler sets where the definition is clear, namely sets of intervals.

The generalization of length we want is called *Lebesgue measure*. Ideally, we would like it to work for *any* subset of the interval $I = [0,1]$, but it turns out that this is not possible to achieve.

There are several properties which we want any notion of "generalized length" to satisfy. These are analogous to the basic properties we required for a definition of integral in Chapter 1. For each bounded subset $A$ of $\mathbb{R}$ we would like to be able to assign a non-negative real number $\mu(A)$ that satisfies the following:

**I. Length:** If $A = (a,b)$ or $[a,b]$, then $\mu(A) = \mathrm{len}(A) = b - a$, i.e., the measure of an open or closed interval is its length.

**II. Translation invariance:** If $A \subset \mathbb{R}$ is a bounded subset of $\mathbb{R}$ and $c \in \mathbb{R}$, then $\mu(A + c) = \mu(A)$, where $A + c$ denotes the set $\{x + c \mid x \in A\}$.

**III. Countable additivity:** If $\{A_n\}_{n=1}^{\infty}$ is a countable collection of bounded subsets of $\mathbb{R}$, then

$$\mu\Big(\bigcup_{n=1}^{\infty} A_n\Big) \le \sum_{n=1}^{\infty} \mu(A_n)$$

and if the sets are *pairwise disjoint,* then

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

Note that for a finite collection $\{A_n\}_{n=1}^{m}$ of bounded sets the same conclusion applies (just let $A_i = \emptyset$ for $i > m$).

**IV. Monotonicity:** If $A \subset B$, then $\mu(A) \le \mu(B)$. Actually, this property is a consequence of additivity since $A$ and $B \backslash A$ are disjoint and their union is $B$.

It should be fairly clear why most of these properties are absolutely necessary for any sensible notion of length. The only exception is property III, which deserves some comment. We might ask that additivity holds only for finite collections of sets, but that is too weak. For example, if we had a collection of pairwise disjoint intervals of length $1/2, 1/4, 1/8, \dots 1/2^n, \dots$, etc., then we would certainly like to be able say that the measure of their union is the sum $\sum 1/2^n = 1$ which would not follow from finite additivity. Alternatively, one might wonder why additivity is only for *countable* collections of pairwise disjoint sets. But it is easy to see why it would lead to problems if we allowed uncountable collections. Suppose $A_x = \{x\}$ is the set consisting of a single point $x \in [0,1]$. Then $\mu(A_x) = 0$ by property I. But $[a,b]$ is an uncountable set and hence an uncountable union of pairwise disjoint sets each containing a single point, namely each of the sets $A_x$ for $x \in [a,b]$. Hence, "uncountable additivity" would imply that $\mu([a,b]) = b-a$ is an *uncountable* sum of zeroes. This is the main reason the concept of uncountable sums isn't very useful. Indeed, we will see that the concept of countability is intimately related to the concept of measure.

Unfortunately, as mentioned above, it turns out that it is impossible to find a $\mu$ which satisfies I–IV and which is defined for *all* bounded subsets of the reals; but we can do it for a very large collection which includes all the open sets and all the closed sets. The measure we are interested in using is called *Lebesgue measure.* Its actual construction is slightly technical and we have relegated that to an appendix. Instead, we will focus on some of the properties of Lebesgue measure and how it is used.

## 2.2. Null Sets

One of our axioms for the regulated integral was, "Finite sets don't matter." Now we want to generalize that to say that a set doesn't matter if its "generalized length," or measure, is zero. It is a somewhat surprising fact that even without defining Lebesgue measure in general we can easily define those sets whose measure must be 0 and investigate the properties of these sets.

**Definition 2.2.1. (Null set).** *A set $X \subset \mathbb{R}$ is called a* null set *if for every $\varepsilon > 0$ there is a collection of open intervals $\{U_n\}_{n=1}^{\infty}$ such that*

$$\sum_{n=1}^{\infty} \operatorname{len}(U_n) < \varepsilon \quad \textit{and} \quad X \subset \bigcup_{n=1}^{\infty} U_n.$$

Notice that this definition makes no use of the measure $\mu$. Indeed, we have not yet defined the measure $\mu$ for *any* set $X$! However, it is clear that if we can do so in a way that satisfies properties I-IV above, then if $X$ is a null set, $\mu(X) < \varepsilon$ for every positive $\varepsilon$. This, of course, implies $\mu(X) = 0$.

If $X$ is a null set, we will say that its complement $X^c$ has *full measure.*

---

**Exercise 2.2.2.**

(1) Prove that a finite set is a null set.

(2) Prove that a countable union of null sets is a null set (and hence, in particular, countable sets are null sets).

(3) Assuming that a measure $\mu$ has been defined and satisfies properties I-IV above, find the numerical value of the integral of the function $f(x)$ defined in equation (2.1.1). Prove that the Riemann integral of this function does not exist.

(4) Prove that if $X$ is a countable compact subset of $\mathbb{R}$, then for any $\varepsilon > 0$ there is a *finite* collection of pairwise disjoint open intervals $\{U_k\}_{k=1}^{n}$ such that

$$\sum_{k=1}^{n} \operatorname{len}(U_k) < \varepsilon \quad \text{and} \quad X \subset \bigcup_{k=1}^{n} U_k.$$

Use this to prove that any closed interval $[a, b]$ with $b > a$ is uncountable.

It is not true that countable sets are the only sets which are null sets. We give an example in Exercise 2.5.4 below, namely, the Cantor middle third set, which is an uncountable null set.

## 2.3. Sigma Algebras

As mentioned before there does not exist a function $\mu$ satisfying properties I-IV from Section 2.1 and which is defined for every subset of $I = [0, 1]$. In this section we want to consider what is the best we can do. Is there a collection of subsets of $I$ for which we can define a "generalized length" or *measure* $\mu$ which satisfies properties I–IV and which is large enough for our purposes? And what properties would such a collection need to have?

Suppose we have somehow defined $\mu$ for all the sets in some collection $\mathcal{A}$ of subsets of $I$ and it satisfies properties I–IV. Property I only makes sense if $\mu$ is defined for open and closed intervals, i.e., we need open and closed intervals to be in $\mathcal{A}$. For property III to make sense we will need that any countable union of sets in $\mathcal{A}$ is also in $\mathcal{A}$. Finally, it seems reasonable that if $A$ is a set in the collection $\mathcal{A}$, then the set $A^c$, its complement in $I$, should also be in $\mathcal{A}$.

All of this motivates the following definition.

**Definition 2.3.1. (Sigma algebra).** *Suppose $X$ is a set and $\mathcal{A}$ is a collection of subsets of $X$. $\mathcal{A}$ is called a $\sigma$-*algebra *of subsets of $X$ provided it contains the set $X$ and is closed under taking complements (with respect to $X$), countable unions, and countable intersections.*

In other words, if $\mathcal{A}$ is a $\sigma$-algebra of subsets of $X$, then any complement (with respect to $X$) of a set in $\mathcal{A}$ is also in $\mathcal{A}$, any countable union of sets in $\mathcal{A}$ is in $\mathcal{A}$, and any countable intersection of sets in $\mathcal{A}$ is in $\mathcal{A}$. In fact, the property concerning countable intersections follows from the other two and Proposition A.5.3 which says that the intersection of a family of sets is the complement of the union of the complements of the sets. Also note that, if $A, B \in \mathcal{A}$, then their set difference $A \setminus B = \{x \in A \mid x \notin B\}$ is in $\mathcal{A}$ because $A \setminus B = A \cap B^c$.

Since $X$ is in any $\sigma$-algebra of subsets of $X$ (by definition), so is its complement, the empty set. A trivial example of a $\sigma$-algebra of subsets of $X$ is $\mathcal{A} = \{X, \emptyset\}$, i.e., it consists of only the whole set $X$ and the empty set. Another example, at the other extreme, is $\mathcal{A} = \mathcal{P}(X)$, the power set of $X$, i.e., the collection of all subsets of $X$. Several more interesting examples are given in the exercises below. Also, in these exercises we ask you to show that any intersection of

$\sigma$-algebras is a $\sigma$-algebra. Thus, for any collection $\mathcal{C}$ of subsets of $\mathbb{R}$ there is a smallest $\sigma$-algebra of subsets of $\mathbb{R}$ which contains all sets in $\mathcal{C}$, namely the intersection of all $\sigma$-algebras containing $\mathcal{C}$ (there is a least one such $\sigma$-algebra, namely the power set $\mathcal{P}(\mathbb{R})$).

**Definition 2.3.2. (Borel sets).** *If $\mathcal{C}$ is a collection of subsets of $\mathbb{R}$ and $\mathcal{A}$ is the the smallest $\sigma$-algebra of subsets of $\mathbb{R}$ which contains all the sets of $\mathcal{C}$, then $\mathcal{A}$ is called the $\sigma$-algebra generated by $\mathcal{C}$. Let $\mathcal{B}$ be the $\sigma$-algebra of subsets of $\mathbb{R}$ generated by the collection of all open intervals. $\mathcal{B}$ is called the* Borel $\sigma$-algebra *and elements of $\mathcal{B}$ are called* Borel sets.

In other words, $\mathcal{B}$ is the collection of subsets of $\mathbb{R}$ which can be formed from open intervals by any finite sequence of countable unions, countable intersections, or complements. The $\sigma$-algebra $\mathcal{B}$ can also be described as the $\sigma$-algebra generated by open subsets of $\mathbb{R}$, or by closed intervals, or by closed subsets of $\mathbb{R}$ (see part (5) of Exercise 2.3.3 below).

---

**Exercise 2.3.3.**

(1) Let $\mathcal{A} = \{A \subset I \mid A \text{ is countable, or } A^c \text{ is countable}\}$. Prove that $\mathcal{A}$ is a $\sigma$-algebra.

(2) Let $\mathcal{A} = \{A \subset I \mid A \text{ is a null set, or } A^c \text{ is a null set}\}$. Prove that $\mathcal{A}$ is a $\sigma$-algebra.

(3) Suppose $\mathcal{A}_\lambda$ is a $\sigma$-algebra of subsets of $X$ for each $\lambda$ in some indexing set $\Lambda$. Prove that
$$\mathcal{A} = \bigcap_{\lambda \in \Lambda} \mathcal{A}_\lambda$$
is a $\sigma$-algebra of subsets of $X$.

(4) Let $\mathcal{A}$ be a $\sigma$-algebra of subsets of $\mathbb{R}$ and suppose $I$ is a closed interval which is in $\mathcal{A}$. Let $\mathcal{A}(I)$ denote the collection of all subsets of $I$ which are in $\mathcal{A}$. Prove that $\mathcal{A}(I)$ is a $\sigma$-algebra of subsets of $I$.

(5) Suppose $\mathcal{C}_1$ is the collection of closed intervals in $\mathbb{R}$, $\mathcal{C}_2$ is the collection of all open subsets of $\mathbb{R}$, and $\mathcal{C}_3$ is the collection of all closed subsets of $\mathbb{R}$. Let $\mathcal{B}_i$ be the $\sigma$-algebra generated

by $\mathcal{C}_i$. Prove that $\mathcal{B}_1, \mathcal{B}_2$, and $\mathcal{B}_3$ are all equal to the Borel $\sigma$-algebra $\mathcal{B}$.

## 2.4. Lebesgue Measure

The $\sigma$-algebra of primary interest to us is the one generated by Borel sets and null sets. Alternatively, as a consequence of Exercise 2.3.3 (5), it is the $\sigma$-algebra of subsets of $\mathbb{R}$ generated by open intervals, and null sets, or the one generated by closed intervals and null sets.

**Definition 2.4.1. (Lebesgue measurable set).** *The $\sigma$-algebra of subsets of $\mathbb{R}$ generated by open intervals and null sets will be denoted by $\mathcal{M}$. Sets in $\mathcal{M}$ will be called Lebesgue measurable, or measurable for short. If $I$ is a closed interval, then $\mathcal{M}(I)$ will denote the Lebesgue measurable subsets of $I$.*

For simplicity we will focus on subsets of $I = [0, 1]$ though we could just as well use any other interval. Notice that it is a consequence of part (4) of Exercise 2.3.3 that $\mathcal{M}(I)$ is a $\sigma$-algebra of subsets of $I$. It is by no means obvious that $\mathcal{M}$ is not the $\sigma$-algebra of all subsets of $\mathbb{R}$. However, in Appendix C we will construct a subset of $I$ which is not in $\mathcal{M}$.

We are now ready to state the main theorem of this chapter.

**Theorem 2.4.2. (Existence of Lebesgue measure).** *There exists a unique function $\mu$, called* Lebesgue measure*, from $\mathcal{M}(I)$ to the non-negative real numbers satisfying:*

- **I. Length:** *If $A = (a, b)$, then $\mu(A) = \operatorname{len}(A) = b - a$, i.e., the measure of an open interval is its length.*
- **II. Translation invariance:** *Suppose $A \in \mathcal{M}(I)$, $c \in \mathbb{R}$ and $A + c \subset I$ where $A + c$ denotes the set $\{x + c \mid x \in A\}$. Then $(A + c) \in \mathcal{M}(I)$ and $\mu(A + c) = \mu(A)$.*
- **III. Countable additivity:** *If $\{A_n\}_{n=1}^{\infty}$ is a countable collection of elements of $\mathcal{M}(I)$, then*

$$\mu\Big(\bigcup_{n=1}^{\infty} A_n\Big) \le \sum_{n=1}^{\infty} \mu(A_n)$$

*and if the sets are pairwise disjoint, then*

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

**IV. Monotonicity:** *If $A, B \in \mathcal{M}(I)$ and $A \subset B$, then $\mu(A) \leq \mu(B)$.*

**V. Null sets:** *If a subset $A \subset I$ is a null set, then $A \in \mathcal{M}(I)$ and $\mu(A) = 0$. Conversely, if $A \in \mathcal{M}(I)$ and $\mu(A) = 0$, then $A$ is a null set.*

**VI. Regularity:** *If $A \in \mathcal{M}(I)$, then*

$$\mu(A) = \inf\{\mu(U) \mid U \textit{ is open and } A \subset U\}.$$

Note that the countable additivity of property III implies the analogous statements about finite additivity. Given a finite collection $\{A_n\}_{n=1}^m$ of sets just let $A_i = \emptyset$ for $i > m$ and the analogous conclusions follow.

We have relegated the proof of this theorem to Appendix A, because it is somewhat technical and it is a diversion from our main task of developing a theory of integration. However, it is worth noting that properties I, III and VI imply the other three and we have included this as an exercise in this section.

Recall that *set difference* $A \setminus B = \{x \in A \mid x \notin B\}$. Since we are focusing on subsets of $I$ complements are with respect to $I$, so $A^c = I \setminus A$.

**Proposition 2.4.3.** *If $A$ and $B$ are in $\mathcal{M}(I)$, then $A \setminus B$ is in $\mathcal{M}(I)$ and $\mu(A \cup B) = \mu(A \setminus B) + \mu(B)$. In particular, if $I = [0, 1]$, then $\mu(I) = 1$, so $\mu(A^c) = 1 - \mu(A)$.*

**Proof.** Note that $A \setminus B = A \cap B^c$ which is in $\mathcal{M}(I)$. Also, $A \setminus B$ and $B$ are disjoint and their union is $A \cup B$. So additivity implies that $\mu(A \setminus B) + \mu(B) = \mu(A \cup B)$. Since $A^c = I \setminus A$ this implies $\mu(A \setminus I) + \mu(A) = \mu(A^c \cup A) = \mu(I) = 1$. □

We have already discussed properties I-IV and null sets, but property VI is new and it is worth discussing. It is extremely useful because it allows us to approximate arbitrary measurable sets by sets we understand better. In fact, it gives us a way to approximate any measurable set $A$ "from the outside" by a countable union of pairwise disjoint open intervals and "from the inside" by a closed set. More precisely, we have the following:

**Proposition 2.4.4. (Regularity).** *If $A \in \mathcal{M}(I)$ and $\varepsilon > 0$, then there is a closed set $C \subset A$ such that*

$$\mu(C) > \mu(A) - \varepsilon$$

*and a countable union of pairwise disjoint open intervals $U = \bigcup U_n$ such that*

$$A \subset U \quad \textit{and} \quad \mu(U) < \mu(A) + \varepsilon.$$

**Proof.** Given $\varepsilon > 0$ the existence of an open set $U$ with $A \subset U$ and $\mu(U) < \mu(A) + \varepsilon$ is exactly a restatement of property VI. Any open set $U$ is a countable union of pairwise disjoint open intervals by Theorem A.6.3.

To see the existence of $C$ let $V$ be an open set containing $A^c$ with $\mu(V) < \mu(A^c) + \varepsilon$. Then $C = V^c$ is closed and a subset of $A$. Also, $\mu(C) = 1 - \mu(V) > 1 - \mu(A^c) - \varepsilon = \mu(A) - \varepsilon$. □

If we have a countable increasing family of measurable sets, then the measure of the union can be expressed as a limit.

**Proposition 2.4.5.** *If $A_1 \subset A_2 \subset \cdots \subset A_n \ldots$ is an increasing sequence of measurable subsets of $I$, then*

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

*If $B_1 \supset B_2 \supset \cdots \supset B_n \ldots$ is a decreasing sequence of measurable subsets of $I$, then*

$$\mu(\bigcap_{n=1}^{\infty} B_n) = \lim_{n\to\infty} \mu(B_n).$$

**Proof.** Let $F_1 = A_1$ and $F_n = A_n \setminus A_{n-1}$ for $n > 1$. Then $\{F_n\}_{n=1}^{\infty}$ are pairwise disjoint measurable sets, $A_n = \bigcup_{i=1}^{n} F_i$ and

$$\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} F_i.$$

Hence, by countable additivity we have

$$\begin{aligned}\mu\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \mu\Big(\bigcup_{i=1}^{\infty} F_i\Big) &= \sum_{i=1}^{\infty} \mu(F_i) \\ &= \lim_{n\to\infty} \sum_{i=1}^{n} \mu(F_i) = \lim_{n\to\infty} \mu\Big(\bigcup_{i=1}^{n} F_i\Big) \\ &= \lim_{n\to\infty} \mu(A_n).\end{aligned}$$

For the decreasing sequence we define $E_n = B_n^c$. Then $\{E_n\}_{n=1}^{\infty}$ is an increasing sequence of measurable functions and

$$\Big(\bigcap_{n=1}^{\infty} B_n\Big)^c = \bigcup_{n=1}^{\infty} E_n.$$

Hence,

$$\begin{aligned}\mu\Big(\bigcap_{n=1}^{\infty} B_n\Big) &= 1 - \mu\Big(\bigcup_{i=1}^{\infty} E_i\Big) \\ &= 1 - \lim_{n\to\infty} \mu(E_n) \\ &= \lim_{n\to\infty} (1 - \mu(E_n)) \\ &= \lim_{n\to\infty} \mu(B_n).\end{aligned}$$

□

---

**Exercise 2.4.6.**

(1) Prove for $a, b \in I$ that $\mu([a,b]) = \mu((a,b]) = \mu([a,b)) = b-a$.

(2) Let $X$ be the subset of irrational numbers in $I$. Prove $\mu(X) = 1$. Prove that if $Y \subset I$ is a closed set and $\mu(Y) = 1$, then $Y = I$.

(3) If $A$ and $B$ are measurable subsets of $[0,1]$, prove that

$$\mu(A) + \mu(B) = \mu(A \cup B) + \mu(A \cap B).$$

(4) Prove that if $X \subset I$ is measurable, then for any $\varepsilon > 0$ there is an open set $U$ containing $X$ such that $\mu(U \setminus X) < \varepsilon$. This is sometimes referred to as the first of *Littlewood's three principles.*

(5) Suppose $a < b$ and let $\mathcal{M}([a,b])$ denote the Lebesgue measurable subsets of $[a,b]$. Define the function $f : [0,1] \to [a,b]$ by $f(x) = mx + a$ where $m = b - a$. Show that the correspondence $A \mapsto f(A)$ is a bijection from $\mathcal{M}([0,1])$ to $\mathcal{M}([a,b])$. Define the function $\mu_0 : \mathcal{M}([a,b]) \to \mathbb{R}$ by $\mu_0(A) = m\mu(f^{-1}(A))$. Prove that Theorem 2.4.2 remains valid if $I$ is replaced by $[a,b]$ and $\mu$ is replaced by $\mu_0$.

(6) The *symmetric difference* between two sets $A$ and $B$ is defined to be $(A \setminus B) \cup (B \setminus A)$. It is denoted $A \Delta B$. Suppose $A_n \subset [a,b]$ for $n \in \mathbb{N}$ is measurable and $B$ is also. Prove that if $\lim_{n\to\infty} \mu(A_n \Delta B) = 0$, then $\lim_{n\to\infty} (\mu(A_n)) = \mu(B)$.

(7) In this exercise we show that properties I, III and VI of Theorem 2.4.2 actually imply the other three properties. Let $\mu$ be a function from $\mathcal{M}(I)$ to the non-negative real numbers satisfying properties I, III and VI.
   (a) Prove that if $A, B \in \mathcal{M}(I)$ and $A \subset B$, then $\mu(A) \leq \mu(B)$, i.e., property IV is satisfied. (This only requires property III.)
   (b) Prove that if $X \subset I$ is a null set, then $X \in \mathcal{M}(I)$ and $\mu(X) = 0$. (This only requires properties I and III.)
   (c) Conversely, prove that if $X \in \mathcal{M}(I)$ and $\mu(X) = 0$, then $X$ is a null set.
   (d) Prove that $\mu$ satisfies property II.

## 2.5. The Lebesgue Density Theorem

The following theorem asserts that if a subset of an interval $I$ is "equally distributed" throughout the interval, then it must be a null set or a set of full measure, i.e., the complement of a null set. For example, it is not possible to have a set $A \subset [0,1]$ which contains half of each subinterval, i.e., it is impossible to have

$$\mu(A \cap [a,b]) = \mu([a,b])/2$$

for all $0 < a < b < 1$. There will always be small intervals with a "high concentration" of points of $A$ and other subintervals with a low concentration. Put another way, it asserts that given any $p < 1$ there is an interval $U$ such that a point in $U$ has "probability" at least $p$ of being in $A$.

**Theorem 2.5.1.** *If $A$ is a Lebesgue measurable set and $\mu(A) > 0$ and if $0 < p < 1$, then there is an open interval $U = (a, b)$ such that $\mu(A \cap U) \geq p\mu(U) = p(b - a)$.*

**Proof.** Let $p \in (0, 1)$ be given. We know from Proposition 2.4.4 that for any $\varepsilon > 0$ there is an open set $V$ which contains $A$ such that $\mu(V) < \mu(A) + \varepsilon$ and that we can express $V$ as $V = \bigcup_{n=1}^{\infty} U_n$ where $\{U_n\}_{n=1}^{\infty}$ is a countable collection of pairwise disjoint open intervals.

Then

$$\mu(A) \leq \mu(V) = \sum_{n=1}^{\infty} \operatorname{len}(U_n) < \mu(A) + \varepsilon.$$

Choosing $\varepsilon = (1 - p)\mu(A)$ we get

$$\begin{aligned}\sum_{n=1}^{\infty} \operatorname{len}(U_n) &< \mu(A) + (1 - p)\mu(A) \\ &< \mu(A) + (1 - p)\sum_{n=1}^{\infty} \operatorname{len}(U_n),\end{aligned}$$

so

$$p\sum_{n=1}^{\infty} \operatorname{len}(U_n) < \mu(A) \leq \sum_{n=1}^{\infty} \mu(A \cap U_n), \tag{2.5.1}$$

where the last inequality follows from subadditivity. Since these infinite series have finite sums, there is at least one $n_0$ such that $p\mu(U_{n_0}) \leq \mu(A \cap U_{n_0})$. This is because if it were the case that $p\mu(U_n) > \mu(A \cap U_n)$ for all $n$, then it would follow that

$$p\sum_{n=1}^{\infty} \operatorname{len}(U_n) > \sum_{n=1}^{\infty} \mu(A \cap U_n),$$

contradicting equation (2.5.1). The interval $U_{n_0}$ is the $U$ we want. □

We have shown that given $p \in (0, 1)$ as close to 1 as we like, there is an open interval in which the "relative density" of $A$ is at least $p$. It is often useful to have these intervals all centered at a particular point called a density point.

**Definition 2.5.2. (Density point).** *If $A$ is a Lebesgue measurable set and $x \in A$, then $x$ is called a* Lebesgue density point *if*

$$\lim_{\varepsilon \to 0} \frac{\mu(A \cap [x-\varepsilon, x+\varepsilon])}{\mu([x-\varepsilon, x+\varepsilon])} = 1.$$

There is a much stronger result than Theorem 2.5.1 above, which we now state, but do not prove. A proof can be found in Section 9.2 of [**T**].

**Theorem 2.5.3. (Lebesgue density theorem).** *If $A$ is a Lebesgue measurable set, then there is a subset $E \subset A$ with $\mu(E) = 0$ such that every point of $A \setminus E$ is a Lebesgue density point.*

---

**Exercise 2.5.4.**

(1) Prove that if $A \subset I = [0,1]$ has measure $\mu(A) < 1$ and $\varepsilon > 0$, then there is an interval $[a,b] \subset I$ such that $\mu(A \cap [a,b]) < \varepsilon(b-a)$.

(2) Let $A$ be a measurable set with $\mu(A) > 0$ and let

$$\Delta = \{x_1 - x_2 \mid x_1, x_2 \in A\}$$

be the set of differences of elements of $A$. Then for some $\varepsilon > 0$ the set $\Delta$ contains the interval $(-\varepsilon, \varepsilon)$.

## 2.6. Lebesgue Measurable Sets – Summary

In this section we provide a summary outline of the key properties of the collection $\mathcal{M}$ of Lebesgue measurable sets which have been developed in this chapter. If $I$ is a closed interval, then $\mathcal{M}(I)$ denotes the subsets of $I$ which are in $\mathcal{M}$.

(1) The collection of Lebesgue measurable sets $\mathcal{M}$ is a *$\sigma$-algebra*, which means:
- If $A \in \mathcal{M}$, then $A^c \in M$.
- If $A_n \in \mathcal{M}$ for $n \in \mathbb{N}$, then $\bigcup_{n=1}^{\infty} A_n \in M$.
- If $A_n \in \mathcal{M}$ for $n \in \mathbb{N}$, then $\bigcap_{n=1}^{\infty} A_n \in M$.

(2) All open sets and all closed sets are in $\mathcal{M}$. Any null set is in $\mathcal{M}$.

(3) If $I = [0,1]$ and $A \in \mathcal{M}(I)$, then there is a non-negative real number $\mu(A)$ called its Lebesgue measure which satisfies:
   - The Lebesgue measure of an interval is its length.
   - Lebesgue measure is translation invariant.
   - If $A \in \mathcal{M}(I)$, then $\mu(A^c) = 1 - \mu(A)$.
   - A set $A \in \mathcal{M}(I)$ is a null set if and only if $\mu(A) = 0$.
   - *Countable subadditivity:* If $A_n \in \mathcal{M}(I)$ for $n \in \mathbb{N}$, then
   $$\mu\big(\bigcup_{n=1}^{\infty} A_n\big) \leq \sum_{n=1}^{\infty} \mu(A_n).$$
   - *Countable additivity:* If $A_n \in \mathcal{M}(I)$ for $n \in \mathbb{N}$ are pairwise disjoint sets, then
   $$\mu\big(\bigcup_{n=1}^{\infty} A_n\big) = \sum_{n=1}^{\infty} \mu(A_n).$$
   - *Regularity:* If $A \in \mathcal{M}(I)$, then
   $$\mu(A) = \inf\{\mu(U) \mid U \text{ is open and } \ A \subset U\}.$$
   - *Increasing sequences*: If $A_n \in \mathcal{M}(I)$ for $n \in \mathbb{N}$ satisfy $A_n \subset A_{n+1}$, then
   $$\mu\big(\bigcup_{n=1}^{\infty} A_n\big) = \lim_{n\to\infty} \mu(A_n).$$
   - *Decreasing sequences*: If $A_n \in \mathcal{M}(I)$ for $n \in \mathbb{N}$ satisfy $A_n \supset A_{n+1}$, then
   $$\mu\big(\bigcap_{n=1}^{\infty} A_n\big) = \lim_{n\to\infty} \mu(A_n).$$

---

**Exercise 2.6.1. (The Cantor middle third set).** Recursively define a nested sequence $\{J_n\}_{n=0}^{\infty}$ of closed subsets of $I = [0,1]$. Each $J_n$ consists of a finite union of closed intervals. We define $J_0$ to be $I$ and let $J_n$ be the union of the closed intervals obtained by deleting the open middle third interval from each of the intervals in $J_{n-1}$.

Thus

$$\begin{aligned}
J_0 &= [0, 1],\\
J_1 &= \left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right],\\
J_2 &= \left[0, \frac{1}{9}\right] \cup \left[\frac{2}{9}, \frac{1}{3}\right] \cup \left[\frac{2}{3}, \frac{7}{9}\right] \cup \left[\frac{8}{9}, 1\right], \text{ etc.}
\end{aligned}$$

We define the *Cantor middle third set* $C$ by

$$C = \bigcap_{n=0}^{\infty} J_n.$$

(1) When the open middle thirds of the intervals in $J_{n-1}$ are removed we are left with two sets of closed intervals: the left thirds of the intervals in $J_{n-1}$ and the right thirds of these intervals. We denote the union of the left thirds by $L_n$ and the right thirds by $R_n$, and we note that $J_n = L_n \cup R_n$. Prove that $L_n$ and $R_n$ each consist of $2^{n-1}$ intervals of length $1/3^n$ and hence $J_n$ contains $2^n$ intervals of length $1/3^n$.

(2) (*Topological properties*)
   (a) Prove $C$ is compact.
   (b) A closed subset of $\mathbb{R}$ is called *nowhere dense* if it contains no non-empty open interval. Prove that $C$ is nowhere dense.
   (c) A closed subset $A$ of $\mathbb{R}$ is called *perfect* if for every $\varepsilon > 0$ and every $x \in A$ there is $y \in A$ with $x \neq y$ and $|x - y| < \varepsilon$. Prove that $C$ is perfect.

(3) Let $\mathcal{D}$ be the uncountable set of all infinite sequences

$$d_1 d_2 d_3 \dots d_n \dots,$$

where each $d_n$ is either 0 or 1 (see part (4) of Exercise A.5.12) and define a function $\psi : C \to \mathcal{D}$ by $\psi(x) = d_1 d_2 d_3 \dots d_n \dots$, where each $d_n = 0$ if $x \in L_n$ and $d_n = 1$ if $x \in R_n$. Prove that $\psi$ is surjective and hence by Corollary A.5.7 the set $C$ is uncountable. *Hint:* You will need to use Theorem A.7.3. Prove that $\psi$ is also injective and hence a bijection.

(4) Prove that $C$ is Lebesgue measurable and that $\mu(C) = 0$. *Hint:* Consider $C^c$, the complement of $C$ in $I$. Show it

is measurable and calculate $\mu(C^c)$. *Alternative hint:* Show directly that $C$ is a null set by finding for each $\varepsilon > 0$ a collection of open intervals $\{U_n\}_{n=1}^{\infty}$ such that

$$\sum_{n=1}^{\infty} \operatorname{len}(U_n) < \varepsilon \quad \text{and} \quad C \subset \bigcup_{n=1}^{\infty} U_n.$$

(5) Prove that $C$ is the subset of elements of $[0,1]$ which can be represented in base three using only the digits 0 and 2. More precisely, prove that $x \in C$ if and only if it can be expressed in the form

$$x = \sum_{n=1}^{\infty} \frac{c_n}{3^n}$$

where each $c_n$ is either 0 or 2.

Chapter 3

# The Lebesgue Integral

## 3.1. Measurable Functions

In this chapter we want to define the Lebesgue integral in a fashion which is analogous to our definitions of regulated integral and Riemann integral from Chapter 1. The difference is that we will no longer use step functions to approximate a function we want to integrate, but instead will use a much more general class called *simple functions*.

**Definition 3.1.1. (Characteristic function).** *If $A \subset [0,1]$, its* characteristic function $\mathfrak{X}_A(x)$ *(sometimes called the* indicator function*) is defined by*

$$\mathfrak{X}_A(x) = \begin{cases} 1, & \textit{if } x \in A; \\ 0, & \textit{otherwise.} \end{cases}$$

**Definition 3.1.2. (Measurable partition).** *A finite* measurable partition *of* $[0,1]$ *is a collection* $\{A_i\}_{i=1}^n$ *of measurable subsets which are pairwise disjoint and whose union is* $[0,1]$.

We can now define *simple functions.* Like step functions these functions have only finitely many values, but unlike step functions the set on which a simple function assumes a given value is no longer an interval. Instead, a simple function is constant on each subset of a finite measurable partition of $[0,1]$.

**Definition 3.1.3. (Simple function).** *A function $f : [0,1] \to \mathbb{R}$ is called* Lebesgue simple *or* simple*, for short, provided there exist a finite measurable partition of* $[0,1]$, $\{A_i\}_{i=1}^n$ *and real numbers $r_i$ such that $f(x) = \sum_{i=1}^n r_i \mathfrak{X}_{A_i}$. The* Lebesgue integral *of a simple function is defined by $\int f \ d\mu = \sum_{i=1}^n r_i \mu(A_i)$.*

Notice that the statement $f(x) = \sum_{i=1}^n r_i \mathfrak{X}_{A_i}$ just says $f(x) = r_i$ if $x \in A_i$. The definition of the integral of a simple function should come as no surprise. The fact that $\int \mathfrak{X}_A(x) \ d\mu$ is defined to be $\mu(A)$ is the generalization of the fact that the Riemann integral $\int_a^b 1 \ dx = \operatorname{len}([a,b])$. The value of $\int f \ d\mu$ for a simple function $f$ is then forced if we want our integral to have the linearity property.

**Lemma 3.1.4. (Properties of simple functions).** *The set of simple functions is a vector space and the Lebesgue integral of simple functions satisfies the following properties:*

(1) **Linearity:** *If $f$ and $g$ are simple functions and $c_1, c_2 \in \mathbb{R}$, then*
$$\int c_1 f + c_2 g \ d\mu = c_1 \int f \ d\mu + c_2 \int g \ d\mu.$$

(2) **Monotonicity:** *If $f$ and $g$ are simple and $f(x) \leq g(x)$ for all $x$, then $\int f \ d\mu \leq \int g \ d\mu$.*

(3) **Absolute value:** *If $f$ is simple, then $|f|$ is simple and $|\int f \ d\mu| \leq \int |f| \ d\mu$.*

**Proof.** If $f$ is simple, then clearly $c_1 f$ is simple. Hence, to show that simple functions form a vector space it suffices to show that the sum of two simple functions is simple.

Suppose $\{A_i\}_{i=1}^n$ and $\{B_j\}_{j=1}^m$ are measurable partitions of $[0,1]$ and that $f(x) = \sum_{i=1}^n r_i \mathfrak{X}_{A_i}$ and $g(x) = \sum_{j=1}^m s_j \mathfrak{X}_{B_j}$ are simple functions. We consider the measurable partition $\{C_{i,j}\}$ with $C_{i,j} = A_i \cap B_j$. Then $A_i = \bigcup_{j=1}^m C_{i,j}$ and $B_j = \bigcup_{i=1}^n C_{i,j}$, so
$$f(x) = \sum_{i=1}^n r_i \mathfrak{X}_{A_i} = \sum_{i=1}^n r_i \sum_{j=1}^m \mathfrak{X}_{C_{i,j}}(x) = \sum_{i,j} r_i \mathfrak{X}_{C_{i,j}}.$$

Likewise,

$$g(x) = \sum_{j=1}^{m} s_j \mathfrak{X}_{B_j} = \sum_{j=1}^{m} s_j \sum_{i=1}^{n} \mathfrak{X}_{C_{i,j}}(x) = \sum_{i,j} s_j \mathfrak{X}_{C_{i,j}}.$$

Hence, $f(x) + g(x) = \sum_{i,j}(r_i + s_j)\mathfrak{X}_{C_{i,j}}(x)$ is simple and the set of simple functions forms a vector space.

It follows immediately from the definition that if $f$ is simple and $a \in \mathbb{R}$, then $\int af\ d\mu = a \int f\ d\mu$. So to prove linearity we need only show that if $f$ and $g$ are simple functions as above, then $\int (f+g)\ d\mu = \int f\ d\mu + \int g\ d\mu$. But this follows because

$$\begin{aligned}
\int (f+g)\ d\mu &= \sum_{i,j}(r_i + s_j)\mu(C_{i,j}) \\
&= \sum_{i,j} r_i \mu(C_{i,j}) + \sum_{i,j} s_j \mu(C_{i,j}) \\
&= \sum_{i=1}^{n} r_i \sum_{j=1}^{m} \mu(C_{i,j}) + \sum_{j=1}^{m} s_j \sum_{i=1}^{n} \mu(C_{i,j}) \\
&= \sum_{i=1}^{n} r_i \mu(A_i) + \sum_{j=1}^{m} s_j \mu(B_j) \\
&= \int f\ d\mu + \int g\ d\mu.
\end{aligned}$$

Monotonicity follows from the fact that if $f$ and $g$ are simple functions with $f(x) \leq g(x)$, then $g(x) - f(x)$ is a non-negative simple function. Clearly, from the definition of the integral of a simple function, if the function is non-negative, then its integral is $\geq 0$. Thus $\int g\ d\mu - \int f\ d\mu = \int g - f\ d\mu \geq 0$.

If $f(x) = \sum r_i \mathfrak{X}_{A_i}$, the absolute value property follows from the fact that

$$\Big| \int f\ d\mu \Big| = \Big| \sum r_i \mu(A_i) \Big| \leq \sum |r_i| \mu(A_i) = \int |f|\ d\mu.$$

□

We would like to consider the measure $\mu(A)$ for any set $A \in \mathcal{M}$, not just subsets of $I$. Mostly this is straightforward, but there is one notational issue. Some subsets of $\mathbb{R}$ have infinite measure, for example, the open interval $(0, \infty)$ or $\mathbb{R}$ itself. Hence, we cannot describe $\mu$ as a real-valued function defined for any set $A \in \mathcal{M}$, because $\mu(A)$ might be infinite. There are other instances also when we want to allow the value of a function to be $+\infty$ or $-\infty$. The conventional solution is to introduce the symbols $\infty$ and $-\infty$ and to agree that a statement like $\mu(A) = \infty$ means that $A$ contains subsets of arbitrarily large finite measure.

**Definition 3.1.5. (Extended real-valued function).** *The set* $\mathbb{R} \cup \{\infty\} \cup \{-\infty\}$ *is called the* extended real numbers. *We denote it by* $[-\infty, +\infty]$. *A function* $f : X \to [-\infty, +\infty]$ *is called an* extended real-valued function.

For $a \in \mathbb{R}$ we will denote the set $(-\infty, a] \cup \{-\infty\}$ by $[-\infty, a]$ and the set $[a, \infty) \cup \{\infty\}$ by $[a, \infty]$. We will sometimes want to compare elements of $[-\infty, +\infty]$ and write inequalities relating them. Thus, for example, by convention $-\infty \leq x \leq \infty$ for every $x \in [-\infty, +\infty]$ and $y + \infty = \infty$ for every $y \in \mathbb{R}$. Likewise, $y - \infty = -\infty$ for every $y \in \mathbb{R}$, but $\infty - \infty$ is undefined.

There is an extremely important class of functions which are well behaved with respect to Lebesgue measure. Functions in this class are called *measurable functions.* Our next task is to characterize them in several ways. We will then be able to define them as the functions satisfying any one of these characterizations.

**Proposition 3.1.6. (Measurable functions).** *If* $X \subset \mathbb{R}$ *and* $f : X \to [-\infty, \infty]$ *is an extended real-valued function, then the following are equivalent:*

(1) *For any* $a \in \mathbb{R}$ *the set* $f^{-1}([-\infty, a])$ *is Lebesgue measurable.*
(2) *For any* $a \in \mathbb{R}$ *the set* $f^{-1}([-\infty, a))$ *is Lebesgue measurable.*
(3) *For any* $a \in \mathbb{R}$ *the set* $f^{-1}([a, \infty])$ *is Lebesgue measurable.*
(4) *For any* $a \in \mathbb{R}$ *the set* $f^{-1}((a, \infty])$ *is Lebesgue measurable.*

**Proof.** We will show (1) $\Rightarrow$ (2) $\Rightarrow$ (3) $\Rightarrow$ (4) $\Rightarrow$ (1).

First assume (1), then $[-\infty, a) = \bigcup_{n=1}^{\infty} [-\infty, a - 2^{-n}]$. So

$$f^{-1}([-\infty, a)) = \bigcup_{n=1}^{\infty} f^{-1}([-\infty, a - 2^{-n}]),$$

which is measurable since it is a countable union of measurable sets. Hence (2) holds.

Now assume (2), then $[a, \infty] = [-\infty, a)^c$, so

$$f^{-1}([a, \infty]) = f^{-1}([-\infty, a)^c) = (f^{-1}([-\infty, a)))^c.$$

Hence (3) holds since the complement of a measurable set is measurable.

Assuming (3) we note, $(a, \infty] = \bigcup_{n=1}^{\infty} [a + 2^{-n}, \infty]$. So

$$f^{-1}((a, \infty]) = \bigcup_{n=1}^{\infty} f^{-1}([a + 2^{-n}, \infty]),$$

which is measurable since it is a countable union of measurable sets. Hence (4) holds.

Finally, assume (4), then $[-\infty, a] = (a, \infty]^c$, so

$$f^{-1}([-\infty, a]) = f^{-1}((a, \infty]^c) = (f^{-1}((a, \infty]))^c.$$

Hence (1) holds. □

We are now ready for one of the most important definitions in this chapter.

**Definition 3.1.7. (Measurable function).** *An extended real-valued function $f$ is called Lebesgue measurable if it satisfies one (and hence all) of the properties of Proposition 3.1.6.*

It is surprising how many functions turn out to be measurable. Indeed, it is hard to find a non-measurable function! Of course, that does not relieve us of the task of proving the functions we want to make use of are measurable.

**Proposition 3.1.8.** *Suppose $f$ and $g$ are extended real-valued functions defined on $[a, b]$.*

(1) *If there is a null set $A \subset [a, b]$ such that $f(x) = 0$ if $x \notin A$, then $f$ is measurable.*

(2) *If $f = g$ except on a null set $A$, then $f$ is measurable if and only if $g$ is.*

**Proof.** For part (1) we observe that by hypothesis,

$$A \supset f^{-1}([-\infty, 0)) \cup f^{-1}((0, \infty]).$$

For $a < 0$ the set $U_a = f^{-1}([-\infty, a])$ is a subset of $A$ so $U_a$ is a null set and hence measurable. For $a \geq 0$ the set $U_a = f^{-1}([-\infty, a])$ is the complement of the null set $f^{-1}((a, \infty])$ and hence measurable. In either case $U_a$ is measurable, so $f$ is a measurable function. This proves (1).

To prove (2) we will assume $f$ is measurable and prove that $g$ is also. The other case is similar. Suppose $a \in \mathbb{R}$. We must show that the set

$$g^{-1}([a, \infty]) = (g^{-1}([a, \infty]) \cap A) \cup (g^{-1}([a, \infty]) \cap A^c)$$

is measurable. Since $f(x) = g(x)$ for all $x \in A^c$ the set

$$g^{-1}([a, \infty]) \cap A^c = f^{-1}([a, \infty]) \cap A^c$$

which is measurable. Also, the set $g^{-1}([a, \infty]) \cap A) \subset A$ is a null set, so it is measurable. Hence, $g^{-1}([a, \infty])$ is measurable. $\square$

**Theorem 3.1.9.** *Let $\{f_n\}_{n=1}^{\infty}$ be a sequence of measurable functions. Then the extended real-valued functions*

$$\begin{aligned} g_1(x) &= \sup_{n \in \mathbb{N}} f_n(x), \\ g_2(x) &= \inf_{n \in \mathbb{N}} f_n(x), \\ g_3(x) &= \limsup_{n \to \infty} f_n(x), \\ g_4(x) &= \liminf_{n \to \infty} f_n(x), \end{aligned}$$

*are all measurable. In particular, the* max *or* min *of a finite set of measurable functions is measurable.*

**Proof.** If $a \in \mathbb{R}$, then

$$\{x \mid g_1(x) > a\} = \bigcup_{n=1}^{\infty} \{x \mid f_n(x) > a\}.$$

Each of the sets on the right is measurable, so $\{x \mid g_1(x) > a\}$ is also since it is a countable union of measurable sets. Hence, $g_1$ is measurable.

Since $g_2(x) = \inf_{n \in \mathbb{N}} f_n(x) = -\sup_{n \in \mathbb{N}} -f_n(x)$ it follows that $g_2$ is also measurable.

Since the limit of a decreasing sequence is the inf of the terms,

$$g_3(x) = \limsup_{n \to \infty} f_n(x) = \inf_{m \in \mathbb{N}} \sup_{n \geq m} f_n(x).$$

It follows that $g_3$ is measurable and since

$$g_4(x) = \liminf_{n \to \infty} f_n(x) = -\limsup_{n \to \infty} -f_n(x),$$

it follows that $g_4$ is measurable. □

For the following result we need to use honest real-valued functions, i.e., not extended. The reason for this is that there is no way to define the sum of two extended real-valued functions if one has the value $+\infty$ at a point and the other has the value $-\infty$ at the same point.

**Theorem 3.1.10.** *The set of Lebesgue measurable functions from* $[0, 1]$ *to* $\mathbb{R}$ *is a vector space. The set of bounded Lebesgue measurable functions is a vector subspace. Moreover, if* $f$ *and* $g$ *are measurable, then their product* $fg$ *is measurable.*

**Proof.** It is immediate from the definition that for $c \in \mathbb{R}$ the function $cf$ is measurable when $f$ is. Suppose $f$ and $g$ are measurable. We need to show that $f + g$ is also measurable, i.e., that for any $a \in \mathbb{R}$ the set $U_a = \{x \mid f(x) + g(x) > a\}$ is measurable.

Let $\{r_n\}_{n=1}^{\infty}$ be an enumeration of the rationals. If $x_0 \in U_a$, i.e., if $f(x_0) + g(x_0) > a$, then $f(x_0) > a - g(x_0)$. Since the rationals are dense there is an $r_m$ such that $f(x_0) > r_m > a - g(x_0)$. Hence, if we define

$$V_m = \{x \mid f(x) > r_m\} \cap \{x \mid g(x) > a - r_m\},$$

then $x_0 \in V_m$. So every point of $U_a$ is in some $V_m$. Conversely, if $y_0 \in V_m$ for some $m$, then $f(y_0) > r_m > a - g(y_0)$, so $f(y_0) + g(y_0) > a$ and $y_0 \in U_a$. Thus, $U_a = \bigcup_{m=1}^{\infty} V_m$ and since each $V_m$ is measurable, we conclude that $U_a$ is measurable. This shows that $f+g$ is a measurable function and hence the measurable functions form a vector space.

Clearly, if $f$ and $g$ are bounded measurable functions and $c \in \mathbb{R}$, then $cf$ and $f+g$ are bounded. We have just shown that they are also measurable, so the bounded measurable functions are a vector subspace.

The proof that the product of measurable functions is measurable is left as an exercise (see part (5) of Exercise 3.1.11 below). □

---

**Exercise 3.1.11.**

(1) Prove that if $f$ and $g$ are simple functions, then so is $fg$. In particular, if $E \subset [0,1]$ is measurable, then $f\mathfrak{X}_E$ is a simple function.

(2) Prove that if $f$ is a continuous function, then $f$ is measurable.

(3) Prove that if $f$ is a measurable extended real-valued function, then $f^{-1}(\infty)$ and $f^{-1}(-\infty)$ are measurable.

(4) Prove that if $f$ is a measurable function, then so is $f^2$.

(5) Prove that if $f$ and $g$ are measurable functions, then so is $fg$. *Hint:* $2fg = (f+g)^2 - f^2 - g^2$.

(6) Suppose that $f : I \to \mathbb{R}$ is a simple function and $\varepsilon > 0$. Prove that there is a step function $g : I \to \mathbb{R}$ such that $\mu(E) < \varepsilon$, where $E = \{x \mid f(x) \neq g(x)\}$. *Hint:* Use part (4) of Exercise 2.4.6.

## 3.2. The Lebesgue Integral of Bounded Functions

In this section we want to define the Lebesgue integral and characterize the bounded integrable functions. In the case of the regulated integral, the integrable functions are the uniform limits of step functions. In the case of the Riemann integral a function $f$ is integrable if the *infimum* of the integrals of step functions larger than $f$ equals the *supremum* of the integrals of step functions less than $f$. It is natural to alter both of these definitions, replacing step functions with simple functions. It turns out that when we do this for bounded functions we get the *same class of integrable functions* whether we use the analog of regulated integral or the analog of Riemann integral. Moreover, this class is precisely the bounded measurable functions!

**Theorem 3.2.1.** *If $f : [0,1] \to \mathbb{R}$ is a bounded function, then the following are equivalent:*

(1) *The function $f$ is Lebesgue measurable.*

(2) *There is a sequence of simple functions $\{f_n\}_{n=1}^\infty$ which converges uniformly to $f$.*

(3) *If $\mathcal{U}_\mu(f)$ denotes the set of all simple functions $u(x)$ such that $f(x) \leq u(x)$ for all $x$ and if $\mathcal{L}_\mu(f)$ denotes the set of all simple functions $v(x)$ such that $v(x) \leq f(x)$ for all $x$, then*

$$\sup_{v \in \mathcal{L}_\mu(f)} \left\{ \int v \, d\mu \right\} = \inf_{u \in \mathcal{U}_\mu(f)} \left\{ \int u \, d\mu \right\}.$$

**Proof.** We will show (1) $\Rightarrow$ (2) $\Rightarrow$ (3) $\Rightarrow$ (1). To show (1) $\Rightarrow$ (2), assume $f$ is a bounded measurable function, say $a \leq f(x) \leq b$ for all $x \in [0,1]$.

Let $\varepsilon_n = (b-a)/n$. We will partition the *range* $[a,b]$ of $f$ by intervals as follows: Let $c_k = a + k\varepsilon_n$, so $a = c_0 < c_1 < \cdots < c_n = b$. Now define a measurable partition of $[0,1]$ by $A_k = f^{-1}([c_{k-1}, c_k))$ for $k < n$ and $A_n = f^{-1}([c_{n-1}, b])$. Then clearly $f_n(x) = \sum_{i=1}^n c_k \mathfrak{X}_{A_k}$ is a simple function. Moreover, we note that for any $x \in [0,1]$ we have $|f(x) - f_n(x)| \leq \varepsilon_n$. This is because $x$ must lie in one of the $A$'s, say $x \in A_j$. So $f_n(x) = c_j$ and $f(x) \in [c_{j-1}, c_j[$. Hence, $|f(x) - f_n(x)| \leq c_j - c_{j-1} = \varepsilon_n$. This implies that the sequence of simple functions $\{f_n\}_{n=1}^\infty$ converges uniformly to $f$.

To show (2) $\Rightarrow$ (3), assume $f$ is the uniform limit of the sequence of simple functions $\{f_n\}_{n=1}^\infty$. This means if $\delta_n = \sup\limits_{x \in [0,1]} |f(x) - f_n(x)|$, then $\lim \delta_n = 0$. We define simple functions $v_n(x) = f_n(x) - \delta_n$ and $u_n(x) = f_n(x) + \delta_n$, so $v_n(x) \leq f(x) \leq u_n(x)$.

Then

$$\begin{aligned}
\inf_{u\in\mathcal{U}_\mu(f)}\Big\{\int u\ d\mu\Big\} &\le \liminf_{n\to\infty}\int u_n\ d\mu\\
&= \liminf_{n\to\infty}\int (f_n+\delta_n)\ d\mu\\
&= \liminf_{n\to\infty}\int f_n\ d\mu\\
&\le \limsup_{n\to\infty}\int f_n\ d\mu\\
&= \limsup_{n\to\infty}\int (f_n-\delta_n)\ d\mu\\
&\le \limsup_{n\to\infty}\int v_n\ d\mu\\
&\le \sup_{v\in\mathcal{L}_\mu(f)}\Big\{\int v\ d\mu\Big\}.
\end{aligned}$$

For any $v \in \mathcal{L}_\mu(f)$ and any $u \in \mathcal{U}_\mu(f)$ we have $\int v\ d\mu \le \int u\ d\mu$, so

$$\sup_{v\in\mathcal{L}_\mu(f)}\Big\{\int v\ d\mu\Big\} \le \inf_{u\in\mathcal{U}_\mu(f)}\Big\{\int u\ d\mu\Big\}.$$

Combining this with the inequality above we conclude that

$$\sup_{v\in\mathcal{L}_\mu(f)}\Big\{\int v\ d\mu\Big\} = \inf_{u\in\mathcal{U}_\mu(f)}\Big\{\int u\ d\mu\Big\}.$$

All that remains is to show that (3) $\Rightarrow$ (1). For this we note that if (3) holds, then for any $n > 0$ there are simple functions $v_n$ and $u_n$ such that $v_n(x) \le f(x) \le u_n(x)$ for all $x$ and such that

$$\int u_n\ d\mu - \int v_n\ d\mu < 2^{-n}. \tag{3.2.1}$$

By Theorem 3.1.9 the functions

$$g_1(x) = \sup_{n\in\mathbb{N}}\Big\{v_n(x)\Big\} \quad\text{and}\quad g_2(x) = \inf_{n\in\mathbb{N}}\Big\{u_n(x)\Big\}$$

are measurable. They are also bounded and satisfy $g_1(x) \le f(x) \le g_2(x)$. We want to show that $g_1(x) = g_2(x)$ except on a set of measure zero, which we do by contradiction. Let $B = \{x \mid g_1(x) < g_2(x)\}$ and suppose $\mu(B) > 0$. Then since $B = \bigcup_{i=1}^{\infty} B_m$ where

$$B_m = \{x \mid g_1(x) < g_2(x) - \frac{1}{m}\}$$

we conclude that $\mu(B_{m_0}) > 0$ for some $m_0$. This implies that for every $n$ and every $x \in B_{m_0}$ we have $v_n(x) \leq g_1(x) < g_2(x) - \frac{1}{m_0} \leq u_n(x) - \frac{1}{m_0}$. So $u_n(x) - v_n(x) > \frac{1}{m_0}$ for all $x \in B_{m_0}$ and hence $u_n(x) - v_n(x) > \frac{1}{m_0}\mathfrak{X}_{B_{m_0}}(x)$ for all $x$. But this would mean that

$$\begin{aligned}\int u_n \; d\mu - \int v_n \; d\mu &= \int u_n - v_n \; d\mu \\ &\geq \int \frac{1}{m_0}\mathfrak{X}_{B_{m_0}} \; d\mu \\ &= \frac{1}{m_0}\mu(B_{m_0})\end{aligned}$$

for all $n$ which contradicts equation (3.2.1) above.

Hence it must be the case that $\mu(B) = 0$ so $g_1(x) = g_2(x)$ except on a set of measure zero. But since $g_1(x) \leq f(x) \leq g_2(x)$ this means that if we define $h(x) = f(x) - g_1(x)$, then $h(x)$ is zero except on a subset of $B$ which is a set of measure 0. It then follows from Proposition 3.1.8 that $h$ is a measurable function. Consequently, $f(x) = g_1(x) + h(x)$ is also measurable and we have completed the proof that (3) $\Rightarrow$ (1). $\square$

We can now provide the (long awaited) definition of the Lebesgue integral, at least for bounded functions.

**Definition 3.2.2. (Lebesgue integral of a bounded function).** *If $f : [0,1] \to \mathbb{R}$ is a bounded measurable function, then we define its Lebesgue integral by*

$$\int f \; d\mu = \sup_{v \in \mathcal{L}_\mu(f)} \Big\{ \int v \; d\mu \Big\},$$

*or equivalently (by Theorem 3.2.1),*

$$\int f \; d\mu = \inf_{u \in \mathcal{U}_\mu(f)} \Big\{ \int u \; d\mu \Big\}.$$

Alternatively, as the following proposition shows, we could have defined it to be the limit of the integrals of a sequence of simple functions converging uniformly to $f$.

**Proposition 3.2.3.** *If $\{g_n\}_n^\infty$ is any sequence of simple functions converging uniformly to a bounded measurable function $f$, then*

$$\lim_{n\to\infty} \int g_n \, d\mu = \int f \, d\mu.$$

**Proof.** If we let $\delta_n = \sup_{x\in[0,1]} |f(x) - g_n(x)|$, then $\lim \delta_n = 0$ and

$$g_n(x) - \delta_n \le f(x) \le g_n(x) + \delta_n.$$

So $g_n - \delta_n \in \mathcal{L}_\mu(f)$ and $g_n + \delta_n \in \mathcal{U}_\mu(f)$. Hence,

$$\begin{aligned}
\int f \, d\mu &= \inf_{u\in\mathcal{U}_\mu(f)} \int u \, d\mu \\
&\le \liminf_{n\to\infty} \int (g_n + \delta_n) \, d\mu \\
&= \liminf_{n\to\infty} \int g_n \, d\mu \\
&\le \limsup_{n\to\infty} \int g_n \, d\mu \\
&\le \limsup_{n\to\infty} \int (g_n - \delta_n) \, d\mu \\
&\le \sup_{v\in\mathcal{L}_\mu(f)} \left\{ \int v \, d\mu \right\} \\
&= \int f \, d\mu.
\end{aligned}$$

Hence, these inequalities must be equalities and

$$\lim_{n\to\infty} \int g_n \, d\mu = \int f \, d\mu.$$

□

We proved in Theorem 3.1.10 that the bounded Lebesgue measurable functions are a vector space and we have now defined the Lebesgue integral as a function from this vector space to the real numbers. Next we want to check some of the “Basic Properties” listed in Section 1.2 of Chapter 1 as necessary for any well-behaved integral. One difference is that we have improved the "finite sets don't matter" property to "null sets don't matter."

**Theorem 3.2.4.** *Suppose $f$ and $g$ are bounded Lebesgue measurable functions defined on $[0,1]$. Then the Lebesgue integral satisfies the following properties:*

- **I. Linearity:** *If $c_1, c_2 \in \mathbb{R}$, then*
$$\int c_1 f + c_2 g \ d\mu = c_1 \int f \ d\mu + c_2 \int g \ d\mu.$$
- **II. Monotonicity:** *If $f(x) \leq g(x)$ for all $x$, then $\int f \ d\mu \leq \int g \ d\mu$.*
- **III. Absolute value:** *The fact that $f$ is Lebesgue measurable implies that $|f|$ is also, and $|\int f \ d\mu| \leq \int |f| \ d\mu$.*
- **IV. Null sets:** *If $f(x) = g(x)$ except on a set of measure zero, then $\int f \ d\mu = \int g \ d\mu$.*

**Proof.** If $f$ and $g$ are measurable, there exist sequences of simple functions $\{f_n\}_{n=1}^{\infty}$ and $\{g_n\}_{n=1}^{\infty}$ converging uniformly to $f$ and $g$ respectively. This implies that the sequence $\{c_1 f_n + c_2 g_n\}_{n=1}^{\infty}$ converges uniformly to the bounded measurable function $c_1 f + c_2 g$. The fact that

$$\begin{aligned}\int c_1 f + c_2 g \ d\mu &= \lim_{n\to\infty} \int (c_1 f_n + c_2 g_n) \ d\mu \\ &= c_1 \lim_{n\to\infty} \int f_n \ d\mu \ + \ c_2 \lim_{n\to\infty} \int g_n \ d\mu \\ &= c_1 \int f \ d\mu \ + \ c_2 \int g \ d\mu\end{aligned}$$

implies the linearity property.

Similarly, the absolute value property follows from Lemma 3.1.4 because

$$\left| \int f \ d\mu \right| = \lim_{n\to\infty} \left| \int f_n \ d\mu \right| \leq \lim_{n\to\infty} \int |f_n| \ d\mu = \int |f| \ d\mu.$$

To show monotonicity we use the definition of the Lebesgue integral. If $f(x) \leq g(x)$, then

$$\int f \ d\mu = \sup_{v \in \mathcal{L}_\mu(f)} \left\{ \int v \ d\mu \right\} \leq \inf_{u \in \mathcal{U}_\mu(g)} \left\{ \int u \ d\mu \right\} = \int g \ d\mu.$$

To show the null set property let $h(x) = f(x) - g(x)$. Then $h$ is a bounded measurable function which is 0 except on a set $E$ with $\mu(E) = 0$. Hence,

$$\Big| \int f \, d\mu - \int g \, d\mu \Big| = \Big| \int h \, d\mu \Big| \leq \int |h| \, d\mu.$$

But the function $h$ is bounded, say $|h(x)| \leq M$. So $|h(x)| \leq M\mathfrak{X}_E(x)$ and by monotonicity

$$\int |h| \, d\mu \leq \int M\mathfrak{X}_E \, d\mu = M\mu(E) = 0.$$

It follows that $|\int f \, d\mu - \int g \, d\mu| = 0$, so $\int f \, d\mu = \int g \, d\mu$. □

With the regulated and Riemann integrals we could integrate over subintervals of $[0, 1]$. We can do that also with the Lebesgue integral. Indeed, we can do better and integrate over any measurable subset of $[0, 1]$.

**Definition 3.2.5.** *If $E \subset [0, 1]$ is a measurable set and $f$ is a bounded measurable function, we define the* Lebesgue integral *of $f$ over $E$ by*

$$\int_E f \, d\mu = \int f\mathfrak{X}_E \, d\mu.$$

**Proposition 3.2.6. (Additivity).** *Suppose $\{E_n\}_{n=1}^N$ is a collection of pairwise disjoint measurable subsets of $I$, and $f$ is a bounded measurable function. If $E = \bigcup_{n=1}^N E_n$, then*

$$\int_E f \, d\mu = \sum_{n=1}^N \int_{E_n} f \, d\mu.$$

In fact, this result is true for a countable collection $\{E_n\}_{n=1}^\infty$, but we postpone the proof of that until we can do it in the more general setting of not necessarily bounded functions.

**Proof.** Note that

$$\mathfrak{X}_E = \sum_{n=1}^N \mathfrak{X}_{E_n}, \quad \text{so} \quad f\mathfrak{X}_E = \sum_{n=1}^{\mathbb{N}} f\mathfrak{X}_{E_n}$$

and the result follows by linearity of the integral. □

It should be clear that there are many Lebesgue integrable functions which are not Riemann integrable (for example, if $A = \mathbb{Q} \cap I$, then $\mathfrak{X}_A$ is not Riemann integrable). We also need to show that in passing to the Lebesgue integral we do not lose anything.

**Proposition 3.2.7. (Riemann integrable implies Lebesgue integrable).** *Every bounded Riemann integrable function $f : [0,1] \to \mathbb{R}$ is measurable and hence Lebesgue integrable. The values of the Riemann and Lebesgue integrals coincide.*

**Proof.** The set $\mathcal{U}(f)$ of step functions greater than $f$ is a subset of the set $\mathcal{U}_\mu(f)$ of simple functions greater than $f$. Likewise, the set $\mathcal{L}(f) \subset \mathcal{L}_\mu(f)$. Hence,

$$\begin{aligned}
\sup_{v \in \mathcal{L}(f)} \left\{ \int_0^1 v(t)\, dt \right\} &\leq \sup_{v \in \mathcal{L}_\mu(f)} \left\{ \int v \, d\mu \right\} \\
&\leq \inf_{u \in \mathcal{U}_\mu(f)} \left\{ \int u \, d\mu \right\} \\
&\leq \inf_{u \in \mathcal{U}(f)} \left\{ \int_0^1 u(t)\, dt \right\}.
\end{aligned}$$

The fact that $f$ is Riemann integrable asserts that the first and last of these values are equal. Hence they are all equal. By Theorem 3.2.1 $f$ is measurable. The fact that the Riemann and Lebesgue integrals coincide now follows from their definitions and the equality of all the values in the inequality above. □

---

**Exercise 3.2.8.**

(1) Since simple functions are themselves bounded measurable functions, we have actually given two definitions of their Lebesgue integral: the one in Definition 3.1.3 and the one above in Definition 3.2.2. Prove that these definitions give the same value.

(2) Suppose $f$ and $g$ are measurable functions defined on $[0,1]$ and suppose $|f(x)| < M$ and $|g(x)| < M$ for all $x$. Let $E$ denote the set of $x$ such that $|f(x) - g(x)| \geq \varepsilon$. Prove that

$$\int |f(x) - g(x)| \, d\mu < \varepsilon + 2M\mu(E).$$

## 3.3. The Bounded Convergence Theorem

We want to investigate the question, "When does the fact that a sequence of functions $\{f_n\}_{n=1}^\infty$ converges pointwise to a function $f$ imply that their Lebesgue integrals converge to the integral of $f$?" This general question is an important one which will occupy much of the remainder of this chapter and a substantial part of the next chapter. It is straightforward to prove that if a sequence of bounded measurable functions converges uniformly to $f$, then their integrals converge to the integral of $f$. We will not do this now, because we prove a stronger result later; but first we consider an example which shows what can go wrong when we only have pointwise convergence.

**Example 3.3.1.** Let

$$f_n(x) = \begin{cases} n, & \text{if } x \in [\frac{1}{n}, \frac{2}{n}]; \\ 0, & \text{otherwise.} \end{cases}$$

Then $f_n$ is a step function equal to $n$ on an interval of length $\frac{1}{n}$ and 0 elsewhere. Thus $\int f_n \, d\mu = n\frac{1}{n} = 1$; but, for any $x \in [0,1]$ we have $f_n(x) = 0$ for all sufficiently large $n$. Thus, the sequence $\{f_n\}_{n=1}^\infty$ converges *pointwise* to the constant function 0. Hence

$$\int (\lim_{n\to\infty} f_n(x)) \, d\mu = 0 \quad \text{and} \quad \lim_{n\to\infty} \int f_n \, d\mu = 1.$$

In this example each $f_n$ is a bounded step function, but there is no single bound which works for all $f_n$ since the maximum value of $f_n$ is $n$. It turns out that any example of this sort must be a sequence of functions which is not uniformly bounded.

**Theorem 3.3.2. (Bounded convergence theorem).** *Let $\{f_n\}_{n=1}^\infty$ be a sequence of measurable functions defined on $[0,1]$ which converges pointwise to a function $f : [0,1] \to \mathbb{R}$. Suppose there is a constant $M > 0$ such that $|f_n(x)| \leq M$ for all $n$ and all $x \in [0,1]$. Then $f$ is a bounded measurable function and*

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

**Proof.** For each $x \in [0,1]$ we know that $\lim_{m\to\infty} f_m(x) = f(x)$. This implies that $|f(x)| \leq M$. Since $f$ is a limit of measurable functions, it is measurable by Theorem 3.1.9.

We must show that

$$\lim_{n\to\infty} \left| \int f_n \, d\mu - \int f \, d\mu \right| = 0,$$

but

$$\begin{aligned} \lim_{n\to\infty} \left| \int f_n \, d\mu - \int f \, d\mu \right| &= \lim_{n\to\infty} \left| \int (f_n - f) \, d\mu \right| \\ &\leq \lim_{n\to\infty} \int |f_n - f| \, d\mu. \end{aligned} \tag{3.3.1}$$

So we need to estimate the integral of $|f_n - f|$.

Given $\varepsilon > 0$, define

$$E_n = \big\{ x \mid |f_m(x) - f(x)| < \varepsilon/2 \text{ for all } m \geq n \big\}.$$

Notice that if for some $n$ the set $E_n$ were all of $[0,1]$ we would be able to estimate $\int |f_m - f| \, d\mu \leq \int \varepsilon/2 \, d\mu = \varepsilon/2$ for all $m \geq n$; but we don't know that. Instead, we know that for any $x$ the limit $\lim_{m\to\infty} f_m(x) = f(x)$, which means that each $x$ is in some $E_n$ (where $n$ depends on $x$). In other words, $\bigcup_{n=1}^{\infty} E_n = [0,1]$.

Since $E_n \subset E_{n+1}$ we know that $\lim_{n\to\infty} \mu(E_n) = \mu([0,1]) = 1$ by Proposition 2.4.5. Thus, there is an $n_0$ such that

$$\mu(E_{n_0}) > 1 - \frac{\varepsilon}{4M}, \quad \text{and hence} \quad \mu(E_{n_0}^c) < \frac{\varepsilon}{4M}.$$

Now for any $n > n_0$ we have

$$\begin{aligned} \int |f_n - f| \, d\mu &= \int_{E_{n_0}} |f_n - f| \, d\mu + \int_{E_{n_0}^c} |f_n - f| \, d\mu \\ &\leq \int_{E_{n_0}} \frac{\varepsilon}{2} \, d\mu + \int_{E_{n_0}^c} 2M \, d\mu \\ &\leq \frac{\varepsilon}{2} \mu(E_{n_0}) + 2M\mu(E_{n_0}^c) \\ &\leq \frac{\varepsilon}{2} + 2M \frac{\varepsilon}{4M} = \varepsilon. \end{aligned}$$

Thus, we have shown $\lim_{n\to\infty} \int |f_n - f| \, d\mu = 0$. Putting this together with the inequality from equation (3.3.1) we see that

$$\lim_{n\to\infty} \left| \int f_n \, d\mu - \int f \, d\mu \right| = 0,$$

as desired. □

**Definition 3.3.3. (Almost everywhere).** *If a property holds for all $x$ except for a set of measure zero, we say that it holds* almost everywhere *or for* almost all *values of $x$.*

For example, we say that two functions $f$ and $g$ defined on $[0,1]$ are equal *almost everywhere* if the set of $x$ with $f(x) \neq g(x)$ has measure zero. The last part of Theorem 3.2.4 asserted that if $f(x) = g(x)$ almost everywhere, then $\int f \, d\mu = \int g \, d\mu$. As another example, we say $\lim_{n\to\infty} f_n(x) = f(x)$ for *almost all $x$* if the set of $x$, where the limit does not exist or is not equal to $f(x)$, is a set of measure zero.

We can now state an improved version of Theorem 3.3.2, the bounded convergence theorem. In this version the hypotheses are weaker in that we require them to hold only almost everywhere. It is important to understand both when and why "almost everywhere" is as good as everywhere. In the following theorem we provide all the details in this transition, but through most of the remainder of the text we will assume the reader is facile in justifying why a hypothesis of almost everywhere is sufficient.

**Theorem 3.3.4. (Better bounded convergence theorem).** *Suppose $\{f_n\}_{n=1}^\infty$ is a sequence of bounded measurable functions defined on $[0,1]$ and $f : [0,1] \to \mathbb{R}$ is a bounded function such that*

$$\lim_{n\to\infty} f_n(x) = f(x)$$

*for almost all $x$. Suppose also there is a constant $M > 0$ such that for each $n > 0$, $|f_n(x)| \leq M$ almost everywhere. Then $f$ is a measurable function and*

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

**Proof.** Let $A = \{x \mid \lim_{n\to\infty} f_n(x) \neq f(x)\}$, then $\mu(A) = 0$. Define the set $D_n = \{x \mid |f_n(x)| > M\}$. Then $\mu(D_n) = 0$, so if

$$E = A \cup \bigcup_{n=1}^{\infty} D_n,$$

then $\mu(E) = 0$. Let

$$\begin{aligned} g_n(x) &= f_n(x)\mathfrak{X}_{E^c}(x) \\ &= \begin{cases} f_n(x), & \text{if } x \notin E; \\ 0, & \text{if } x \in E. \end{cases} \end{aligned}$$

The function $g_n$ is a product of measurable functions, so by Theorem 3.1.10 it is measurable.

Define the function $g$ by $g(x) = f(x)\mathfrak{X}_{E^c}(x)$. Note that $|g_n(x)| \leq M$ for *all* $x \in [0,1]$ and $\lim_{n\to\infty} g_n(x) = g(x)$ for *all* $x \in [0,1]$. This is true because if $x \in E$, both $g_n(x)$ and $g(x)$ are 0 and if $x \in E^c$, then

$$\lim_{n\to\infty} g_n(x) = \lim_{n\to\infty} f_n(x) = f(x) = g(x).$$

We conclude that $|g(x)| \leq M$ since $|g_n(x)| \leq M$. Since $g$ is a limit of measurable functions it is measurable by Theorem 3.1.9.

Finally, note from their definitions that $g(x) = f(x)$ almost everywhere and $g_n(x) = f_n(x)$ almost everywhere. It follows that $f$ is measurable by Proposition 3.1.8. By the null set property from Theorem 3.2.4

$$\int f\,d\mu = \int g\,d\mu \quad \text{and} \quad \int f_n\,d\mu = \int g_n\,d\mu.$$

Hence, it will suffice to show that

$$\lim_{n\to\infty} \int g_n\,d\mu = \int g\,d\mu,$$

but this is true by the bounded convergence theorem, Theorem 3.3.2. $\square$

---

**Exercise 3.3.5.**

(1) Suppose $\{f_n\}$ is a sequence of measurable functions and $\lim_{n\to\infty} f_n(x) = f(x)$ almost everywhere. Prove that $f$ is measurable.

(2) A function $f$ is called *essentially bounded* provided there exists $M > 0$ such that $|f(x)| \leq M$ almost everywhere. The number $M$ is called an *essential bound.* Prove that essentially bounded measurable functions are a vector space. Prove that the product of essentially bounded functions is essentially bounded.

(3) Formulate a definition of the Lebesgue integral of an essentially bounded function in terms of the integral of a bounded function. Prove that Theorem 3.2.4 remains valid for essentially bounded functions with your definition.
*Hint:* Show that if $M$ is an essential bound for $f$ and

$$f_M(x) = \begin{cases} M, \text{ if } f(x) > M; \\ f(x), \text{ if } -M \leq f(x) \leq M; \\ -M \text{ if } f(x) < -M, \end{cases}$$

then $\int f_M \, d\mu$ is independent of $M$.

(4) A sequence $\{f_n\}$ with the property that

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

is said to *converge in the mean* to the function $f$. Suppose $\{f_n\}$ is a sequence of measurable functions satisfying $|f_n(x)| < M$ for all $n$ and all $x \in [0, 1]$ and suppose that this sequence converges pointwise to the function $f$. Prove that the sequence converges in the mean to $f$.

(5) In the previous exercise we showed that a uniformly bounded sequence of measurable functions which converges pointwise to $f$ also converges in the mean to $f$. This exercise is intended to show that the converse of this statement is not true.

Let $P$ be the countable set $\{(p, q) \mid p, q \in \mathbb{N},\ p < q\}$ and define the function

$$f_{(p,q)}(x) = \begin{cases} 1, & \text{if } x \in [\frac{p-1}{q}, \frac{p+1}{q}]; \\ 0, & \text{otherwise.} \end{cases}$$

Let $\phi : \mathbb{N} \to P$ be a bijection and define $g_n = f_{\phi(n)}$ for $n \in \mathbb{N}$.

(a) Prove that

$$\lim_{n\to\infty} \int |g_n| \, d\mu = 0.$$

Hence, the sequence $\{g_n\}$ converges in the mean to the constant function 0.

(b) Prove that for any $x \in [0,1]$ there is a subsequence $\{n_i\}_{i=1}^{\infty}$ of the natural numbers such that $g_{n_i}(x) = 1$. Hence there is no $x \in [0,1]$ for which

$$\lim_{n\to\infty} g_n(x) = 0.$$

Therefore, $\{g_n\}$ clearly fails to converge pointwise to 0.

(c) Prove there is a subsequence of $\{g_n\}$ which does converge pointwise to 0.

Chapter 4

# The Integral of Unbounded Functions

In this section we wish to define and investigate the Lebesgue integral of functions which are not necessarily bounded and even the integral of extended real-valued functions. In fact, henceforth, we will use the term "measurable function" to refer to extended real-valued measurable functions. If a function is unbounded both above and below it is more complicated than if it is only unbounded above. Hence, we first focus our attention on the simpler case.

## 4.1. Non-negative Functions

**Definition 4.1.1. (Integrable function).** *If $f : [0,1] \to [0,\infty]$ is a non-negative Lebesgue measurable function, we let*

$$f_n(x) = \min\{f(x), n\}.$$

*Then $f_n$ is a bounded measurable function and we define*

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

*If $\int f \, d\mu < \infty$, we say $f$ is* integrable.

Notice that the sequence $\{\int f_n \, d\mu\}_{n=1}^{\infty}$ is a monotonic increasing sequence of numbers, so the limit $\lim_{n\to\infty} \int f_n \, d\mu$ either exists or is $+\infty$.

**Proposition 4.1.2.** *If $f$ is a non-negative integrable function and $A = \{x \mid f(x) = +\infty\}$, then $\mu(A) = 0$.*

**Proof.** For $x \in A$ we observe that $f_n(x) = n$ and hence $f_n(x) \geq n\mathfrak{X}_A(x)$ for all $x$. Thus, $\int f_n \, d\mu \geq \int n\mathfrak{X}_A \, d\mu = n\mu(A)$. If $\mu(A) > 0$, then

$$\int f \, d\mu = \lim_{n\to\infty} \int f_n \, d\mu \geq \lim_{n\to\infty} n\mu(A) = +\infty.$$

□

**Example 4.1.3.** Let $f(x) = 1/\sqrt{x}$ for $x \in (0, 1]$ and let $f(0) = +\infty$, so $f$ is a non-negative measurable function. Then the function

$$f_n(x) = \begin{cases} n, & \text{if } 0 \leq x < \frac{1}{n^2}; \\ \frac{1}{\sqrt{x}}, & \text{if } \frac{1}{n^2} \leq x \leq 1. \end{cases}$$

Hence, if $E_n = [0, 1/n^2)$, then

$$\begin{aligned} \int f_n \, d\mu &= \int_{E_n} f_n \, d\mu + \int_{E_n^c} f_n \, d\mu \\ &= \int n\mathfrak{X}_{E_n} \, d\mu + \int_{\frac{1}{n^2}}^{1} \frac{1}{\sqrt{x}} \, dx \\ &= n\mu(E_n) + \left(2 - \frac{2}{n}\right) \\ &= \frac{n}{n^2} + 2 - \frac{2}{n} = 2 - \frac{1}{n}. \end{aligned}$$

Hence,

$$\int f \, d\mu = \lim_{n\to\infty} \int f_n \, d\mu = 2.$$

So $f$ is integrable.

**Proposition 4.1.4.** *Suppose $f$ and $g$ are non-negative measurable functions with $g(x) \leq f(x)$ for almost all $x$. If $f$ is integrable, then $g$ is integrable and $\int g \, d\mu \leq \int f \, d\mu$. In particular, if $g = 0$ almost everywhere, then $\int g \, d\mu = 0$.*

**Proof.** If $f_n(x) = \min\{f(x), n\}$ and $g_n(x) = \min\{g(x), n\}$, then $f_n$ and $g_n$ are bounded measurable functions and satisfy $g_n(x) \leq f_n(x)$ for almost all $x$. It follows that $\int g_n \, d\mu \leq \int f_n \, d\mu \leq \int f \, d\mu$. Since the sequence of numbers $\{\int g_n \, d\mu\}_{n=1}^{\infty}$ is monotonic increasing and bounded above by $\int f \, d\mu$ it has a finite limit. By definition this limit is $\int g \, d\mu$. Since for each $n$ we have $\int g_n \, d\mu \leq \int f \, d\mu$, the limit is also bounded by $\int f \, d\mu$. That is,

$$\int g \, d\mu = \lim_{n\to\infty} \int g_n \, d\mu \leq \int f \, d\mu.$$

If $g = 0$ almost everywhere, then $0 \leq g(x) \leq 0$ for almost all $x$, so we have $\int g \, d\mu = 0$. □

**Corollary 4.1.5.** *If $f : [0, 1] \to [0, \infty]$ is a non-negative integrable function and $\int f \, d\mu = 0$, then $f(x) = 0$ for almost all $x$.*

**Proof.** Let $E_n = \{x \mid f(x) \geq 1/n\}$. Then $f(x) \geq \frac{1}{n}\mathfrak{X}_{E_n}(x)$, so

$$\frac{1}{n}\mu(E_n) = \int \frac{1}{n}\mathfrak{X}_{E_n} \, d\mu \leq \int f \, d\mu = 0.$$

Hence, $\mu(E_n) = 0$, but if $E = \{x \mid f(x) > 0\}$, then $E = \bigcup_{n=1}^{\infty} E_n$, so $\mu(E) = 0$. □

For a fixed non-negative integrable function $f$ the Lebesgue integral $\int_E f \, d\mu$ assigns a number to each measurable subset $E$ of $[0, 1]$. This assignment satisfies a strong continuity property which we will need in the next section.

**Theorem 4.1.6. (Absolute continuity).** *If $f$ is a non-negative integrable function, then for any $\varepsilon > 0$ there exists a $\delta > 0$ such that $\int_A f \, d\mu < \varepsilon$ for every measurable $A \subset [0, 1]$ with $\mu(A) < \delta$.*

**Proof.** Let $f_n(x) = \min\{f(x), n\}$ so $\lim \int f_n \, d\mu = \int f \, d\mu$. Let

$$E_n = \{x \in [0, 1] \mid f(x) \geq n\}$$

so

$$f_n(x) = \begin{cases} n, & \text{if } x \in E_n; \\ f(x), & \text{if } x \in E_n^c. \end{cases}$$

Therefore

$$f(x) - f_n(x) = \begin{cases} f(x) - n, & \text{if } x \in E_n; \\ 0, & \text{if } x \in E_n^c. \end{cases}$$

Consequently,

$$\int f \, d\mu - \int f_n \, d\mu = \int f - f_n \, d\mu = \int_{E_n} (f - n) \, d\mu.$$

Integrability of $f$ implies

$$\lim_{n\to\infty} \Big( \int f \, d\mu - \int f_n \, d\mu \Big) = 0,$$

so

$$\lim_{n\to\infty} \int_{E_n} (f - n) \, d\mu = 0.$$

Hence, we may choose $N$ such that $\int_{E_N} (f - N) \, d\mu < \varepsilon/2$. Now pick $\delta < \varepsilon/2N$. Then if $\mu(A) < \delta$, we have

$$\begin{aligned} \int_A f \, d\mu &= \int_{A\cap E_N} f \, d\mu + \int_{A\cap E_N^c} f \, d\mu \\ &\leq \int_{A\cap E_N} (f - N) \, d\mu + \int_{A\cap E_N} N \, d\mu + \int_{A\cap E_N^c} N \, d\mu \\ &\leq \int_{E_N} (f - N) \, d\mu + \int_A N \, d\mu \\ &< \frac{\varepsilon}{2} + N\mu(A) < \frac{\varepsilon}{2} + N\delta < \varepsilon. \end{aligned}$$

□

Theorem 4.1.6 is labeled “Absolute Continuity” for reasons that will become clear later in Section 4.3. But as a nearly immediate consequence we have the following continuity result which is a generalization of a result from Exercise 1.6.3 about regulated functions.

**Corollary 4.1.7. (Continuity of the integral).** *If $f : [0,1] \to [0,\infty]$ is a non-negative integrable function and we define $F(x) = \int_{[0,x]} f \, d\mu$, then $F(x)$ is continuous.*

**Proof.** Given $\varepsilon > 0$ let $\delta > 0$ be the corresponding value guaranteed by Theorem 4.1.6. Now suppose $x < y$ and $|y - x| < \delta$. Then $\mu([x, y]) < \delta$, so

$$\big|F(y) - F(x)\big| = \left|\int_{[0,y]} f\, d\mu - \int_{[0,x]} f\, d\mu\right| = \left|\int_{[x,y]} f\, d\mu\right| < \varepsilon$$

by Theorem 4.1.6. In fact, we have proven that $F$ is uniformly continuous. □

---

**Exercise 4.1.8.**

(1) Define $f_p(x) = \frac{1}{x^p}$ for $x \in (0, 1]$ and $f_p(0) = +\infty$. Prove that $f$ is integrable if and only if $p < 1$. For $p < 1$ calculate the value of $\int f_p\, d\mu$.

(2) Give an example of a non-negative extended real-valued function $g : [0, 1] \to [0, \infty]$ which is integrable and which has the value $+\infty$ at infinitely many points of $[0, 1]$.

(3) Suppose $f$ and $g$ are non-negative integrable functions on $[0, 1]$. Suppose that for every measurable $E \subset [0, 1]$ the integrals $\int_E f\, d\mu$ and $\int_E g\, d\mu$ are equal. Prove $f$ and $g$ are equal almost everywhere.

## 4.2. Convergence Theorems

The following result is very similar to the bounded convergence theorem (see Theorem 3.3.2 and Theorem 3.3.4). The difference is that instead of having a constant bound on the functions $f_n$ we have them bounded by an integrable function $g$. This is enough to make essentially the same proof work.

**Theorem 4.2.1. (Non -negative Lebesgue convergence theorem).** *Suppose $f_n$ is a sequence of non-negative measurable functions defined on $[0, 1]$ and $g$ is a non-negative integrable function such that $f_n(x) \le g(x)$ for all $n$ and almost all $x$. If $\lim f_n(x) = f(x)$ for almost all $x$, then $f$ is integrable and*

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

**Proof.** If we let $h_n = f_n \mathfrak{X}_E$ and $h = f\mathfrak{X}_E$, where

$$E = \{x \mid \lim f_n(x) = f(x)\},$$

then $f = h$ almost everywhere and $f_n = h_n$ almost everywhere. So it suffices to prove

$$\int h \, d\mu = \lim_{n\to\infty} \int h_n \, d\mu,$$

and we now have the stronger property that $\lim h_n(x) = h(x)$ for *all* $x$, instead of almost all. Since $h_n(x) = f_n(x)\mathfrak{X}_E(x) \leq g(x)$ for almost all $x$ we know that $h(x) \leq g(x)$ for almost all $x$ and hence by Proposition 4.1.4 that $h$ is integrable.

The remainder of the proof is very similar to the proof of Theorem 3.3.2. We must show that

$$\lim_{n\to\infty} \left| \int h_n \, d\mu - \int h \, d\mu \right| = 0;$$

but

$$\left| \int h_n \, d\mu - \int h \, d\mu \right| = \left| \int (h_n - h) \, d\mu \right| \leq \int |h_n - h| \, d\mu. \tag{4.2.1}$$

So we need to estimate the integral of $|h_n - h|$.

Given $\varepsilon > 0$ define $E_n = \{x \mid |h_m(x) - h(x)| < \varepsilon/2 \text{ for all } m \geq n\}$. We know by Theorem 4.1.6 that there is a $\delta > 0$ such that $\int_A g \, d\mu < \varepsilon/4$ whenever $\mu(A) < \delta$.

We also know that for any $x$ the limit $\lim_{m\to\infty} h_m(x) = h(x)$, which means that each $x$ is in some $E_n$ (where $n$ depends on $x$). In other words, $\bigcup_{n=1}^{\infty} E_n = [0,1]$. Since $E_n \subset E_{n+1}$ by Proposition 2.4.5 we know that $\lim_{n\to\infty} \mu(E_n) = \mu([0,1]) = 1$. Thus, there is an $n_0$ such that $\mu(E_{n_0}) > 1 - \delta$, so $\mu(E_{n_0}^c) < \delta$.

Now

$$|h_n(x) - h(x)| \leq |h_n(x)| + |h(x)| \leq 2g(x)$$

for almost all $x$, so for any $n > n_0$ we have

$$\begin{aligned}\int |h_n - h| \, d\mu &= \int_{E_{n_0}} |h_n - h| \, d\mu + \int_{E_{n_0}^c} |h_n - h| \, d\mu \\ &\le \int_{E_{n_0}} \frac{\varepsilon}{2} \, d\mu + \int_{E_{n_0}^c} 2g \, d\mu \\ &\le \frac{\varepsilon}{2}\mu(E_{n_0}) + 2\int_{E_{n_0}^c} g \, d\mu \\ &\le \frac{\varepsilon}{2} + 2\frac{\varepsilon}{4} = \varepsilon.\end{aligned}$$

Thus, we have shown $\lim_{n\to\infty} \int |h_n - h| \, d\mu = 0$. Putting this together with equation (4.2.1) we see that

$$\lim_{n\to\infty} \left| \int h_n \, d\mu - \int h \, d\mu \right| = 0$$

as desired. □

**Theorem 4.2.2. (Fatou's lemma).** *Suppose $g_n$ is a sequence of non-negative measurable functions defined on $[0,1]$. If $\lim g_n(x) = f(x)$ for almost all $x$, then*

$$\int f \, d\mu \le \liminf_{n\to\infty} \int g_n \, d\mu.$$

*In particular, if $\liminf \int g_n \, d\mu < +\infty$, then $f$ is integrable.*

**Proof.** The function $f(x)$ is measurable by Theorem 3.1.9. Let $h(x)$ be a bounded measurable function with $h(x) \le f(x)$ for all $x$ and define $h_n(x) = \min\{h(x), g_n(x)\}$, so $h_n$ is bounded and measurable and $\lim_{n\to\infty} h_n(x) = h(x)$. Then

$$\int h \, d\mu = \lim_{n\to\infty} \int h_n \, d\mu \le \liminf_{n\to\infty} \int g_n \, d\mu$$

where the equality follows from the bounded convergence theorem, Theorem 3.3.4, and the inequality comes from the fact that $h_n(x) \le g_n(x)$ for all $x$.

Since this inequality holds for any bounded measurable $h$ which is less than $f$, it holds when $h$ is $f_m(x) = \min\{f(x), m\}$, so for all

$m \in \mathbb{N}$ we have

$$\int f_m \, d\mu \le \liminf_{n\to\infty} \int g_n \, d\mu.$$

Taking the limit as $m$ tends to infinity and recalling the definition of the integral of $f$ we get

$$\int f \, d\mu = \lim_{m\to\infty} \int f_m \, d\mu \le \liminf_{n\to\infty} \int g_n \, d\mu.$$

□

In the case of a monotone sequence of measurable functions the inequality of the previous result can be strengthened to an equality.

**Theorem 4.2.3. (Monotone convergence theorem).** *Suppose $g_n$ is an increasing sequence of non-negative measurable functions defined on $[0,1]$. If $\lim g_n(x) = f(x)$ for almost all $x$, then*

$$\int f \, d\mu = \lim_{n\to\infty} \int g_n \, d\mu.$$

*In particular, $f$ is integrable if and only if $\lim \int g_n \, d\mu < +\infty$.*

**Proof.** The function $f$ is measurable by Theorem 3.1.9. If it is integrable, then the fact that $f(x) \ge g_n(x)$ for almost all $x$ allows us to apply the Lebesgue convergence theorem, Theorem 4.2.1, to conclude the desired result.

Hence, we need only show that if $\int f \, d\mu = +\infty$, then

$$\lim_{n\to\infty} \int g_n \, d\mu = +\infty.$$

This follows immediately from Fatou's lemma. □

**Corollary 4.2.4. (Integral of infinite series).** *Suppose $u_n$ is a non-negative measurable function and $f$ is a non-negative function such that $\sum_{n=1}^{\infty} u_n(x) = f(x)$ for almost all $x$. Then*

$$\int f \, d\mu = \sum_{n=1}^{\infty} \int u_n \, d\mu.$$

**Proof.** Define

$$f_N(x) = \sum_{n=1}^{N} u_n(x).$$

Now the result follows from Theorem 4.2.3. □

**Corollary 4.2.5. (Countable additivity of the Lebesgue integral).** *Suppose $\{E_n\}_{n=1}^{\infty}$ is a countable collection of pairwise disjoint measurable subsets of $I$, and $f$ is a non-negative integrable function. If*

$$E = \bigcup_{n=1}^{\infty} E_n \quad \textit{then} \quad \int_E f \, d\mu = \sum_{n=1}^{\infty} \int_{E_n} f \, d\mu.$$

**Proof.** Note that

$$\mathfrak{X}_E = \sum_{n=1}^{\infty} \mathfrak{X}_{E_n}.$$

Define $u_n(x) = f\mathfrak{X}_{E_n}$. Then

$$f\mathfrak{X}_E = f\sum_{n=1}^{\infty} \mathfrak{X}_{E_n} = \sum_{n=1}^{\infty} u_n.$$

Now the result follows from the previous corollary. □

---

**Exercise 4.2.6.**

(1) Give an example to show the inequality in Fatou's lemma can be strict.

(2) Give an example of a a decreasing sequence of non-negative measurable functions which are defined on the interval $[0, 1]$ and which converge pointwise to a bounded function $f$ such that
$$\int f_n \, d\mu = +\infty \quad \text{and} \quad \int f \, d\mu = 0.$$
This shows that the monotone convergence theorem does not hold for decreasing sequences.

(3) Suppose $g_n$ is a sequence of non-negative measurable functions defined on $[0, 1]$. Prove the following slightly stronger version of Fatou's Lemma:
$$\int \liminf_{n\to\infty} g_n \, d\mu \le \liminf_{n\to\infty} \int g_n \, d\mu.$$

(4) Let $f$ be a non-negative integrable extended real-valued function defined on $[0, 1]$ and let $\mathcal{L}_\mu(f)$ denote the set of simple functions which are less than or equal to $f$. Let $\mathcal{U}_\mu(f)$ be

the set of simple functions greater than or equal to $f$. Prove that
$$\int f\, d\mu = \sup_{v \in \mathcal{L}_\mu(f)} \int v\, d\mu.$$
Give an example of non-negative (non-extended) real-valued function $f$ for which
$$\int f\, d\mu \neq \inf_{u \in U_\mu(f)} \int u\, d\mu.$$
*Hint:* The function $f$ can be integrable and have $U_\mu(f) = \emptyset$.

(5) ($\star$) *Egorov's theorem*: If $\{f_n : I \to \mathbb{R}\}$ is a sequence of measurable functions converging pointwise to $f : [0,1] \to \mathbb{R}$ prove that for any $\varepsilon > 0$ there is a set $A \subset I$ with $\mu(A) < \varepsilon$ such that $\{f_n\}$ converges uniformly to $f$ on $A^c$. This is sometimes referred to as the third of *Littlewood's three principles*.
*Hint:* Consider the sets
$$E(n,m) = \bigcup_{k=n}^{\infty} \left\{ x \;\middle|\; |f_k(x) - f(x)| \geq \frac{1}{m} \right\}.$$
For each $m$ show there is $n_m$ such that $\mu(E(n_m, m)) < \dfrac{\varepsilon}{2^m}$. Then take $A = \bigcup_{m=1}^{\infty} E(n_m, m)$.

## 4.3. Other Measures

There are other measures besides Lebesgue and indeed measures on other spaces besides $[0,1]$ or $\mathbb{R}$. We will mostly limit our attention to measures defined on $I = [0,1]$, for simplicity, but it is worth formulating some definitions for a general $\sigma$-algebra of subsets of a set $X$.

Recall from Definition 2.3.1 that a collection $\mathcal{A}$ of subsets of a set $X$ is called a $\sigma$-algebra provided it contains the set $X$ and is closed under taking complements, countable unions, and countable intersections.

**Examples 4.3.1.** The following are examples of $\sigma$-algebras of subsets of $I = [0,1]$:

(1) *The trivial $\sigma$-algebra.* $\mathcal{A} = \{\emptyset, I\}$.

(2) $\mathcal{A} = \{A \subset I \mid A$ is countable, or $A^c$ is countable$\}$.

(3) $\mathcal{A} = \mathcal{M}(I)$ the Lebesgue measurable sets.

(4) $\mathcal{A}$ is the algebra of Borel subsets of $I$, the smallest $\sigma$-algebra containing the open intervals.

**Definition 4.3.2. (Measure, measurable).** *If $\mathcal{A}$ is a $\sigma$-algebra of subsets of a set $X$, then a function $\nu : \mathcal{A} \to [0, \infty]$ is called a* measure *provided*

- $\nu(A) \geq 0$ *for every* $A \in \mathcal{A}$,
- $\nu(\emptyset) = 0$, *and*
- $\nu$ *is countably additive, i.e., if* $\{A_n\}_{n=1}^{\infty}$ *are pairwise disjoint sets in* $\mathcal{A}$, *then*

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

*A measure $\nu$ is finite if $\nu(X) < \infty$. An extended real-valued function $f$ defined on $X$ is called measurable with respect to the $\sigma$-algebra $\mathcal{A}$ if for any $a \in \mathbb{R}$ the set $f^{-1}([-\infty, a])$ is in $\mathcal{A}$.*

**Proposition 4.3.3.** *Suppose $\mu$ and $\nu$ are finite measures defined for sets in the $\sigma$-algebra $\mathcal{A}$ and $a$ and $b$ are non-negative numbers, not both* 0. *If we define $\rho(A) = a\mu(A) + b\nu(A)$ for all $A \in \mathcal{M}(I)$, then $\rho$ is a finite measure.*

The (easy) proof is left as an exercise (see Exercise 4.3.9 problem (1) below).

The integral of a measurable function with respect to a measure $\nu$ is defined analogously to Lebesgue measure. The concept of simple function is the same and we let $\mathcal{L}_\mu(g)$ denote the set of simple functions $v : X \to \mathbb{R}$ such that $v(x) \leq g(x)$ for all $x \in X$.

**Definition 4.3.4. (Integrable function).** *Let $\nu$ be a finite measure defined on the $\sigma$-algebra $\mathcal{A}$. If $f(x) = \sum_{i=1}^n r_i \mathfrak{X}_{A_i}$ is a simple function, then its* integral with respect to $\nu$ *is defined by $\int f \, d\nu = \sum_{i=1}^n r_i \nu(A_i)$. If $g : [0, 1] \to \mathbb{R}$ is a bounded measurable function, then* *we define its integral with respect to $\nu$ by*

$$\int g\ d\nu = \sup_{v \in \mathcal{L}_\mu(g)} \Big\{ \int v\ d\nu \Big\}.$$

*If $h$ is a non-negative extended measurable function we define*

$$\int h\ d\nu = \lim_{n\to\infty} \int \min\{h, n\}\ d\nu.$$

We will now turn our attention back to measures defined on the $\sigma$-algebra of Lebesgue measurable sets.

**Definition 4.3.5. (Absolutely continuous measure).** *If $\nu$ is a measure defined on $\mathcal{M}(I)$, the Lebesgue measurable subsets of $I$, then we say $\nu$ is* absolutely continuous *with respect to Lebesgue measure $\mu$ if $\mu(A) = 0$ implies $\nu(A) = 0$. We write $\nu \ll \mu$.*

The following result motivates the name "absolute continuity."

**Theorem 4.3.6.** *If $\nu$ is a measure defined on $\mathcal{M}(I)$ which is absolutely continuous with respect to Lebesgue measure, then for any $\varepsilon > 0$ there is a $\delta > 0$ such that $\nu(A) < \varepsilon$ whenever $\mu(A) = \delta$.*

**Proof.** We assume there is a counterexample and show this leads to a contradiction. If the measure $\nu$ does not satisfy the conclusion of the theorem, then there is an $\varepsilon > 0$ for which it fails, i.e., there is no $\delta > 0$ which works for this $\varepsilon$. In particular, for any positive integer $m$ there is a set $B_m$ such that $\nu(B_m) \geq \varepsilon$ and $\mu(B_m) < 1/2^m$. Hence, if we define $A_n = \bigcup_{m=n+1}^\infty B_m$, then

$$\mu(A_n) \leq \sum_{m=n+1}^\infty \mu(B_m) \leq \sum_{m=n+1}^\infty \frac{1}{2^m} = \frac{1}{2^n}.$$

The sets $A_n$ are nested, i.e., $A_n \supset A_{n+1}$. Let

$$C = \bigcap_n^\infty A_n.$$

It follows from Proposition 2.4.5 that

$$\mu(C) = \mu(\bigcap_{n=1}^\infty A_n) = \lim_{n\to\infty} \mu(A_n) \leq \lim_{n\to\infty} \frac{1}{2^n} = 0. \tag{4.3.1}$$

The proof of Proposition 2.4.5 made use only of the countable additivity of the measure. Hence, it is also valid for $\nu$, i.e.,

$$\nu(C) = \nu(\bigcap_{n=1}^{\infty} A_n) = \lim_{n\to\infty} \nu(A_n) \geq \varepsilon$$

since $\nu(A_n) \geq \nu(B_{n+1}) \geq \varepsilon$ for every $n$. This contradicts the absolute continuity of $\nu$ with respect to $\mu$. We have proven the contrapositive of the result we desire. □

**Proposition 4.3.7.** *If $f$ is a non-negative integrable function on $I$ and we define*

$$\nu_f(A) = \int_A f\ d\mu,$$

*then $\nu_f$ is a measure with $\sigma$-algebra $\mathcal{M}(I)$ which is absolutely continuous with respect to Lebesgue measure $\mu$.*

**Proof.** Clearly, $\nu_f(A) = \int_A f\ d\mu \geq 0$ for all $A \in \mathcal{M}$ since $f$ is non-negative. Also, $\nu_f(\emptyset) = 0$. We need to check countable additivity.

Suppose $\{A_n\}_{n=1}^{\infty}$ is a sequence of pairwise disjoint measurable subsets of $[0,1]$ and $A$ is their union. Then for all $x \in [0,1]$,

$$f(x)\mathfrak{X}_A(x) = \sum_{n=1}^{\infty} f(x)\mathfrak{X}_{A_n}(x).$$

Hence, by Theorem 4.2.4

$$\int f\mathfrak{X}_A\ d\mu = \sum_{n=1}^{\infty} \int f\mathfrak{X}_{A_n}\ d\mu,$$

and so

$$\nu_f(A) = \sum_{n=1}^{\infty} \nu_f(A_n).$$

Thus, $\nu$ is a measure.

If $\mu(A) = 0$, then by Proposition 4.1.4

$$\nu_f(A) = \int f\mathfrak{X}_A\ d\mu = 0,$$

so $\nu$ is absolutely continuous with respect to $\mu$. □

The converse to Proposition 4.3.7 is called the Radon-Nikodym theorem. Its proof is beyond the scope of this text. A proof can be found in Chapter 11, Section 6 of Royden's book [**Ro**] or in Chapter 11, Section 2 of [**L**].

**Theorem 4.3.8. (Radon-Nikodym).** *Suppose that $\nu$ is a finite measure with $\sigma$-algebra $\mathcal{M}(I)$ and that $\nu$ is absolutely continuous with respect to Lebesgue measure $\mu$. Then there is a non-negative integrable function $f$ on $[0,1]$ such that*

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

*The function $f$ is unique up to measure $0$, i.e., if $g$ is another function with these properties, then $f = g$ almost everywhere.*

The function $f$ is called the *Radon-Nikodym derivative* of $\nu$ with respect to $\mu$. In fact, the Radon-Nikodym Theorem is more general than we have stated, since it applies to any two finite measures $\nu$ and $\mu$ defined on a $\sigma$-algebra $\mathcal{A}$ with the property that $\nu$ is absolutely continuous with respect to $\mu$.

---

**Exercise 4.3.9.**

(1) Suppose $\mu$ and $\nu$ are finite measures defined for sets in the $\sigma$-algebra $\mathcal{M}(I)$ and $a$ and $b$ are non-negative numbers, not both 0. If we define $\rho(A) = a\mu(A) + b\nu(A)$ for all $A \in \mathcal{M}(I)$, prove that $\rho$ is a finite measure.

(2) Let $X$ be a countable set and let $\mathcal{A}$ be the $\sigma$-algebra of all subsets of $X$. Prove there is no finite measure on $X$ other than the trivial measure which assigns measure 0 to every set.

(3) Define the function $\nu : \mathcal{M} \to [0, \infty]$ by $\nu(A) = \infty$ if 0 is in the closure of $A$ and $\nu(A) = 0$ otherwise. Prove that $\nu$ is finitely additive but not countably additive.

(4) Show that Fatou's lemma and the monotone convergence theorem are valid for an arbitrary finite measure $\nu$ defined on a $\sigma$-algebra $\mathcal{A}$.

(5) Given a point $x_0 \in [0,1]$ define the function $\delta_{x_0} : \mathcal{M}(I) \to \mathbb{R}$ by $\delta_{x_0}(A) = 1$ if $x_0 \in A$ and $\delta_{x_0}(A) = 0$ if $x_0 \notin A$. Let $\nu(A) = \delta_{x_0}(A)$.
   (a) Prove that $\nu$ is a measure on the $\sigma$-algebra $\mathcal{M}(I)$.
   (b) Prove that if $f$ is a measurable function $\int f\, d\nu = f(x_0)$.
   (c) Give examples of three other $\sigma$-algebras on which $\nu$ defines a measure. The measure $\nu$ is called the *Dirac $\delta$-measure.*

(6) If $\{\nu_n\}$, $n \in \mathbb{N}$ is a sequence of finite measures defined on the $\sigma$-algebra $\mathcal{M}(I)$, then we say that this sequence *converges weakly* to the finite measure $\nu$ if for every continuous function $f : [0,1] \to \mathbb{R}$,
$$\lim_{n\to\infty} \int f\, d\nu_n = \int f\, d\nu.$$
Let
$$\nu_n = \frac{1}{n}\sum_{k=1}^{n} \delta_k$$
where $\delta_k$ is the Dirac $\delta$-measure, $\delta_{x_0}$, with $x_0 = \frac{k}{n}$ (with $\sigma$-algebra the Lebesgue measurable sets). Prove that $\{\nu_n\}$ converges weakly to Lebesgue measure $\mu$. with respect to $\mu$.

## 4.4. General Measurable Functions

In this section we return to Lebesgue measure $\mu$ on the interval $[0,1]$ but, consider extended measurable functions which may be unbounded both above and below. We define
$$f^+(x) = \max\{f(x), 0\}$$
and
$$f^-(x) = -\min\{f(x), 0\}.$$
These are both non-negative measurable functions and
$$f(x) = f^+(x) - f^-(x).$$

**Definition 4.4.1. (Lebesgue integrable).** *If $f : [0,1] \to [-\infty,\infty]$ is a measurable function, then we say $f$ is* Lebesgue integrable *provided both $f^+$ and $f^-$ are integrable (as non-negative functions). If*

*$f$ is integrable, we define*

$$\int f\ d\mu = \int f^+\ d\mu - \int f^-\ d\mu.$$

**Proposition 4.4.2.** *Suppose $f$ and $g$ are measurable functions on $[0,1]$ and $f = g$ almost everywhere. Then if $f$ is integrable, so is $g$ and $\int f\ d\mu = \int g\ d\mu$. In particular, if $f = 0$ almost everywhere, $\int f\ d\mu = 0$.*

**Proof.** If $f$ and $g$ are measurable functions on $[0,1]$ and $f = g$ almost everywhere, then $f^+ = g^+$ almost everywhere, $f^- = g^-$ almost everywhere, and $f^+$ and $f^-$ are integrable. It then follows from Proposition 4.1.4 that $g^+$ and $g^-$ are integrable and that $\int f^+\ d\mu \geq \int g^+\ d\mu$ and $\int f^-\ d\mu \geq \int g^-\ d\mu$. Switching the roles of $f$ and $g$ this same proposition gives the reverse inequalities, so we have $\int f^+\ d\mu = \int g^+\ d\mu$ and $\int f^-\ d\mu = \int g^-\ d\mu$. □

**Proposition 4.4.3.** *A measurable function $f : [0,1] \to [-\infty, \infty]$ is integrable if and only if the the function $|f|$ is integrable.*

**Proof.** Notice that $|f(x)| = f^+(x) + f^-(x)$ so $|f(x)| \geq f^+(x)$ and $|f(x)| \geq f^-(x)$. Thus, if $|f|$ is integrable, it follows from Proposition 4.1.4 that both $f^+$ and $f^-$ are integrable. Conversely, if $f^+$ and $f^-$ are integrable, then so is their sum $|f|$. □

**Theorem 4.4.4. (Lebesgue convergence theorem).** *Suppose $f_n$ is a sequence of measurable functions defined on $[0,1]$ and $g$ is an integrable function such that $|f_n(x)| \leq g(x)$ for all $n$ and almost all $x$. If $\lim f_n(x) = f(x)$ for almost all $x$, then $f$ is integrable and*

$$\int f\ d\mu = \lim_{n\to\infty} \int f_n\ d\mu.$$

**Proof.** The functions

$$f_n^+(x) = \max\{f_n(x), 0\} \quad \text{and} \quad f_n^-(x) = -\min\{f_n(x), 0\}$$

satisfy

$$\lim_{n\to\infty} f_n^+(x) = f^+(x) \quad \text{and} \quad \lim_{n\to\infty} f_n^-(x) = f^-(x)$$

for almost all $x$. Also, $g(x) \geq f_n^+(x)$ and $g(x) \geq f_n^-(x)$ for almost all $x$. Hence, by Theorem 4.2.1

$$\int f^+ \, d\mu = \lim_{n\to\infty} \int f_n^+ \, d\mu \quad \text{and} \quad \int f^- \, d\mu = \lim_{n\to\infty} \int f_n^- \, d\mu.$$

Thus, $f = f^+ - f^-$ is integrable and

$$\begin{aligned}\int f \, d\mu &= \int f^+ \, d\mu - \int f^- \, d\mu \\ &= \lim_{n\to\infty} \int f_n^+ \, d\mu - \lim_{n\to\infty} \int f_n^- \, d\mu \\ &= \lim_{n\to\infty} \int f_n^+ - f_n^- \, d\mu \\ &= \lim_{n\to\infty} \int f_n \, d\mu.\end{aligned}$$

□

The following theorem says that for any $\varepsilon > 0$ any integrable function can be approximated within $\varepsilon$ by a step function if we are allowed to exclude a set of measure $\varepsilon$.

**Theorem 4.4.5.** *If $f : [0,1] \to [-\infty,\infty]$ is an integrable function, then given $\varepsilon > 0$ there is a step function $g : [0,1] \to \mathbb{R}$ and a measurable subset $A \subset [0,1]$ such that $\mu(A) < \varepsilon$ and*

$$|f(x) - g(x)| < \varepsilon \text{ for all } x \notin A.$$

*Moreover, if $|f(x)| \leq M$ for all $x$, then we may choose $g$ with this same bound.*

**Proof.** We first prove the result for the special case of $f(x) = \mathfrak{X}_E(x)$ for some measurable set $E$. This follows because there is a countable cover of $E$ by open intervals $\{U_i\}_{i=1}^\infty$ such that

$$\mu(E) \leq \sum_{i=1}^\infty \operatorname{len}(U_i) \leq \mu(E) + \frac{\varepsilon}{2},$$

and hence

$$\mu\Big(\big(\bigcup_{i=1}^\infty U_i\big) \setminus E\Big) < \frac{\varepsilon}{2}. \tag{4.4.1}$$

Also, we may choose $N > 0$ such that

$$\mu\Big(\bigcup_{i=N}^{\infty} U_i\Big) \leq \sum_{i=N}^{\infty} \operatorname{len}(U_i) < \frac{\varepsilon}{2}. \tag{4.4.2}$$

Let $V_N = \bigcup_{i=1}^{N} U_i$. It is a finite union of intervals, so the function $g(x) = \mathfrak{X}_{V_N}$ is a step function and if $A = \{x \mid f(x) \neq g(x)\}$, then

$$A \subset \Big(V_N \setminus E\Big) \cup \Big(E \setminus V_N\Big) \subset \Big(\big(\bigcup_{i=1}^{\infty} U_i\big) \setminus E\Big) \cup \Big(\bigcup_{i=N}^{\infty} U_i\Big),$$

so it follows from equations (4.4.1) and (4.4.2) that $\mu(A) < \varepsilon$. This proves the result for $f = \mathfrak{X}_E$.

From this the result follows for simple functions $f = \sum r_i \mathfrak{X}_{E_i}$ because if $g_i$ is the approximating step function for $\mathfrak{X}_{E_i}$, then $g = \sum r_i g_i$ approximates $f$ (with a suitably adjusted $\varepsilon$).

If $f$ is a bounded measurable function by Theorem 3.2.1 there is a simple function $h$ such that $|f(x) - h(x)| < \varepsilon/2$ for all $x$. Let $g$ be a step function such that $|h(x) - g(x)| < \varepsilon/2$ for all $x \notin A$ with $\mu(A) < \varepsilon$. Then

$$|f(x) - g(x)| \leq |f(x) - h(x)| + |h(x) - g(x)| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

for all $x \notin A$. That is, the result is true if $f$ is a bounded measurable function.

Suppose $f$ is a non-negative integrable function and let $A_n = \{x \mid f(x) > n\}$. Then

$$n\mu(A_n) = \int n\mathfrak{X}_{A_n}\ d\mu \leq \int f\ d\mu < \infty.$$

It follows that

$$\lim_{n\to\infty} \mu(A_n) = \lim_{n\to\infty} \frac{1}{n} \int f\ d\mu = 0.$$

Hence, there is an $N > 0$ such that $\mu(A_N) < \varepsilon/2$.

If $f_N = \min\{f, N\}$, then $f_N$ is a bounded measurable function. So we may choose a step function $g$ such that $|f_N(x) - g(x)| < \varepsilon/2$ for all $x \notin B$ with $\mu(B) < \varepsilon/2$. It follows that if $A = A_N \cup B$, then

$\mu(A) < \varepsilon$. Also, if $x \notin A$, then $f(x) = f_N(x)$, so

$$\begin{aligned}|f(x) - g(x)| &\le |f(x) - f_N(x)| + |f_N(x) - g(x)| \\ &= |f_N(x) - g(x)| < \varepsilon.\end{aligned}$$

Hence, the result holds for non-negative $f$.

For a general integrable $f$ we have $f = f^+ - f^-$. The fact that the result holds for $f^+$ and $f^-$ easily implies it holds for $f$.

Suppose now that $f$ is bounded, say $|f(x)| \le M$ for all $x$, and $g$ satisfies the conclusion of our theorem, then we define

$$g_1(x) = \begin{cases} M, \text{ if } g(x) > M; \\ g(x), \text{ if } -M \le g(x) \le M; \\ -M, \text{ if } g(x) < -M. \end{cases}$$

The function $g_1$ is a step function with $|g_1(x)| \le M$ and $g_1(x) = g(x)$ except when $|g(x)| > M$. Note that if $g(x) > M$ and $x \notin A$, then

$$f(x) \le M = g_1(x) < g(x),$$

so $|g_1(x) - f(x)| < \varepsilon$. The case $g(x) < -M$ is similar. □

We can now generalize Theorem 3.2.4 to cover Lebesgue integrable functions and not just bounded functions. The proof is left as an exercise.

**Theorem 4.4.6.** *The Lebesgue integral satisfies the following properties:*

**I. Linearity:** *If $f$ and $g$ are Lebesgue integrable functions and $c_1, c_2 \in \mathbb{R}$, then*

$$\int c_1 f + c_2 g \; d\mu = c_1 \int f \; d\mu + c_2 \int g \; d\mu.$$

**II. Monotonicity:** *If $f$ and $g$ are Lebesgue measurable and $f(x) \le g(x)$ for all $x$, then $\int f \; d\mu \le \int g \; d\mu$.*

**III. Absolute value:** *If $f$ is Lebesgue integrable, then $|f|$ is also and $|\int f \; d\mu| \le \int |f| \; d\mu$.*

**IV. Null sets:** *If $f$ and $g$ are equal except on a set of measure zero and $f$ is integrable, then $g$ is also integrable and $\int f \; d\mu = \int g \; d\mu$.*

**Exercise 4.4.7.**

(1) Prove that if $f, g, h$ are measurable functions and $f = g$ almost everywhere and $g = h$ almost everywhere, then $f = h$ almost everywhere.

(2) Suppose $\nu$ is a finite measure defined on a $\sigma$-algebra of subsets of a set $X$. Define $f : X \to [-\infty, \infty]$ to be an *integrable function* analogously to Definition 4.4.1. Prove that the Lebesgue convergence theorem remains valid in this more general setting. *Hint:* You will need part (4) of Exercise (4.3.9).

(3) Prove that if $f : [0, 1] \to [-\infty, \infty]$ is an integrable function, then given $\varepsilon > 0$ there exists a continuous function $g : [0, 1] \to \mathbb{R}$ and a set $A$ such that $\mu(A) < \varepsilon$ and $|f(x) - g(x)| < \varepsilon$ for all $x \notin A$, and $g(0) = g(1)$. *Hint:* Use part (6) of Exercise 3.1.11.

(4) Prove that the, not necessarily bounded, integrable functions from $[0, 1]$ to $\mathbb{R}$ form a vector space.

(5) Prove that if $g$ is an integrable function on $[0, 1]$, then there is a sequence of simple functions $\{g_n\}$ such that $g_n$ converges to $g$ pointwise and $|g_n(x)| \leq |g(x)|$ for all $x$. Conclude from the Lebesgue convergence theorem that $\lim \int g_n \, d\mu = \int g \, d\mu$ and $\lim \int |g_n| \, d\mu = \int |g| \, d\mu$.

(6) Suppose $\{E_n\}_{n=1}^{\infty}$ is a countable collection of pairwise disjoint measurable subsets of $I$ and $f$ is an integrable function. Prove that if

$$E = \bigcup_{n=1}^{\infty} E_n, \text{ then } \int_E f \, d\mu = \sum_{n=1}^{\infty} \int_{E_n} f \, d\mu.$$

This was proved in Corollary 4.2.5 with the additional assumption that $f$ is non-negative.

(7) Proposition 4.4.2 proves the null set part of Theorem 4.4.6. Prove the remaining parts of this theorem, namely linearity, monotonicity, and the absolute value property. (You may use Theorem 3.2.4).

Chapter 5

# The Hilbert Space $L^2$

## 5.1. Square Integrable Functions

In this chapter we will develop the beginnings of a theory of function spaces with many properties analogous to the basic properties of $\mathbb{R}^n$. To motivate these developments we first take a look at $\mathbb{R}^n$ in a different way. We let $X = \{1, 2, 3, \ldots, n\}$, the first $n$ natural numbers, and we define a measure $\nu$ on $X$ which is called the "counting measure".

More precisely, we take as $\sigma$-algebra the family of all subsets of $X$ and for any $A \subset X$ we define $\nu(A)$ to be the number of elements in the set $A$. It is easy to see that this is a measure and that *any* function $f : X \to \mathbb{R}$ is measurable. In fact, any function is a simple function. This is because there is a partition of $X$ given by $A_i = \{i\}$ and clearly $f$ is constant on each $A_i$, so $f = \sum_{i=1}^n r_i \mathfrak{X}_{A_i}$ where $r_i = f(i)$.

Consequently, we have

$$\int f \, d\nu = \sum_{i=1}^{n} r_i \nu(A_i) = \sum_{i=1}^{n} f(i).$$

For reasons that will be clear below we will denote the collection of all functions from $X$ to $\mathbb{R}$ by $L^2(X)$. The important thing to note is that this is just another name for $\mathbb{R}^n$. More formally, $L^2(X)$ is the vector space of all functions from $X$ to $\mathbb{R}$. This vector space can also be seen as the set of all finite sequences $(x_1, x_2, \ldots, x_n)$ with the correspondence between the two given by $f \longleftrightarrow (x_1, x_2, \ldots, x_n)$ where $x_i = f(i)$.

It is useful to view the usual inner product on $\mathbb{R}^n$ (the "dot" product given by $\langle x, y\rangle = \sum_{i=1}^n x_i y_i$) in light of this correspondence. If $f, g \in L^2(X)$ are the functions corresponding to vectors $x$ and $y$, respectively, then $x_i = f(i)$ and $y_i = g(i)$, so

$$\langle x, y\rangle \;=\; \sum_{i=1}^n x_i y_i \;=\; \sum_{i=1}^n f(i)g(i) \;=\; \int fg\ d\nu.$$

Also, the norm (or length) of a vector is given by

$$\|x\|^2 \;=\; \langle x, x\rangle \;=\; \sum_{i=1}^n x_i^2 \;=\; \sum_{i=1}^n f(i)^2 \;=\; \int f^2\ d\nu.$$

It is this way of viewing the inner product and norm on $\mathbb{R}^n$ which generalizes nicely to a space of real-valued functions on the interval. In order to make this generalization work we have to restrict our attention to functions which are not just integrable, but whose square is integrable. In fact, if the square of a function is integrable, then the function is integrable, but the converse is not true. (See Exercise 5.1.10 below.)

In this chapter it will be convenient to consider functions on the more general closed interval $[a, b]$ rather than $[0, 1]$. Of course, all of our results about measurable functions and their integrals remain valid on this different interval.

**Definition 5.1.1. (Square integrable).** *A measurable function $f : [a, b] \to [-\infty, \infty]$ is called* square integrable *if $f^2$ is integrable. We denote the set of all square integrable functions by $L^2[a, b]$. We define the* norm *of $f \in L^2[a, b]$ by*

$$\|f\| = \Big(\int f^2\ d\mu\Big)^{\frac{1}{2}}.$$

Notice that the function $\|\ \|$ is not quite a norm in the usual sense (as defined in Definition A.9.4 and described in Proposition A.9.6). The one way in which it fails strictly to be a norm is that $\|f\| = 0$ implies $f(x) = 0$ *almost everywhere* rather than everywhere. The pedantic way to overcome this problem is to define $L^2[a, b]$ as the vector space of equivalence classes of square integrable functions, where

$f$ and $g$ are considered "equivalent" if they are equal almost everywhere. It is customary, however, to overlook this infelicity and simply consider $L^2[a,b]$ as a vector space of functions rather than equivalence classes of functions. In doing this we should keep in mind that we are generally considering two functions the same if they agree almost everywhere.

**Proposition 5.1.2.** *The norm $\| \ \|$ on $L^2[a,b]$ satisfies $\|cf\| = |c|\|f\|$ for all $c \in \mathbb{R}$ and all $f \in L^2[a,b]$. Moreover, for all $f$, $\|f\| \geq 0$ with equality only if $f = 0$ almost everywhere.*

**Proof.** We see

$$\|cf\| = \Big( \int c^2 f^2 \, d\mu \Big)^{\frac{1}{2}} = \sqrt{c^2} \Big( \int f^2 \, d\mu \Big)^{\frac{1}{2}} = |c|\|f\|.$$

Since $\int f^2 \, d\mu \geq 0$, clearly $\|f\| \geq 0$. Also, if

$$\|f\| = 0, \text{ then } \int f^2 \, d\mu = 0.$$

So by Corollary 4.1.5 $f^2 = 0$ almost everywhere and hence $f = 0$ almost everywhere. □

**Lemma 5.1.3.** *If $f, g \in L^2[a,b]$, then $fg$ is integrable and*

$$2 \int |fg| \, d\mu \leq \|f\|^2 + \|g\|^2.$$

*Equality holds if and only if $|f| = |g|$ almost everywhere.*

**Proof.** Since

$$0 \leq (|f(x)| - |g(x)|)^2 = f(x)^2 - 2|f(x)g(x)| + g(x)^2$$

we have $2|f(x)g(x)| \leq f(x)^2 + g(x)^2$. Hence, by Proposition 4.1.4, we conclude that $|fg|$ is integrable and that

$$2 \int |fg| \, d\mu \leq \|f\|^2 + \|g\|^2.$$

Equality holds if and only if $\int (|f(x)| - |g(x)|)^2 \, d\mu = 0$ and we may conclude by Corollary 4.1.5 that this happens if and only if $(|f(x)| - |g(x)|)^2 = 0$ almost everywhere and hence that $|f| = |g|$ almost everywhere. □

**Theorem 5.1.4.** *$L^2[a,b]$ is a vector space.*

**Proof.** We must show that if $f, g \in L^2[a,b]$ and $c \in \mathbb{R}$, then $cf \in L^2[a,b]$ and $(f+g) \in L^2[a,b]$. The first of these is clear since $f^2$ is integrable implies that $c^2 f^2$ is integrable.

To check the second we observe that

$$(f+g)^2 = f^2 + 2fg + g^2 \leq f^2 + 2|fg| + g^2.$$

Since $f^2$, $g^2$ and $|fg|$ are all integrable, it follows from Proposition 4.1.4 that $(f+g)^2$ is also. Hence, $(f+g) \in L^2[a,b]$. □

**Theorem 5.1.5. (Hölder inequality).** *If $f, g \in L^2[a,b]$, then*

$$\int |fg|\; d\mu \leq \|f\|\; \|g\|.$$

*Equality holds if and only if there is a constant $c$ such that $|f(x)| = c|g(x)|$ or $|g(x)| = c|f(x)|$ almost everywhere.*

**Proof.** If either $\|f\|$ or $\|g\|$ is 0, the result is trivial so assume they are both non-zero. In that case the functions $f_0 = f/\|f\|$ and $g_0 = g/\|g\|$ satisfy $\|f_0\| = \|g_0\| = 1$.

Then by Lemma 5.1.3

$$2 \int |f_0 g_0|\; d\mu \leq \|f_0\|^2 + \|g_0\|^2 = 2,$$

so

$$\int |f_0 g_0|\; d\mu \leq 1,$$

and equality holds if and only if $|f_0| = |g_0|$ almost everywhere. So

$$\frac{1}{\|f\|\; \|g\|} \int |fg|\; d\mu = \int |f_0 g_0|\; d\mu \leq 1,$$

and hence

$$\int |fg|\; d\mu \leq \|f\|\; \|g\|.$$

Equality holds if and only if $|f_0| = |g_0|$ almost everywhere, which implies there is a constant $c$ with $|f(x)| = c|g(x)|$ almost everywhere. □

**Corollary 5.1.6. (Cauchy-Schwarz).** *If $f, g \in L^2[a,b]$, then*

$$\left| \int fg\; d\mu \right| \leq \|f\|\; \|g\|.$$

*Equality holds if and only if there is a constant $c$ such that $f(x) = cg(x)$ or $g(x) = cf(x)$ almost everywhere.*

**Proof.** The inequality follows from Hölder's inequality and the absolute value inequality since

$$\left|\int fg\ d\mu\right| \leq \int |fg|\ d\mu \leq \|f\|\ \|g\|.$$

Equality holds when both of these inequalities are equalities. In this case, suppose first that $\int fg\ d\mu \geq 0$. Then $\int |fg|\ d\mu = \int fg\ d\mu$, so $\int |fg| - fg\ d\mu = 0$ and hence $|fg| = fg$ almost everywhere. This says that $f$ and $g$ have the same sign almost everywhere. Since the second inequality is an equality we know from Hölder that there is a constant $c$ such that $|f(x)| = c|g(x)|$ or $|g(x)| = c|f(x)|$ almost everywhere. This togther with the fact that $f$ and $g$ have the same sign almost everywhere implies $f(x) = cg(x)$ or $g(x) = cf(x)$ almost everywhere. For the case that $\int fg\ d\mu \leq 0$ we can replace $f$ with $-f$ and conclude that $f(x) = -cg(x)$ or $g(x) = -cf(x)$. Conversely, it is easy to see that if $f(x) = cg(x)$ or $g(x) = cf(x)$ almost everywhere, then the inequality above is an equality. □

The following result, called the Minkowski inequality, is the triangle inequality for the normed vector space $L^2[a,b]$.

**Theorem 5.1.7. (Minkowski's inequality).** *If $f, g \in L^2[a,b]$, then*

$$\|f + g\| \leq \|f\| + \|g\|.$$

**Proof.** We observe that

$$\begin{aligned}
\|f+g\|^2 &= \int (f+g)^2\ d\mu \\
&= \int (f^2 + 2fg + g^2)\ d\mu \\
&\leq \int f^2 + 2|fg| + g^2\ d\mu \\
&\leq \|f\|^2 + 2\|f\|\ \|g\| + \|g\|^2 \quad \text{by Hölder's inequality} \\
&= (\|f\| + \|g\|)^2.
\end{aligned}$$

Taking square roots of both sides of this equality gives the triangle inequality. □

**Definition 5.1.8. (Inner product on $L^2[a,b]$).** *If $f, g \in L^2[a,b]$, then we define their* inner product *by*

$$\langle f, g \rangle = \int fg \; d\mu.$$

Of course, we have to prove that $\langle \ , \ \rangle$ actually satisfies the properties required of an inner product on a real vector space as defined in Definition A.9.3.

**Theorem 5.1.9. (Inner product on $L^2[a,b]$).** *For any $f_1, f_2, g \in L^2[a,b]$ and any $c_1, c_2 \in \mathbb{R}$ the inner product on $L^2[a,b]$ satisfies the following properties:*

(1) Commutativity: $\langle f_1, g \rangle = \langle g, f_1 \rangle$.

(2) Bilinearity: $\langle c_1 f_1 + c_2 f_2, g \rangle = c_1 \langle f_1, g \rangle + c_2 \langle f_2, g \rangle$.

(3) Positive Definiteness: $\langle g, g \rangle = \|g\|^2 \geq 0$ *with equality if and only if $g = 0$ almost everywhere.*

(4) *The norm associated with the inner product $\langle \ , \rangle$ is the norm $\| \ \|$ on $L^2[a,b]$. I.e. $\|g\|^2 = \langle g, g \rangle$ for every $g \in L^2[a,b]$.*

**Proof.** Clearly, $\langle f, g \rangle = \int fg \; d\mu = \int gf \; d\mu = \langle g, f \rangle$. Bilinearity holds because of the linearity of the integral. Also, $\langle g, g \rangle = \int g^2 \; d\mu \geq 0$. Corollary 4.1.5 implies that equality holds only if $g^2 = 0$ almost everywhere. Finally, from the definitions

$$\|g\|^2 = \int g^2 \; d\mu = \langle g, g \rangle.$$

□

---

**Exercise 5.1.10.**

(1) Give an example of a function $f : [a,b] \to \mathbb{R}$ such that $f^2$ is integrable but $f$ is not. Hence, not all integrable functions are in $L^2[a,b]$. (See part (1) of Exercise 4.1.8.)

(2) Prove that if $f \in L^2[a,b]$, then $f$ is integrable. Together with (1) this proves $L^2[a,b]$ is a proper subset of the integrable functions on $[a,b]$.

(3) Prove that if $f \in L^2[a,b]$, then

$$\left(\int |f|\,d\mu\right)^2 \le (b-a)\int f^2\,d\mu.$$

## 5.2. Convergence in $L^2$

We have discussed uniform convergence and pointwise convergence and now we wish to discuss convergence in the $L^2[a,b]$ norm $\|\ \|$. Note that again we are adopting the customary convention by which $\|\ \|$ is called a norm even though $\|f\| = 0$ implies only that $f(x) = 0$ almost everywhere. So again if we were to be pedantic this is a norm on the vector space of equivalence classes of functions which are equal almost everywhere. With this *caveat* the vector space $L^2[a,b]$ is a metric space with distance function given by $\operatorname{dist}(f,g) = \|f - g\|$.

**Definition 5.2.1. (Convergence in $L^2$).** *If $\{f_n\}_{n=1}^\infty$ is a sequence in $L^2[a,b]$, then it is said to converge to $f$ in $L^2[a,b]$ if*

$$\lim_{n\to\infty} \|f - f_n\| = 0.$$

Functions which are bounded form a dense subset of $L^2[a,b]$. In fact, the following result shows that a good bounded function approximating $f \in L^2[a,b]$ can be obtained by taking $n$ large and defining $f_n(x)$ to be $n$ if $f(x) > n$, or $-n$ if $f(x) < -n$, and $f(x)$ otherwise.

**Lemma 5.2.2. (Density of bounded functions).** *Suppose $f \in L^2[a,b]$. If we define*

$$f_n(x) = \begin{cases} n, & \text{if } f(x) > n; \\ f(x), & \text{if } -n \le f(x) \le n; \\ -n & \text{if } f(x) < -n, \end{cases}$$

*then*

$$\lim_{n\to\infty} \|f - f_n\| = 0.$$

**Proof.** We will show that for any $\varepsilon > 0$ there is an $n$ such that $\|f - f_n\|^2 < \varepsilon$. First we note that $|f_n(x)| \le |f(x)|$ so

$$\begin{aligned} |f(x) - f_n(x)|^2 &\le |f(x)|^2 + 2|f(x)|\,|f_n(x)| + |f_n(x)|^2 \\ &\le 4|f(x)|^2. \end{aligned}$$

Let $E_n = \{x \mid |f(x)| > n\} = \{x \mid |f(x)|^2 > n^2\}$ and let $C = \int |f|^2 \, d\mu$. Then

$$C = \int |f|^2 \, d\mu \geq \int_{E_n} |f|^2 \, d\mu \geq \int_{E_n} n^2 \, d\mu = n^2 \mu(E_n)$$

and we conclude that $\mu(E_n) \leq C/n^2$.

We know from absolute continuity, Theorem 4.1.6, that there is a $\delta > 0$ such that $\int_A |f|^2 \, d\mu < \varepsilon/4$ whenever $\mu(A) < \delta$. Since $|f(x) - f_n(x)| = 0$ when $x \notin E_n$, we have

$$\begin{aligned} \|f - f_n\|^2 &= \int |f - f_n|^2 \, d\mu \\ &= \int_{E_n} |f - f_n|^2 \, d\mu \\ &\leq \int_{E_n} 4|f|^2 \, d\mu \\ &< 4\frac{\varepsilon}{4} = \varepsilon, \end{aligned}$$

whenever $n$ is sufficiently large that $\mu(E_n) \leq C/n^2 < \delta$. □

In fact, we can do better than approximating a function $f \in L^2[a,b]$ by bounded functions; we can approximate it by continuous functions or even step functions. But this is an approximation only in the $L^2$ norm! That is, we can show $\|f - g\|$ is small, but that does not imply anything about the size of $|f(x) - g(x)|$ for a particular $x$.

**Proposition 5.2.3. (Density of step functions and continuous functions).** *The step functions are dense in $L^2[a,b]$. That is, for any $\varepsilon > 0$ and any $f \in L^2[a,b]$, there is a step function $g : [a,b] \to \mathbb{R}$ such that $\|f - g\| < \varepsilon$. Likewise, there is a continuous function $h : [a,b] \to \mathbb{R}$ such that $\|f - h\| < \varepsilon$. The function $h$ may be chosen so $h(a) = h(b)$.*

**Proof.** By the preceding result we may choose $n$ so that $\|f - f_n\| < \frac{\varepsilon}{2}$. Note that $|f_n(x)| \leq n$ for all $x$. Suppose now that $\delta$ is any given small positive number. According to Theorem 4.4.5 there is a step function $g$ with $|g| \leq n$ and a measurable set $A$ with $\mu(A) < \delta$ such that

$|f_n(x) - g(x)| < \delta$ if $x \notin A$. Hence,

$$\begin{aligned}\|f_n - g\|^2 &= \int |f_n - g|^2 \, d\mu \\ &= \int_A |f_n - g|^2 \, d\mu + \int_{A^c} |f_n - g|^2 \, d\mu \\ &\le \int_A 4n^2 \, d\mu + \int_{A^c} \delta^2 \, d\mu \\ &\le 4n^2\mu(A) + \delta^2\mu(A^c) \le 4n^2\delta + 2\delta^2.\end{aligned}$$

Clearly, if we choose $\delta$ sufficiently small, then

$$\|f_n - g\| \le \sqrt{4n^2\delta + 2\delta^2} < \frac{\varepsilon}{2}.$$

It follows that $\|f - g\| \le \|f - f_n\| + \|f_n - g\| < \varepsilon$.

The proof for continuous functions is the same, except Exercise 4.4.7 3 is used in place of Theorem 4.4.5. The details are left as an exercise. ☐

In any vector space with a norm $\| \ \|$ we can talk about convergent sequences. The meaning is precisely what you would expect. As stated in Definition A.11.1 of our "background" appendix, if $\{v_n\}$ is a sequence in $\mathcal{V}$, then $\lim_{n\to\infty} v_n = w$ means $\lim_{n\to\infty} \|w - v_n\| = 0$.

This is equivalent to the usual definition in $\mathbb{R}$ except we use the norm $\| \ \|$ in place of absolute value. Similarly, we can define the concept of Cauchy sequence in a normed vector space. There are several items from Definition A.11.2 which we will now make use of. The reader may wish to review the material of Section A.9

**Definition 5.2.4. (Cauchy sequence, complete).** *Let $\mathcal{V}$ be a real vector space with inner product $\langle \ , \ \rangle$ and associated norm $\| \ \|$.*

- *A sequence $\{v_n\}$ in $\mathcal{V}$ is said to converge to $w \in \mathcal{V}$ provided* $\lim_{n\to\infty} \|w - v_n\| = 0$.
- *A sequence is called Cauchy provided for every $\varepsilon > 0$ there is an $N > 0$ such that $\|v_n - v_m\| < \varepsilon$ when $n, m \ge N$.*
- *If all Cauchy sequences in $\mathcal{V}$ converge, then $\mathcal{V}$ is called* complete.

An inner product space with a complete norm is one of the basic objects of analysis. It is commonly given the name *Hilbert space* after the German mathematician David Hilbert.

**Definition 5.2.5. (Hilbert space).** *A* Hilbert space *is a vector space with an inner product whose associated norm is complete.*

As with $\mathbb{R}$, in a Hilbert space a sequence is a Cauchy sequence if and only if it converges.

**Examples 5.2.6.**

(1) The vector space $\mathbb{R}^n$ with the usual dot product is a Hilbert space.

(2) The space of square summable sequences $\ell^2$ with the inner product of Exercise A.9.9 is a Hilbert space.

(3) In general, a vector space of functions with an inner product may not be complete. Both step functions and continuous functions are vector subspaces of $L^2[a, b]$ and inherit its inner product, but they are not complete. Indeed, Proposition 5.2.3 shows for any $f \in L^2[a, b]$ which is not continuous there is a Cauchy sequence of continuous functions converging to $f$ and hence *not* converging in the subspace of continuous functions.

Our next result asserts that $L^2[a, b]$ *is* complete and hence is a Hilbert space.

**Theorem 5.2.7. ($L^2[a, b]$ is complete).** *The vector space $L^2[a, b]$ with inner product $\langle\ ,\ \rangle$ is a Hilbert space.*

**Proof.** We have already shown that $L^2[a, b]$ is an inner product space. All that remains is to prove that the norm $\|\ \|$ is complete, i.e., that Cauchy sequences converge (in the $L^2$ norm).

Let $\{f_n\}_{n=1}^{\infty}$ be a Cauchy sequence. Then we may choose positive integers $n_i$ such that

$$\|f_m - f_n\| < \frac{1}{2^i}$$

whenever $m, n \geq n_i$. If we define $g_0 = 0$ and $g_i = f_{n_i}$ for $i > 0$, then

$$\|g_{i+1} - g_i\| < \frac{1}{2^i}$$

for $i \geq 1$ so, in particular, $\sum_{i=0}^{\infty} \|g_{i+1} - g_i\|$ converges, say to $S$.

Consider the function $h_n(x)$ defined by

$$h_n(x) = \sum_{i=0}^{n-1} |g_{i+1}(x) - g_i(x)|.$$

For any fixed $x$ the sequence $\{h_n(x)\}$ is monotone increasing so we may define the extended real-valued function $h$ by $h(x) = \lim_{n\to\infty} h_n(x)$. Note that by the Minkowski inequality

$$\|h_n\| \leq \sum_{i=0}^{n-1} \|g_{i+1} - g_i\| \leq S.$$

Hence, $\int h_n^2 \, d\mu = \|h_n\|^2 \leq S^2$. Since $h_n(x)^2$ is a monotonic increasing sequence of non-negative measurable functions converging to $h^2$, we conclude from the Monotone convergence theorem (Theorem 4.2.3) that $\displaystyle\int h^2 \, d\mu = \lim_{n\to\infty} \int h_n^2 \, d\mu \leq S^2$, so $h^2$ is integrable.

Since $h^2$ is integrable, $h(x)$ is finite almost everywhere. For each $x$ with finite $h(x)$ the series of real numbers $\sum_{i=0}^{\infty}(g_{i+1}(x) - g_i(x))$ converges absolutely and hence converges (see Theorem A.3.7). We denote its sum by $g(x)$. For $x$ in the set of measure 0 where $h(x) = +\infty$ we define $g(x) = 0$.

Notice that

$$g_n(x) = \sum_{i=0}^{n-1}(g_{i+1}(x) - g_i(x))$$

because it is a telescoping series. Hence,

$$\lim_{n\to\infty} g_n(x) = \lim_{n\to\infty} \sum_{i=0}^{n-1}(g_{i+1}(x) - g_i(x)) = g(x)$$

for almost all $x$. Moreover,

$$\begin{aligned} |g(x)| &= \lim_{n\to\infty} |g_n(x)| \\ &\leq \lim_{n\to\infty} \sum_{i=0}^{n-1} |g_{i+1}(x) - g_i(x)| \\ &= \lim_{n\to\infty} h_n(x) = h(x) \end{aligned}$$

for almost all $x$, so $|g(x)|^2 \leq h(x)^2$ and hence $|g(x)|^2$ is integrable and $g \in L^2[a,b]$.

We also observe that

$$|g(x) - g_n(x)|^2 \leq (|g(x)| + |g_n(x)|)^2 \leq (2h(x))^2.$$

Since $\lim_{n\to\infty} |g(x) - g_n(x)|^2 = 0$ for almost all $x$ the Lebesgue convergence theorem, Theorem 4.4.4, tells us $\lim_{n\to\infty} \int |g(x) - g_n(x)|^2 \, d\mu = 0$. This implies $\lim_{n\to\infty} \|g - g_n\| = 0$.

Hence, given $\varepsilon > 0$ there is an $i$ such that $\|g - g_i\| < \varepsilon/2$ and $1/2^i < \varepsilon/2$. Recalling that $g_i = f_{n_i}$ we see that whenever $m \geq n_i$ we have

$$\begin{aligned} \|g - f_m\| &\leq \|g - g_i\| + \|g_i - f_m\| \\ &< \varepsilon/2 + \varepsilon/2 = \varepsilon. \end{aligned}$$

Hence $\lim_{m\to\infty} \|g - f_m\| = 0$. □

We have shown that if $\{f_n\}$ is a Cauchy sequence in $L^2[a,b]$, then there is a function $g \in L^2[a,b]$ such that $\lim_{m\to\infty} \|g - f_m\| = 0$. This, however, does not immediately tell us for any particular $x \in [a,b]$ that $\lim_{m\to\infty} |g(x) - f_m(x)| = 0$, i.e., it implies nothing about pointwise convergence. On the other hand, our proof of convergence in $L^2[a,b]$ does imply the existence of a *subsequence* $\{f_{n_i}\}$ which converges pointwise almost everywhere. We formalize this as follows

**Corollary 5.2.8.** *If the sequence $\{f_n\}$ converges in $L^2[a,b]$ to $f$, then there exists a subsequence $\{f_{n_i}\}_{i=0}^{\infty}$ such that*

$$\lim_{i\to\infty} f_{n_i}(x) = f(x)$$

*for almost all $x \in [a,b]$.*

**Proof.** Since $\{f_n\}$ converges in $L^2[a,b]$ it is Cauchy (see part (1) of Exercise A.11.3). In the proof of Theorem 5.2.7 we constructed a function $g$, associated to the Cauchy sequence $\{f_n\}$, and a subsequence $g_i = f_{n_i}$ with the property that $\lim_{i\to\infty} \|g - g_i\| = 0$ in $L^2[a,b]$ and

$$\lim_{i\to\infty} g_i(x) = g(x)$$

for almost all $x \in [a,b]$. Since $\{g_i\}$ converges to $f$ and to $g$ in $L^2[a,b]$ we conclude that $f(x) = g(x)$ for almost all $x$. $\square$

## 5.3. Hilbert Space

In any Hilbert space we can, of course, talk about convergent series as well as sequences. The meaning is precisely what you would expect. In particular, if $\mathcal{H}$ is a Hilbert space and $\{u_n\}$ is a sequence in $\mathcal{H}$, then

$$\sum_{m=1}^{\infty} u_m = s$$

means $\lim s_n = s$ where

$$s_n = \sum_{m=1}^{n} u_m.$$

We will say a series $\sum_{m=1}^{\infty} u_m$ *converges absolutely* provided the sequence $\sum_{m=1}^{\infty} \|u_m\|$ converges.

**Proposition 5.3.1. (Absolutely convergent series).** *If a series in a Hilbert space converges absolutely, then it converges.*

**Proof.** Given $\varepsilon > 0$ there is an $N > 0$ such that whenever $n > m \geq N$,

$$\sum_{i=m}^{n} \|u_m\| \leq \sum_{i=m}^{\infty} \|u_m\| < \varepsilon.$$

Let $s_n = \sum_{i=1}^{n} u_i$, then $\|s_n - s_m\| \leq \sum_{i=m}^{n} \|u_m\| < \varepsilon$. It follows that $\{s_n\}$ is a Cauchy sequence. Hence it converges. $\square$

We will also talk about perpendicularity in $\mathcal{H}$. We say $x, y \in \mathcal{H}$ are *perpendicular* (written $x \perp y$) if $\langle x, y \rangle = 0$.

**Theorem 5.3.2. (Pythagorean theorem).** *If $x_1, x_2, \ldots, x_n$ are mutually perpendicular elements of a Hilbert space, then*

$$\Big\| \sum_{i=1}^{n} x_i \Big\|^2 = \sum_{i=1}^{n} \|x_i\|^2.$$

**Proof.** Consider the case $n = 2$. If $x \perp y$, then

$$\|x + y\|^2 = \langle x + y, x + y\rangle = \langle x, x\rangle + 2\langle x, y\rangle + \langle y, y\rangle = \|x\|^2 + \|y\|^2$$

since $\langle x, y\rangle = 0$. The general case follows by induction on $n$. $\square$

**Definition 5.3.3. (Bounded linear functional).** *A* bounded linear functional *on a Hilbert space $\mathcal{H}$ is a function $L : \mathcal{H} \to \mathbb{R}$ such that for all $v, w \in \mathcal{H}$ and $c_1, c_2 \in \mathbb{R}$, $L(c_1v + c_2w) = c_1L(u) + c_2L(w)$ and such that there is a constant $M$ satisfying $|L(v)| \leq M\|v\|$ for all $v \in \mathcal{H}$.*

The Cauchy-Schwarz inequality is a standard result for any real vector space with an inner product. A proof has been provided in Proposition A.9.5 of Appendix A. In the case of the Hilbert space $L^2[a, b]$ this is just the corollary to Hölder's inequality, given in Corollary 5.1.6.

**Proposition 5.3.4. (Cauchy-Schwarz inequality).** *If $(\mathcal{H}, \langle\ ,\ \rangle)$ is a Hilbert space and $v, w \in \mathcal{H}$, then*

$$|\langle v, w\rangle| \leq \|v\|\ \|w\|,$$

*with equality if and only if $v$ and $w$ are multiples of a single vector.*

For any fixed $x \in \mathcal{H}$ we may define $L : \mathcal{H} \to \mathbb{R}$ by $L(v) = \langle v, x\rangle$. Then $L$ is a linear function and as a consequence of the Cauchy-Schwarz inequality it is bounded. Indeed, $\|L(v)\| \leq M\|v\|$ where $M = \|x\|$. Our next goal is to prove that these are the *only* bounded linear functionals on $\mathcal{H}$.

**Lemma 5.3.5.** *Suppose $\mathcal{H}$ is a Hilbert space and $L : \mathcal{H} \to \mathbb{R}$ is a bounded linear functional which is not identically $0$. If $\mathcal{V} = L^{-1}(1)$, then there is a unique $x \in \mathcal{V}$ such that*

$$\|x\| = \inf_{v \in \mathcal{V}} \|v\|.$$

*That is, there is a unique vector in $\mathcal{V}$ closest to $0$. Moreover, the vector $x$ is perpendicular to every element of $L^{-1}(0)$, i.e., if $v \in \mathcal{H}$ and $L(v) = 0$ then $\langle x, v\rangle = 0$.*

**Proof.** We first observe that $\mathcal{V}$ is closed, i.e., that any convergent sequence in $\mathcal{V}$ has its limit in $\mathcal{V}$. To see this suppose $\lim x_n = x$ and

$x_n \in \mathcal{V}$. Then $|L(x) - L(x_n)| = |L(x - x_n)| \le M\|x - x_n\|$ for some $M$. Hence since $L(x_n) = 1$ for all $n$, we have

$$|L(x) - 1| \le \lim_{n\to\infty} M\|x - x_n\| = 0.$$

Therefore, $L(x) = 1$ and $x \in \mathcal{V}$.

Now let $d = \inf_{v\in\mathcal{V}} \|v\|$ and choose a sequence $\{x_n\}_{n=1}^{\infty}$ in $\mathcal{V}$ such that $\lim \|x_n\| = d$. We will show that this sequence is Cauchy and hence converges.

Notice that $(x_n + x_m)/2$ is in $\mathcal{V}$, so

$$\|(x_n + x_m)/2\| \ge d \quad \text{and} \quad \|x_n + x_m\| \ge 2d.$$

By the parallelogram law (Proposition A.9.6)

$$\|x_n - x_m\|^2 + \|x_n + x_m\|^2 = 2\|x_n\|^2 + 2\|x_m\|^2.$$

Hence,

$$\begin{aligned}\|x_n - x_m\|^2 &= 2\|x_n\|^2 + 2\|x_m\|^2 - \|x_n + x_m\|^2 \\ &\le 2\|x_n\|^2 + 2\|x_m\|^2 - 4d^2.\end{aligned}$$

As $m$ and $n$ tend to infinity the right side of this inequality goes to 0. Hence the left side does also and $\lim \|x_n - x_m\| = 0$. That is, the sequence $\{x_n\}_{n=1}^{\infty}$ is Cauchy. Let $x \in \mathcal{V}$ be the limit of this sequence. Since $\|x\| \le \|x - x_n\| + \|x_n\|$ for all $n$, we have

$$\|x\| \le \lim_{n\to\infty} \|x - x_n\| + \lim_{n\to\infty} \|x_n\| = d;$$

but $x \in \mathcal{V}$ implies $\|x\| \ge d$, so $\|x\| = d$.

To see that $x$ is unique suppose that $y$ is another element of $\mathcal{V}$ and $\|y\| = d$. Then $(x + y)/2$ is in $\mathcal{V}$ so $\|x + y\| \ge 2d$. Hence using the parallelogram law again,

$$\|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2 - \|x + y\|^2 \le 4d^2 - 4d^2 = 0.$$

We conclude that $x = y$.

Suppose that $v \in L^{-1}(0)$. We wish to show it is perpendicular to $x$. Note that for all $t \in \mathbb{R}$ the vector $x + tv \in L^{-1}(1)$, so $\|x + tv\|^2 \ge \|x\|^2$. Hence

$$\|x\|^2 + 2t\langle x, v\rangle + t^2\|v\|^2 \ge \|x\|^2, \text{ so}$$

$2t\langle x, v\rangle + t^2\|v\|^2 \geq 0$ for all $t \in \mathbb{R}$. This is possible only if $\langle x, v\rangle = 0$. □

In the following theorem we characterize all the bounded linear functionals on a Hilbert space. Each of them is obtained by taking the inner product with some fixed vector.

**Theorem 5.3.6.** *If $\mathcal{H}$ is a Hilbert space and $L : \mathcal{H} \to \mathbb{R}$ is a bounded linear functional, then there is a unique $x \in \mathcal{H}$ such that $L(v) = \langle v, x\rangle$.*

**Proof.** If $L(v) = 0$ for all $v$, then $x = 0$ has the property we want, so suppose $L$ is not identically 0. Let $x_0 \in \mathcal{H}$ be the unique point in $L^{-1}(1)$ with smallest norm, guaranteed by Lemma 5.3.5.

Suppose first that $v \in \mathcal{H}$ and $L(v) = 1$. Then $L(v - x_0) = L(v) - L(x_0) = 1 - 1 = 0$, so by Lemma 5.3.5 $\langle v - x_0, x_0\rangle = 0$. It follows that the vector $x = x_0/\|x_0\|^2$ is also perpendicular to $v - x_0$, so

$$\langle v, x\rangle = \langle v, \frac{x_0}{\|x_0\|^2}\rangle = \langle v - x_0, \frac{x_0}{\|x_0\|^2}\rangle + \langle x_0, \frac{x_0}{\|x_0\|^2}\rangle = 1 = L(v).$$

Hence, for any $v$ with $L(v) = 1$ we have $L(v) = \langle v, x\rangle$. Also, for any $v$ with $L(v) = 0$ we have $L(v) = 0 = \langle v, x\rangle$ by Lemma 5.3.5.

Finally, for an arbitrary $w \in \mathcal{H}$ with $L(w) = c \neq 0$ we define $v = w/c$, so $L(v) = L(w)/c = 1$. Hence,

$$L(w) = L(cv) = cL(v) = c\langle v, x\rangle = \langle cv, x\rangle = \langle w, x\rangle.$$

To see that $x$ is unique, suppose that $y \in \mathcal{H}$ has the same properties. Then for every $v \in \mathcal{H}$ we have $\langle v, x\rangle = L(v) = \langle v, y\rangle$. Thus, $\langle v, x - y\rangle = 0$ for all $v$ and, in particular, for $v = x - y$. We conclude that $\|x - y\|^2 = \langle x - y, x - y\rangle = 0$, so $x = y$. □

---

**Exercise 5.3.7.** If $\mathcal{V}_1$ and $\mathcal{V}_2$ are Hilbert spaces, we say a function $f : \mathcal{V}_1 \to \mathcal{V}_2$ is linear if for every $x, y \in \mathcal{V}_1$ and every $a, b \in \mathbb{R}$,

$$f(ax + by) = af(x) + bf(y).$$

The function $f$ is continuous at $x_0$ if for every $\varepsilon > 0$ there is a $\delta > 0$ such that $\|f(x) - f(x_0)\| < \varepsilon$ for all $x$ with $\|x - x_0\| < \delta$.

(1) Prove that a linear function $f : \mathcal{V}_1 \to \mathcal{V}_2$ is continuous (i.e., continuous at every $x_0 \in \mathcal{V}_1$) if and only if it is continuous at $x_0 = 0 \in \mathcal{V}_1$.

(2) Prove that a linear function $f : \mathcal{V}_1 \to \mathcal{V}_2$ is continuous if and only if there is an $M > 0$ such that for all $x \in \mathcal{V}_1$,

$$\|f(x)\| \leq M\|x\|.$$

(3) The *kernel* of a continuous linear function $f : \mathcal{V}_1 \to \mathcal{V}_2$ is defined to be

$$K = \{v \in \mathcal{V}_1 \mid f(v) = 0\}.$$

Prove that $K$ is a vector subspace of $\mathcal{H}$. Prove $K$ is closed; i.e., if $v_n \in K$ and $\lim v_n = w$, then $w \in K$.

(4) Suppose $\mathcal{H}$ is a Hilbert space, $L : \mathcal{H} \to \mathbb{R}$ is a bounded linear functional, and $\mathcal{V} = L^{-1}(0)$. Prove there is a linear function $P : \mathcal{H} \to \mathcal{V}$ such that $P(P(x)) = P(x)$ for all $x \in \mathcal{H}$ and such that $x - P(x)$ is perpendicular to $v$ for all $x \in \mathcal{H}$ and all $v \in \mathcal{V}$. The function $P$ is called an *orthogonal projection* onto $\mathcal{V}$.

## 5.4. Fourier Series

As always we would like to reproduce for a Hilbert space as many properties of $\mathbb{R}^n$ with its usual "dot product" as possible. In $\mathbb{R}^n$ it is very useful to have an orthonormal basis, i.e., a set of $n$ unit vectors $\{u_i\}_{i=1}^n$ which are pairwise perpendicular (and which necessarily then span $\mathbb{R}^n$ and are necessarily linearly independent). If we have such an orthonormal basis then it is not difficult to show (see Proposition A.9.8) that if we denote the dot product of $u$ and $v$ by $\langle u, v\rangle$, then for any $v \in \mathbb{R}^n$,

$$v = \sum_{i=1}^{n} \langle v, u_i\rangle u_i.$$

Moreover, this expression is unique, i.e., if

$$v = \sum_{i=1}^{n} a_i u_i$$

for some real numbers $a_i$, then $a_i = \langle v, u_i\rangle$.

It is these properties we wish to generalize to (infinite dimensional) Hilbert spaces. It is not generally possible to find vectors $\{u_i\}_{i=1}^{\infty}$ in a Hilbert space $\mathcal{H}$ such that any $v \in \mathcal{H}$ can be expressed as a *finite* linear combination of the $u_i$'s. Instead we want to write $v \in \mathcal{H}$ as an infinite series

$$v = \sum_{i=1}^{\infty} a_i u_i,$$

by which, of course, we will mean the series $\sum a_i u_i$ converges in $\mathcal{H}$ to $v$.

**Definition 5.4.1. (Orthonormal family).** *A family of vectors $\{u_n\}$ in a Hilbert space $\mathcal{H}$ is called* orthonormal *provided for each $n$, $\|u_n\| = 1$ and $\langle u_n, u_m \rangle = 0$ if $n \neq m$.*

**Theorem 5.4.2.** *If $\{u_n\}_{n=0}^{N}$ is a finite orthonormal family of vectors in a Hilbert space $\mathcal{H}$ and $w \in \mathcal{H}$, then the minimum value of*

$$\left\| w - \sum_{n=0}^{N} c_n u_n \right\|$$

*for all choices of $c_n \in \mathbb{R}$ occurs when $c_n = \langle w, u_n \rangle$.*

**Proof.** Let $c_n$ be arbitrary real numbers and define $a_n = \langle w, u_n \rangle$. Let

$$u = \sum_{n=0}^{N} a_n u_n, \quad \text{and} \quad v = \sum_{n=0}^{N} c_n u_n.$$

Notice that by the Pythagorean theorem (Theorem 5.3.2),

$$\langle u, u \rangle = \sum_{n=0}^{N} |a_n|^2 \text{ and } \langle v, v \rangle = \sum_{n=0}^{N} |c_n|^2.$$

Also,

$$\langle w, v \rangle = \sum_{n=0}^{N} c_n \langle w, u_n \rangle = \sum_{n=0}^{N} a_n c_n.$$

Hence,

$$\begin{aligned}
\|w-v\|^2 &= \langle w-v, w-v\rangle \\
&= \|w\|^2 - 2\langle w, v\rangle + \|v\|^2 \\
&= \|w\|^2 - 2\sum_{n=0}^{N} a_n c_n + \sum_{n=0}^{N} |c_n|^2 \\
&= \|w\|^2 - \sum_{n=0}^{N} |a_n|^2 + \sum_{n=0}^{N} (a_n - c_n)^2 \\
&= \|w\|^2 - \|u\|^2 + \sum_{n=0}^{N} |a_n - c_n|^2.
\end{aligned}$$

It follows that

$$\|w-v\|^2 \geq \|w\|^2 - \|u\|^2$$

for any choices of the $c_n$'s and we have equality if and only if $c_n = a_n$. That is, for all choices of $v$, the minimum value of $\|w-v\|^2$ occurs precisely when $v = u$. □

**Definition 5.4.3. (Complete orthonormal family).** *If $\{u_n\}_{n=0}^{\infty}$ is an orthonormal family of vectors in a Hilbert space $\mathcal{H}$, it is called* complete *if every $w \in \mathcal{H}$ can be written as an infinite series*

$$w = \sum_{n=0}^{\infty} c_n u_n$$

*for some choice of the numbers $c_n \in \mathbb{R}$.*

Theorem 5.4.2 suggests that the only reasonable choice for $c_n$ is $c_n = \langle w, u_n\rangle$ and we will show that this is the case. These numbers are sufficiently frequently used that they have a name.

**Definition 5.4.4. (Fourier series).** *The $n^{th}$ Fourier coefficient of $w$ with respect to an orthonormal family $\{u_n\}_{n=0}^{\infty}$ is the number $\langle w, u_n\rangle$. The infinite series*

$$\sum_{n=0}^{\infty} \langle w, u_n\rangle u_n$$

*is called the* Fourier series *of $w$.*

**Theorem 5.4.5. (Bessel's inequality).** *If $\{u_i\}_{i=0}^{\infty}$ is an orthonormal family of vectors in a Hilbert space $\mathcal{H}$ and $w$ is any element of $\mathcal{H}$, then*

$$\sum_{i=0}^{\infty} |\langle w, u_i \rangle|^2 \leq \|w\|^2.$$

*In particular this series converges.*

**Proof.** Let $s_n$ be the partial sum for the Fourier series. That is, $s_n = \sum_{i=0}^{n} \langle w, u_n \rangle u_n$. Then since the family is orthogonal, we know by the Pythagorean theorem (Theorem 5.3.2) that

$$\|s_n\|^2 = \sum_{i=0}^{n} \|\langle w, u_i \rangle u_i\|^2 = \sum_{i=0}^{n} |\langle w, u_i \rangle|^2. \tag{5.4.1}$$

This implies that $s_n \perp (w - s_n)$ because

$$\langle w - s_n, s_n \rangle = \langle w, s_n \rangle - \langle s_n, s_n \rangle = \sum_{i=0}^{n} |\langle w, u_n \rangle|^2 - \|s_n\|^2 = 0.$$

Since $s_n \perp (w - s_n)$ we know

$$\|w\|^2 = \|s_n\|^2 + \|w - s_n\|^2 \tag{5.4.2}$$

by the Pythagorean theorem again. Hence, by equation (5.4.1),

$$\sum_{i=0}^{n} |\langle w, u_n \rangle|^2 = \|s_n\|^2 \leq \|w\|^2.$$

Since $\|s_n\|^2$ is an increasing sequence it follows that the series

$$\sum_{i=0}^{\infty} |\langle w, u_n \rangle|^2 = \lim_{n \to \infty} \|s_n\|^2 \leq \|w\|^2$$

converges. □

**Proposition 5.4.6. (Fourier series converge).** *If $\{u_n\}_{n=0}^{\infty}$ is an orthonormal family of vectors in a Hilbert space $\mathcal{H}$ and $w \in \mathcal{H}$, then the Fourier series $\sum_{i=0}^{\infty} \langle w, u_i \rangle u_i$ with respect to $\{u_i\}_{i=0}^{\infty}$ converges. If the orthonormal family is complete, then it converges to $w$. Moreover, it is unique in the sense that if $\sum_{i=0}^{\infty} c_i u_i = w$, then $c_i = \langle w, u_i \rangle$.*

**Proof.** Let $s_n$ be the partial sum for the Fourier series. That is, $s_n = \sum_{i=0}^{n} \langle w, u_i \rangle u_i$. So if $n > m$, $s_n - s_m = \sum_{i=m+1}^{n} \langle w, u_i \rangle u_i$.

Then since the family is orthogonal, we know by Theorem 5.3.2 that

$$\|s_n - s_m\|^2 = \sum_{i=m+1}^{n} \|\langle w, u_i \rangle u_i\|^2 = \sum_{i=m+1}^{n} |\langle w, u_i \rangle|^2.$$

Since the series $\sum_{i=0}^{\infty} |\langle w, u_i \rangle|^2$ converges we conclude that given $\varepsilon > 0$, there is an $N > 0$ such that $\|s_n - s_m\|^2 < \varepsilon^2$ whenever $n, m \geq N$. In other words, the sequence $\{s_n\}$ is Cauchy, so it converges.

If the family $\{u_n\}_{n=0}^{\infty}$ is complete there exist numbers $c_i$ such that the series $\sum_{i=0}^{\infty} c_i u_i$ converges to $w$. So the partial sums $S_n = \sum_{i=0}^{n} c_i u_i$ satisfy

$$w = \lim_{n \to \infty} S_n.$$

Hence, since the linear functional $L(x) = \langle x, u_j \rangle$ is continuous

$$\langle w, u_j \rangle = \langle \lim_{n \to \infty} S_n, u_j \rangle = \lim_{n \to \infty} \langle S_n, u_j \rangle = c_j.$$

Thus, the sequence $\sum_{i=0}^{\infty} c_i u_i$ which converges to $w$ is the Fourier series of $w$. □

If Bessel's inequality is actually an equality, then the Fourier series for $w$ must converge to $w$ in $\mathcal{H}$. This result is called Parseval's theorem.

**Theorem 5.4.7. (Parseval's theorem).** *If $\{u_n\}_{n=0}^{\infty}$ is an orthonormal family of vectors in a Hilbert space $\mathcal{H}$ and $w \in \mathcal{H}$, then*

$$\sum_{i=0}^{\infty} |\langle w, u_i \rangle|^2 = \|w\|^2$$

*if and only if the Fourier series with respect to $\{u_n\}_{n=0}^{\infty}$ converges to $w$, i.e.,*

$$\sum_{i=0}^{\infty} \langle w, u_i \rangle u_i = w.$$

**Proof.** As above let $s_n$ be the partial sum for the Fourier series. We showed in equation (5.4.2) that $\|w\|^2 = \|s_n\|^2 + \|w - s_n\|^2$.

Clearly, $\lim \|w - s_n\| = 0$ if and only if $\lim \|s_n\|^2 = \|w\|^2$. Equivalently (using equation (5.4.1)), $\sum_{n=0}^{\infty} \langle w, u_n \rangle u_n = w$ if and only if $\sum_{n=0}^{\infty} |\langle w, u_n \rangle|^2 = \|w\|^2$. □

---

**Exercise 5.4.8.** Suppose $\mathcal{H}$ is a Hilbert space and $\{u_i\}_{i=0}^{\infty}$ is an orthonormal family. Define $P : \mathcal{H} \to \mathcal{H}$ by

$$P(w) = \sum_{i=0}^{\infty} \langle w, u_i \rangle u_i,$$

which we know converges by Proposition 5.4.6.

(1) Prove that $P$ and $(I - P)$ are continuous and linear, where $I : \mathcal{H} \to \mathcal{H}$ is the identity function. (See Exercise 5.3.7.)

(2) The *kernel* of $P$ is defined to be $K = \{u \in \mathcal{H} \mid P(u) = 0\}$. Prove that $K$ and $P(\mathcal{H})$ are closed subspaces of $\mathcal{H}$. (See Exercise 5.3.7.)

(3) Prove that every element $w \in \mathcal{H}$ can be written uniquely as $w = u + v$ with $u \in K$ and $v \in P(\mathcal{H})$.

## 5.5. Complex Hilbert Space

Up until now we have focused on real vector spaces and real-valued functions. However, much of modern mathematics and physics (especially quantum mechanics) is based on the study of complex vector spaces and, in particular, complex Hilbert spaces which we now investigate. In Section 6.2 we will discuss the integration of complex-valued functions and then in Section 6.3 we will consider a very important complex Hilbert space of square integrable functions.

However, our first task is to replicate the results of the previous section for complex vector spaces instead of real vector spaces. We must replace the inner product with a Hermitian form whose definition is repeated here from Definition A.10.3. The reader unfamiliar with complex vector spaces may wish to consult Section A.10.

**Definition 5.5.1. (Hermitian form, associated norm).** *A* Hermitian form *on a complex vector space $\mathcal{V}$ is a function $\langle\ ,\ \rangle : \mathcal{V} \times \mathcal{V} \to \mathbb{C}$ which satisfies:*

(1) **Conjugate symmetry:** $\langle v, w\rangle = \overline{\langle w, v\rangle}$ *for all* $v, w \in \mathcal{V}$.

(2) **Sesquilinearity:**

$$\langle c_1 v_1 + c_2 v_2, w\rangle = c_1\langle v_1, w\rangle + c_2\langle v_2, w\rangle \quad \text{and}$$
$$\langle v, c_1 w_1 + c_2 w_2, \rangle = \bar{c}_1\langle v, w_1\rangle + \bar{c}_2\langle v, w_2\rangle$$

*for all* $v_1, v_2, w_1, w_2 \in \mathcal{V}$ *and all* $c_1, c_2 \in \mathbb{C}$.

(3) **Positive definiteness:** *For all* $w \in \mathcal{V}$, $\langle w, w\rangle$ *is real and* $\geq 0$ *with equality only if* $w = 0$.

*The norm* $\|\ \|$ *associated to* $\langle\ ,\ \rangle$ *is defined by* $\|v\|^2 = \langle v, v\rangle$.

Note that, as is customary, we will use the same notation, namely $\langle\ ,\ \rangle$, for a Hermitian form on a complex vector space and an inner product on a real vector space and will also write $\|\ \|$ for the associated norm in both cases. The reader must know which is intended by the context. This is similar to our use of the same notation $|\ |$ for absolute value of real numbers and modulus of complex numbers. These notational conventions have the useful effect of making the proofs of some results about real inner product spaces and the proofs of their complex vector space analogs *identical* word for word and symbol for symbol. In other words, the syntax of the two proofs is identical, but the semantics is different with $\langle\ ,\ \rangle$ meaning a real inner product in one and a Hermitian form in the other. When this occurs as in Theorem 5.5.11, Proposition 5.5.12, and Theorem 5.5.13 below, rather than repeat the proof of the real version word for word we will simply refer the reader to that proof.

We will also need the following results which are analogous to results with the same names for real vector spaces.

**Proposition 5.5.2.** *If* $\mathcal{V}$ *is a complex vector space with Hermitian form* $\langle\ ,\ \rangle$ *and associated norm* $\|\ \|$, *then*

(1) **(Pythagorean theorem).** *If the vectors* $x_1, x_2, \ldots, x_n$ *are mutually perpendicular, then*

$$\Big\|\sum_{i=1}^n x_i\Big\|^2 = \sum_{i=1}^n \|x_i\|^2.$$

(2) **(Cauchy-Schwarz inequality).** *For all $v, w \in \mathcal{V}$,*

$$|\langle v, w\rangle| \leq \|v\| \, \|w\|,$$

*with equality if and only if $v$ and $w$ are multiples of a single vector.*

(3) **(Triangle inequality).** *For all $v, w \in \mathcal{V}$, $\|v+w\| \leq \|v\| + \|w\|$.*

**Proof.** This proposition combines the results of Proposition A.10.7, Proposition A.10.8, and Proposition A.10.9 from the discussion of complex vector spaces in Appendix A. The proofs can be found in Section A.10. □

**Definition 5.5.3. (Orthonormal family).** *A family of vectors $\{u_n\}$ in a complex Hilbert space $\mathcal{H}$ is called* orthonormal *provided for each $n$, $\|u_n\| = 1$ and $\langle u_n, u_m\rangle = 0$ if $n \neq m$.*

As with real vector spaces we have the notion of a complex Hilbert space in which Cauchy sequences converge.

**Definition 5.5.4. (Complex Hilbert space).** *If $\mathcal{H}$ is a complex vector space with Hermitian form $\langle\ ,\ \rangle$ and associated norm $\|\ \|$, then it is called a* complex Hilbert space *provided it is complete, i.e., sequences which are Cauchy with respect to the norm $\|\ \|$ converge.*

In a complex Hilbert space linear functionals take values in the complex numbers.

**Definition 5.5.5. (Bounded linear functional).** *If $\mathcal{H}$ is a complex Hilbert space, a* bounded linear functional *on $\mathcal{H}$ is a function $L : \mathcal{H} \to \mathbb{C}$ such that for all $v, w \in \mathcal{H}$ and $c_1, c_2 \in \mathbb{C}$, $L(c_1 v + c_2 w) = c_1 L(u) + c_2 L(w)$ and such that there is a real constant $M \geq 0$ satisfying $|L(v)| \leq M\|v\|$ for all $v \in \mathcal{H}$.*

The following theorem is the complex version of Theorem 5.3.6. The proof is very similar and is left as an exercise.

**Theorem 5.5.6.** *If $\mathcal{H}$ is a complex Hilbert space and $L : \mathcal{H} \to \mathbb{C}$ is a bounded linear functional, then there is a unique $x \in \mathcal{H}$ such that $L(v) = \langle v, x\rangle$.*

**Exercise 5.5.7.**

(1) If $\{u_n\}_{n=1}^{\infty}$ are elements of a complex Hilbert space $\mathcal{H}$, we say the series $\sum_{n=1}^{\infty} u_n$ converges *absolutely* if $\sum_{n=1}^{\infty} \|u_n\|$ converges. Prove that if a series converges absolutely, then it converges in $\mathcal{H}$.

(2) State an analog of Lemma 5.3.5 for complex Hilbert spaces and give its proof.

(3) Give the proof of Theorem 5.5.6.

(4) Suppose $\mathcal{H}$ is a complex Hilbert space, $L : \mathcal{H} \to \mathbb{C}$ is a bounded linear functional, and $\mathcal{V} = L^{-1}(0)$. Prove there is a linear function $P : \mathcal{H} \to \mathcal{V}$ such that $P(P(x)) = P(x)$ for all $x \in \mathcal{H}$ and such that $x - P(x)$ is perpendicular to $v$ for all $x \in \mathcal{H}$ and all $v \in \mathcal{V}$. The function $P$ is called an *orthogonal projection* onto $\mathcal{V}$.

**Theorem 5.5.8.** *If $\{u_n\}_{n=0}^{N}$ is a finite orthonormal family of vectors in a complex Hilbert space $\mathcal{H}$ and $w \in \mathcal{H}$, then the minimum value of*

$$\left\| w - \sum_{n=0}^{N} c_n u_n \right\|$$

*for all choices of $c_n \in \mathbb{C}$ occurs when $c_n = \langle w, u_n \rangle$.*

**Proof.** Let $c_n$ be arbitrary complex numbers and define $a_n = \langle w, u_n \rangle$. Let

$$u = \sum_{n=0}^{N} a_n u_n \quad \text{and} \quad v = \sum_{n=0}^{N} c_n u_n.$$

Notice that the Pythagorean theorem and Proposition 5.5.2 imply that

$$\langle u, u \rangle = \sum_{n=0}^{N} |a_n|^2 \quad \text{and} \quad \langle v, v \rangle = \sum_{n=0}^{N} |c_n|^2.$$

Also,

$$\begin{aligned} \langle w, v \rangle &= \sum_{n=0}^{N} \bar{c}_n \langle w, u_n \rangle \\ &= \sum_{n=0}^{N} a_n \bar{c}_n. \end{aligned}$$

Hence,

$$\begin{aligned}
\|w-v\|^2 &= \langle w-v, w-v\rangle \\
&= \|w\|^2 - \langle w, v\rangle - \langle v, w\rangle + \|v\|^2 \\
&= \|w\|^2 - \sum_{n=0}^{N} a_n\bar{c}_n - \sum_{n=0}^{N} \bar{a}_n c_n + \sum_{n=0}^{N} |c_n|^2 \\
&= \|w\|^2 - \sum_{n=0}^{N} |a_n|^2 + \sum_{n=0}^{N}\langle a_n - c_n, a_n - c_n\rangle \\
&= \|w\|^2 - \|u\|^2 + \sum_{n=0}^{N} |a_n - c_n|^2.
\end{aligned}$$

It follows that

$$\|w-v\|^2 \geq \|w\|^2 - \|u\|^2$$

for any choices of the $c_n$'s and we have equality if only if $c_n = a_n$. That is, for all choices of $v$, the minimum value of $\|w-v\|^2$ occurs precisely when $v = u$. □

**Definition 5.5.9. (Complete orthonormal family).** *If $\{u_n\}_{n=0}^{\infty}$ is an orthonormal family of vectors in a complex Hilbert space $\mathcal{H}$, it is called* complete *if every $w \in \mathcal{H}$ can be written as an infinite series*

$$w = \sum_{n=0}^{\infty} c_n u_n$$

*for some choice of the numbers $c_n \in \mathbb{C}$.*

Theorem 5.5.8 suggests that the only reasonable choice for $c_n$ is $c_n = \langle w, u_n\rangle$ and we will show that this is the case. As in the real case these numbers are called Fourier coefficients.

**Definition 5.5.10. (Complex Fourier series).** *If $\mathcal{H}$ is a complex Hilbert space, the* complex Fourier coefficient *of $w \in \mathcal{H}$ with respect to an orthonormal family $\{u_n\}_{n=0}^{\infty}$ is the number $\langle w, u_n\rangle$. The infinite series*

$$\sum_{n=0}^{\infty}\langle w, u_n\rangle u_n$$

*is called the* Fourier series.

**Theorem 5.5.11. (Bessel's inequality).** *If $\{u_i\}_{i=0}^{\infty}$ is an orthonormal family of vectors in a complex Hilbert space $\mathcal{H}$ and $w$ is any element of $\mathcal{H}$, then*

$$\sum_{i=0}^{\infty} |\langle w, u_i \rangle|^2 \leq \|w\|^2.$$

*In particular, this series converges.*

**Proof.** The proof is identical to that of Theorem 5.4.5 when $\langle\ ,\ \rangle$ is interpreted as the Hermitian form on $\mathcal{H}$ rather than an inner product. □

**Proposition 5.5.12. (Fourier series converge).** *If $\{u_n\}_{n=0}^{\infty}$ is an orthonormal family of vectors in a complex Hilbert space $\mathcal{H}$ and $w \in \mathcal{H}$, then the Fourier series $\sum_{i=0}^{\infty} \langle w, u_i \rangle u_i$ with respect to $\{u_i\}_{i=0}^{\infty}$ converges. If the orthonormal family is complete, then it converges to $w$. Moreover, it is unique in the sense that if $\sum_{i=0}^{\infty} c_i u_i = w$, then $c_i = \langle w, u_i \rangle$.*

**Proof.** The proof is identical to the proof of Proposition 5.4.6. □

As in the real case, if Bessel's inequality is actually an equality, then the Fourier series for $w$ must converge to $w$ in $\mathcal{H}$.

**Theorem 5.5.13. (Parseval's theorem).** *Suppose $\{u_n\}_{n=0}^{\infty}$ is an orthonormal family of vectors in a complex Hilbert space $\mathcal{H}$ and $w \in \mathcal{H}$, then*

$$\sum_{i=0}^{\infty} \langle w, u_i \rangle^2 = \|w\|^2$$

*if and only if the Fourier series with respect to $\{u_n\}_{n=0}^{\infty}$ converges to $w$, i.e.,*

$$\sum_{i=0}^{\infty} \langle w, u_i \rangle u_i = w.$$

**Proof.** The proof is identical to the proof of Theorem 5.4.7. □

Chapter 6

# Classical Fourier Series

In this chapter we describe the classical expansion of a square integrable function as a trigonometric series and then its analogue for square integrable complex-valued functions. We will work on the interval $[a, b] = [-\pi, \pi]$ because it greatly simplifies the formulas.

## 6.1. Real Fourier Series

In addition to using the domain $[-\pi, \pi]$ for our functions it simplifies notation to alter the definition of the inner product by a factor of $1/\pi$.

**Definition 6.1.1. (Inner product on $L^2[-\pi, \pi]$).** *We define the inner product $\langle\ ,\ \rangle$ on $L^2[-\pi, \pi]$, the vector space of square integrable functions on $[-\pi, \pi]$, by*

$$\langle f, g\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} fg\ d\mu.$$

Clearly, the factor $1/\pi$ does not alter the property of $\langle\ ,\ \rangle$ being an inner product or of the associated norm being a norm. The advantage of this definition is we have a fairly simple orthonormal family for the Hilbert space $L^2[-\pi, \pi]$.

**Theorem 6.1.2.** *The family of functions*

$$\mathcal{F} = \left\{ \frac{1}{\sqrt{2}},\ \cos nx,\ \sin nx \right\}_{n=1}^{\infty}$$

*is an orthonormal family in* $L^2[-\pi, \pi]$.

**Proof.** Recall that Euler's formula says $e^{i\theta} = \cos\theta + i\sin\theta$ and hence

$$e^{inx} = \cos nx + i \sin nx$$

and

$$e^{-inx} = \cos nx - i \sin nx.$$

From this it is straightforward to calculate

$$\cos nx = \frac{e^{inx} + e^{-inx}}{2} \quad \text{and} \quad \sin nx = \frac{e^{inx} - e^{-inx}}{2i}.$$

Multiplying the expressions for $\cos nx$ and $\cos mx$ we see

$$\cos nx \cos mx = \frac{e^{i(n+m)x} + e^{i(n-m)x} + e^{i(m-n)x} + e^{-i(n+m)x}}{4}. \tag{6.1.1}$$

According to part (3) of Exercise1.6.3 for any integer $k \neq 0$,

$$\int_{-\pi}^{\pi} e^{ikx}\ dx = \frac{e^{ik\pi} - e^{-ik\pi}}{ik} = 0.$$

Therefore, since $n, m > 0$, the expression (6.1.1) for $\cos nx \cos mx$ implies that when $n \neq m$,

$$\langle \cos nx, \cos mx \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} \cos nx \cos mx\ dx = 0.$$

Also, when $n = m$ we conclude

$$\begin{aligned}
\| \cos nx \|^2 &= \langle \cos nx, \cos nx \rangle \\
&= \frac{1}{\pi} \int_{-\pi}^{\pi} \cos^2 nx\ dx \\
&= \frac{1}{\pi} \int_{-\pi}^{\pi} \frac{e^{2inx} + 2 + e^{-2inx}}{4}\ dx \\
&= 1.
\end{aligned}$$

A similar argument shows $\langle \sin nx, \cos mx \rangle = 0$ and also that $\langle \sin nx, \sin mx \rangle = 1$ if $n = m$ and 0 otherwise.

Finally, it is clear that $\| \frac{1}{\sqrt{2}} \| = 1$ and

$$\langle \frac{1}{\sqrt{2}}, \sin nx \rangle = \langle \frac{1}{\sqrt{2}}, \cos nx \rangle = 0.$$

□

We can now define the classical Fourier coefficients. With one exception these are just the special case for the orthonormal family $\mathcal{F}$ of the general Fourier coefficients defined in Definition 5.4.4. The one exception is the Fourier coefficient $A_0$, corresponding to the element $\frac{1}{\sqrt{2}}$ in the orthonormal family $\mathcal{F}$. It is not the one given by Definition 5.4.4. According to that definition the value should be

$$\langle f, \frac{1}{\sqrt{2}} \rangle = \frac{1}{\pi\sqrt{2}} \int f \; d\mu.$$

However, in the Fourier series that coefficient should be multiplied by the (constant) function $\frac{1}{\sqrt{2}}$ giving the value

$$\frac{1}{2\pi} \int f \; d\mu.$$

For simplicity it is conventional to combine the coefficient and the constant function $\frac{1}{\sqrt{2}}$ in this way and define $A_0$ as below, simplifying the expression of the Fourier series.

**Definition 6.1.3. (Classical Fourier coefficients).** *If $f$ is an element of $L^2[-\pi, \pi]$, then its* classical Fourier coefficients *are*

$$A_0 = \frac{1}{2\pi} \int f(x) \; d\mu,$$
$$A_n = \frac{1}{\pi} \int f(x) \cos nx \; d\mu,$$
$$B_n = \frac{1}{\pi} \int f(x) \sin nx \; d\mu,$$

*for $n > 0$. The* Fourier series *of $f$ is*

$$A_0 + \sum_{n=1}^{\infty} A_n \cos nx + \sum_{n=1}^{\infty} B_n \sin nx.$$

The goal of the remainder of this section is to show that this Fourier series with respect to the orthonormal family

$$\mathcal{F} = \left\{ \frac{1}{\sqrt{2}},\ \cos nx,\ \sin nx \right\}_{n=1}^{\infty}$$

converges to $f$ in $L^2[-\pi, \pi]$. Hence, the orthonormal family $\mathcal{F}$ is complete.

**Definition 6.1.4. (Algebra of functions).** *An* algebra of functions *is a vector space $\mathcal{A}$ of real-valued functions defined on some set $X$ with the additional property that if $f, g \in \mathcal{A}$, then $fg \in \mathcal{A}$.*

We will be particularly interested in the set which is the unit circle in $\mathbb{R}^2$ and which we denote by $\mathbb{T}$.[1] Note that $\mathbb{T} = \{(\cos\theta, \sin\theta) \mid -\pi \leq \theta \leq \pi\}$. We will identify points in $\mathbb{T}$ by the parameter $\theta$ noting that $\theta = -\pi$ and $\theta = \pi$ correspond to the same point of $\mathbb{T}$. The vector space of all real-valued continuous functions defined on $\mathbb{T}$ will be denoted $C(\mathbb{T})$. Note that $C(\mathbb{T})$ is an algebra of functions. We will sometimes abuse notation slightly and consider $C(\mathbb{T})$ as the set of continuous functions $h : [-\pi, \pi] \to \mathbb{R}$ which satisfy $h(-\pi) = h(\pi)$.

**Theorem 6.1.5. (Stone-Weierstrass).** *Suppose $\mathcal{A} \subset C(\mathbb{T})$ is an algebra satisfying*

(1) *the constant function* 1 *is in $\mathcal{A}$, and*

(2) *$\mathcal{A}$ separates points; i.e., for any distinct $\theta_0$ and $\theta_1$ in $\mathbb{T}$ there is $p \in \mathcal{A}$ such that $p(\theta_0) \neq p(\theta_1)$.*

*Then given any $\varepsilon > 0$ and any $f \in C(\mathbb{T})$ there is $p \in \mathcal{A}$ such that $|f(\theta) - p(\theta)| < \varepsilon$ for all $\theta \in \mathbb{T}$.*

A proof can be found in Chapter III of [**L**] or in Chapter 9, Section 9 of [**Ro**]. This result is usually stated in greater generality than we do here. For example, the set $\mathbb{T}$ can be replaced by any compact metric space, but the special case above suffices for our purposes. Indeed, what we will use is the fact that any $f \in C(\mathbb{T})$ can be approximated by a so-called "trigonometric polynomial". The following corollary to the Stone-Weierstrass theorem makes this precise.

[1] It is denoted by $\mathbb{T}$ because it is the one-dimensional torus $\mathbb{T}^1$.

**Corollary 6.1.6.** *Suppose that $g$ is a continuous function defined on $[-\pi, \pi]$ such that $g(-\pi) = g(\pi)$. If $\varepsilon > 0$, then there are $N > 0$ and $a_n, b_n \in \mathbb{R},\ 1 \le n \le N$ such that $|g(x) - p(x)| < \varepsilon$, for all $x$, where*

$$p(x) = a_0 + \sum_{n=1}^{N} a_n \cos nx + \sum_{n=1}^{N} b_n \sin nx.$$

**Proof.** Let $\mathcal{A}$ be the collection of all functions on $\mathbb{T}$ of the form

$$q(\theta) = a_0 + \sum_{n=1}^{N} a_n \cos n\theta + \sum_{n=1}^{N} b_n \sin n\theta$$

for some choices of $N$, $a_n$, and $b_n$ and $-\pi \le \theta \le \pi$. Then $\mathcal{A}$ is a vector space and contains the constant function 1. It is an algebra as a consequence of the following trigonometric identities (see part (1) of Exercise 6.1.10):

$$\begin{aligned}
\sin(\theta_1)\cos(\theta_2) &= \frac{1}{2}\big(\sin(\theta_1+\theta_2) + \sin(\theta_1-\theta_2)\big),\\
\cos(\theta_1)\cos(\theta_2) &= \frac{1}{2}\big(\cos(\theta_1+\theta_2) + \cos(\theta_1-\theta_2)\big),\\
\sin(\theta_1)\sin(\theta_2) &= \frac{1}{2}\big(\cos(\theta_1+\theta_2) - \cos(\theta_1-\theta_2)\big).
\end{aligned}$$

It is also the case that $\mathcal{A}$ separates points of $\mathbb{T}$. To see this, note that if $-\pi < \theta_1, \theta_2 \le \pi$ are real numbers with $\sin\theta_1 = \sin\theta_2$ and $\cos\theta_1 = \cos\theta_2$ then $\theta_1 = \theta_2$. Therefore, if $\theta_1 \ne \theta_2$, then either $\sin\theta_1 \ne \sin\theta_2$ or $\cos\theta_1 \ne \cos\theta_2$.

We may consider the function $g$ given in the hypothesis as a function $g : \mathbb{T} \to \mathbb{R}$. It is well defined and continuous at the point of $\mathbb{T}$ where $\theta = \pi$ (which is also the point where $\theta = -\pi$) because $g(-\pi) = g(\pi)$.

Given $\varepsilon > 0$, we apply the Stone-Weierstrass theorem to the function $g$ and the algebra $\mathcal{A}$. We conclude that there is an $N > 0$ and real numbers $a_n$ and $b_n,\ 1 \le n \le N$ such that the function

$$p(\theta) = a_0 + \sum_{n=1}^{N} a_n \cos n\theta + \sum_{n=1}^{N} b_n \sin n\theta$$

satisfies $|g(\theta) - p(\theta)| < \varepsilon$ for all $\theta \in [-\pi, \pi]$. Note that $p(-\pi) = p(\pi)$. The function $p(x)$ for $x \in [-\pi, \pi]$ then satisfies the conclusion of the theorem. □

**Theorem 6.1.7. (Fourier series converge in $L^2$).** *Suppose that $f \in L^2[-\pi, \pi]$. Then the Fourier series for $f$ with respect to the orthonormal family $\mathcal{F}$ converges to $f$ in $L^2[-\pi, \pi]$. In particular, the orthonormal family $\mathcal{F}$ is complete.*

**Proof.** Given $\varepsilon > 0$, we know by Proposition 5.2.3 there is a continuous function $g \in L^2[-\pi, \pi]$ such that $g(-\pi) = g(\pi)$ and $\|f - g\| < \varepsilon/2$.

By Corollary 6.1.6 to the Stone-Weierstrass theorem there is a function

$$p(x) = a_0 + \sum_{n=1}^{N} a_n \cos nx + \sum_{n=1}^{N} b_n \sin nx$$

with $|g(x) - p(x)| < \sqrt{\varepsilon}/2$ for all $x$. So $\|g - p\|^2 = \int (g - p)^2 \, d\mu \leq \int \varepsilon/4 \, d\mu = \varepsilon/2\pi$. Hence, $\|f - p\| \leq \|f - g\| + \|g - p\| < \varepsilon/2 + \varepsilon/2\pi < \varepsilon$.

Let

$$S_N(x) = A_0 + \sum_{n=1}^{N} A_n \cos nx + \sum_{n=1}^{N} B_n \sin nx$$

where $A_n$ and $B_n$ are the Fourier coefficients for $f$ with respect to $\mathcal{F}$. Then $S_N(x)$ is the partial sum of the Fourier series of $f$. According to Theorem 5.4.2 for every $m \geq N$, $\|f - S_m\| \leq \|f - p\|$, so $\|f - S_m\| < \varepsilon$. This proves $\lim_{n \to \infty} \|f - S_m\| = 0$. □

It is important to note that while we have proved that the Fourier series of a function $f \in L^2[-\pi, \pi]$ converges to $f$ in the Hilbert space $L^2[-\pi, \pi]$ this is not at all the same thing as saying that for a fixed $x_0$ the Fourier series

$$A_0 + \sum_{n=1}^{\infty} A_n \cos nx_0 + \sum_{n=1}^{\infty} B_n \sin nx_0$$

converges to $f(x_0)$. However, if we define the partial sum

$$S_m = A_0 + \sum_{n=1}^{m} A_n \cos nx_0, + \sum_{n=1}^{m} B_n \sin nx_0$$

then according to Corollary 5.2.8 the fact that $S_m$ converges to $f$ in $L^2[-\pi, \pi]$ implies there is a *subsequence* $\{n_i\}$ such that $\lim_{i\to\infty} S_{n_i}(x) = f(x)$ for almost all $x \in [-\pi, \pi]$.

It turns out, however, it is not necessary to take a subsequence! In what is considered one of the most remarkable achievements of twentieth century mathematics Lennart Carleson proved the following result in [**C**].

**Theorem 6.1.8. (Carleson's theorem).** *Suppose $f \in L^2[-\pi, \pi]$ and*

$$A_0 + \sum_{n=1}^{\infty} A_n \cos nx + \sum_{n=1}^{\infty} B_n \sin nx$$

*is its classical Fourier series. Then this series converges to $f(x)$ for almost all values of $x \in [-\pi, \pi]$.*

The proof is considered difficult and delicate and is well beyond the scope of this text. In fact, Carleson proved the more general analog of this result for complex Fourier series (see Theorem 6.3.5 below).

In general, if a function $f$ is continuous, its Fourier series may not converge pointwise to $f$ at every point. But if a function $f$ is differentiable the situation is better.

**Theorem 6.1.9.** *If $f : [-\pi, \pi] \to \mathbb{R}$ is differentiable at $x_0 \in (-\pi, \pi)$, then the Fourier series of $f$ at $x_0$,*

$$A_0 + \sum_{n=1}^{\infty} A_n \cos nx_0 + \sum_{n=1}^{\infty} B_n \sin nx_0,$$

*converges to $f(x_0)$. If the right and left derivatives of $f$ exist at $-\pi$ and $\pi$ respectively, then the Fourier series evaluated at either $-\pi$ or $\pi$ converges to*

$$\frac{f(-\pi) + f(\pi)}{2}.$$

Again the proof of this result is outside the scope of this text. A more general result (with proof) can be found as Theorem 1.2.24 of the text [**P**] by Pinsky.

**Exercise 6.1.10.**

(1) Use Euler's formula which says $e^{i\theta} = \cos\theta + i\sin\theta$ to derive the trigonometric identities:

$$\sin\theta_1\cos\theta_2 = \frac{1}{2}\big(\sin(\theta_1+\theta_2) + \sin(\theta_1-\theta_2)\big),$$
$$\cos\theta_1\cos\theta_2 = \frac{1}{2}\big(\cos(\theta_1+\theta_2) + \cos(\theta_1-\theta_2)\big),$$
$$\sin\theta_1\sin\theta_2 = \frac{1}{2}\big(\cos(\theta_1+\theta_2) - \cos(\theta_1-\theta_2)\big).$$

(2) A function $f$ is called *even* provided $f(-x) = f(x)$ for all $x$. If $f \in L^2[-\pi,\pi]$ is even, prove that its classical Fourier coefficients $B_n$ are 0 for all $n > 0$.

(3) If $f \in L^2[-\pi,\pi]$, prove that for each $n \in \mathbb{N}$,

$$\Big(\int_{-\pi}^{\pi} |x^n f(x)|\ d\mu\Big)^2 \le \frac{2\pi^{2n+1}}{2n+1}\int_{-\pi}^{\pi} f^2\ d\mu.$$

(4) Find the Fourier series for the function $f(x) = x$ on $[-\pi,\pi]$. Use it with a suitable value of $x$ to show

$$\frac{\pi}{4} = \sum_{n=0}^{\infty}(-1)^n\frac{1}{2n+1}.$$

(5) Find the Fourier series for the function $f(x) = x^2$ on $[-\pi,\pi]$. Use it with a suitable value of $x$ to evaluate

$$\sum_{n=1}^{\infty}\frac{1}{n^2}.$$

(6) Let $\mathcal{P} = \{c + (x^2-1)p(x) \mid p(x) \text{ is a polynomial and } c \in \mathbb{R}\}$. Prove that for any $\varepsilon > 0$ and any continuous $g : [-1,1] \to \mathbb{R}$ with $g(-1) = g(1)$ there is a $p_0 \in \mathcal{P}$ such that $|g(x) - p_0(x)| < \varepsilon$ for all $x \in [-1,1]$. You may use the Stone-Weierstrass theorem.

(7) Give an example of a sequence of functions $\{f_n\}$ which converge in $L^2[a,b]$ to a function $f$, but for which the sequence $\{f_n(x)\}$ fails to have a limit for every $x \in [a,b]$. *Hint:* See part (5) of Exercise 3.3.5.

## 6.2. Integrable Complex-Valued Functions

Now we want to consider functions $f : [a, b] \to \mathbb{C}$, i.e., functions whose domain is an interval of reals, but whose values are complex numbers.

**Definition 6.2.1. (Integrable complex function).** *Suppose $f : [a, b] \to \mathbb{C}$ is a complex-valued function and $u(x) = \Re(f(x))$, the real part of $f$, and $v(x) = \Im(f(x))$, the imaginary part of $f$. Then $f$ is called measurable if $u$ and $v$ are measurable. We say $f$ is integrable if $u$ and $v$ are integrable and in that case we define*

$$\int_a^b f \, d\mu = \int_a^b u \, d\mu + i \int_a^b v \, d\mu.$$

**Proposition 6.2.2.** *The integral is a complex linear functional, i.e., if $c_1, c_2 \in \mathbb{C}$ and $f, g$ are integrable, then $c_1 f + c_2 g$ is integrable and*

$$\int_a^b c_1 f + c_2 g \, d\mu = c_1 \int_a^b f \, d\mu + c_2 \int_a^b g \, d\mu.$$

*It also commutes with complex conjugation, i.e., the function $f$ is integrable if and only if its complex conjugate $\bar{f}$ is integrable and*

$$\int_a^b \bar{f} \, d\mu = \overline{\int_a^b f \, d\mu}.$$

The proof follows easily from the linearity of the integral of real-valued functions and the definition of complex multiplication. It is left as an exercise.

**Proposition 6.2.3.** *Suppose $f : [a, b] \to \mathbb{C}$ is measurable. Then:*

(1) *The function $f$ is integrable if and only if the real-valued function $|f(x)|$ is integrable. Moreover,*

$$\Big| \int_a^b f \, d\mu \Big| \le \int_a^b |f| \, d\mu.$$

(2) *The function $|f(x)|^2$ is integrable if and only if $u(x) = \Re(f(x))$ and $v(x) = \Im(f(x))$ are in $L^2[a, b]$.*

**Proof.** To show part (1) we note that since $|f(x)| \ge |u(x)|$ for all $x$, it follows that if $|f(x)|$ is integrable, then $|u(x)|$ is integrable and hence $u(x)$ is integrable. The same argument applies to $v(x)$.

Conversely, if $u$ and $v$ are integrable, then so are $|u|$ and $|v|$. Since

$$|f(x)| = \sqrt{u(x)^2 + v(x)^2} \leq \sqrt{\Big(|u(x)| + |v(x)|\Big)^2} = |u(x)| + |v(x)|$$

for all $x$, it follows that $|f(x)|$ is integrable.

To prove the inequality from (1) we first show it for simple functions. Suppose

$$f(x) = \sum_{i=1}^{m} c_i \mathfrak{X}_{A_i}$$

where $\{A_i\}_{i=1}^m$ is a measurable partition of $[a, b]$ and $c_i \in \mathbb{C}$. Then

$$\begin{aligned}
\Big| \int_a^b f \, d\mu \Big| &= \Big| \sum_{i=1}^{m} c_i \mu(A_i) \Big| \\
&\leq \sum_{i=1}^{m} |c_i \mu(A_i)| \\
&= \sum_{i=1}^{m} |c_i| \mu(A_i) \\
&= \int_a^b |f| \, d\mu
\end{aligned}$$

where the inequality is a consequence of the triangle inequality for $\mathbb{C}$.

To prove the general result we let $u$ and $v$ be the real and complex parts of $f$. By part (5) of Exercise 4.4.7 there is a sequence of simple functions $\{u_n\}$ converging pointwise to $u$ such that $|u_n(x)| \leq |u(x)|$ for all $x$. Likewise, there is a sequence of simple functions $\{v_n\}$ converging pointwise to $v$ such that $|v_n(x)| \leq |v(x)|$ for all $x$. Let $f_n = u_n + iv_n$ and observe that

$$|f_n(x)|^2 = |u_n(x)|^2 + |v_n(x)|^2 \leq |u(x)|^2 + |v(x)|^2 = |f(x)|^2.$$

Since $f_n$ converges pointwise to $f$, the functions $|f_n|$ converge pointwise to $|f|$. The Lebesgue convergence theorem applied to the sequences $\{u_n\}$ and $\{v_n\}$ (which are bounded by $|u|$ and $|v|$ respectively) then tells us that

$$\begin{aligned}\lim_{n\to\infty}\int f_n\, d\mu &= \lim_{n\to\infty}\int (u_n + iv_n)\, d\mu\\ &= \lim_{n\to\infty}\int u_n\, d\mu + i\lim_{n\to\infty}\int v_n\, d\mu\\ &= \int u\, d\mu + i\int v\, d\mu\\ &= \int f\, d\mu.\end{aligned}$$

Also applying Lebesgue convergence to $\{|f_n|\}$ which is bounded by $|f|$ we see

$$\lim_{n\to\infty}\int |f_n|\, d\mu = \int |f|\, d\mu.$$

Since $f_n$ is a simple function for which we have proved the desired inequality we conclude that

$$\begin{aligned}\Big|\int f\, d\mu\Big| &= \lim_{n\to\infty}\Big|\int f_n\, d\mu\Big|\\ &\le \lim_{n\to\infty}\int |f_n|\, d\mu\\ &= \int |f|\, d\mu.\end{aligned}$$

To prove (2) we observe that since $|f(x)|^2 \ge u(x)^2$ for all $x$, it follows that integrability of $|f(x)|^2$ implies integrability of $u(x)^2$. The same argument applies to $v(x)$. Conversely, if $u^2$ and $v^2$ are integrable, then so is $|f(x)|^2 = u(x)^2 + v(x)^2$. □

**Definition 6.2.4. (Square integrable complex functions).** *The set of all complex functions $f : [a,b] \to \mathbb{C}$ such that $|f(x)|^2$ is integrable are called the square integrable complex functions and will be denoted $L^2_{\mathbb{C}}[a,b]$.*

**Proposition 6.2.5.** *The set $L^2_{\mathbb{C}}[a,b]$ is a complex vector space. Moreover, if $f, g \in L^2_{\mathbb{C}}[a,b]$, then $f\bar{g}$ is integrable.*

**Proof.** Suppose $f, g \in L^2_{\mathbb{C}}[a,b]$, so the four functions $u_1(x) = \Re(f(x))$, $u_2(x) = \Re(g(x))$, $v_1(x) = \Im(f(x))$, and $v_2(x) = \Im(g(x))$ are all in $L^2[a,b]$. It follows that $(u_1 + u_2)$ and $(v_1 + v_2)$ are in $L^2[a,b]$, so

$f + g = (u_1 + u_2) + i(v_1 + v_2)$ is in $L^2_{\mathbb{C}}[a, b]$. Clearly, if $c \in \mathbb{C}$, then $cf \in L^2_{\mathbb{C}}[a, b]$, so $L^2_{\mathbb{C}}[a, b]$ is a complex vector space.

Also, $f\bar{g} = (u_1u_2 - v_1v_2) + i(u_1v_2 + v_1u_2)$. Since the product of two functions in $L^2[a, b]$ is integrable it follows that $f\bar{g}$ is integrable. □

---

**Exercise 6.2.6.**

(1) Prove Proposition 6.2.2.

(2) Suppose $f_n : [a, b] \to \mathbb{C}$, $n \in \mathbb{N}$ is a sequence of measurable functions and $u_n(x) = \Re(f_n(x))$ and $v_n(x) = \Im(f_n(x))$. Prove that $\lim_{n\to\infty} f_n = f$ in $L^2_{\mathbb{C}}[-\pi, \pi]$ if and only if

$$\lim_{n\to\infty} u_n = \Re(f) \text{ and } \lim_{n\to\infty} v_n = \Im(f) \text{ in } L^2[-\pi, \pi].$$

## 6.3. The Complex Hilbert Space $L^2_{\mathbb{C}}[-\pi, \pi]$

As we want to discuss complex Fourier series it is again convenient to use the interval $[a, b] = [-\pi, \pi]$. Also, we will again scale the Hermitian form in order to obtain a simple orthonormal family. This time we scale by a factor of $1/2\pi$.

**Proposition 6.3.1. (Hermitian form).** *The function*

$$\langle\ ,\ \rangle : L^2_{\mathbb{C}}[-\pi, \pi] \times L^2_{\mathbb{C}}[-\pi, \pi] \to \mathbb{C}$$

*defined by*

$$\langle f, g\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} f\bar{g}\ d\mu$$

*is a Hermitian form.*

**Proof.** The sesquilinearity is straightforward and will be left as an exercise.

We note

$$\langle f,\ f\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} |f|^2\ d\mu \geq 0,$$

with equality only if $|f(x)|^2 = 0$ almost everywhere. Hence, $\langle f,\ f\rangle = 0$ only if $f(x) = 0$ almost everywhere.

The fact that $\langle f,g\rangle = \overline{\langle g,f\rangle}$ follows from the fact that

$$\langle f,g\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} f\bar{g}\,d\mu = \frac{1}{2\pi}\int_{-\pi}^{\pi} \overline{\bar{f}g}\,d\mu = \frac{1}{2\pi}\overline{\int_{-\pi}^{\pi} g\bar{f}\,d\mu} = \overline{\langle g,f\rangle}.$$

□

In the complex Hilbert space $L^2_{\mathbb{C}}[-\pi,\pi]$ there is an even simpler orthonormal family than the trigonometric polynomials we used for $L^2[-\pi,\pi]$.

**Theorem 6.3.2.** *The family of functions $\mathcal{F}_{\mathbb{C}} = \{e^{inx}\}_{n=-\infty}^{\infty}$ is an orthonormal family in $L^2_{\mathbb{C}}[-\pi,\pi]$.*

**Proof.** According to part (3) of Exercise 1.6.3 for any integer $k \neq 0$,

$$\int_{-\pi}^{\pi} e^{ikx}\,dx = \frac{e^{ik\pi} - e^{-ik\pi}}{ik} = 0,$$

so it follows that if $n \neq m$,

$$\langle e^{inx}, e^{imx}\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i(n-m)x}\,dx = 0.$$

Also,

$$\langle e^{inx}, e^{inx}\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} 1\,dx = 1,$$

so the family $\mathcal{F}_{\mathbb{C}}$ is an orthonormal family. □

Following Definition 5.5.10 we define the Fourier series with respect to $\mathcal{F}_{\mathbb{C}}$.

**Definition 6.3.3. (Fourier series).** *If $f \in L^2_{\mathbb{C}}[-\pi,\pi]$, its Fourier coefficients with respect to $\mathcal{F}_{\mathbb{C}}$ are defined to be*

$$C_n = \langle f, e^{inx}\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-inx}\,dx$$

*for each $n \in \mathbb{Z}$. The* Fourier series *of $f$ is defined to be*

$$\sum_{n=-\infty}^{\infty} C_n e^{inx}.$$

**Theorem 6.3.4. (Fourier series converge in $L^2_{\mathbb{C}}[-\pi,\pi]$).** *The classical Fourier series of $f \in L^2_{\mathbb{C}}[-\pi,\pi]$ converges to $f$ in $L^2_{\mathbb{C}}[-\pi,\pi]$. Hence,*

$$\mathcal{F}_{\mathbb{C}} = \left\{ e^{inx} \right\}_{n=-\infty}^{\infty}$$

*is a complete orthonormal family in $L^2_{\mathbb{C}}[-\pi,\pi]$.*

**Proof.** To see that the complex Fourier series converges to $f$ we use the fact that the *real* classical Fourier series of a real function $u \in L^2[-\pi,\pi]$ converges to $u$. Let $f$ be an element of $L^2_{\mathbb{C}}[-\pi,\pi]$ and suppose $u(x)$ and $v(x)$ are its real and imaginary parts, so $f(x) = u(x) + iv(x)$.

By Theorem 6.1.7 there is a trigonometric polynomial

$$p_m(x) = a_0 + \sum_{n=1}^{m} a_n \cos nx + \sum_{n=1}^{m} b_n \sin nx$$

such that in the real Hilbert space $L^2[-\pi,\pi]$,

$$\lim_{m\to\infty} p_m = u.$$

We recall from the proof of Theorem 6.1.2 that

$$\cos nx = \frac{e^{inx} + e^{-inx}}{2} \quad \text{and} \quad \sin nx = \frac{e^{inx} - e^{-inx}}{2i}.$$

Hence, there are complex numbers $c_n$, $-m \le n \le m$, such that

$$p_m(x) = \sum_{n=-m}^{m} c_n e^{inx}.$$

Similarly, there are complex numbers $d_n$, $-m \le n \le m$, such that

$$q_m(x) = \sum_{n=-m}^{m} d_n e^{inx}$$

is real and

$$\lim_{m\to\infty} q_m = v$$

in $L^2[-\pi,\pi]$.

If we now define $A_n = c_n + id_n$ and let

$$T_m(x) = p_m(x) + iq_m(x) = \sum_{n=-m}^{m} A_n e^{inx},$$

then $T_m \in L^2_{\mathbb{C}}[-\pi, \pi]$ and

$$\|f - T_m\| = \|u - p_m + i(v - q_m)\| \leq \|u - p_m\| + \|v - q_m\|.$$

So $\lim \|f - T_m\| = 0$.

If $S_m = \sum_{n=-m}^{m} C_n e^{inx}$, then by Theorem 5.5.8 $\|f - S_m\| \leq \|f - T_m\|$ and hence $\lim \|f - S_m\| = 0$ and we conclude $\lim S_m = f$ in $L^2_{\mathbb{C}}[-\pi, \pi]$. $\square$

As in the real case it is important to note that while we have proved that the Fourier series of a function $f \in L^2_{\mathbb{C}}[-\pi, \pi]$ converges to $f$ in the Hilbert space $L^2_{\mathbb{C}}[-\pi, \pi]$, this is not at all the same thing as saying that for a fixed $x_0$ the Fourier series

$$\sum_{n=-\infty}^{\infty} C_n e^{inx_0}$$

converges to $f(x_0)$. That is, we have not shown anything about pointwise convergence. In fact, however, the series converges pointwise almost everywhere. In what is considered one of the most remarkable achievements of twentieth century mathematics Lennart Carleson proved the following result in [**C**].

**Theorem 6.3.5. (Carleson's theorem).** *Suppose $f \in L^2_{\mathbb{C}}[-\pi, \pi]$ and*

$$\sum_{n=-\infty}^{\infty} C_n e^{inx}$$

*is its classical Fourier series. Then this series converges to $f(x)$ for almost all values of $x \in [-\pi, \pi]$.*

The proof is considered difficult and delicate and is well beyond the scope of this text. One indication of the delicacy is that there is an example due to Andrey Kolmogorov of a function which is integrable on $[-\pi, \pi]$, but for which the Fourier series does not converge at any point!

## 6.4. The Hilbert Space $L^2_{\mathbb{C}}[\mathbb{T}]$

We described the set $\mathbb{T}$ as the unit length vectors in $\mathbb{R}^2$, but using the standard correspondence $(x, y) \leftrightarrow x + iy$ between $\mathbb{R}^2$ and the complex plane $\mathbb{C}$ we can also consider it to be the complex numbers with modulus 1, i.e., $\mathbb{T} = \{z \in \mathbb{C} \mid |z| = 1\} = \{x + iy \mid x^2 + y^2 = 1\}$. There is then a natural parametrization of $\mathbb{T}$ given by $\exp : [-\pi, \pi) \to \mathbb{T}$ where $\exp(x) = e^{ix} = \cos x + i \sin x$. The parametrizing function is a bijection from $[-\pi, \pi)$ to $\mathbb{T}$ and maps intervals to arcs preserving their length. We will say $A \subset \mathbb{T}$ is Lebesgue measurable if $\exp^{-1}(A)$ is measurable. We also define the Lebesgue measure $\mu(A)$ of $A$ to be $\mu(\exp^{-1}(A))$.

**Definition 6.4.1.** *We will denote by $L^2_{\mathbb{C}}[\mathbb{T}]$ the set of measurable functions $f : \mathbb{T} \to \mathbb{C}$ such that $\int_{\mathbb{T}} |f|^2 \, d\mu < \infty$.*

**Proposition 6.4.2.** *There is a correspondence between elements of $L^2_{\mathbb{C}}[\mathbb{T}]$ and $L^2_{\mathbb{C}}[-\pi, \pi]$ given by $f \mapsto f \circ \exp$ which satisfies:*

(1) *A function $f$ is in $L^2_{\mathbb{C}}[\mathbb{T}]$ if and only if $f \circ \exp \in L^2_{\mathbb{C}}[-\pi, \pi]$.*
(2) *The set $L^2_{\mathbb{C}}[\mathbb{T}]$ is a Hilbert space with Hermitian form given by $\langle f, g\rangle = \int_{\mathbb{T}} f\bar{g} \, d\mu$.*
(3) *The function $E^* : L^2_{\mathbb{C}}[\mathbb{T}] \to L^2_{\mathbb{C}}[-\pi, \pi]$ given by $E^*(f) = f \circ \exp$ is an isomorphism, i.e., it is an invertible linear transformation and preserves the Hermitian forms, so*

$$\langle E^*(f), E^*(g)\rangle = \langle f, g\rangle.$$

The proof is straightforward and is left as an exercise. The one slightly subtle point is that if $f \in L^2_{\mathbb{C}}[\mathbb{T}]$, then $E^*(f) : [-\pi, \pi] \to \mathbb{C}$ will always have the same value at $\pi$ and $-\pi$ so it might seem that $E^*$ is not surjective. However, this is not a problem when we recall that we consider two functions in $L^2_{\mathbb{C}}[-\pi, \pi]$ to be equal if they agree almost everywhere. More precisely, $L^2_{\mathbb{C}}[-\pi, \pi]$ is really the equivalence classes of almost everywhere equal functions and $E^*$ is a bijection on these classes.

It is common to use $e^{ix}$ (rather than $x$) as the independent variable for a function $f : \mathbb{T} \to \mathbb{C}$. So we might write $f(e^{ix}) = e^{2ix}$ rather than $f(z) = z^2$, which is the same function. This slight notational abuse is really just identifying $f$ with $E^*(f)$ since $f(e^{ix})$ considered as a function with independent variable $x$ is just $E^*(f)$.

With this notational convention our results about Fourier series carry over nicely to $L^2_{\mathbb{C}}[\mathbb{T}]$. The function $e^{inx}$ is easily considered a function of $e^{ix}$, namely $e^{inx} = (e^{ix})^n$. Hence, the family $\mathcal{F}_{\mathbb{C}} = \{e^{inx}\}_{n=-\infty}^{\infty}$ thought of as functions defined on $\mathbb{T}$ is an orthonormal family (as an immediate consequence of Proposition 6.4.2).

**Definition 6.4.3. (Fourier series).** *If $f \in L^2_{\mathbb{C}}[\mathbb{T}]$, its Fourier coefficients with respect to $\mathcal{F}_{\mathbb{C}}$ are defined to be*

$$C_n = \langle f, e^{inx} \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(e^{ix}) e^{-inx}\, dx$$

*for each $n \in \mathbb{Z}$. The* Fourier series *of $f$ is defined to be*

$$\sum_{n=-\infty}^{\infty} C_n e^{inx}.$$

**Theorem 6.4.4. (Fourier series in $L^2_{\mathbb{C}}[\mathbb{T}]$).** *The Fourier series of $f \in L^2_{\mathbb{C}}[\mathbb{T}]$ converges to $f$ in $L^2_{\mathbb{C}}[\mathbb{T}]$. So $\mathcal{F}_{\mathbb{C}} = \{e^{inx}\}_{n=-\infty}^{\infty}$ is a complete orthonormal family in $L^2_{\mathbb{C}}[\mathbb{T}]$.*

**Proof.** Considering $g(e^{ix}) = e^{inx}$ as an element of $L^2_{\mathbb{C}}[\mathbb{T}]$ we observe $E^*(g)(x) = e^{inx}$, as a function in $L^2_{\mathbb{C}}[-\pi, \pi]$. Hence, since $E^*$ preserves the Hermitian product it preserves Fourier coefficients. Therefore, if $f \in L^2_{\mathbb{C}}[\mathbb{T}]$, then $E^*$ of its Fourier series is the Fourier series of $E^*(f)$.

By Proposition 6.4.2 a series in $L^2_{\mathbb{C}}[\mathbb{T}]$ converges to $f$ if and only if $E^*$ of that series converges in $L^2_{\mathbb{C}}[-\pi, \pi]$ to $E^*(f)$. The result then follows from Theorem 6.3.4. □

One advantage of $L^2_{\mathbb{C}}[\mathbb{T}]$ is that it is sometimes useful in computation to consider parametrizations by $\exp(x) = e^{ix}$ on other intervals, e.g., $\exp : [0, 2\pi] \to \mathbb{T}$ and it is not difficult to show that integrals $\int_{\mathbb{T}} f\, d\mu$ can be calculated using a parametrization by exp on any interval of length $2\pi$ (see part (2) of Exercise 6.4.5).

---

**Exercise 6.4.5.**

(1) Prove that if $f \in L^2_{\mathbb{C}}[-\pi, \pi]$, then for all $n \in \mathbb{Z},\ n \neq 0$,

$$\left| \int_{-\pi}^{\pi} e^{nx} f(x)\, d\mu \right|^2 \leq \frac{\sinh(2\pi n)}{n} \int_{-\pi}^{\pi} |f|^2\, d\mu.$$

*Hint:* Recall $\sinh(x) = (e^x - e^{-x})/2$.

(2) Prove that if $f : \mathbb{T} \to \mathbb{C}$ is integrable, then

$$\int_{\mathbb{T}} f \, d\mu = \int_a^{a+2\pi} f(e^{ix}) \, dx$$

for any $a \in \mathbb{R}$.

(3) Prove Proposition 6.4.2.

(4) Prove a complex version of the Stone-Weierstrass theorem as a consequence of the real version, Theorem 6.1.5. More precisely, let $C_{\mathbb{C}}(\mathbb{T})$ denote the continuous complex-valued functions $f : \mathbb{T} \to \mathbb{C}$ defined on $\mathbb{T}$, the unit circle. Suppose $\mathcal{A} \subset C_{\mathbb{C}}(\mathbb{T})$ is an algebra satisfying:
 (a) The constant function 1 is in $\mathcal{A}$.
 (b) The algebra $\mathcal{A}$ separates points.
 (c) The function $p \in \mathcal{A}$ if and only if its complex conjugate $\bar{p} \in \mathcal{A}$.

Then given any $\varepsilon > 0$ and any $f \in C_{\mathbb{C}}(\mathbb{T})$ prove there is $p \in \mathcal{A}$ such that $|f(\theta) - p(\theta)| < \varepsilon$ for all $\theta \in \mathbb{T}$.

(5) Use the complex Stone-Weierstrass theorem above to give an alternate proof that complex Fourier series converge (Theorem 6.3.4). You should adapt the proof of Theorem 6.1.7 to cover complex functions.

Chapter 7

# Two Ergodic Transformations

In this final chapter we will apply some of the results on measure theory and Fourier series to the study of two measurable dynamical systems. In Section 4.3 we discussed measures other than Lebesgue and we will frame some of the results in this section in that generality though the examples we explore will focus on Lebesgue measure.

A measurable dynamical system is a function $T : X \to X$ from a set $X$ of "states" to itself. The set of states $X$ is assumed to have a measure $\nu$ defined on a $\sigma$-algebra of its subsets. The function $T$ is thought of as representing time evolution of the states. So starting at an initial state $x_0 \in X$ after one unit of time the system is in state $T(x_0)$, after two it is in state $T^2(x_0) = T(T(x_0))$, etc. It is natural to define $T^0(x) = x$. Note that here $T^n(x)$ means the result of $n$ applications of the function $T$, not the $n^{th}$ power of $T(x)$. The set $\{T^k(x) \mid k \geq 0\}$ is called the *forward orbit* of $x$.

There are two general dynamical questions we will address in special cases. The first question we ask concerns a measurable subset $A$ of $X$. We ask if states in $A$ return again and again (infinitely often) to $A$. Such states are called *recurrent* with respect to $A$. If $X$ has a notion of distance and we choose $A$ to have small diameter this means that a recurrent state returns infinitely often to a state close to its starting point.

The second question is what fraction of the forward orbit of a state lies in a given measurable set $A$. More precisely, if we consider $\{T^k(x) \mid 0 \leq k \leq m-1\}$, the first $m$ points in the forward orbit of $x$, and let $N_m(x)$ denote the number of those points which lie in $A$ then we would like to know if the limit

$$\lim_{m \to \infty} \frac{N_m(x)}{m}$$

exists. If so, how is the value of the limit related to $\nu(A)$?

As we did in the last chapter we will consider functions defined on the unit circle $\mathbb{T}$ in the complex plane. So $\mathbb{T} = \{e^{i\theta} \mid \theta \in \mathbb{R}\} = \{\cos\theta + i\sin\theta \mid \theta \in \mathbb{R}\}$. It is clear that $\theta$ and $\theta + 2\pi n$ correspond to the same point in $\mathbb{T}$ so to have a single value for each point we parameterize $\mathbb{T}$ by a half open interval of length $2\pi$. In the last chapter we used $\theta \in [-\pi, \pi)$. However, in this chapter it is more convenient to use $\theta \in [0, 2\pi)$.

## 7.1. Measure Preserving Transformations

We will primarily be interested in Lebesgue measure $\mu$ on $\mathbb{T}$ but many of the basic definitions and results are valid for any finite measure $\nu$. When this is the case we will make the definition or state the result for more general measures than Lebesgue. In particular, we will say that $X$ is a *finite measure space* with measure $\nu$ if $X$ is a set, $\mathcal{A}$ is a $\sigma$-algebra of subsets of $X$ and $\nu$ is a finite measure defined for all $A \in \mathcal{A}$. Sets in the $\sigma$-algebra $\mathcal{A}$ will be called $\nu$-measurable. If some property holds for all $x$ except a set of $\nu$ measure 0, we will say it holds for *$\nu$-almost all $x$.*

**Definition 7.1.1. (Measure preserving).** *Suppose that $\mathcal{A}$ is a $\sigma$-algebra of subsets of $X$ and $\nu$ is a finite measure defined on $\mathcal{A}$.*

- *A function $T : X \to X$ is called a* measure preserving transformation *provided for each $A \in \mathcal{A}$ the set $T^{-1}(A) \in \mathcal{A}$ and $\nu(T^{-1}(A)) = \nu(A)$.*
- *A function $f : X \to \mathbb{C}$ which satisfies $f(x) = f(T(x))$ for $\nu$-almost all $x$ is called $T$-invariant.*

- *A set $A \in \mathcal{A}$ is called $T$-invariant if $\mathfrak{X}_A(x)$ is a $T$-invariant function.*

It may at first seem strange that we require $\nu(T^{-1}(A)) = \nu(A)$ rather than $\nu(T(A)) = \nu(A)$. These hypotheses are equivalent if $T$ is a bijection, but otherwise may differ. The reason our choice is the natural one is that we are interested in integrating the composition $f \circ T$ when $f$ is integrable. In particular, we need this composition to be measurable. Notice that $(f \circ T)^{-1}([a, \infty]) = T^{-1}(f^{-1}([a, \infty]))$. So the hypothesis that $T^{-1}(A)$ is measurable for every measurable $A$ will imply that $f \circ T$ is measurable whenever $f$ is. Moreover, we will see in the proof of the following proposition that $\nu(T^{-1}(A)) = \nu(A)$ is exactly what is needed to show that composing with a measure preserving transformation $T$ does not change the value of the integral of a function.

**Proposition 7.1.2. (Invariance of the integral).** *Suppose $T$ is a measure preserving transformation on a finite measure space $X$ with measure $\nu$. If $f : X \to \mathbb{C}$ is integrable, then so is $f \circ T : X \to \mathbb{C}$ and*

$$\int f \, d\nu = \int f \circ T \, d\nu.$$

**Proof.** It suffices to prove the result for real-valued functions since then it will hold for both the real and imaginary parts of $f$, and hence for $f$. So assume $f$ is real. The result holds for $f = \mathfrak{X}_A$, for a $\nu$-measurable set $A$, because $\mathfrak{X}_A \circ T = \mathfrak{X}_{T-1(A)}$. It follows that the result holds for simple functions since both sides of the equality are linear functions of $f$. If $f$ is a bounded real measurable function, then it is a uniform limit of a sequence of simple functions $\{\phi_n\}$. It is clear that $\phi \circ T$ is the uniform limit of $\{\phi_n \circ T\}$. We know that the integral of a uniform limit of simple functions is the limit of the integrals (see Proposition 3.2.3), so

$$\int f \, d\nu = \lim_{n \to \infty} \int \phi_n \, d\nu = \lim_{n \to \infty} \int \phi_n \circ T \, d\nu = \int f \circ T \, d\nu.$$

Finally, if $f$ is integrable we let $f^+(x) = \max\{f(x), 0\}$ and define $f_n^+(x) = \min\{f^+(x), n\}$. Then the definition of the integral of a non-negative function (Definition 4.3.4) says that the integral of $f^+$ is the limit of the integrals of $f_n^+$ and the same holds for $f^+ \circ T$ and $f_n^+ \circ T$. Hence,

$$\int f^+ \, d\nu = \lim_{n \to \infty} \int f_n^+ \, d\nu = \lim_{n \to \infty} \int f_n^+ \circ T \, d\nu = \int f^+ \circ T \, d\nu.$$

Similarly, if $f^-(x) = -\min\{f(x), 0\}$, then $\int f^- \, d\nu = \int f^- \circ T \, d\nu$ and $f = f^+ - f^-$, so $\int f \, d\nu = \int f \circ T \, d\nu$. $\square$

**Proposition 7.1.3.** *Let $\alpha$ be an element of $\mathbb{R}$. The following transformations preserve Lebesgue measure on $\mathbb{T}$ :*

(1) *The function $T_\alpha : \mathbb{T} \to \mathbb{T}$ given by $T_\alpha(e^{ix}) = e^{i(x+\alpha)}$.*

(2) *The function $D : \mathbb{T} \to \mathbb{T}$ given by $T(e^{ix}) = e^{2ix}$.*

**Proof.** We will parameterize $\mathbb{T}$ by the interval $[0, 2\pi)$ with the correspondence $x \mapsto e^{ix}$. To show (1) we observe that $T_\alpha = T_{\alpha+2\pi n}$ for any integer $n$, so by altering $\alpha$ by an integer multiple of $2\pi$ we may assume $\alpha \in [0, 2\pi)$. In terms of our parametrization we have

$$T(x) = \begin{cases} x + \alpha, & \text{if } x + \alpha < 2\pi; \\ x + \alpha - 2\pi, & \text{otherwise.} \end{cases}$$

Thus, $T$ is translation by $\alpha$ or by $\alpha - 2\pi$. The result now follows because Lebesgue measure is translation invariant (see Theorem 2.4.2).

For (2) we also parameterize $\mathbb{T}$ by $x \in [0, 2\pi)$. Then, if $A \subset \mathbb{T}$, the set $D^{-1}(A)$ will consist of two disjoint pieces

$$B_1 = D^{-1}(A) \cap [0, \pi) \quad \text{and} \quad B_2 = D^{-1}(A) \cap [\pi, 2\pi).$$

The function $D$ on $[0, \pi)$ is given (in terms of our parametrization) by $D(x) = 2x$, so $D(B_1) = A$. Hence, $\mu(B_1) = \frac{1}{2}\mu(A)$ (see part (5) of Exercise 2.4.6). Likewise, on $[\pi, 2\pi)$ the function $D$ is given by $D(x) = 2x - \pi$ and $D(B_2) = A$. Hence $\mu(B_2) = \frac{1}{2}\mu(A)$. It follows that $\mu(D^{-1}(A)) = \mu(B_1) + \mu(B_2) = \mu(A)$. $\square$

**Definition 7.1.4. (Recurrent).** *Suppose $T : X \to X$ is a transformation of a finite measure space which preserves the measure $\nu$. A point $x \in X$ is said to be* recurrent *for $T$ with respect to a $\nu$-measurable set $A$ provided $x \in A$ and the set of return times,*

$$R(x) = \{n \mid T^n(x) \in A, \; n \in \mathbb{N}\},$$

*is infinite.*

**Theorem 7.1.5. (Poincaré recurrence).** *Suppose $T : X \to X$ is a transformation of a finite measure space, preserving the measure $\nu$ and suppose $A \subset X$ is $\nu$-measurable. Then $\nu$-almost all $x \in A$ are recurrent for $T$ with respect to $A$.*

**Proof.** Let $B_m$ be the subset of $A$ consisting of points whose return times are all less than $m$, i.e., $B_m = \{x \mid T^n(x) \notin A \text{ if } n \geq m\}$. We note that $T^{-km}(B_m) \cap B_m = \emptyset$ for $k \in \mathbb{N}$ because if $x$ is in this intersection, then $x \in B_m$ and $T^{km}(x) \in B_m$, contradicting the definition of $B_m$. It follows that the family of sets $\{T^{-km}(B_m)\}_{k=0}^{\infty}$ are pairwise disjoint, since if $j > k$ and $T^{-km}(B_m) \cap T^{-jm}(B_m) \neq \emptyset$, then applying $T^{km}$ would imply $B_m \cap T^{(k-j)m}(B_m) \neq \emptyset$.

We note that every set in the family $\{T^{-km}(B_m)\}$ has $\nu$ measure equal to $\nu(B_m)$ because $T$ preserves $\nu$, so $\nu(T^{-1}(E)) = \nu(E)$ for any $\nu$-measurable set $E$. This implies that $\nu(B_m) = 0$, since otherwise countable additivity implies the set $\bigcup_k T^{-km}(B_m)$ has infinite measure, but $\nu(X)$ is finite.

Finally, we note $B = \bigcup_{m=1}^{\infty} B_m$ has $\nu$ measure 0 and $B$ is the subset of $A$ whose points have only finitely many return times. □

---

**Exercise 7.1.6.**

(1) Prove that a $\nu$-measurable set $A$ is $T$-invariant if and only if $\nu(A \setminus T^{-1}(A)) = 0$ and $\nu(T^{-1}(A) \setminus A) = 0$.

(2) Prove that if $\alpha/2\pi$ is rational, then every point of $\mathbb{T}$ is periodic for $T_\alpha$, i.e., for each $x \in \mathbb{T}$ there is an $n > 0$ such that $T_\alpha^n(x) = x$.

(3) Prove that there are non-periodic points of $D$, but the $D$-periodic points are dense in $\mathbb{T}$.

(4) Define $T : [0,1] \to [0,1]$ by

$$T(x) = \begin{cases} 2x, & \text{if } x \in [0, \frac{1}{2}]; \\ 2 - 2x, & \text{if } x \in (\frac{1}{2}, 1]. \end{cases}$$

Prove that Lebesgue measure is $T$-invariant.

(5) Suppose $T : [0,1] \to [0,1]$ preserves Lebesgue measure. Prove that for almost all $x \in [0,1]$ there is a sequence of positive integers $\{n_i\}$ such that
$$\lim_{i\to\infty} T^{n_i}(x) = x.$$

## 7.2. Ergodicity

We now turn to the second question mentioned in the introduction to this chapter, namely, what fraction of the forward orbit of a state $x_0$ lies in a set $A$. What we would like to show is sometimes phrased as "the time average equals the space average." What this means is that the fraction of the points in a forward orbit that lie in $A$ (i.e., the fraction of the "time" spent in $A$) is equal to the fraction of the measure of $X$ represented by $A$ (i.e., $\nu(A)/\nu(X)$). This certainly does not hold for all measure preserving transformations. Think about the identity transformation, for example. Indeed, this property will fail any time there is a set $A$ with $0 < \nu(A) < \nu(X)$ and $A = T^{-1}(A)$ because then, if $x_0 \in A$, the entire forward orbit of $x_0$ lies in $A$, but $A$ does not contain 100% of the measure of $X$. Clearly, the condition that $A = T^{-1}(A)$ implies that $A$ is $T$-invariant. In fact, $A$ is $T$-invariant if and only if $A$ and $T^{-1}(A)$ differ only by a set of $\nu$ measure 0 (see part (1) of Exercise 7.1.6). We conclude that a necessary condition for the time average to equal the space average is that no $T$-invariant set $A$ has measure strictly between 0 and $\nu(X)$. Remarkably we will see below that this condition is sufficient as well as necessary. It should be clear that this is a property worthy of a name (albeit a strange one).

**Definition 7.2.1. (Ergodic transformation).** *Suppose $T : X \to X$ is a measure preserving transformation for the finite measure $\nu$ defined on a $\sigma$-algebra $\mathcal{A}$ of subsets of $X$. Then $T$ is called* ergodic *if every $T$-invariant set $A \in \mathcal{A}$ satisfies either $\nu(A) = 0$ or $\nu(A^c) = 0$.*

For an ergodic transformation $T$ the only $T$-invariant sets are sets of measure zero or the complements of sets of measure zero. So we expect $T$-invariant functions also to be very restricted. This is indeed the case.

**Proposition 7.2.2. (Invariant functions).** *Suppose $T : X \to X$ preserves the finite measure $\nu$. Then $T$ is ergodic if and only if every $\nu$-measurable function $f : X \to \mathbb{C}$ which is $T$-invariant is constant except on a set of $\nu$ measure 0.*

**Proof.** Suppose the only $T$-invariant functions are $\nu$-almost everywhere constant. A set $A$ is $T$-invariant only if the function $\mathfrak{X}_A$ is $T$-invariant. Since $\mathfrak{X}_A(x)$ has only two possible values, 0 and 1, the function $\mathfrak{X}_A$ must either be equal to 0, except on a set of $\nu$ measure 0, or $\mathfrak{X}_A$ must be 1, except on a set of $\nu$ measure 0. Therefore, $\nu(A) = 0$ or $\nu(A^c) = 0$.

Conversely, if $f$ is a $T$-invariant measurable function that is not $\nu$-almost everywhere constant, then there is a $c \in \mathbb{R}$ such that if $A = f^{-1}([0, c))$, then $\nu(A) > 0$ and $\nu(A^c) > 0$. The set $A$ is $T$-invariant. ☐

We are now prepared to prove that two interesting transformations of $\mathbb{T}$ that preserve Lebesgue measure are actually ergodic. Our primary tool in doing this is the Fourier series for functions in $L^2_{\mathbb{C}}[\mathbb{T}]$ which we investigated in Section 6.4.

**Proposition 7.2.3. (Ergodicity of $T_\alpha$).** *If $\alpha/2\pi$ is irrational, then the transformation $T_\alpha : \mathbb{T} \to \mathbb{T}$ given by $T_\alpha(e^{ix}) = e^{i(x+\alpha)}$ is ergodic.*

**Proof.** For $f \in L^2_{\mathbb{C}}[\mathbb{T}]$ and $T$ a measure preserving transformation define $U_T(f) = f \circ T$. Proposition 7.1.2 implies

$$\int_{\mathbb{T}} |f|^2 \, d\mu = \int_{\mathbb{T}} |U_T(f)|^2 \, d\mu,$$

so $U_T(f) \in L^2_{\mathbb{C}}[\mathbb{T}]$ and $\|U_T(f)\| = \|f\|$. Similarly, if $g \in L^2_{\mathbb{C}}[\mathbb{T}]$, the same proposition implies that

$$\langle f, g \rangle = \langle U_T(f), U_T(g) \rangle.$$

The function $U_T : L^2_{\mathbb{C}}[\mathbb{T}] \to L^2_{\mathbb{C}}[\mathbb{T}]$ is also linear.

It follows that if $\sum_{n\in\mathbb{Z}} v_n$ is any series which converges in $L^2_{\mathbb{C}}[\mathbb{T}]$ to $f$, then $\sum_{n\in\mathbb{Z}} U_T(v_n)$ is a series which converges to $U_T(f) = f \circ T$ in $L^2_{\mathbb{C}}[\mathbb{T}]$.

Suppose $f \in L^2_{\mathbb{C}}[\mathbb{T}]$ and let $\sum_{n\in\mathbb{Z}} C_n e^{inx}$ be its Fourier series. If $f$ is $T_\alpha$-invariant, then $f = U_{T_\alpha}(f)$ almost everywhere and by the remarks above, the series $\sum_{n\in\mathbb{Z}} C_n U_{T_\alpha}(e^{inx})$ converges to $f$ in $L^2_{\mathbb{C}}[\mathbb{T}]$. Since $U_{T_\alpha}(e^{inx}) = e^{in(x+\alpha)} = e^{in\alpha}e^{inx}$, we conclude that the series $\sum_{n\in\mathbb{Z}} e^{in\alpha} C_n e^{inx}$ converges in $L^2_{\mathbb{C}}[-\pi,\pi]$ to $f$. The uniqueness of the Fourier series then implies $C_n = e^{in\alpha}C_n$. So for each $n \in \mathbb{Z}$ either $C_n = 0$ or $e^{in\alpha} = 1$; but $e^{in\alpha} = 1$ only if $n\alpha = 2\pi m$ for some integer $m$. Since $\alpha/2\pi$ is irrational this happens only if $n = 0$. We conclude that $C_n = 0$ whenever $n \neq 0$. Thus, the Fourier series for $f$ reduces to the constant function $f(e^{ix}) = C_0$. Hence, the only $T_\alpha$ invariant functions are constant almost everywhere and $T_\alpha$ is ergodic. □

**Proposition 7.2.4. (Ergodicity of $D$).** *The transformation $D : \mathbb{T} \to \mathbb{T}$ given by $D(e^{ix}) = e^{2ix}$ is ergodic.*

**Proof.** As in the previous proposition we observe that if $U_D(f)$ is defined by $U_D(f) = f \circ D$, then $U_D(f) \in L^2_{\mathbb{C}}[\mathbb{T}]$ for all $f \in L^2_{\mathbb{C}}[\mathbb{T}]$ and

$$\langle f, g\rangle = \langle U_D(f), U_D(g)\rangle$$

whenever $f, g \in L^2_{\mathbb{C}}[\mathbb{T}]$ by Proposition 7.1.2.

Suppose $f \in L^2_{\mathbb{C}}[\mathbb{T}]$ is $D$-invariant and let $\sum_{n\in\mathbb{Z}} C_n e^{inx}$ be its Fourier series. Then the series $\sum_{n\in\mathbb{Z}} C_n U_D(e^{inx})$ converges to $U_D(f)$ which equals $f$ almost everywhere. Since $U_D(e^{inx}) = e^{2inx}$ we conclude $\sum_{n\in\mathbb{Z}} C_n e^{2inx}$ converges in $L^2_{\mathbb{C}}[\mathbb{T}]$ to $f$. By the uniqueness of Fourier series the coefficient of $e^{inx}$ must be the same in the two series $\sum_{n\in\mathbb{Z}} C_n e^{inx}$ and $\sum_{n\in\mathbb{Z}} C_n e^{2inx}$. If $n$ is odd this coefficient is 0 in the second series and $C_n$ in the first. We conclude that $C_n = 0$ if $n$ is odd. Likewise, the coefficient of $e^{2inx}$ is $C_{2n}$ in one series and $C_n$ in the other. Hence $C_{2n} = C_n$ for all $n \in Z$. It follows that $C_{2n} = 0$ if $n$ is odd. A simple induction shows that $C_n = 0$ for all $n \neq 0$.

Thus, again the Fourier series for $f$ reduces to the constant function $f(e^{ix}) = C_0$. □

---

**Exercise 7.2.5.**

(1) Suppose $\alpha/2\pi$ is irrational. Prove that every point of $\mathbb{T}$ has a forward orbit for the transformation $T_\alpha$ which is dense in $\mathbb{T}$.

(2) Prove that there exists an $x \in \mathbb{T}$ whose forward orbit for the transformation $D$ is dense in $\mathbb{T}$. But give an explicit example of a point in $\mathbb{T}$ whose forward orbit for $D$ is not dense.

(3) Prove that if $m \in \mathbb{N}$ and $m > 1$, then the function $M : \mathbb{T} \to \mathbb{T}$ given by $M(e^{ix}) = e^{imx}$ preserves Lebesgue measure and is ergodic. That is, there is nothing special about the role of the natural number 2 in the ergodicity of $D$.

## 7.3. The Birkhoff Ergodic Theorem

We are now prepared to state (though not prove) what is by far the most important theorem of measurable dynamics. A proof can be found in [**J**], [**M**] or [**Z**].

**Theorem 7.3.1. (Birkhoff ergodic theorem).** *Suppose $T : X \to X$ is a transformation of a finite measure space preserving the measure $\nu$ and suppose $f : X \to \mathbb{C}$ is integrable. Then there is a $T$-invariant integrable function $f^* : X \to \mathbb{C}$ with the property that for $\nu$-almost all $x \in X$,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x)) = f^*(x)$$

*and $\int f^*\, d\nu = \int f\, d\nu$.*

A particularly important special case is when $T$ is ergodic so the only $T$-invariant functions are constants. In that case we conclude that for almost all $x$ the limit of the sequence of orbit averages

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x))$$

is equal to the constant function whose value is

$$\frac{1}{\nu(X)} \int f\, d\nu,$$

the average value of $f$.

We can now explain why for ergodic transformations we say, "*The time average equals the space average.*" If $f$ is an integrable function defined on $X$, its "space average" is

$$\frac{1}{\nu(X)} \int f \, d\nu.$$

On the other hand, the "time average" of $f$ over $n$ units of time starting at an initial state $x_0$ is

$$\frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x_0)).$$

So the Birkhoff ergodic theorem says the limit of the time averages of $f$ as time goes to infinity is equal to the space average of $f$ — at least for all initial states except a set of $\nu$ measure 0.

If in the Birkhoff ergodic theorem we let $f$ be $\mathfrak{X}_A$ for some measurable set $A$, we can answer the question raised at the beginning of this chapter, at least in the special case that $T$ is an ergodic transformation.

**Corollary 7.3.2.** *Suppose $T : X \to X$ is an ergodic transformation preserving a finite measure $\nu$ on a measure space $X$ and $A \subset X$ is $\nu$-measurable. Let $N_m(x)$ denote the number of points in the set $A \cap \{f^k(x)\}_{k=0}^{m-1}$, i.e., the number of points in $A$ from the orbit segment of length $m$ starting with $x$. Then for $\nu$-almost all $x \in X$,*

$$\lim_{m \to \infty} \frac{N_m(x)}{m} = \frac{\nu(A)}{\nu(X)}.$$

**Proof.** Let $f = \mathfrak{X}_A$ and observe that

$$\lim_{m \to \infty} \frac{N_m(x)}{m} = \lim_{m \to \infty} \frac{1}{m} \sum_{k=0}^{m-1} \mathfrak{X}_A(T^k(x)) = \frac{1}{\nu(X)} \int \mathfrak{X}_A \, d\nu = \frac{\nu(A)}{\nu(X)}$$

where the second equality follows from the Birkhoff ergodic theorem. □

The Poincaré recurrence theorem (Theorem 7.1.5) asserts that for $\nu$-almost all points $x \in A$ the forward orbit of $x$ for a measure preserving transformation $T$ returns to $A$ infinitely often. For an ergodic transformation $T$ we can do much better and consider how often the forward orbit of a point $x$ not necessarily in $A$ visits the set $A$. In fact, if $\nu(A) > 0$, then for $\nu$-almost all $x \in X$ the forward orbit of $x$ visits $A$ infinitely often.

**Corollary 7.3.3.** *Suppose $T : X \to X$ is an ergodic transformation preserving a finite measure $\nu$ on a measure space $X$ and $\nu(A) > 0$ for a subset $A \subset X$. Then for $\nu$-almost all $x \in X$ the set $\{n \mid T^n(x) \in A\}$ of times when the forward orbit of $x$ is in $A$ is an infinite set.*

**Proof.** By the previous corollary, for $\nu$-almost all $x$,

$$\lim_{m\to\infty} \frac{N_m(x)}{m} = \frac{\nu(A)}{\nu(X)} > 0.$$

Since the sequence $\{N_m(x)\}$ is monotonic increasing it cannot be bounded as that would imply the limit is 0. $\square$

Even this relatively weak corollary of the Birkhoff ergodic theorem has some remarkably surprising consequences. Part (2) of the exercise below illustrates one of them.

---

**Exercise 7.3.4.**

(1) Suppose $T : [0,1] \to [0,1]$ is defined by

$$T(x) = \begin{cases} 2x, & \text{if } x \in [0, \frac{1}{2}]; \\ 2x - 1, & \text{if } x \in (\frac{1}{2}, 1]. \end{cases}$$

Prove that $T$ preserves Lebesgue measure and is ergodic. *Hint:* How is $T$ related to $D$?

(2) Consider your favorite movie $\mathbb{M}$ in the digital format of your choice. So $\mathbb{M}$ is a very long (but finite) string of 0's and 1's. Prove that for almost all $x \in [0,1]$ your movie's encoding $\mathbb{M}$ occurs infinitely often in the digits of the binary expansion of $x$. Indeed, the binary expansion of almost all $x \in [0,1]$ will contain the digital encoding of every movie which has ever been made or ever will be made. Each will occur infinitely often. *Hint:* Find an expression for $T(x)$ in the previous exercise in terms of the binary expansion of $x$.

Appendix A

# Background and Foundations

This appendix gives a very terse summary of the properties of the real numbers which are used throughout the text. It is intended as a review and reference for standard facts about the real numbers rather than an introduction to these concepts.

**Notation. (Numbers).** *We will denote the set of* real numbers *by* $\mathbb{R}$*, the* complex numbers *by* $\mathbb{C}$*, the* rational numbers *by* $\mathbb{Q}$*, the* integers *by* $\mathbb{Z}$ *and the* natural numbers *by* $\mathbb{N}$.

## A.1. The Completeness of $\mathbb{R}$

In addition to the standard properties of being an ordered field (i.e., the properties of arithmetic) the real numbers $\mathbb{R}$ satisfy a property which makes analysis as opposed to algebra possible.

**The Completeness Axiom.** *Suppose $A$ and $B$ are non-empty subsets of $\mathbb{R}$ such that $x \leq y$ for every $x \in A$ and every $y \in B$. Then there exists at least one real number $z$ such that $x \leq z$ for all $x \in A$ and $z \leq y$ for all $y \in B$.*

**Example A.1.1.** The rational numbers, $\mathbb{Q}$, fail to satisfy this property. If $A = \{x \mid x^2 < 2\}$ and $B = \{y \mid y > 0 \text{ and } y^2 > 2\}$, then there is no $z \in \mathbb{Q}$ such that $x \leq z$ for all $x \in A$ and $z \leq y$ for all $y \in B$.

**Definition A.1.2. (Infimum, supremum).** *If $A \subset \mathbb{R}$, then $b \in \mathbb{R}$ is called an* upper bound *for $A$ if $b \geq x$ for all $x \in A$. The number $\beta$ is called the* least upper bound *or* supremum *of the set $A$ if $\beta$ is an upper bound and $\beta \leq b$ for every upper bound $b$ of $A$. A number $a \in \mathbb{R}$ is called a* lower bound *for $A$ if $a \leq x$ for all $x \in A$. The number $\alpha$ is called the* greatest lower bound *or* infimum *of the set $A$ if $\alpha$ is a lower bound and $\alpha \geq a$ for every lower bound $a$ of $A$.*

**Theorem A.1.3.** *If a non-empty set $A \subset \mathbb{R}$ has an upper bound, then it has a unique supremum $\beta$. If $A$ has a lower bound, then it has a unique infimum $\alpha$.*

**Proof.** Let $B$ denote the non-empty set of upper bounds for $A$. Then $x \leq y$ for every $x \in A$ and every $y \in B$. The Completeness Axiom tells us there is a $\beta$ such that $x \leq \beta \leq y$ for every $x \in A$ and every $y \in B$. This implies that $\beta$ is an upper bound of $A$ and that $\beta \leq y$ for every upper bound $y$. Hence, $\beta$ is a supremum or least upper bound of $A$. It is unique, because any $\beta'$ with the same properties must satisfy $\beta \leq \beta'$ (since $\beta$ is a least upper bound) and $\beta' \leq \beta$ (since $\beta'$ is a least upper bound). This, of course implies $\beta = \beta'$.

The proof for the *infimum* is similar. □

We will denote the *supremum* of a set $A$ by $\sup A$ and the *infimum* by $\inf A$. If $A$ is not bounded above, we will write $\sup A = +\infty$ and if it is not bounded below we write $\inf A = -\infty$. The supremum of a bounded set $A$ may or may not be in the set $A$. However, as the following result shows, there is always an element of $A$ which is arbitrarily close to the supremum of $A$.

**Proposition A.1.4.** *If $A$ has an upper bound and $\beta = \sup A$, then for any $\varepsilon > 0$ there is an $x \in A$ with $\beta - \varepsilon < x \leq \beta$. Moreover, $\beta$ is the only upper bound for $A$ with this property. If $A$ has a lower bound its infimum satisfies the analogous property.*

**Proof.** If $\beta = \sup A$ and there is no $x \in (\beta - \varepsilon, \beta]$, then every $x \in A$ satisfies $x \leq \beta - \varepsilon$. It follows that $\beta - \varepsilon$ is an upper bound for $A$ and is smaller than $\beta$ contradicting the definition of $\beta$ as the least upper bound. Hence there must be an $x \in A$ with $x \in (\beta - \varepsilon, \beta]$.

If $\beta' \neq \beta$ is another upper bound for $A$, then $\beta' > \beta$. There is no $x \in A$ with $x \in (\beta, \beta']$, since such an $x$ would be greater than $\beta$ and hence $\beta$ would not be an upper bound for $A$.

The proof for the *infimum* is similar. □

## A.2. Functions and Sequences

At the risk of being somewhat pedantic we will give the definition of a function as a set of ordered pairs and the definition of a sequence as a function defined on the natural numbers. The reader should note, however, that after these definitions, when referring to a function $\phi$ we will always opt for the notation $b = \phi(a)$ rather than $(a, b) \in \phi$ and treat sequences as an indexed set of values rather than a function.

If $A$ and $B$ are sets their *Cartesian product*, denoted $A \times B$, is defined to be the set of ordered pairs $A \times B = \{(a, b) \mid a \in A,\ b \in B\}$.

**Definition A.2.1. (Function).** *Suppose $A$ and $B$ are sets. A function $\phi : A \to B$ is a subset of the Cartesian product $A \times B$ such that:*

(1) *If $a \in A$, there exists $b \in B$ such that $(a, b) \in \phi$.*
(2) *If $(a, b) \in \phi$ and $(a, b') \in \phi$, then $b = b'$.*

It is worth noting that this definition makes sense even if $A$ or $B$ is the empty set. In that case the set $\phi$ is necessarily empty since $A \times B = \emptyset$. The set $A$ is called the *domain* of $\phi$, the set $B$ is called the *codomain* of $\phi$, and the set of all $b \in B$ such that $(a, b) \in \phi$ is called the *range* or *image* of $\phi$. As noted above we will almost always use the notation $\phi : A \to B$ to indicate that $\phi$ is a function with domain $A$ and codomain $B$, and write $b = \phi(a)$ to indicate $(a, b) \in \phi$. We will denote the range of $\phi$ by $\phi(A)$.

**Definition A.2.2. (Injective, surjective, bijective).** *Suppose $A$ and $B$ are sets and $\phi : A \to B$ is a function. Then*

(1) *The function $\phi$ is called* injective *(or* one-to-one*) if $\phi(x) = \phi(y)$ implies $x = y$.*

(2) *The function* $\phi$ *is called* surjective *(or* onto*) if for every* $b \in B$ *there exists* $x \in A$ *such that* $\phi(x) = b$. *Equivalently,* $\phi$ *is surjective if its codomain equals its range.*

(3) *The function* $\phi$ *is called* bijective *if it is both injective and surjective.*

(4) *If* $C \subset B$, *the* set inverse $\phi^{-1}(C)$ *is defined to be* $\{a \mid a \in A \text{ and } \phi(a) \in C\}$. *If* $C$ *consists of a single element* $c$, *we write* $\phi^{-1}(c)$ *instead of the more cumbersome* $\phi^{-1}(\{c\})$.

Injective, surjective and bijective as defined above are adjectives which may be applied to a function $\phi$. The corresponding nouns referring to a function with the given property are *injection*, *surjection*, and *bijection*.

**Definition A.2.3. (Finite, infinite).** *A set* $A$ *is* finite *if it is empty or there is an* $n \in \mathbb{N}$ *and a bijection from* $A$ *to* $\{1, 2, \ldots, n\}$. *A set is* infinite *if it is not finite.*

One important special class of functions, called sequences, are treated in a notationally different way.

**Definition A.2.4. (Sequence).** *A sequence of elements in a set* $A$ *is a function* $\phi : \mathbb{N} \to A$.

However, we typically do not name the function defining a sequence, but write only the indexed set of values $\{a_n\}_{n=1}^{\infty}$ where $a_n = \phi(n)$.

---

**Exercise A.2.5.**

(1) Prove that if $f : A \to B$ is injective, then $f : A \to f(A)$ is a bijection.

(2) Suppose $A$ and $B$ are sets. Prove that if $A$ is infinite and there is an injective function $f : A \to B$, then $B$ is infinite. Prove that if $A$ is infinite and there is a surjective function $f : B \to A$, then $B$ is infinite.

(3) (*Inverse function*)
If $f : A \to B$, then $g : B \to A$ is called the *inverse function*

of $f$ provided $g(f(a)) = a$ for all $a \in A$ and $f(g(b)) = b$ for all $b \in B$.

(a) Prove that if the inverse function exists it is unique (and hence it is appropriate to refer to it as *the* inverse).

(b) Prove that $f$ has an inverse if and only if $f$ is a bijection.

(c) If it exists, we denote the inverse function of $f$ by $f^{-1}$. This is a slight abuse of notation since we denote the set inverse (see part (4) of Definition A.2.2) the same way. To justify this abuse somewhat prove that if $f$ has an inverse $g$, then for each $b \in B$ the set inverse $f^{-1}(\{b\})$ is the set consisting of the single element $g(b)$. Conversely, show that if for every $b \in B$ the set inverse $f^{-1}(\{b\})$ contains a single element, then $f$ has an inverse $g$ defined by letting $g(b)$ be that single element.

## A.3. Limits

There are a number of equivalent formulations we could have chosen for the Completeness Axiom. For example, we could have taken the existence of the supremum for bounded sets (Theorem A.1.3) as an axiom and then proved the Completeness Axiom as a theorem following from this axiom. In this section we prove several more theorems which we will derive from the Completeness Axiom, but which are in fact equivalent to it in the sense that, if we assumed any one as an axiom, we could prove the others as consequences. Results of this type include Theorem A.3.2, Corollary A.3.3, and Theorem A.3.5.

We recall the definition of limit of a sequence in $\mathbb{R}$.

**Definition A.3.1. (Limit, converge).** *Suppose $\{x_n\}_{n=1}^{\infty}$ is a sequence in $\mathbb{R}$. We say the sequence* converges *to $L \in \mathbb{R}$ and write*

$$\lim_{n\to\infty} x_n = L$$

*if for every $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that*

$$|x_m - L| < \varepsilon$$

*for all $m \geq N$.*

Let $\{x_n\}_{n=1}^{\infty}$ be a sequence in $\mathbb{R}$. We will say it is *monotone increasing* if $x_{n+1} \geq x_n$ for all $n$ and *monotone decreasing* if $x_{n+1} \leq x_n$ for all $n$.

**Theorem A.3.2. (Bounded monotone sequences converge).** *If $\{x_n\}_{n=1}^{\infty}$ is a bounded* monotone *sequence, then $\lim\limits_{n\to\infty} x_n$ exists.*

**Proof.** If $\{x_n\}_{n=1}^{\infty}$ is a bounded monotone increasing sequence, let $L = \sup\{x_n\}_{n=1}^{\infty}$. Given any $\varepsilon > 0$ there is an $N$ such that

$$L - \varepsilon < x_N \leq L$$

by Proposition A.1.4.

For any $n > N$ we have $x_N \leq x_n \leq L$ and hence $|L - x_n| < \varepsilon$. Thus, $\lim\limits_{n\to\infty} x_n = L$.

If $\{x_n\}_{n=1}^{\infty}$ is a monotone decreasing sequence, then we note that the sequence $\{-x_n\}_{n=1}^{\infty}$ is increasing and $\lim\limits_{n\to\infty} x_n = -\lim\limits_{n\to\infty} -x_n$. ☐

**Corollary A.3.3.** *If $\{x_n\}_{n=1}^{\infty}$ is a bounded sequence, then*

$$\lim_{m\to\infty} \sup\{x_n\}_{n=m}^{\infty} \quad \textit{and} \quad \lim_{m\to\infty} \inf\{x_n\}_{n=m}^{\infty}$$

*both exist. The sequence $\{x_n\}_{n=1}^{\infty}$ has limit $L$, i.e., $\lim x_n = L$, if and only if both limits equal $L$.*

**Notation. (lim sup and lim inf).** *We will denote*

$$\lim_{m\to\infty} \sup\{x_n\}_{n=m}^{\infty} \quad \textit{by} \quad \limsup_{n\to\infty} x_n$$

*and*

$$\lim_{m\to\infty} \inf\{x_n\}_{n=m}^{\infty} \quad \textit{by} \quad \liminf_{n\to\infty} x_n$$

**Proof.** If $y_m = \sup\{x_n\}_{n=m}^{\infty}$, then $\{y_m\}_{m=1}^{\infty}$ is a monotone decreasing sequence, so $\lim\limits_{m\to\infty} y_m$ exists. The proof that $\liminf x_n$ exists is similar.

The fact that $\inf\{x_n\}_{n=m}^{\infty} \leq x_m \leq \sup\{x_n\}_{n=m}^{\infty}$ implies that if

$$\liminf_{n\to\infty} x_n = \limsup_{n\to\infty} x_n = L,$$

then $\lim\limits_{n\to\infty} x_n$ exists and equals $L$. ☐

**Definition A.3.4. (Cauchy sequence).** *A sequence* $\{x_n\}_{n=1}^{\infty}$ *in* $\mathbb{R}$ *is called a* Cauchy sequence *if for every* $\varepsilon > 0$ *there is an* $N > 0$ *(depending on* $\varepsilon$*) such that* $|x_n - x_m| < \varepsilon$ *for all* $n, m \geq N$.

The reason Cauchy sequences are important is that they converge, which is the content of our next theorem. In fact, the converse is also true, so convergent sequences are Cauchy sequences (see part (1) of Exercise A.3.8). Proving that a sequence is Cauchy is an extremely useful technique for proving convergence.

**Theorem A.3.5. (Cauchy sequences converge).** *If* $\{x_n\}_{n=1}^{\infty}$ *is a Cauchy sequence in* $\mathbb{R}$, *then* $\lim\limits_{n\to\infty} x_n$ *exists.*

**Proof.** First we show that if $\{x_n\}_{n=1}^{\infty}$ is a Cauchy sequence, then it is bounded. For $\varepsilon = 1$ there is an $N_1$ such that $|x_n - x_m| < 1$ for all $n, m \geq N_1$. Hence, for any $n \geq N_1$ we have $|x_n| \leq |x_n - x_{N_1}| + |x_{N_1}| \leq |x_{N_1}| + 1$. It follows that if $M = 1 + max\{x_n\}_{n=1}^{N_1}$, then $|x_n| \leq M$ for all $n$. Hence $\limsup\limits_{n\to\infty} x_n$ exists.

Since the sequence is Cauchy, given $\varepsilon > 0$ there is an $N$ such that that $|x_n - x_m| < \varepsilon/2$ for all $n, m \geq N$. Let

$$L = \limsup_{n\to\infty} x_n = \lim_{n\to\infty} \sup\{x_m\}_{m=n}^{\infty}.$$

Then by Proposition A.1.4 there is an $M \geq N$ such that $|x_M - L| < \varepsilon/2$. It follows that for any $n > M$ we have $|x_n - L| \leq |x_n - x_M| + |x_M - L| < \varepsilon/2 + \varepsilon/2 = \varepsilon$. So $\lim\limits_{n\to\infty} x_n = L$. □

**Definition A.3.6. (Convergent, absolutely convergent).** *Suppose* $\sum_{n=1}^{\infty} x_n$ *is an infinite series of real numbers and* $S_m = \sum_{n=1}^{m} x_n$ *is its* $m^{th}$ *partial sum. The series* $\sum_{n=1}^{\infty} x_n$ *is said to* converge *with limit* $L$ *provided* $\lim\limits_{m\to\infty} S_m = L$. *It is said to* converge absolutely *provided the series* $\sum_{n=1}^{\infty} |x_n|$ *converges.*

**Theorem A.3.7. (Absolutely convergent series).** *If the series* $\sum_{n=1}^{\infty} x_n$ *converges absolutely, then it converges.*

**Proof.** Let $S_m = \sum_{i=1}^{m} x_i$ be the partial sum. We must show that $\lim\limits_{m\to\infty} S_m$ exists. We will do this by showing it is a Cauchy sequence.

Since the series $\sum_{i=1}^{\infty} |x_i|$ converges, given $\varepsilon > 0$, there is an $N > 0$ such that $\sum_{i=N}^{\infty} |x_i| < \varepsilon$. Hence, if $m > n \geq N$,

$$|S_m - S_n| = \left| \sum_{i=n+1}^{m} x_i \right| \leq \sum_{i=n+1}^{m} |x_i| \leq \sum_{i=N}^{\infty} |x_i| < \varepsilon.$$

Hence, $\{S_n\}$ is a Cauchy sequence and converges. □

---

**Exercise A.3.8.**

(1) Prove that a sequence in $\mathbb{R}$ is Cauchy if and only if it converges.

(2) (a) Prove that if $|r| < 1$, the sequence $\sum_{n=0}^{\infty} ar^n$ converges to $a/(1-r)$.
(b) Prove that if

$$\limsup_{n\to\infty} \left| \frac{x_{n+1}}{x_n} \right| < 1,$$

then $\sum_{n=0}^{\infty} x_n$ converges.

(3) A sequence $\{x_n\}$ in $\mathbb{R}$ is called *square summable* if $\sum_{n=1}^{\infty} x_n^2$ converges. The set of all square summable sequences is denoted $\ell^2$ (pronounced "little ell two"). Suppose $\{x_n\}$ and $\{y_n\}$ are in $\ell^2$.
   (a) Prove that $\sum_{n=1}^{\infty} x_n y_n$ converges.
   (b) Prove that if $z_n = x_n + y_n$, then $\{z_n\} \in \ell^2$.

## A.4. Complex Limits

We will denote the complex numbers by $\mathbb{C}$, so

$$\mathbb{C} = \{a + bi \mid a, b \in \mathbb{R}\}$$

where $i^2 = -1$.

Recall the following elementary properties of complex numbers:

(1) The *complex conjugate* of $z = a + bi$ is denoted by $\bar{z}$ and is defined to be $a - bi$. Note that if $z \in \mathbb{C}$, then $z \in \mathbb{R}$ if and only if $z = \bar{z}$.

(2) The *modulus* of $z = a + bi$ is denoted $|z|$ and is defined to be $\sqrt{a^2 + b^2}$ so $|z|^2 = z\bar{z} = a^2 + b^2$. Note that $|z| = |\bar{z}|$ and $|zw| = |z||w|$.

(3) We define the *real part* of $z = a + bi$ to be $a$ and denote it $\Re(z)$. Likewise, the *imaginary part* of $z$ is $b$ and is denoted $\Im(z)$, so $z = \Re(z) + i\Im(z)$ for all $z \in \mathbb{C}$.

(4) *Euler's formula*: For every real number $x$,

$$e^{ix} = \cos x + i \sin x.$$

**Definition A.4.1. (Complex limit and convergence).** *Suppose* $\{z_n\}_{n=1}^{\infty}$ *is a sequence in* $\mathbb{C}$ *and* $L \in \mathbb{C}$. *We say the sequence* converges *to* $L$ *and write*

$$\lim_{n\to\infty} z_n = L$$

*if for every* $\varepsilon > 0$ *there exists* $N \in \mathbb{N}$ *such that*

$$|z_m - L| < \varepsilon$$

*for all* $m \geq N$.

Note that this definition is identical to the definition in $\mathbb{R}$ except that $|\ |$ here means modulus instead of absolute value. Notice that if $z$ is a real number (and hence also a complex number), then its modulus and its absolute value coincide, so our notation is consistent. We could have defined complex limits in terms of real limits as the following proposition shows.

**Proposition A.4.2.** *If* $\{z_n\}_{n=1}^{\infty}$ *is a sequence in* $\mathbb{C}$, *then*

$$\lim_{n\to\infty} z_n = L$$

*if and only if*

$$\lim_{n\to\infty} \Re(z_n) = \Re(L) \text{ and } \lim_{n\to\infty} \Im(z_n) = \Im(L).$$

**Proof.** Notice that $\lim\limits_{n\to\infty} z_n = L$ is equivalent to the *real* limit

$$\lim_{n\to\infty} |L - z_n| = 0.$$

If $u_n = \Re(z_n),\ u = \Re(L),\ v_n = \Im(z_n)$, and $v = \Im(L)$, then

$$|L - z_n|^2 = |u + iv - u_n - iv_n|^2 = |u - u_n|^2 + |v - v_n|^2.$$

So $\lim |L - z_n|^2 = 0$ if and only if

$$\lim_{n\to\infty} |u - u_n|^2 = 0 \quad \text{and} \quad \lim_{n\to\infty} |v - u_v|^2 = 0.$$

$\square$

**Definition A.4.3. (Complex Cauchy sequence).** *A sequence of complex numbers* $\{z_n\}_{n=1}^{\infty}$ *is called a* Cauchy sequence *if for every* $\varepsilon > 0$ *there is an* $N > 0$ *(depending on* $\varepsilon$*) such that* $|x_n - x_m| < \varepsilon$ *for all* $n, m \geq N$.

**Theorem A.4.4. (Cauchy sequences converge).** *If* $\{z_n\}_{n=1}^{\infty}$ *is a Cauchy sequence in* $\mathbb{C}$, *then* $\lim_{n\to\infty} z_n$ *exists.*

**Proof.** We notice that $|z_n - z_m| \geq |\Re(z_n - z_m)| = |\Re(z_n) - \Re(z_m)|$, so $\{\Re(z_n)\}_{n=1}^{\infty}$ is a Cauchy sequence in $\mathbb{R}$ and hence converges. Similarly, $\{\Im(z_n)\}_{n=1}^{\infty}$ converges. Therefore, by Proposition A.4.2 the sequence $\{z_n\}_{n=1}^{\infty}$ converges. $\square$

**Definition A.4.5. (Convergent series, absolutely convergent).** *Suppose* $\sum_{n=1}^{\infty} z_n$ *is an infinite series of complex numbers and* $S_m = \sum_{n=1}^{m} z_n$ *is its* $m^{th}$ *partial sum. The series* $\sum_{n=1}^{\infty} z_n$ *is said to* converge *with limit* $L \in \mathbb{C}$ *provided* $\lim_{m\to\infty} S_m = L$. *It is said to* converge absolutely *provided the real series* $\sum_{n=1}^{\infty} |z_n|$ *converges.*

**Theorem A.4.6. (Absolutely convergent series).** *If the series* $\sum_{n=1}^{\infty} z_n$ *in* $\mathbb{C}$ *converges absolutely, then it converges.*

**Proof.** Suppose $u_n = \Re(z_n)$ and $v_n = \Im(z_n)$. Then $|z_n| \geq |u_n|$ and $|z_n| \geq |v_n|$. Hence if $\sum_{n=1}^{\infty} z_n$ converges absolutely, then so do $\sum_{n=1}^{\infty} u_n$ and $\sum_{n=1}^{\infty} v_n$. It follows that

$$\sum_{n=1}^{\infty} z_n = \sum_{n=1}^{\infty} (u_n + iv_n) = \sum_{n=1}^{\infty} u_n + i \sum_{n=1}^{\infty} v_n$$

converges. $\square$

## A.5. Set Theory and Countability

**Proposition A.5.1. (Distributivity of $\cap$ and $\cup$).** *If for each $j$ in some index set $J$ there is a set $B_j$ and $A$ is an arbitrary set, then*

$$A \cap \bigcup_{j \in J} B_j = \bigcup_{j \in J} (A \cap B_j) \quad \textit{and} \quad A \cup \bigcap_{j \in J} B_j = \bigcap_{j \in J} (A \cup B_j).$$

The proof is straightforward and is left to the reader.

**Definition A.5.2. (Set difference, complement).** *We define the* set difference *of sets $A$ and $B$ by*

$$A \setminus B = \{x \mid x \in A \quad \textit{and} \quad x \notin B\}.$$

*If all the sets under discussion are subsets of some fixed larger set $E$, then we can define the* complement *of $A$ with respect to $E$ to be $A^c = E \setminus A$.*

We will normally just speak of the complement $A^c$ of $A$ when it is clear what the larger set $E$ is. Note the obvious facts that $(A^c)^c = A$ and that $A \setminus B = A \cap B^c$.

**Proposition A.5.3.** *If for each $j$ in some index set $J$ there is a set $B_j \subset E$, then*

$$\bigcap_{j \in J} B_j^c = \Big( \bigcup_{j \in J} B_j \Big)^c \quad \textit{and} \quad \bigcup_{j \in J} B_j^c = \Big( \bigcap_{j \in J} B_j \Big)^c.$$

Again the elementary proof is left to the reader.

**Proposition A.5.4. (Well ordering of $\mathbb{N}$).** *Every non-empty subset $A$ of $\mathbb{N}$ has a least element which we will denote $min(A)$.*

**Proof.** Every finite subset of $\mathbb{N}$ clearly has a greatest element and a least element. Suppose $A \subset \mathbb{N}$ is non-empty. Let $B = \{n \in \mathbb{N} \mid n < a \text{ for all } a \in A\}$. Then $B$ is finite since it is a subset of $\{1, 2, 3, \dots, a_0\}$ for any $a_0 \in A$.

If $B = \emptyset$, then 1 must be in $A$ since otherwise it would be an element of $B$. Clearly, in this case $1 \in A$ is the least element of $A$.

If $B$ is non-empty, let $m$ be the greatest element of $B$ (which exists since $B$ is finite). The element $m_0 = m + 1$ must be in $A$ (since otherwise it would be in $B$ and greater than $m$). Clearly, then $m_0$ is the least element of $A$ since $m = m_0 - 1$ is less than every element of $A$. □

The notion of *countability*, which we now define, turns out to be a crucial ingredient in the concept of measure which is the main focus of this text.

**Definition A.5.5. (Countable, uncountable).** *A set $A$ is called* countable *if it is finite or there is a bijection from $A$ to the natural numbers $\mathbb{N}$, (i.e., a one-to-one correspondence between elements of $A$ and elements of $\mathbb{N}$). A set which is not countable is called* uncountable*.*

It should be noted that there is no universal agreement about whether finite sets should be called countable. Some authors reserve this term for infinite countable sets. In this text we will follow what seems to be the most common usage and finite sets will be called countable. In particular, the empty set will be called countable. It is a good exercise to understand why the proof of Propostion A.5.8 below remains valid when any of the countable sets in its statement are empty.

The next three propositions give terse proofs of some standard properties of countable sets which we will need.

**Proposition A.5.6. (Countable sets).**

(1) *A set $A$ is countable if and only if there is an injective function $f : A \to \mathbb{N}$. Hence, any subset of a countable set is countable.*

(2) *A set $A$ is countable if and only if there is a surjective function $f : \mathbb{N} \to A$.*

**Proof.** One direction of part (1) is easy: namely assuming $A$ is countable and constructing the injection $f$. If $A$ is infinite, there is a bijection $f : A \to \mathbb{N}$ which is, of course, an injection. If $A$ is finite, there is an $n > 0$ and a bijection from $A$ to $\{1, 2, \dots, n\}$ which can be considered an injection from $A$ to $\mathbb{N}$.

Conversely, if $f : A \to \mathbb{N}$ is an injection, it is a bijection from $A$ to $f(A) \subset \mathbb{N}$, so it suffices to show any subset of $\mathbb{N}$ is countable.

To prove this suppose $B$ is a subset of $\mathbb{N}$. If $B$ is finite, it is countable, so assume it is infinite. Define $\phi : \mathbb{N} \to B$ by $\phi(1) = min(B)$, and

$$\phi(k) = min(B \setminus \{\phi(1), \dots, \phi(k-1)\}).$$

The function $\phi$ is injective and defined for all $k \in \mathbb{N}$. Suppose $m \in B$ and let $c$ be the number of elements in the finite set $\{n \in B \mid n \leq m\}$. Then $\phi(c) = m$ and hence $\phi$ is surjective. This shows $\phi$ is a bijection. To observe that any subset of a countable set $A$ is countable note that an injection $f : A \to \mathbb{N}$ defines an injection from any subset of $A$ by restriction.

To prove (2) suppose $f : \mathbb{N} \to A$ is surjective. Define $\psi : A \to \mathbb{N}$ by $\psi(x) = min(f^{-1}(x))$. This is a bijection from $A$ to $\psi(A)$. Since $\psi(A)$ is a subset of $\mathbb{N}$ it is countable by (1). This proves one direction of (2). The converse is nearly obvious. If $A$ is countably infinite, then there is a bijection (and hence a surjection) $f : \mathbb{N} \to A$. But if $A$ is finite one can easily define a surjection $f : \mathbb{N} \to A$. $\square$

It is sometimes useful to think of the statements of Proposition A.5.6 in what amounts to their contrapositive form.

**Corollary A.5.7. (Uncountable sets).** *Suppose $f : A \to B$.*

(1) *If $f$ is injective and $A$ is uncountable, then $B$ is uncountable.*

(2) *If $f$ is surjective and $B$ is uncountable, then $A$ is uncountable.*

**Proof.** The first assertion follows from Proposition A.5.6, since if $B$ were countable, the set $A$ would also have to be countable. The second assertion is also a consequence of Proposition A.5.6, since if $A$ were countable, the set $B$ would also have to be countable. $\square$

**Proposition A.5.8. (Countable products and unions).**

(1) *If $A$ and $B$ are countable, then their* Cartesian product $A \times B$ *is a countable set.*

(2) *A countable union of countable sets is countable. That is, if $A_x$ is countable for each $x \in B$ and $B$ is countable, then $\bigcup_{x \in B} A_x$ is countable.*

**Proof.** We first prove part (1) in the case that $A = B = \mathbb{N}$. We define a function $\psi : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ by $\psi(m, n) = 2^m 3^n$. It is injective since $\psi(m, n) = \psi(p, q)$ implies $2^m 3^n = 2^p 3^q$, or $2^{m-p} = 3^{n-q}$. This is only possible if $m - p = 0$ and $n - q = 0$. Hence, $\psi$ is an injective function from $\mathbb{N}\times\mathbb{N}$ to $\mathbb{N}$, so $\mathbb{N}\times\mathbb{N}$ is countable by part (1) of Proposition A.5.6.

For the general case of part (1) we let $f_A : \mathbb{N} \to A$ and $f_B : \mathbb{N} \to B$ be surjections onto the countable sets $A$ and $B$. Then

$$f : \mathbb{N} \times \mathbb{N} \to A \times B$$

defined by $f(a, b) = (f_A(a), f_B(b))$ is a surjection and $A \times B$ is countable by part (2) of Proposition A.5.6.

To prove part (2) note that if $A_x$ is countable, there is a surjection $\Psi_x : \mathbb{N} \to A_x$ by part (2) of Proposition A.5.6. Likewise, there is a surjection $\phi : \mathbb{N} \to B$. The function

$$\Phi : \mathbb{N} \times \mathbb{N} \to \bigcup_{x \in B} A_x$$

given by

$$\Phi(n, m) = \Psi_{\phi(n)}(m)$$

is a surjection. Since $\mathbb{N} \times \mathbb{N}$ is countable it follows that $\bigcup_{x\in B} A_x$ is countable by part (2) of Proposition A.5.6. $\square$

**Corollary A.5.9. ($\mathbb{Q}$ is countable).** *The rational numbers $\mathbb{Q}$ are countable.*

**Proof.** The set $\mathbb{Z}$ is countable since it is the union of the countable sets $\mathbb{N}$, $-\mathbb{N}$, and $\{0\}$. So $\mathbb{Z} \times \mathbb{N}$ is countable and the function $\phi : \mathbb{Z} \times \mathbb{N} \to \mathbb{Q}$ given by $\phi(n, m) = n/m$ is surjective. Hence, the set of rationals $\mathbb{Q}$ is countable by part (2) of Proposition A.5.6. $\square$

It is not obvious that *any* uncountable sets exist. Our next task is to prove that they do, in fact, exist. We will also see in the exercises below that $\mathbb{R}$ is uncountable. For an arbitrary set $A$ we will denote by $\mathcal{P}(A)$ its *power set*, which is the set of all subsets of $A$.

**Proposition A.5.10.** *Suppose $A$ is a non-empty set and*

$$f : A \to \mathcal{P}(A)$$

*is a function from $A$ to its power set. Then $f$ is not surjective.*

**Proof.** This proof is short and elegant, but slightly tricky. For $a \in A$ either $a \in f(a)$ or $a \notin f(a)$. Let $B = \{a \in A \mid a \notin f(a)\}$.

Let $x$ be any element of $A$. From the definition of $B$ we observe that $x$ is in $B$ if and only if $x$ is not in the set $f(x)$. Or, equivalently, $x \notin B$ if and only if it is in the set $f(x)$. Hence, the sets $B$ and $f(x)$ can never be equal since one of them contains $x$ and the other does not. Therefore, there is no $x$ with $f(x) = B$, so $f$ is not surjective. $\square$

As an immediate consequence we have the existence of an uncountable set.

**Corollary A.5.11.** *The set $\mathcal{P}(\mathbb{N})$ is uncountable.*

**Proof.** This follows immediately from Proposition A.5.10. Since there is no surjection from $\mathbb{N}$ to $\mathcal{P}(\mathbb{N})$ there can be no bijection. $\square$

Later we will give an easy proof using measure theory that the set of irrationals is not countable (see Corollary B.2.8 and also part (4) of Exercise 2.2.2). But an elementary proof of this fact is outlined in the exercises below.

The next axiom asserts that there is a way to pick an element from each non-empty subset of $A$.

**The Axiom of Choice.** *For any non-empty set A there is a choice function*

$$\phi : \mathcal{P}(A) \setminus \{\emptyset\} \to A,$$

*i.e., a function such that for every non-empty subset $B \subset A$ we have $\phi(B) \in B$.*

---

**Exercise A.5.12.**

(1) Prove Propositions A.5.1 and A.5.3.

(2) Find an explicit bijection $f : \mathbb{Z} \to \mathbb{N}$ and conclude that $\mathbb{Z}$ is countable. Similarly, find explicit bijections from the even integers to $\mathbb{N}$ and the odd integers to $\mathbb{N}$.

(3) Prove that any infinite set contains a countable infinite subset.

(4) (*Uncountabilitity of* $\mathbb{R}$) Let $\mathcal{D}$ be the set of all infinite sequences $d_1d_2d_3\ldots d_n\ldots$ where each $d_n$ is either 0 or 1.

(a) Prove that $\mathcal{D}$ is uncountable. *Hint:* Consider the function $f : \mathcal{P}(\mathbb{N}) \to \mathcal{D}$ defined as follows. If $A \subset \mathbb{N}$, then $f(A) = d_1d_2d_3\ldots d_n\ldots$ where $d_n = 1$ if $n \in A$ and 0 otherwise.

(b) Define $h : \mathcal{D} \to [0,1]$ by letting $h(d_1d_2d_3\ldots d_n\ldots)$ be the real number whose decimal expansion is

$$0.d_1d_2d_3\ldots d_n\ldots.$$

Prove that $h$ is injective and hence by Corollary A.5.7 the interval $[0,1]$ is uncountable.

(c) Prove that if $a < b$, the closed interval $[a,b] = \{x \mid a \leq x \leq b\}$, the open interval $(a,b) = \{x \mid a < x < b\}$, the ray $[a,\infty) = \{x \mid a \leq x < \infty\}$, and $\mathbb{R}$ are all uncountable, by exhibiting an injective function from $[0,1]$ to each of them. Prove that the set of irrational numbers in $\mathbb{R}$ is uncountable.

(5) Let $f : [0,1] \to \mathbb{R}$ be a function such that $f(x) > 0$ for all $x \in [0,1]$, but which is otherwise arbitrary. Prove that there is a sequence $\{x_n\}_{n=1}^{\infty}$ of distinct elements $x_n \in [0,1]$ such that

$$\sum_{n=1}^{\infty} f(x_n) = \infty.$$

## A.6. Open and Closed Sets

We will denote the *closed interval* $\{x \mid a \leq x \leq b\}$ by $[a,b]$ and the *open interval* $\{x \mid a < x < b\}$ by $(a,b)$. We will also have occasion to refer to the *half open* intervals $(a,b] = \{x \mid a < x \leq b\}$ and $[a,b) = \{x \mid a \leq x < b\}$. Note that the interval $[a,a]$ is the set consisting of the single point $a$ and $(a,a)$ is the empty set.

**Definition A.6.1. (Open, closed, dense).** *A subset $A \subset \mathbb{R}$ is called an* open set *if for every $x \in A$ there is an open interval $(a,b) \subset A$ such that $x \in (a,b)$. A subset $B \subset \mathbb{R}$ is called* closed *if $\mathbb{R} \setminus B$ is open. A set $A \subset \mathbb{R}$ is said to be* dense *in $\mathbb{R}$ if every non-empty open subset contains a point of $A$.*

**Proposition A.6.2. ($\mathbb{Q}$ is dense in $\mathbb{R}$).** *The rational numbers $\mathbb{Q}$ are a dense subset of $\mathbb{R}$.*

**Proof.** Let $U$ be a non-empty open subset of $\mathbb{R}$. By the definition of open set there is a non-empty interval $(a, b) \subset U$. Choose an integer $n$ such that $\frac{1}{n} < b-a$. Then every point of $\mathbb{R}$ is in one of the intervals $[\frac{i-1}{n}, \frac{i}{n})$. In particular, for some integer $i_0$, $\frac{i_0-1}{n} \leq a < \frac{i_0}{n}$. Since $\frac{1}{n} < b - a$ it follows that

$$\frac{i_0 - 1}{n} \leq a < \frac{i_0}{n} \leq a + \frac{1}{n} < b.$$

Hence, the rational number $i_0/n$ is in $(a, b)$ and therefore in $U$. □

**Theorem A.6.3. (Open sets).** *A non-empty open set $U \subset \mathbb{R}$ is a countable union of pairwise disjoint open intervals $\bigcup_{n=1}^{\infty}(a_n, b_n)$.*

**Proof.** Let $x \in U$. Define $a_x = \inf\{y \mid [y, x] \subset U\}$ and $b_x = \sup\{y \mid [x, y] \subset U\}$ and let $U_x = (a_x, b_x)$. Then $U_x \subset U$, but $a_x \notin U$ since otherwise for some $\varepsilon > 0$, $[a_x - \varepsilon, a_x + \varepsilon] \subset U$ and hence $[a_x - \varepsilon, x] \subset [a_x - \varepsilon, a_x + \varepsilon] \cup [a_x + \varepsilon, x] \subset U$ and this would contradict the definition of $a_x$. Similarly, $b_x \notin U$. It follows that if $z \in U_x$, then $a_z = a_x$ and $b_z = b_x$. Hence, if $U_z \cap U_x \neq \emptyset$, then $U_z = U_x$ or equivalently, if $U_z \neq U_x$, then they are disjoint.

Thus, $U$ is a union of open intervals, namely the set of all the open intervals $U_x$ for $x \in U$. Any two such intervals are either equal or disjoint, so the collection of distinct intervals is pairwise disjoint.

To see that this is a countable collection observe that the rationals $\mathbb{Q}$ are countable, so $U \cap \mathbb{Q}$ is countable and the function $\phi$ which assigns to each $r \in U \cap \mathbb{Q}$ the interval $U_r$ is a surjective map onto this collection. By Proposition A.5.6 this collection must be countable. □

---

**Exercise A.6.4.**

(1) Prove that the complement of a closed subset of $\mathbb{R}$ is open.

(2) Prove that an arbitrary union of open sets is open and an arbitrary intersection of closed sets is closed.

(3) Prove that if $S$ is a closed bounded set, then both $\inf S$ and $\sup S$ are elements of $S$.

(4) The set $\mathcal{D} = \{m/2^n \mid m \in \mathbb{Z},\ n \in \mathbb{N}\}$ is called the *dyadic rationals*. Prove that $\mathcal{D}$ is dense in $\mathbb{R}$.

(5) A point $x$ is called a *limit point* of a set $S \subset \mathbb{R}$ if every open interval containing $x$ contains points of $S$ other than $x$. Prove that a set $S \subset \mathbb{R}$ is closed if and only if it contains all of its limit points. Show this implies that if $S$ is closed, $x_n \in S$ and $\lim x_n = z$, then $z \in S$.

## A.7. Compact Subsets of $\mathbb{R}$

One of the most important concepts for analysis is the notion of compactness.

**Definition A.7.1. (Compact).** *A set $X \subset \mathbb{R}$ is called* compact *provided every open cover of $X$ has a finite subcover.*

Less tersely, $X$ is compact if for every collection $\mathcal{V}$ of open sets with the property that

$$X \subset \bigcup_{U \in \mathcal{V}} U$$

there is a finite collection $U_1, U_2, \dots, U_n$ of open sets in $\mathcal{V}$ such that

$$X \subset \bigcup_{k=1}^{n} U_k.$$

For our purposes the key property is that closed and bounded subsets of $\mathbb{R}$ are compact.

**Theorem A.7.2. (The Heine-Borel theorem).** *A subset $X$ of $\mathbb{R}$ is compact if and only if it is closed and bounded.*

**Proof.** To see that a compact set is bounded observe that if $U_n = (-n, n)$, then $\{U_n\}$ is an open cover of any subset $X$ of $\mathbb{R}$. If this cover has a finite subcover, then $X \subset U_m$ for some $m$ and hence $X$ is bounded. To show a compact set $X$ is closed observe that if $y \notin X$, then

$$U_n = (-\infty, y - \frac{1}{n}) \ \cup \ (y + \frac{1}{n}, \infty)$$

defines an open cover of $\mathbb{R} \setminus \{y\}$ and hence of $X$. Since this cover of $X$ has a finite subcover there is an $m > 0$ such that $X \subset U_m$. It follows that $(y - 1/m, y + 1/m)$ is in the complement of $X$. Since $y$ is an arbitrary point of the complement of $X$, this complement is open and $X$ is closed.

To show the converse we first consider the special case that $X$ is the closed interval $[a, b]$. Let $\mathcal{V}$ be an open cover of $X$ and define

$$Y = \{x \in [a, b] \mid \text{the cover } \mathcal{V} \text{ of } [a, x] \text{ has a finite subcover}\}.$$

The set $Y$ is non-empty since $a \in Y$, and bounded since it is a subset of $[a, b]$. We define $z = \sup Y$ and note that $z \in [a, b]$.

Hence, there is an open set $U_0 \in \mathcal{V}$ with $z \in U_0$. From the definition of open sets we know there are points $z_0, z_1 \in U_0$ satisfying $z_0 < z < z_1$. From the definition of $z$ the cover $\mathcal{V}$ of $[a, z_0]$ has a finite subcover $U_1, U_2, \dots, U_n$. Then the finite subcover $U_0, U_1, U_2, \dots, U_n$ of $\mathcal{V}$ is a cover of $[a, x]$ for any $x \in [a, z_1]$. Since $z < z_1$, this contradicts the definition of $z$ unless $z_1 \geq b$. Hence, we conclude $z_1 \geq b$ and the finite cover of $[a, z_1]$ is also a cover of $[a, b]$.

For an arbitrary closed and bounded set $X$ we choose $a, b \in \mathbb{R}$ such that $X \subset [a, b]$. If $\mathcal{V}$ is any open cover of $X$ and we define $U_0 = \mathbb{R} \setminus X$, then $\mathcal{V} \cup \{U_0\}$ is an open cover of $[a, b]$ which must have a finite subcover, say $U_0, U_1, U_2, \dots, U_n$. Then $U_1, U_2, \dots, U_n$ must be a cover of $X$. □

There is a very important property of nested families of compact sets which we will use.

**Theorem A.7.3. (Nested families of compact sets).** *If $\{A_n\}_{n=1}^{\infty}$ is a nested family of non-empty compact subsets of $\mathbb{R}$, (i.e., $A_{n+1} \subset A_n$ for all $n$), then $\bigcap_{n=1}^{\infty} A_n$ is non-empty.*

**Proof.** Let $U_n = A_n^c$ be the complement of $A_n$. Then each $U_n$ is open and $U_n \subset U_{n+1}$. If $\bigcap_{n=1}^{\infty} A_n$ is empty, then its complement $\cup_{n=1}^{\infty} U_n$ is all of $\mathbb{R}$. Therefore, $\{U_n\}$ is an open cover of the compact set $A_1$, but it has no finite subcover since the union of the sets in any finite subcover is $U_N$ for some $N \in \mathbb{N}$ and $U_N$ does not contain the non-empty set $A_N \subset A_1$. We have contradicted the assumption that $\bigcap_{n=1}^{\infty} A_n$ is empty. □

**Exercise A.7.4.**

(1) Give an example of a nested family of non-empty open intervals $U_1 \supset U_2 \supset \cdots \supset U_n \ldots$, such that $\bigcap_{n=1}^{\infty} U_n$ is empty.

(2) A collection $\mathcal{F}$ of subsets of $\mathbb{R}$ is said to have the *finite intersection property* if the intersection of the sets in any finite subcollection of $\mathcal{F}$ is non-empty. Prove that if $\mathcal{F}$ is a collection of compact subsets of $\mathbb{R}$ and it has the finite intersection property, then the intersection of *all* the sets in $\mathcal{F}$ is non-empty.

(3) Prove that a subset $A \subset \mathbb{R}$ is compact if and only if it has the property that every sequence in $A$ has a subsequence which converges to an element of $A$.

## A.8. Continuous and Differentiable Functions

**Definition A.8.1. (Continuity and uniform continuity).** *Suppose $X \subset \mathbb{R}$. A function $f : X \to \mathbb{R}$ is* continuous *if for every $x \in X$ and every $\varepsilon > 0$ there is a $\delta(x)$ (depending on $x$) such that $|f(y) - f(x)| < \varepsilon$ whenever $y \in X$ and $|y - x| < \delta(x)$. A function $f : X \to \mathbb{R}$ is* uniformly continuous *if for every $\varepsilon > 0$ there is a $\delta$ (independent of $x$ and $y$) such that $|f(y) - f(x)| < \varepsilon$ whenever $x, y \in X$ and $|y - x| < \delta$.*

**Theorem A.8.2.** *If $f$ is defined and continuous on a compact subset $X \subset \mathbb{R}$, then it is uniformly continuous.*

**Proof.** Suppose $\varepsilon > 0$ is given. For any $x \in X$ and any positive number $\delta$ let $U(x, \delta) = (x - \delta, x + \delta)$. From the definition of continuity it follows that for each $x$ there is a $\delta(x) > 0$ such that for every $y \in U(x, \delta(x)) \cap X$ we have $|f(x) - f(y)| < \varepsilon/2$. Therefore, if $y_1$ and $y_2$ are both in $U(x, \delta(x)) \cap X$ we note

$$|f(y_1) - f(y_2)| \leq |f(y_1) - f(x)| + |f(x) - f(y_2)| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

The collection $\{U(x, \delta(x)/2) \mid x \in [a, b]\}$ is an open cover of the compact set $X$ so it has a finite subcover $\{U(x_i, \delta(x_i)/2) \mid 1 \leq i \leq n\}$. Let

$$\delta = \frac{1}{2} \min\{\delta(x_i) \mid 1 \leq i \leq n\}.$$

Suppose now $y_1, y_2 \in X$ and $|y_1 - y_2| < \delta$. Then

$$y_1 \in U\Big(x_j, \frac{\delta(x_j)}{2}\Big)$$

for some $1 \leq j \leq n$ and

$$|y_2 - x_j| \leq |y_2 - y_1| + |y_1 - x_j| < \delta + \frac{\delta(x_j)}{2} \leq \delta(x_j).$$

So both $y_1$ and $y_2$ are in $U(x_j, \delta(x_j))$ and hence

$$|f(y_1) - f(y_2)| < \varepsilon.$$

□

We will also make use of the following result from elementary calculus.

**Theorem A.8.3. (Mean value theorem).** *If $f$ is differentiable on the interval $[a, b]$, then there is $c \in (a, b)$ such that*

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

**Corollary A.8.4.** *If $f$ and $g$ are differentiable functions on $[a, b]$ and $f'(x) = g'(x)$ for all $x$, then there is a constant $C$ such that $f(x) = g(x) + C$.*

**Proof.** Let $h(x) = f(x) - g(x)$, then $h'(x) = 0$ for all $x$ and we wish to show $h$ is constant. But if $a_0, b_0 \in [a, b]$, then the mean value theorem says $h(b_0) - h(a_0) = h'(c)(b_0 - a_0) = 0$ since $h'(c) = 0$. Thus, for arbitrary $a_0, b_0 \in [a, b]$ we have $h(b_0) = h(a_0)$, so $h$ is constant. □

---

**Exercise A.8.5.**

(1) Suppose that $f : \mathbb{R} \to \mathbb{R}$ is a surjective function such that $x \leq y$ implies $f(x) \leq f(y)$ for all $x, y \in \mathbb{R}$. Prove that $f$ is continuous.

(2) (*Characterization of continuity*) Suppose $f$ is a function $f : \mathbb{R} \to \mathbb{R}$.

  (a) Prove that $f$ is continuous if and only if the set inverse $f^{-1}(U)$ is open for every open set $U \subset \mathbb{R}$.

  (b) Prove that $f$ is continuous if and only if the set inverse $f^{-1}((a, b))$ is open for every open interval $(a, b)$.

(c) Prove that $f$ is continuous if and only if the set inverse $f^{-1}(C)$ is closed for every closed set $C \subset \mathbb{R}$.

(3) (*Intermediate value theorem*) Prove that if $f : [a, b] \to \mathbb{R}$ is continuous with $f(a) < 0$ and $0 < f(b)$, then there exists $c \in (a, b)$ such that $f(c) = 0$.

(4) Prove that if $A \subset \mathbb{R}$ is compact and $f : \mathbb{R} \to \mathbb{R}$ is continuous, then $f(A)$ is compact.

(5) (*Complex continuous functions*) If $A \subset \mathbb{R}$ and $f : A \to \mathbb{C}$, we define $f$ to be continuous if $\Re(f(x))$ and $\Im(f(x))$ are continuous functions from $A$ to $\mathbb{R}$. Prove that $f : \mathbb{R} \to \mathbb{C}$ is continuous if and only if for each $a \in \mathbb{R}$,

$$\lim_{x \to a} f(x) = f(a).$$

## A.9. Real Vector Spaces

In this section we describe some basic properties of real vector spaces. A rigorous definition of a real vector space can be found in any linear algebra text. However, in this text we will only consider vector spaces of real-valued functions defined on some fixed domain and these are much simpler to define. For a discussion of general vector spaces in greater depth than we present here see [**K**].

**Definition A.9.1. (Vector space of functions).** *If $A$ is a set, a non-empty collection $\mathcal{V}$ of real-valued functions with domain $A$ is a* vector space of real-valued functions *provided:*

(1) *If $f$ and $g$ are in $\mathcal{V}$, then $f + g$ is in $\mathcal{V}$.*

(2) *If $f \in \mathcal{V}$ and $c \in \mathbb{R}$, then $cf$ is in $\mathcal{V}$.*

**Examples A.9.2.** The following are examples of vector spaces of real-valued functions.

(1) The collection of all functions from a set $A$ to $\mathbb{R}$.

(2) The collection of all continuous functions from $[a, b]$ to $\mathbb{R}$.

(3) The collection of all infinite sequences $\{x_n\}$ in $\mathbb{R}$ (Recall these are just the functions from $\mathbb{N}$ to $\mathbb{R}$.)

(4) The collection of all finite sequences $\{x_i\}_{i=1}^n$ in $\mathbb{R}$ (This is another description of $\mathbb{R}^n$ and is also the collection of all real-valued functions with domain $\{1, 2, \ldots, n\}$.)

Henceforth, we will use the term *real vector space* in its standard linear algebra sense. However, the reader will suffer no great loss of generality in thinking only of vector spaces of real-valued functions as defined above.

**Definition A.9.3. (Inner product space).** *An* inner product space *is a real vector space $\mathcal{V}$ together with a function $\langle\ ,\ \rangle : \mathcal{V}\times\mathcal{V} \to \mathbb{R}$ which for any $v_1, v_2, w \in \mathcal{V}$ and any $a, c_1, c_2 \in \mathbb{R}$ satisfies:*

(1) **Commutativity:** $\langle v_1, v_2\rangle = \langle v_2, v_1\rangle$.
(2) **Bilinearity:** $\langle c_1v_1 + c_2v_2, w\rangle = c_1\langle v_1, w\rangle + c_2\langle v_2, w\rangle$.
(3) **Positive Definiteness:** $\langle w, w\rangle \geq 0$ *with equality only if* $w = 0$.

**Definition A.9.4. (Norm).** *If $\mathcal{V}$ is a real vector space with inner product $\langle\ ,\ \rangle$, we define the associated* norm $\|\ \|$ *by* $\|v\| = \sqrt{\langle v, v\rangle}$.

**Proposition A.9.5. (Cauchy-Schwarz inequality).** *If $(\mathcal{V}, \langle\ ,\ \rangle)$ is an inner product space and $v, w \in \mathcal{V}$, then*

$$|\langle v, w\rangle| \leq \|v\|\ \|w\|,$$

*with equality if and only if $v$ and $w$ are multiples of a single vector.*

**Proof.** First assume that $\|v\| = \|w\| = 1$. Then

$$\begin{aligned}\|\langle v, w\rangle w\|^2 + \|v - \langle v,\ w\rangle w\|^2 & \\ &= \langle v, w\rangle^2\|w\|^2 + \langle v - \langle v, w\rangle w,\ v - \langle v, w\rangle w\rangle \\ &= \langle v, w\rangle^2\|w\|^2 + \|v\|^2 - 2\langle v, w\rangle^2 + \langle v, w\rangle^2\|w\|^2 \\ &= \|v\|^2 = 1,\end{aligned}$$

since $\|v\|^2 = \|w\|^2 = 1$. Hence,

$$\langle v, w\rangle^2 = \|\langle v, w\rangle w\|^2 \leq \|\langle v, w\rangle w\|^2 + \|v - \langle v, w\rangle w\|^2 = 1.$$

This implies the inequality $|\langle v, w\rangle| \leq 1 = \|v\|\ \|w\|$, when $v$ and $w$ are unit vectors. Also this inequality is an equality only if $\|v - \langle v, w\rangle w\| = 0$ or $v = \langle v, w\rangle w$, i.e., since $v$ and $w$ are unit vectors, only if $v = \pm w$.

The general result is trivial if either $v$ or $w$ is 0. Hence, we may assume the vectors are non-zero multiples, $v = av_0$ and $w = bw_0$, of unit vectors $v_0$ and $w_0$. In this case we have $|\langle v, w\rangle| = |\langle av_0, bw_0\rangle| = |ab||\langle v_0, w_0\rangle| \leq |ab| = \|av_0\|\ \|bw_0\| = \|v\|\ \|w\|$.

Observe that we have equality only if $|\langle v_0, w_0\rangle| = 1$, which, as noted above, only happens if $v_0 = \pm w_0$, i.e., only if $v$ and $w$ are multiples of each other. $\square$

**Proposition A.9.6. (Normed linear space).** *If $\mathcal{V}$ is an inner product space and $\| \ \|$ is the norm defined by $\|v\| = \sqrt{\langle v, v\rangle}$, then*

(1) *For all $a \in \mathbb{R}$ and $v \in \mathcal{V}$, $\|av\| = |a|\|v\|$.*
(2) *For all $v \in \mathcal{V}$, $\|v\| \geq 0$ with equality only if $v = 0$.*
(3) **Triangle Inequality:** *For all $v, w \in \mathcal{V}$, $\|v + w\| \leq \|v\| + \|w\|$.*
(4) **Parallelogram Law:** *For all $v, w \in \mathcal{V}$,*
$$\|v - w\|^2 + \|v + w\|^2 = 2\|v\|^2 + 2\|w\|^2.$$

**Proof.** The first two of these properties follow immediately from the definition of inner product. To prove item (3), the triangle inequality, observe:

$$\begin{aligned}
\|v + w\|^2 &= \langle v + w, v + w\rangle \\
&= \langle v, v\rangle + 2\langle v, w\rangle + \langle w, w\rangle \\
&= \|v\|^2 + 2\langle v, w\rangle + \|w\|^2 \\
&\leq \|v\|^2 + 2|\langle v, w\rangle| + \|w\|^2 \\
&\leq \|v\|^2 + 2\|v\|\ \|w\| + \|w\|^2 \text{ by Cauchy-Schwarz,} \\
&= (\|v\| + \|w\|)^2.
\end{aligned}$$

To prove item (4), the parallelogram law, note

$$\|v - w\|^2 = \langle v - w, v - w\rangle = \|v\|^2 - 2\langle v, w\rangle + \|w\|^2$$

and

$$\|v + w\|^2 = \langle v + w, v + w\rangle = \|v\|^2 + 2\langle v, w\rangle + \|w\|^2.$$

Hence, the sum $\|v - w\|^2 + \|v + w\|^2$ equals $2\|v\|^2 + 2\|w\|^2$. $\square$

**Definition A.9.7. (Orthonormal).** *A set of vectors* $\{u_i\}_{i=1}^k$ *in* $\mathbb{R}^n$ *is called* orthonormal *with respect to the inner product* $\langle\ ,\ \rangle$ *provided* $\langle u_i, u_j\rangle = 0$ *if* $i \neq j$ *and* $\|u_i\|^2 = \langle u_i, u_i\rangle = 1$ *for* $1 \leq i \leq k$.

**Proposition A.9.8. (Orthonormal basis).** *Suppose* $\{u_i\}_{i=1}^n$ *is an orthonormal set of in* $\mathbb{R}^n$. *Then any* $v \in \mathbb{R}^n$ *can be uniquely expressed as*

$$v = \sum_{i=1}^n \langle v, u_i\rangle u_i.$$

**Proof.** Suppose $v \in \mathbb{R}^n$ is arbitrary. Let

$$w = \sum_{i=1}^n \langle v, u_i\rangle u_i.$$

We will show $w = v$. Note that for each $i$, $\langle w, u_i\rangle = \langle v, u_i\rangle$. It follows that

$$\langle w - v, u_i\rangle = \langle w, u_i\rangle - \langle v, u_i\rangle = 0$$

for $1 \leq i \leq n$.

By part (1) of Exercise A.9.9 below a vector which is perpendicular to each $u_i$ must be 0. Hence, $w - v = 0$.

To show uniqueness suppose that $v$ also equals $\sum_{i=1}^n a_i u_i$ for some real numbers $a_i$. Then

$$\langle v, u_j\rangle = \langle \sum_{i=1}^n a_i u_i, u_j\rangle = a_j.$$

for each $1 \leq j \leq n$. □

---

**Exercise A.9.9.**

(1) *This exercise requires a knowledge of dimension and/or linear independence in* $\mathbb{R}^n$. Prove there is no orthonormal subset of $\mathbb{R}^n$ with more than $n$ elements. Show, in fact, that if $\{u_i\}_{i=1}^n$ is an orthonormal set in $\mathbb{R}^n$, then $\langle v, u_i\rangle = 0$ for all $1 \leq i \leq n$ only if $v = 0$.

(2) Consider the set $\ell^2$ of square summable sequences in $\mathbb{R}$ defined in part (3) of Exercise A.3.8. Prove that $\ell^2$ is a vector space of real-valued functions.

(3) If $\{x_n\}$ and $\{y_n\}$ are elements of $\ell^2$, define a function $\langle\ ,\ \rangle : \ell^2 \times \ell^2 \to \mathbb{R}$ by

$$\langle \{x_n\}, \{y_n\} \rangle = \sum_{n=1}^{\infty} x_n y_n.$$

Prove that $\langle\ ,\ \rangle$ is an inner product on $\ell^2$. (See part (3) of Exercise A.3.8.)

## A.10. Complex Vector Spaces

In this section we describe some basic properties of complex vector spaces. Most of the properties we describe are analogous to properties of real vector spaces we showed in the previous section and the proofs are often nearly identical. A rigorous definition of a complex vector space can be found in most linear algebra texts (see [**K**], for example). However, here we will only consider vector spaces of complex-valued functions defined on some fixed domain and these are much simpler to define.

**Definition A.10.1. (Complex vector space of functions).** *If $A$ is a set, a non-empty collection $\mathcal{V}$ of complex-valued functions with domain $A$ is a* vector space of complex-valued functions *provided:*

(1) *If $f$ and $g$ are in $\mathcal{V}$, then $f+g$ is in $\mathcal{V}$.*

(2) *If $f \in \mathcal{V}$ and $c \in \mathbb{C}$, then $cf$ is in $\mathcal{V}$.*

**Examples A.10.2.** The following are examples of vector spaces of complex-valued functions.

(1) The collection of all functions from a set $A$ to $\mathbb{C}$.

(2) The collection of all continuous functions from $[a, b]$ to $\mathbb{C}$.

(3) The collection of all infinite sequences $\{x_n\}$ in $\mathbb{C}$. (Recall these are just the functions from $\mathbb{N}$ to $\mathbb{C}$.)

(4) The collection of all finite sequences $\{x_i\}_{i=1}^n$ in $\mathbb{C}$. (This is another description of $\mathbb{C}^n$ and is also the collection of all complex-valued functions with domain $\{1, 2, \dots, n\}$.)

Henceforth, we will use the term *complex vector space* in its standard linear algebra sense. However, the reader will suffer no great loss of generality in thinking only of vector spaces of complex-valued functions as defined above. For more details about complex vector spaces and the properties summarized below see [**K**].

**Definition A.10.3. (Hermitian form).** *A* Hermitian form *on a complex vector space $\mathcal{V}$ is a function $\langle\ ,\ \rangle : \mathcal{V} \times \mathcal{V} \to \mathbb{C}$ which satisfies:*

(1) **Conjugate symmetry:** $\langle v, w\rangle = \overline{\langle w, v\rangle}$ *for all* $v, w \in \mathcal{V}$.

(2) **Sesquilinearity:**

$$\langle c_1 v_1 + c_2 v_2, w\rangle = c_1\langle v_1, w\rangle + c_2\langle v_2, w\rangle \quad \textit{and}$$
$$\langle v, c_1 w_1 + c_2 w_2, \rangle = \bar{c}_1\langle v, w_1\rangle + \bar{c}_2\langle v, w_2\rangle$$

*for all* $v_1, v_2, w_1, w_2 \in \mathcal{V}$ *and all* $c_1, c_2 \in \mathbb{C}$.

(3) **Positive definiteness:** *For all* $w \in \mathcal{V}$, $\langle w, w\rangle$ *is real and* $\geq 0$ *with equality only if* $w = 0$.

Note that the property that $\langle w, w\rangle$ is real for all $w \in \mathcal{V}$ is a consequence of skew symmetry since the fact that $\langle w, w\rangle = \overline{\langle w, w\rangle}$ implies it is real.

**Example A.10.4.** Let $\mathbb{C}^n$ denote the complex vector space of $n$-tuples of complex numbers. If $z = (z_1, \dots, z_n)$ and $w = (w_1, \dots, w_n)$ are elements of $\mathbb{C}^n$, we define

$$\langle z, w\rangle = \sum_{i=1}^{n} z_i \bar{w}_i.$$

This is called the *standard Hermitian form* on $\mathbb{C}^n$.

**Proposition A.10.5. (Hermitian form).** *If $\mathcal{V}$ is a complex vector space with a Hermitian form and $\|\ \|$ is defined by $\|v\| = \sqrt{\langle v, v\rangle}$, then*

(1) *For all $c \in \mathbb{C}$ and $v \in \mathcal{V}$,* $\|cv\| = |c|\|v\|$.

(2) *For all $v \in \mathcal{V}$,* $\|v\| \geq 0$ *with equality only if* $v = 0$.

*So $\|\ \|$ is a norm on $\mathcal{V}$.*

**Proof.** To show property (1) observe that

$$\|cv\|^2 = \langle cv, cv\rangle = c\bar{c}\langle v, v\rangle = |c|^2\|v\|^2.$$

Property (2) is immediate from positive definiteness of the Hermitian form. □

**Definition A.10.6. (Norm).** *If $\mathcal{V}$ is a complex vector space with inner product $\langle\ ,\ \rangle$, we define the associated* norm $\|\ \|$ *by* $\|v\| = \sqrt{\langle v, v\rangle}$.

As in the real case, two vectors $x$ and $y$ are said to be *perpendicular* with respect to the Hermitian form $\langle\ ,\ \rangle$ if $\langle x, y\rangle = 0$.

**Proposition A.10.7. (Pythagorean theorem).** *If $x_1, x_2, \ldots, x_n$ are mutually perpendicular elements of a complex vector space with Hermitian form $\langle\ ,\ \rangle$ and associated norm $\|\ \|$, then*

$$\Big\| \sum_{i=1}^{n} x_i \Big\|^2 = \sum_{i=1}^{n} \|x_i\|^2.$$

**Proof.** Consider the case $n = 2$. If $x$ is perpendicular to $y$, then

$$\|x+y\|^2 = \langle x+y, x+y\rangle = \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle = \|x\|^2 + \|y\|^2$$

since $\langle x, y\rangle = \langle y, x\rangle = 0$. The general case follows by induction on $n$. □

**Proposition A.10.8. (Cauchy-Schwarz inequality).** *If $\mathcal{V}$ is a complex vector space and $\langle\ ,\ \rangle$ is a Hermitian form, then for all $v, w \in \mathcal{V}$,*

$$|\langle v, w\rangle| \leq \|v\|\ \|w\|,$$

*with equality if and only if $v$ and $w$ are multiples of a single vector.*

**Proof.** The proof is *very* similar to the proof for inner products on real vector spaces. One difference to keep in mind is that in this proof $\langle\ ,\ \rangle$ is a Hermitian form rather than an inner product and $|\ |$ denotes the modulus of a complex number rather than absolute value.

First assume $\|v\| = \|w\| = 1$. Then

$$\begin{aligned}
\|\langle v, w\rangle w\|^2 + &\|v - \langle v,\ w\rangle w\|^2 \\
&= |\langle v, w\rangle|^2 \|w\|^2 + \langle v - \langle v, w\rangle w,\ v - \langle v, w\rangle w\rangle \\
&= |\langle v, w\rangle|^2 \|w\|^2 + \|v\|^2 - 2\langle v, w\rangle \overline{\langle v, w\rangle} + |\langle v, w\rangle|^2 \|w\|^2 \\
&= \|v\|^2 = 1,
\end{aligned}$$

since $\|v\|^2 = \|w\|^2 = 1$. Hence,

$$\langle v, w\rangle^2 = \|\langle v, w\rangle w\|^2 \leq \|\langle v, w\rangle w\|^2 + \|v - \langle v, w\rangle w\|^2 = 1$$

This implies the inequality $|\langle v, w\rangle| \leq 1 = \|v\| \, \|w\|$, when $v$ and $w$ are unit vectors. Also this inequality is an equality only if $\|v - \langle v, w\rangle w\| = 0$ or $v = \langle v, w\rangle w$, i.e., since $v$ and $w$ are unit vectors, only if $v = cw$ for some $c \in \mathbb{C}$ with modulus 1.

The general result is trivial if either $v$ or $w$ is 0. Hence, we may assume the vectors are non-zero multiples, $v = av_0$ and $w = bw_0$, of unit vectors $v_0$ and $w_0$. In this case we have $|\langle v, w\rangle| = |\langle av_0, bw_0\rangle| = |a\bar{b}||\langle v_0, w_0\rangle| \leq |a\bar{b}| = |a||b| = \|av_0\| \, \|bw_0\| = \|v\| \, \|w\|$.

Observe that we have equality only if $|\langle v_0, w_0\rangle| = 1$, which, as noted above, only happens if $v_0 = cw_0$ for some $c$ with modulus 1, i.e., only if $v$ and $w$ are multiples of each other. $\square$

**Proposition A.10.9. (Triangle inequality and parallelogram law).** *If $\mathcal{V}$ is a complex vector space with a Hermitian form $\langle \ , \ \rangle$ and $\| \ \|$ is the associated norm, then*

(1) *for all $v, w \in \mathcal{V}$, $\|v + w\| \leq \|v\| + \|w\|$;*

(2) *for all $v, w \in \mathcal{V}$, $\|v - w\|^2 + \|v + w\|^2 = 2\|v\|^2 + 2\|w\|^2$.*

**Proof.** Observe

$$\begin{aligned}
\|v + w\|^2 &= \langle v + w, v + w\rangle \\
&= \langle v, v\rangle + \langle v, w\rangle + \langle w, v\rangle + \langle w, w\rangle \\
&= \|v\|^2 + 2\Re(\langle v, w\rangle) + \|w\|^2 \\
&\leq \|v\|^2 + 2|\langle v, w\rangle| + \|w\|^2 \\
&\leq \|v\|^2 + 2\|v\| \, \|w\| + \|w\|^2 \text{ by Cauchy-Schwarz,} \\
&= (\|v\| + \|w\|)^2.
\end{aligned}$$

The proof of item (2), the parallelogram law, is very similar to the real version. Note $\|v - w\|^2 = \langle v - w, v - w\rangle = \|v\|^2 - \langle v, w\rangle - \langle w, v\rangle + \|w\|^2$. Likewise, $\|v + w\|^2 = \langle v + w, v + w\rangle = \|v\|^2 + \langle v, w\rangle + \langle w, v\rangle + \|w\|^2$. Hence, the sum $\|v - w\|^2 + \|v + w\|^2$ equals $2\|v\|^2 + 2\|w\|^2$. $\square$

**Exercise A.10.10.**

(1) Prove that $\langle\ ,\ \rangle : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$ defined in Example A.10.4 is a Hermitian form.

(2) Consider the set $\ell^2_{\mathbb{C}}$ of square summable sequences in $\mathbb{C}$ defined to be those sequences $\{z_i\}_{i=1}^{\infty}$ such that

$$\sum_{i=1}^{\infty} |z_i|^2 < \infty.$$

Prove that $\ell^2_{\mathbb{C}}$ is a vector space of complex-valued functions.

(3) If $\{z_n\}$ and $\{w_n\}$ are elements of $\ell^2_{\mathbb{C}}$, define a function $\langle\ ,\ \rangle : \ell^2_{\mathbb{C}} \times \ell^2_{\mathbb{C}} \to \mathbb{C}$ by

$$\langle \{z_n\}, \{w_n\} \rangle = \sum_{n=1}^{\infty} z_n \bar{w}_n.$$

Prove that $\langle\ ,\ \rangle$ is a Hermitian form on $\ell^2_{\mathbb{C}}$.

## A.11. Complete Normed Vector Spaces

In this section $\mathcal{V}$ will denote either a real or a complex vector space with a norm $\|\ \|$. The concepts of limit and Cauchy sequence extend naturally to vector spaces with a norm. Indeed, the definitions are essentially identical to those given for $\mathbb{R}$ and $\mathbb{C}$. The only difference is that where before we used $|\ |$ to represent absolute value or modulus of a complex number we now use $\|\ \|$, the norm.

**Definition A.11.1. (Limit and convergence).** *Suppose $\{v_n\}_{n=1}^{\infty}$ is a sequence in a vector space $\mathcal{V}$ with norm $\|\ \|$ and $w \in \mathcal{V}$. We say the sequence* converges *to $w$ and write*

$$\lim_{n\to\infty} v_n = w$$

*if for every $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that*

$$\|v_m - w\| < \varepsilon$$

*for all $m \geq N$.*

As mentioned earlier, the assertion that Cauchy sequences in $\mathbb{R}$ always converge is equivalent to the Completeness Axiom. This formulation is the one which generalizes to $\mathbb{C}$ (as we have seen in Theorem A.4.4). The concept of completeness is also useful in normed vector spaces. However, in normed vector spaces, especially the function spaces considered in this text, it is a very strong property not shared by many spaces.

**Definition A.11.2. (Cauchy sequence, complete).** *A sequence* $\{v_n\}_{n=1}^{\infty}$ *in a normed vector space* $\mathcal{V}$ *is called a* Cauchy sequence *if for every* $\varepsilon > 0$ *there is an* $N > 0$ *(depending on* $\varepsilon$*) such that* $\|v_n - v_m\| < \varepsilon$ *for all* $n, m \geq N$*. The normed vector space* $\mathcal{V}$ *is called* complete *if every Cauchy sequence in* $\mathcal{V}$ *converges to an element of* $\mathcal{V}$*.*

---

**Exercise A.11.3.**

(1) Suppose the sequence $\{v_n\}_{n=1}^{\infty}$ in a normed vector space converges. Prove it is Cauchy.

(2) Prove that $\mathbb{R}^n$ with the standard norm is complete.

(3) Prove that $\mathbb{C}^n$ with the norm associated to the standard Hermitian form from Example A.10.4 is complete.

(4) Let $\| \ \|$ be the norm on $\ell^2$ associated with the inner product $\langle \ , \ \rangle$. Prove $\ell^2$ with this norm is a complete normed linear space. (See Exercise A.9.9.) Similarly, let $\| \ \|$ be the norm on $\ell^2_{\mathbb{C}}$ associated with the inner product $\langle \ , \ \rangle$. Prove $\ell^2_{\mathbb{C}}$ with this norm is a complete normed linear space.

Appendix B

# Lebesgue Measure

## B.1. Introduction

In this appendix we define a generalization of length called *measure* for bounded subsets of the real line or subsets of the interval $[a, b]$. There are several properties which we want it to have. For a bounded subset $A$ of $\mathbb{R}$ we would like to be able to assign a non-negative real number $\mu(A)$ in a way that satisfies the following:

**I. Length:** If $A = (a, b)$ or $[a, b]$, then $\mu(A) = \operatorname{len}(A) = b - a$, i.e., the measure of an open or closed interval is its length.

**II. Translation invariance:** If $A \subset \mathbb{R}$ is a bounded subset of $\mathbb{R}$ and $c \in \mathbb{R}$, then $\mu(A + c) = \mu(A)$, where $A + c$ denotes the set $\{x + c \mid x \in A\}$.

**III. Countable additivity:** If $\{A_n\}_{n=1}^{\infty}$ is a countable collection of bounded subsets of $\mathbb{R}$, then

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

and if the sets are *pairwise disjoint,* then

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

Note the same conclusion applies to $\{A_n\}_{n=1}^m$, a finite collection of bounded sets (just let $A_i = \emptyset$ for $i > m$).

**IV. Monotonicity:** If $A \subset B$, then $\mu(A) \leq \mu(B)$. Actually, this property is a consequence of additivity since $A$ and $B \backslash A$ are disjoint and their union is $B$.

It turns out that it is not possible to find a $\mu$ which satisfies I–IV and which is defined for *all* bounded subsets of the reals; but we can do it for a very large collection including the open sets and the closed sets.

## B.2. Outer Measure

We first describe the notion of "outer measure" which comes close to what we want. It is defined for *all* subsets of the reals and satisfies properties I, II and IV above. It also satisfies the inequality part of the additivity condition, III, which is called *subadditivity*; but it fails to be additive for some choices of disjoint sets. The resolution of this difficulty will be to restrict its definition to a certain large collection of nice sets (the measurable sets) on which we can show the additivity condition holds. Our task for this appendix then is threefold: (1) we must define the concept of measurable sets and develop their properties; (2) we must define the notion of *Lebesgue measure* $\mu$ for such sets; and (3) we must prove that properties I-IV hold, if we restrict our attention to measurable sets.

Suppose $A \subset \mathbb{R}$ is a bounded set and $\{U_n\}$ is a countable covering of $A$ by open intervals, i.e., $A \subset \bigcup_n U_n$ where $U_n = (a_n, b_n)$. Then if we were able to define a function $\mu$ satisfying the properties I-IV above we could conclude from monotonicity and subadditivity that

$$\mu(A) \leq \mu\Big(\bigcup_{n=1}^{\infty} U_n\Big) \leq \sum_{n=1}^{\infty} \mu(U_n) = \sum_{n=1}^{\infty} \operatorname{len}(U_n)$$

and hence that $\mu(A)$ would be less than or equal to the *infimum* of all such sums where we consider all possible coverings of $A$ by a countable collection of open intervals. This turns out to lead to a very useful definition of an extended real-valued function $\mu^*$ defined for every subset $A$ of $\mathbb{R}$.

**Definition B.2.1. (Lebesgue outer measure).** *Suppose $A \subset \mathbb{R}$ and $\mathcal{U}(A)$ is the collection of all countable coverings of $A$ by open intervals. We define the* Lebesgue outer measure $\mu^*(A)$ *by*

$$\mu^*(A) = \inf_{\{U_n\} \in \mathcal{U}(A)} \Big\{ \sum_{n=1}^{\infty} \operatorname{len}(U_n) \Big\},$$

*where the infimum is taken over all possible countable coverings of $A$ by open intervals. If for every cover of $A$ by open intervals*

$$\sum_{n=1}^{\infty} \operatorname{len}(U_n) = +\infty,$$

*we define $\mu^*(A) = +\infty$.*

Notice that this definition plus Definition 2.2.1, the definition of a null set, imply that a set $A \subset I$ is a null set if and only if $\mu^*(A) = 0$.

We can immediately show that property I, the length property, holds for Lebesgue outer measure.

**Proposition B.2.2. (Length property for outer measure).** *For any $a, b \in \mathbb{R}$ with $a \leq b$ we have $\mu^*([a,b]) = \mu^*((a,b)) = b - a$.*

**Proof.** First consider the closed interval $[a,b]$. It is covered by the single interval $U_1 = (a-\varepsilon, b+\varepsilon)$, so $\mu^*([a,b]) \leq \operatorname{len}(U_1) = b - a + 2\varepsilon$. Since $2\varepsilon$ is arbitrary we conclude that $\mu^*([a,b]) \leq b - a$.

On the other hand, by Theorem A.7.2, the Heine-Borel theorem, any open covering of $[a,b]$ has a finite subcovering, so it suffices to prove that for any finite cover $\{U_i\}_{i=1}^n$ we have $\sum \operatorname{len}(U_i) \geq b - a$ as this will imply $\mu^*([a,b]) \geq b - a$. We prove this by induction on $n$, the number of elements in the cover by open intervals. Clearly, the result holds if $n = 1$. If $n > 1$ we note that two of the open intervals must intersect. This is because one of the intervals (say $(c,d)$) contains $b$ and if $c > a$ another interval contains $c$ and hence these two intersect. By renumbering the intervals we can assume that $U_{n-1}$ and $U_n$ intersect.

Now define $V_{n-1} = U_{n-1} \cup U_n$ and $V_i = U_i$ for $i < n-1$. Then $\{V_i\}$ is an open cover of $[a,b]$ containing $n-1$ intervals. By the induction hypothesis,

$$\sum_{i=1}^{n-1} \operatorname{len}(V_i) \geq b - a.$$

But $\operatorname{len}(U_{n-1}) + \operatorname{len}(U_n) > \operatorname{len}(V_{n-1})$ and $\operatorname{len}(U_i) = \operatorname{len}(V_{i-1})$ for $i > 2$. Hence,

$$\sum_{i=1}^{n} \operatorname{len}(U_i) > \sum_{i=1}^{n-1} \operatorname{len}(V_i) \geq b - a.$$

This completes the proof that $\mu^*([a, b]) \geq b - a$ and hence that $\mu^*([a, b]) = b - a$.

For the open interval $(a, b)$ we note that $U = (a, b)$ covers itself, so $\mu^*((a, b)) \leq b - a$. On the other hand, any cover $\{U_i\}_{i=1}^{\infty}$ of $(a, b)$ by open intervals is also a cover of the closed interval $[a + \varepsilon, b - \varepsilon]$ so, as we just showed,

$$\sum_{i=1}^{\infty} \operatorname{len}(U_i) \geq b - a - 2\varepsilon.$$

As $\varepsilon$ is arbitrary, $\sum \operatorname{len}(U_i) \geq b - a$ and hence $\mu^*((a, b)) \geq b - a$ which completes our proof. $\square$

Two special cases are worthy of note:

**Corollary B.2.3.** *The outer measure of a set consisting of a single point is* 0. *The outer measure of the empty set is also* 0.

Lebesgue outer measure satisfies a monotonicity property with respect to inclusion.

**Proposition B.2.4. (Monotonicity of outer measure).** *If $A$ and $B$ are subsets of $\mathbb{R}$ and $A \subset B$, then $\mu^*(A) \leq \mu^*(B)$.*

**Proof.** Since $A \subset B$, every countable cover $\{U_n\} \in \mathcal{U}(B)$ of $B$ by open intervals is also in $\mathcal{U}(A)$ since it also covers $A$. Thus,

$$\inf_{\{U_n\} \in \mathcal{U}(A)} \Big\{ \sum_{n=1}^{\infty} \operatorname{len}(U_n) \Big\} \leq \inf_{\{U_n\} \in \mathcal{U}(B)} \Big\{ \sum_{n=1}^{\infty} \operatorname{len}(U_n) \Big\},$$

so $\mu^*(A) \leq \mu^*(B)$. $\square$

**Corollary B.2.5.** *If $a \in \mathbb{R}$, then the outer measure of each of the sets $(a, \infty)$, $(-\infty, a)$ and $\mathbb{R}$ is $+\infty$.*

**Proof.** Since $\mu^*((-N, N)) = 2N$ and $(-N, N) \subset \mathbb{R}$ we know $\mu^*(\mathbb{R}) \geq 2N$ for all $N$. Likewise, $\mu^*((a, \infty)) \geq \mu^*((a, N)) \geq N - |a|$, and $\mu^*((-\infty, a)) \geq \mu^*((-N, a)) \geq N - |a|$. □

We can now prove the first part of the countable additivity property we want. It turns out that this is the best we can do if we want our measure defined on all subsets of $\mathbb{R}$. Note that the following result is stated in terms of a countably infinite collection $\{A_n\}_{n=1}^{\infty}$ of sets, but it is perfectly valid for a finite collection also.

**Theorem B.2.6. (Countable subadditivity).** *A countable collection $\{A_n\}_{n=1}^{\infty}$ of subsets of $\mathbb{R}$ satisfies*

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

**Proof.** By the definition of outer measure we know that each $A_n$ has a countable cover by open intervals $\{U_i^n\}$ such that

$$\sum_{i=1}^{\infty} \operatorname{len}(U_i^n) \leq \mu^*(A_n) + 2^{-n}\varepsilon.$$

But the union of all these covers $\{U_i^n\}$ is a countable cover of $\bigcup_{n=1}^{\infty} A_n$. So

$$\begin{aligned}
\mu^*\Big(\bigcup_{n=1}^{\infty} A_n\Big) &\leq \sum_{n=1}^{\infty}\sum_{i=1}^{\infty} \operatorname{len}(U_i^n) \\
&\leq \sum_{n=1}^{\infty} \mu^*(A_n) + \sum_{n=1}^{\infty} 2^{-n}\varepsilon \\
&= \sum_{n=1}^{\infty} \mu^*(A_n) + \varepsilon.
\end{aligned}$$

Since this is true for every $\varepsilon$ the result follows. □

**Corollary B.2.7.** *If $A$ is countable, then $\mu^*(A) = 0$*

**Proof.** Suppose $A = \bigcup_{i=1}^{\infty}\{x_i\}$. We saw in Corollary B.2.3 that $\mu^*(\{x_i\}) = 0$, so

$$\mu^*(A) = \mu^*(\bigcup_{i=1}^{\infty}\{x_i\}) \leq \sum_{i=1}^{\infty} \mu^*(\{x_i\}) = 0,$$

which implies $\mu^*(A) = 0$. $\square$

We also immediately obtain the following non-trivial result. (For alternate proofs see part (4) of Exercise A.5.12 and part (4) of Exercise 2.2.2).

**Corollary B.2.8. (Intervals are uncountable).** *If $a < b$, then $[a, b]$ is not countable.*

**Proof.** Countable sets have outer measure 0, but $\mu^*([a, b]) = b - a \neq 0$. $\square$

Outer Lebesgue measure satisfies property II of those we enumerated at the beginning, namely it is translation invariant.

**Theorem B.2.9. (Translation invariance).** *If $c \in \mathbb{R}$ and $A$ is a subset of $\mathbb{R}$, then $\mu^*(A) = \mu^*(A + c)$ where $A + c = \{x + c \mid x \in A\}$.*

We leave the (easy) proof as an exercise.

**Proposition B.2.10. (Regularity of outer measure).** *If $A \subset \mathbb{R}$ and $\mu^*(A)$ is finite, then for any $\varepsilon > 0$ there is an open set $V$ containing $A$ such that $\mu^*(V) < \mu^*(A) + \varepsilon$. As a consequence*

$$\mu^*(A) = \inf\{\mu^*(U) \mid U \textit{ is open and } A \subset U\}.$$

**Proof.** We observe from the definition of $\mu^*$ that if $\varepsilon > 0$, there is a countable cover $\{V_j\}$ of $A$ by open intervals such that

$$\sum_{j=1}^{\infty} \operatorname{len}(V_j) < \mu^*(A) + \varepsilon.$$

Hence, if $V = \bigcup V_j$, then subadditivity implies

$$\mu^*(V) \leq \sum \mu^*(V_j) \leq \sum \operatorname{len}(V_j) < \mu^*(A) + \varepsilon,$$

so $V$ has the desired property. Moreover, it follows that

$$\inf\{\mu^*(U) \mid U \text{ is open and } A \subset U\} \leq \mu^*(A) + \varepsilon.$$

Since this is true for all $\varepsilon > 0$, we conclude

$$\inf\{\mu^*(U) \mid U \text{ is open and } \ A \subset U\} \leq \mu^*(A).$$

The reverse inequality follows from monotonicity of $\mu^*$ since $\mu^*(A) \leq \mu^*(U)$ for any $U$ containing $A$. Combining the two we obtain the desired equality. □

---

**Exercise B.2.11.**

(1) Prove that if $N$ is a null set and $A \subset \mathbb{R}$ then, $\mu^*(A \cup N) = \mu^*(A)$.

(2) Prove Theorem B.2.9.

(3) Suppose $A \subset \mathbb{R}$ and $r > 0$. Let $rA = \{rx \mid x \in A\}$. Prove $\mu^*(rA) = r\mu^*(A)$.

(4) Prove that $\mu^*(A)$ is also the infimum of $\sum \operatorname{len}(U_n)$ for all countable open covers of $A$ by *pairwise disjoint* open intervals $U_n$.

(5) Prove that given $\varepsilon > 0$ there exist a countable collection of open intervals $U_1, U_2, \ldots, U_n, \ldots$ such that $\bigcup_n U_n$ contains all rational numbers in $\mathbb{R}$ and such that $\sum_{n=1}^{\infty} \operatorname{len}(U_n) = \varepsilon$.

(6) Give an example of a subset $A$ of $I = [0, 1]$ such that $\mu^*(A) = 0$, but with the property that if $U_1, U_2, \ldots, U_n$ is a *finite* cover by open intervals, then $\sum_{i=1}^{n} \operatorname{len}(U_i) \geq 1$.

(7) Suppose $\{U_n\}$ is a countable collection of pairwise disjoint open intervals in $I$ and $U = \bigcup_n U_n$. Prove that $\mu^*(U) = \sum \operatorname{len}(U_n)$.

(8) ($\star$) Outer measure $\mu^*$ is not generally additive. Prove, however, that if $U = \bigcup_{n=1}^{\infty} U_n$ is a countable union of pairwise disjoint open intervals $\{U_n\}$ and $A$ is a bounded subset of $\mathbb{R}$, then

$$\mu^*(A \cap U) = \sum_{n=1}^{\infty} \mu^*(A \cap U_n).$$

## B.3. The $\sigma$-algebra of Lebesgue Measurable Sets

In Definition 2.4.1 we defined the $\sigma$-algebra $\mathcal{M}$ to be the $\sigma$-algebra of subsets of $\mathbb{R}$ generated by open intervals and null sets (it is also the $\sigma$-algebra of subsets of $\mathbb{R}$ generated by Borel sets and null sets). We defined a set to be *Lebesgue measurable* if it is in this $\sigma$-algebra. However, now, in order to prove the existence of Lebesgue measure, we want to use a different, but equivalent definition.

Our program for this section and the next is roughly as follows:

- We will define a collection $\mathcal{M}_0$ of subsets of $\mathbb{R}$. The criterion used to define $\mathcal{M}_0$ is often given as the definition of Lebesgue measurable sets. We will first motivate this criterion.
- We will define the Lebesgue measure $\mu(A)$ of a set $A$ in $\mathcal{M}_0$ to be the outer measure $\mu^*(A)$.
- We will show that the collection $\mathcal{M}_0$ is a $\sigma$-algebra of subsets of $\mathbb{R}$ and, in fact, precisely the $\sigma$-algebra $\mathcal{M}$. Hence, $\mu(A)$ will be defined for all $A \in \mathcal{M}$.
- In the next section we will prove that $\mu$ defined in this way satisfies the properties promised in Chapter 2, namely properties I–VI of Theorem 2.4.2. Several of these properties follow from the corresponding properties for outer measure $\mu^*$, which we proved in Section B.2.

Lebesgue outer measure as in Definition B.2.1 has most of the properties we want for the measure $\mu$. There is one serious problem, however; namely, there exist subsets $A$ and $B$ of $I$ such that $A \cap B = \emptyset$ and $A \cup B = I$, but $\mu^*(A) + \mu^*(B) \neq \mu^*(I)$. That is, the additivity property fails even with two disjoint sets whose union is an interval. Fortunately, the sets for which it fails are rather exotic and not too frequently encountered.

When we developed the theory of the Riemann integral in Chapter 1, we saw that once we specified the obvious definition for the integral of a step function and also required the monotonicity property, there was a large class of functions for which the value of the integral was forced. These were the functions $f$ for which the infimum of the integrals of step functions greater than $f$ is equal to the supremum of the integrals of step functions less than $f$. Because of monotonicity there was only one choice for the value of the integral of $f$ — it had to be this common value of the supremum and infimum. And the class of functions for which this value is forced was *defined* to be the Riemann integrable functions.

To carefully develop the theory of Lebesgue measure and measurable sets we will do something very similar. We will prove that if $A$ is any subset of $\mathbb{R}$ and $J = [a, b]$, then there are upper and lower bounds for all possible ways to define the measure of $A \cap J$, the part of $A$ in $J$. Indeed, if $\mu$ is to be defined on a $\sigma$-algebra containing open and closed sets, then the only possibilities for the value of $\mu(A \cap J)$ must be less than $\mu^*(A \cap J)$ and greater than $(\operatorname{len}(J) - \mu^*(A^c \cap J))$. Should these upper and lower bounds be equal, life is good, and the value of $\mu(A \cap J)$ has been determined! In particular, if $A$ had this property for an interval $J$ which contains $A$, then we will have determined $\mu(A \cap J) = \mu(A)$.

**Proposition B.3.1.** *Suppose $\mathcal{A}$ is a $\sigma$-algebra of subsets of $\mathbb{R}$ which contains all Borel sets and $\mu$ is an extended real-valued function defined on $\mathcal{A}$ which satisfies the Countable additivity, Monotonicity and Length properties. Then for any interval $J = [a, b]$ and any set $A \in \mathcal{A}$,*

$$\operatorname{len}(J) - \mu^*(A^c \cap J) \leq \mu(A \cap J) \leq \mu^*(A \cap J).$$

**Proof.** Suppose $A \in \mathcal{A}$ and $\{U_n\}$ is a countable covering of $A \cap J$ by open intervals. Then

$$\mu(A \cap J) \leq \mu\Big(\bigcup_{n=1}^{\infty} U_n\Big) \leq \sum_{n=1}^{\infty} \mu(U_n) = \sum_{n=1}^{\infty} \operatorname{len}(U_n),$$

where the first inequality follows from monotonicity of $\mu$ and the second from countable subadditivity. Hence, $\mu(A \cap J)$ is less than or equal to the *infimum* of all such sums where we consider all possible countable coverings of $A \cap J$ by open intervals. We conclude that

$$\mu(A \cap J) \leq \mu^*(A \cap J). \tag{B.3.1}$$

Since $A^c \cap J$ is also in the $\sigma$-algebra $\mathcal{A}$ the same argument shows $\mu(A^c \cap J) \leq \mu^*(A^c \cap J)$. Since the function $\mu$ satisfies both the Length property and the additivity property for disjoint sets in $\mathcal{A}$, we must also have $\mu(A \cap J) + \mu(A^c \cap J) = \mu(J) = \text{len}(J)$. So

$$\text{len}(J) - \mu(A \cap J) = \mu(A^c \cap J) \leq \mu^*(A^c \cap J)$$

and it follows that

$$\mu(A \cap J) \geq \text{len}(J) - \mu^*(A^c \cap J). \tag{B.3.2}$$

□

Equations (B.3.1) and (B.3.2) tell us that the only possible values of $\mu(A \cap J)$ must be between $\text{len}(J) - \mu^*(A^c \cap J)$ and $\mu^*(A \cap J)$. If these two values are equal, i.e., if $\mu^*(A \cap J) + \mu^*(A^c \cap J) = \text{len}(J)$, then there is only one choice for the value of $\mu(A \cap J)$. It must be equal to the common value $\mu^*(A \cap J) = (\text{len}(J) - \mu^*(A^c \cap J))$.

This certainly suggests that we consider subsets $A$ of $\mathbb{R}$ such that

$$\mu^*(A \cap J) + \mu^*(A^c \cap J) = \text{len}(J). \tag{B.3.3}$$

It turns out (see Exercise B.3.5 below) that if a set $A$ satisfies equation (B.3.3) for every interval $J$, then, in fact, it satisfies it for *every* subset $X \subset \mathbb{R}$, if we replace $J$ by $X$ and $\text{len}(J)$ by $\mu^*(X)$. So we might as well make things easier for ourselves and ask that equation (B.3.3) hold when we replace $J$ with an arbitrary subset $X \subset \mathbb{R}$ if we use $\mu^*(X)$ in place of $\text{len}(X)$. All this should motivate the following definition.

**Definition B.3.2. (Lebesgue measure).** *Let $\mathcal{M}_0$ denote the collection of all $A \subset \mathbb{R}$ with the property that for* any $X \subset \mathbb{R}$,

$$\mu^*(A \cap X) + \mu^*(A^c \cap X) = \mu^*(X).$$

*For any set $A \in \mathcal{M}_0$ we define its* Lebesgue measure, $\mu(A)$, *to be* $\mu^*(A)$.

Once again we have followed the paradigm of listing the properties we want a function (Lebesgue measure in this case) to have, next finding the class of objects where there is only one possible value of that function, and then defining the function on that class to be those uniquely determined values.

Notice that if $\mu^*(X) = +\infty$, then subadditivity implies

$$\mu^*(A \cap X) + \mu^*(A^c \cap X) = \mu^*(X)$$

no matter what $A$ is since at least one of $\mu^*(A \cap X)$ and $\mu^*(A^c \cap X)$ must be infinite. So if we want to check if a particular set $A$ is in $\mathcal{M}_0$, it suffices to check that $\mu^*(A \cap X) + \mu^*(A^c \cap X) = \mu^*(X)$ for all subsets $X \subset \mathbb{R}$ with $\mu^*(X)$ finite.

The goal of the remainder of this section is to prove that, in fact, $\mathcal{M}_0$ is nothing other than the $\sigma$-algebra $\mathcal{M}$ of Lebesgue measurable subsets of $\mathbb{R}$. In the next section we will show the extended real-valued function $\mu$ satisfies the properties I–VI which we claimed for Lebesgue measure in Chapter 2. The defining condition above for a set $A$ to be in $\mathcal{M}_0$ is often taken as the definition of a Lebesgue measurable set because it is what is needed to prove the properties we want for Lebesgue measure. Since we have already given a different definition of Lebesgue measurable sets in Definition 2.4.1 we will instead prove the properties of $\mathcal{M}_0$ and $\mu$ which we want and then show in Corollary B.3.11 that the sets in $\mathcal{M}_0$ are precisely the sets in $\mathcal{M}$, the $\sigma$-algebra generated by Borel subsets and null sets.

As a first step we show that bounded sets in $\mathcal{M}_0$ are in $\mathcal{M}$. We will later see that unbounded ones are also.

**Proposition B.3.3.** *Every bounded set $A \in \mathcal{M}_0$ can be written as*

$$A = B \setminus N$$

*where $B$ is a Borel set and $N = A^c \cap B$ is a null set.*

**Proof.** It follows from Proposition B.2.10, the regularity of outer measure, that for any $k \in \mathbb{N}$ there is a an open set $V_k$ containing $A$ such that $\mu^*(V_k) \leq \mu^*(A) + 1/k$. Let $B = \bigcap_{k=1}^{\infty} V_k$. By monotonicity we have

$$\mu^*(A) \leq \mu^*(B) \leq \mu^*(V_k) \leq \mu^*(A) + \frac{1}{k}.$$

Since this holds for all $k > 0$ we conclude that $\mu^*(B) = \mu^*(A)$.

In the defining equation of $\mathcal{M}_0$ (see Definition B.3.2) we may take $X = B$ and conclude that

$$\mu^*(A \cap B) + \mu^*(A^c \cap B) = \mu^*(B).$$

Since $A \subset B$ we have $A \cap B = A$ and hence $\mu^*(A) + \mu^*(A^c \cap B) = \mu^*(B)$. From the fact that $\mu^*(B) = \mu^*(A)$ it follows that $\mu^*(A^c \cap B) =$

0. (This uses the fact that $\mu^*(A)$ is finite.) Therefore, if $N = A^c \cap B$, then $N$ is a null set. Finally, $A = B \setminus (B \cap A^c)$, so $A = B \setminus N$. □

The definition of $\mathcal{M}_0$ is relatively simple, but to show it has the properties we want requires some work. We begin, however, with two very easy properties.

**Proposition B.3.4.** *Suppose $A \subset \mathbb{R}$, then*

(1) *The set $A$ is in $\mathcal{M}_0$ provided for* every *subset $X \subset \mathbb{R}$,*

$$\mu^*(A \cap X) + \mu^*(A^c \cap X) \leq \mu^*(X).$$

(2) *The set $A$ is in $\mathcal{M}_0$ if and only if $A^c$ is in $\mathcal{M}_0$.*

**Proof.** For part (1) observe $X = (A \cap X) \cup (A^c \cap X)$ so the subadditivity property of outer measure in Theorem B.2.6 tells us that

$$\mu^*(A \cap X) + \mu^*(A^c \cap X) \geq \mu^*(X)$$

always holds. This plus the inequality of our hypothesis gives the equality of the definition of $\mathcal{M}_0$.

For part (2) suppose $A$ is an arbitrary subset of $\mathbb{R}$. The fact that $(A^c)^c = A$ implies immediately from Definition B.3.2 that $A$ is in $\mathcal{M}_0$ if and only if $A^c$ is. □

---

**Exercise B.3.5.** ($\star$) Let $\mathcal{S}$ denote the collection of all subsets $A$ of $\mathbb{R}$ such that

$$\mu^*(A \cap J) + \mu^*(A^c \cap J) = \mu^*(J)$$

for every interval $J = [a, b]$ in $\mathbb{R}$.

(1) Prove that if $A \in \mathcal{S}$ and $U$ is a bounded open subset of $\mathbb{R}$, then

$$\mu^*(A \cap U) + \mu^*(A^c \cap U) = \mu^*(U).$$

*Hint:* Use part (8) of Exercise B.2.11

(2) Use (1) to prove that if $A \in \mathcal{S}$ and $X$ is any bounded subset of $\mathbb{R}$, then

$$\mu^*(A \cap X) + \mu^*(A^c \cap X) = \mu^*(X).$$

**Proposition B.3.6.** *A set $A \subset \mathbb{R}$ is a null set if and only if $A \in \mathcal{M}_0$ and $\mu(A) = 0$.*

**Proof.** Let $X$ be a subset of $\mathbb{R}$. By definition a set $A$ is a null set if and only if $\mu^*(A) = 0$. If $A$ is a null set, then since $A \cap X \subset A$ we know by the monotonicity of outer measure (Proposition B.2.4) that $\mu^*(A \cap X) = 0$. Similarly, since $A^c \cap X \subset X$ we know that $\mu^*(A^c \cap X) \leq \mu^*(X)$. Hence, again using monotonicity of outer measure from Proposition B.2.4, we know that

$$\mu^*(A \cap X) + \mu^*(A^c \cap X) = \mu^*(A^c \cap X) \leq \mu^*(X)$$

and the fact that $A \in \mathcal{M}_0$ follows from part (1) of Proposition B.3.4. $\square$

**Proposition B.3.7.** *If $A$ and $B$ are in $\mathcal{M}_0$, then $A \cup B$ and $A \cap B$ are in $\mathcal{M}_0$.*

**Proof.** To prove the union of two sets, $A$ and $B$, in $\mathcal{M}_0$ is also in $\mathcal{M}_0$ requires some work. Suppose $X \subset I$. Since $(A \cup B) \cap X = (B \cap X) \cup (A \cap B^c \cap X)$, the subadditivity of Theorem B.2.6 tells us

$$\mu^*((A \cup B) \cap X) \leq \mu^*(B \cap X) + \mu^*(A \cap B^c \cap X). \tag{B.3.4}$$

Also, the definition of $\mathcal{M}_0$ tells us

$$\mu^*(B^c \cap X) = \mu^*(A \cap B^c \cap X) + \mu^*(A^c \cap B^c \cap X). \tag{B.3.5}$$

Notice that $(A \cup B)^c = A^c \cap B^c$. So we get

$$\begin{aligned} &\mu^*((A \cup B) \cap X) + \mu^*((A \cup B)^c \cap X) \\ &\qquad = \mu^*((A \cup B) \cap X) + \mu^*(A^c \cap B^c \cap X). \end{aligned}$$

Using the inequality from equation (B.3.4) we get

$$\begin{aligned} &\mu^*((A \cup B) \cap X) + \mu^*((A \cup B)^c \cap X) \\ &\qquad \leq \mu^*(B \cap X) + \mu^*(A \cap B^c \cap X) + \mu^*(A^c \cap B^c \cap X) \\ &\qquad = \mu^*(B \cap X) + \mu^*(B^c \cap X) = \mu^*(X) \end{aligned}$$

where the first equality comes from equation (B.3.5) and the second follows from the fact that $B \in \mathcal{M}_0$.

Hence, we conclude

$$\mu^*((A \cup B) \cap X) + \mu^*((A \cup B)^c \cap X) \leq \mu^*(X)$$

and according to part (1) of Proposition B.3.4 this is sufficient to show that $A \cup B$ is in $\mathcal{M}_0$.

The intersection now follows easily using what we know about the union and complement. More precisely, $A \cap B = (A^c \cup B^c)^c$, so if $A$ and $B$ are in $\mathcal{M}_0$, then so is $(A^c \cup B^c)$ and hence its complement $(A^c \cup B^c)^c$ is also. □

Next we wish to show intervals are in $\mathcal{M}_0$.

**Proposition B.3.8.** *Any interval, open, closed or half open, is in* $\mathcal{M}_0$.

**Proof.** First consider $(-\infty, a]$ with complement $(a, \infty)$. If $X$ is an arbitrary subset of $\mathbb{R}$, we must show $\mu^*((-\infty, a] \cap X) + \mu^*((a, \infty) \cap X) = \mu^*(X)$. Let $X^- = (-\infty, a] \cap X$ and $X^+ = (a, \infty) \cap X$. Given $\varepsilon > 0$, the definition of outer measure tells us we can find a countable cover of $X$ by open intervals $\{U_n\}_{n=1}^\infty$ such that

$$\sum_{n=1}^{\infty} \operatorname{len}(U_n) \leq \mu^*(X) + \varepsilon. \tag{B.3.6}$$

Let $U_n^- = U_n \cap (-\infty, a]$ and $U_n^+ = U_n \cap (a, \infty)$. Then $X^- \subset \bigcup_{n=1}^\infty U_n^-$ and $X^+ \subset \bigcup_{n=1}^\infty U_n^+$. Note that $U_n^+$ is an open interval and $U_n^-$ is either an open interval or a half open interval. Subadditivity of outer measure implies

$$\begin{aligned}\mu^*(X^-) &\leq \mu^*\Big(\bigcup_{n=1}^{\infty} U_n^-\Big) \\ &\leq \sum_{n=1}^{\infty} \mu^*(U_n^-) = \sum_{n=1}^{\infty} \operatorname{len}(U_n^-)\end{aligned}$$

and

$$\begin{aligned}\mu^*(X^+) &\leq \mu^*\Big(\bigcup_{n=1}^{\infty} U_n^+\Big) \\ &\leq \sum_{n=1}^{\infty} \mu^*(U_n^+) = \sum_{n=1}^{\infty} \operatorname{len}(U_n^+).\end{aligned}$$

Adding these inequalities and using equation (B.3.6) we get

$$\begin{aligned}\mu^*(X^-) + \mu^*(X^+) &\leq \sum_{n=1}^{\infty} \operatorname{len}(U_n^-) + \operatorname{len}(U_n^+) \\ &= \sum_{n=1}^{\infty} \operatorname{len}(U_n) \\ &\leq \mu^*(X) + \varepsilon.\end{aligned}$$

Since $\varepsilon$ is arbitrary we conclude that $\mu^*(X^-) + \mu^*(X^+) \leq \mu^*(X)$ which by Proposition B.3.4 implies that $(-\infty, a]$ is in $\mathcal{M}_0$ for any $a \in \mathbb{R}$. A similar argument implies that $[a, \infty)$ is in $\mathcal{M}_0$. Taking complements, unions and intersections it is clear that any interval, open closed or half open, is in $\mathcal{M}_0$. $\square$

**Lemma B.3.9.** *Suppose $A$ and $B$ are disjoint sets in $\mathcal{M}_0$. Then*

$$\mu^*(A \cup B) = \mu^*(A) + \mu^*(B).$$

*The analogous result for a finite union of disjoint measurable sets is also valid.*

**Proof.** It is always true that

$$A \cap (A \cup B) = A.$$

Since $A$ and $B$ are disjoint

$$A^c \cap (A \cup B) = B.$$

Hence, the fact that $A$ is in $\mathcal{M}_0$ (using $A \cup B$ for $X$) tells us

$$\begin{aligned}\mu^*(A \cup B) &= \mu^*(A \cap (A \cup B)) + \mu^*(A^c \cap (A \cup B)) \\ &= \mu^*(A) + \mu^*(B).\end{aligned}$$

The result for a finite collection $A_1, A_2, \ldots, A_n$ follows immediately by induction on $n$. $\square$

**Theorem B.3.10.** *The collection $\mathcal{M}_0$ of subsets of $I$ is closed under countable unions and countable intersections and taking complements. Hence, $\mathcal{M}_0$ is a $\sigma$-algebra.*

**Proof.** We have already observed in part (2) of Proposition B.3.4 that the complement of a set in $\mathcal{M}_0$ is a set in $\mathcal{M}_0$.

We have also shown that the union or intersection of a finite collection of sets in $\mathcal{M}_0$ is a set in $\mathcal{M}_0$.

Suppose $\{A_n\}_{n=1}^{\infty}$ is a countable collection of sets in $\mathcal{M}_0$. We want to construct a countable collection of pairwise disjoint sets $\{B_n\}_{n=1}^{\infty}$ which are in $\mathcal{M}_0$ and have the same union.

To do this we define $B_1 = A_1$ and

$$B_{n+1} = A_{n+1} \setminus \bigcup_{i=1}^{n} A_n = A_{n+1} \cap \Big( \bigcup_{i=1}^{n} A_n \Big)^c.$$

Since finite unions, intersections and complements of sets in $\mathcal{M}_0$ are sets in $\mathcal{M}_0$, it is clear that $B_n$ is in $\mathcal{M}_0$. Also, it follows easily by induction that $\bigcup_{i=1}^{n} B_i = \bigcup_{i=1}^{n} A_i$ for any $n$. Thus, $\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i$.

Hence to prove $\bigcup_{i=1}^{\infty} A_i$ is in $\mathcal{M}_0$ we will prove that $\bigcup_{i=1}^{\infty} B_i$ is in $\mathcal{M}_0$. Let $F_n = \bigcup_{i=1}^{n} B_i$ and $F = \bigcup_{i=1}^{\infty} B_i$. If $X$ is an arbitrary subset of $\mathbb{R}$, then since $F_n$ is in $\mathcal{M}_0$ and $F^c \subset F_n^c$

$$\mu^*(X) = \mu^*(F_n \cap X) + \mu^*(F_n^c \cap X) \geq \mu^*(F_n \cap X) + \mu^*(F^c \cap X).$$

By Lemma B.3.9

$$\mu^*(F_n \cap X) = \sum_{i=1}^{n} \mu^*(B_i \cap X).$$

Putting these together we have

$$\mu^*(X) \geq \sum_{i=1}^{n} \mu^*(B_i \cap X) + \mu^*(F^c \cap X)$$

for all $n > 0$. Hence,

$$\mu^*(X) \geq \sum_{i=1}^{\infty} \mu^*(B_i \cap X) + \mu^*(F^c \cap X).$$

But subadditivity of $\mu^*$ implies

$$\sum_{i=1}^{\infty} \mu^*(B_i \cap X) \geq \mu^*(\bigcup_{i=1}^{\infty} (B_i \cap X)) = \mu^*(F \cap X).$$

Hence,

$$\mu^*(X) \geq \mu^*(F \cap X) + \mu^*(F^c \cap X)$$

and $F$ is in $\mathcal{M}_0$ by Proposition B.3.4.

To see that a countable intersection of sets in $\mathcal{M}_0$ is in $\mathcal{M}_0$ we observe that

$$\bigcap_{n=1}^{\infty} A_n = \Big( \bigcup_{n=1}^{\infty} A_n^c \Big)^c$$

so the desired result follows from the result on unions together with the fact that $\mathcal{M}_0$ is closed under taking complements. ☐

**Corollary B.3.11.** *The $\sigma$-algebra $\mathcal{M}_0$ equals $\mathcal{M}$ the $\sigma$-algebra generated by Borel sets and null sets.*

**Proof.** In Proposition B.3.8 we showed the $\sigma$-algebra $\mathcal{M}_0$ contains open intervals and hence it contains the $\sigma$-algebra they generate, the Borel subsets of $\mathbb{R}$. Also, $\mathcal{M}_0$ contains null sets by Proposition B.3.6. Therefore, $\mathcal{M}_0$ contains $\mathcal{M}$ the $\sigma$-algebra generated by Borel sets and null sets.

On the other hand, by Proposition B.3.3 bounded sets which are in $\mathcal{M}_0$ are also in $\mathcal{M}$. Hence, if $A \in \mathcal{M}_0$, then $A \cap [n, n+1] \in \mathcal{M}_0$, so $A \cap [n, n+1] \in \mathcal{M}$ and it follows that $A = \bigcup (A \cap [n, n+1])$ is in $\mathcal{M}$. Hence $\mathcal{M}_0 = \mathcal{M}$. ☐

## B.4. The Existence of Lebesgue Measure

Since we now know the sets in $\mathcal{M}_0$, i.e., the sets which satisfy Definition B.3.2, coincide with the sets in $\mathcal{M}$, we will refer to them as Lebesgue measurable sets, or simply measurable sets for short. We also no longer need to use outer measure, but can refer to the Lebesgue measure $\mu(A)$ of a measurable set $A$ (which, of course, has the same value as the outer measure $\mu^*(A)$).

**Theorem B.4.1. (Countable additivity).** *If $\{A_n\}_{n=1}^{\infty}$ is a countable collection of measurable subsets of $\mathbb{R}$, then*

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

*If the sets are pairwise disjoint, then*

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

**Proof.** The first inequality is simply a special case of the subadditivity from Theorem B.2.6. If the sets $A_i$ are pairwise disjoint, then by Lemma B.3.9 we know that for each $n$,

$$\mu(\bigcup_{i=1}^{\infty} A_i) \geq \mu(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mu(A_i).$$

Hence,

$$\mu(\bigcup_{i=1}^{\infty} A_i) \geq \sum_{i=1}^{\infty} \mu(A_i).$$

Since the reverse inequality follows from subadditivity we have equality. □

We can now prove the main result of this appendix, which was presented as Theorem 2.4.2 in Chapter 2. In that chapter we restricted our attention to measurable sets in an interval $I$. This was done to gain the simplicity of dealing only with sets of finite measure, because that was sufficient for our purposes. However, we have defined extended real-valued functions (see Definition 3.1.5) and can therefore prove the existence and uniqueness of the extended real-valued function $\mu$ defined on all of $\mathcal{M}$.

**Theorem B.4.2. (Existence of Lebesgue measure).** *There exists a unique function $\mu$, called Lebesgue measure, from $\mathcal{M}$ to the extended real numbers satisfying:*

- **I. Length:** *If $A = (a, b)$, then $\mu(A) = \operatorname{len}(A) = b - a$, i.e., the measure of an open interval is its length*
- **II. Translation Invariance:** *Suppose $A \in \mathcal{M}$, and $c \in \mathbb{R}$, then $\mu(A+c) = \mu(A)$ where $A+c$ is the set $\{x+c \mid x \in A\}$.*
- **III. Countable additivity:** *If $\{A_n\}_{n=1}^{\infty}$ is a countable collection of elements of $\mathcal{M}$, then*

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

*and if the sets are* pairwise disjoint, *then*

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

**IV. Monotonicity:** *If $A, B \in \mathcal{M}$ and $A \subset B$, then $\mu(A) \leq \mu(B)$*

**V. Null sets:** *A subset $A \subset \mathbb{R}$ is a null set if and only if $A \in \mathcal{M}(I)$ and $\mu(A) = 0$.*

**VI. Regularity:** *If $A \in \mathcal{M}$, then for any $\varepsilon > 0$ there is an open set $V$ containing $A$ such that $\mu(V \setminus A) < \varepsilon$. In particular,*

$$\mu(A) = \inf\{\mu(U) \mid U \textit{ is open and } A \subset U\}.$$

**Proof.** The Lebesgue measure, $\mu(A)$, of any set $A \in \mathcal{M}$ is defined to be its outer measure $\mu^*(A)$. Hence, properties I, II, and IV for $\mu$ follow from the corresponding properties of $\mu^*$. These were established in Propostion B.2.2, Theorem B.2.9, and Proposition B.2.4, respectively.

Property III, countable additivity, was proved in Theorem B.4.1, and Property V is a consequence of Proposition B.3.6.

Property VI is an immediate consequence of Proposition B.2.10 (regularity of outer measure) when $\mu(A)$ is finite, because that proposition asserts $\mu(V) < \mu(A) + \varepsilon$. Since $\mu(V \setminus A) = \mu(V) - \mu(A)$ we conclude $\mu(V \setminus A) < \varepsilon$. The case when $\mu(A)$ is infinite is Exercise B.4.3 below.

We are left with the task of showing that $\mu$ is unique. Suppose $\mu_1$ and $\mu_2$ are two functions defined on $\mathcal{M}$ and satisfying properties I–VI. They must agree on any open interval by property I. By Theorem A.6.3 any open set is a countable union of pairwise disjoint open intervals, so countable additivity implies $\mu_1$ and $\mu_2$ agree on open sets.

Finally, if $A$ is an arbitrary set in $\mathcal{M}$, then, since $\mu_1$ and $\mu_2$ agree on any open set $U$ containing $A$, property VI (regularity) implies they agree on $A$. $\square$

**Exercise B.4.3.**

(1) Prove that a measure which satisfies all properties from Theorem B.4.2 except II, must also satisfy property II.

(2) Prove that if $\nu$ is a measure which satisfies properties II–VI of Theorem B.4.2, then there is a $c \in [0, \infty)$ such that $\nu = c\mu$.

(3) If $A \subset \mathbb{R}$ is measurable and $\mu(A)$ is infinite prove that for any $\varepsilon > 0$ there is an open set $V$ containing $A$ such that $\mu(V \setminus A) < \varepsilon$.

Appendix C

# A Non-measurable Set

We are now prepared to prove the existence of a *non-measurable set*. The proof (necessarily) depends on the *Axiom of choice* (see the end of Section A.5).

**Theorem C.1.1. (Non-measurable set).** *There exists a subset $E$ of $[0,1]$ which is not Lebesgue measurable.*

**Proof.** Let $\mathbb{Q} \subset \mathbb{R}$ denote the rational numbers. The rationals are an additive subgroup of $\mathbb{R}$ and we wish to consider the so-called "cosets" of this subgroup. More precisely, we want to consider the sets of the form $\mathbb{Q} + x$ where $x \in \mathbb{R}$.

We observe that two such sets $\mathbb{Q} + x_1$ and $\mathbb{Q} + x_2$ are either equal or disjoint. This is because the existence of one point in their intersection,

$$z \in (\mathbb{Q} + x_1) \cap (\mathbb{Q} + x_2),$$

implies $z = x_1 + r_1 = x_2 + r_2$ with $r_1, r_2 \in \mathbb{Q}$, so $x_1 - x_2 = (r_2 - r_1) \in \mathbb{Q}$. This, in turn implies that

$$\begin{aligned}
\mathbb{Q} + x_2 &= \{x_2 + r \mid r \in \mathbb{Q}\} \\
&= \{x_2 + r + (x_1 - x_2) \mid r \in \mathbb{Q}\} \\
&= \{x_1 + r \mid r \in \mathbb{Q}\} \\
&= \mathbb{Q} + x_1.
\end{aligned}$$

Notice that since $\mathbb{Q}$ is dense in $\mathbb{R}$ so is each coset $\mathbb{Q}+x$. Hence, each coset has a non-empty intersection with the interval $[0,1]$. According to the Axiom of Choice, there is a choice function $\Psi : \mathcal{P}(\mathbb{R}) \to \mathbb{R}$, i.e., a function such that $\Psi(A)$ is an element of $A$ for every non-empty subset $A \subset \mathbb{R}$. We define

$$E = \big\{ y \mid y = \Psi\big([0,1] \cap (\mathbb{Q}+x)\big) \text{ for some } x \in \mathbb{R} \big\}.$$

The set $E$ is a subset of $[0,1]$ and contains one element from each of the cosets $\mathbb{Q}+x$. That is, for any $x_0 \in \mathbb{R}$ the set $E \cap (\mathbb{Q}+x_0)$ contains exactly one point and that point is in $[0,1]$. Now let $\{r_n\}_{n=1}^{\infty}$ be an enumeration of the rational numbers.

We want to show that $\{E+r_n\}_{n=1}^{\infty}$ defines a partition of $\mathbb{R}$. That is, the sets $(E+r_n)$ are pairwise disjoint and

$$\mathbb{R} = \bigcup_{n=1}^{\infty} (E+r_n).$$

To see this first note that if $y \in (E+r_n) \cap (E+r_m)$, then there are $e_1, e_2 \in E$ such that $e_1 + r_n = e_2 + r_m$ and hence $(e_1 - e_2) \in \mathbb{Q}$, so $e_1$ and $e_2$ are in the same coset of $\mathbb{Q}$. Since $E$ contains only one element from each coset we conclude $e_1 = e_2$ which implies $r_n = r_m$. Thus, the sets $E+r_n$ and $E+r_m$ are disjoint unless $r_n = r_m$.

Now let $x \in \mathbb{R}$ be arbitrary and let $\{x_0\} = E \cap (\mathbb{Q}+x)$. Then $x = x_0 + r$ for some $r \in \mathbb{Q}$. Hence, $x \in E+r$, so $x \in \bigcup_{n=1}^{\infty} E+r_n$. We have shown that $\mathbb{R} = \bigcup_{n=1}^{\infty}(E+r_n)$.

Note that $E \subset [0,1]$, so the outer measure $\mu^*(E) \leq 1$. It is not possible that $\mu^*(E) = 0$ since translation invariance of outer measure would imply $\mu^*(E+r_n) = \mu^*(E) = 0$ and hence (by subadditivity) that

$$\mu^*(\mathbb{R}) = \mu^*\Big( \bigcup_{n=1}^{\infty} E+r_n \Big) \leq \sum_{n=1}^{\infty} \mu^*(E+r_n) = 0.$$

We conclude that $\mu^*(E) > 0$.

Let $\{s_n\}_{n=1}^{N}$ be a set of $N$ distinct rational numbers in $[0,1]$. Each set $E+s_n$ is a subset of $[0,2]$, so

$$\mu^*\Big( \bigcup_{n=1}^{N} E+s_n \Big) \leq \mu^*([0,2]).$$

Since $\mu^*$ equals $\mu$ on measurable sets and $\mu$ is additive, if $E$ were measurable, we could conclude

$$\begin{aligned}\mu^*\Big(\bigcup_{n=1}^{N}(E+s_n)\Big) &= \mu\Big(\bigcup_{n=1}^{N}(E+s_n)\Big)\\ &= \sum_{n=1}^{N}\mu(E+s_n)\\ &= N\mu(E) = N\mu^*(E).\end{aligned}$$

Since $\mu^*(E) > 0$ it is clearly impossible to have $N\mu^*(E) \leq \mu^*([0,2]) = 2$ for all $N \in \mathbb{N}$. Hence, we conclude that $E$ is not measurable. $\square$

---

**Exercise C.1.2.**

(1) Prove that there exist subsets $A$ and $B$ of $I = [0,1]$ such that $A \cap B = \emptyset$ and $A \cup B = I$, but $\mu^*(A) + \mu^*(B) \neq 1$.

(2) Prove that if $r \in (0,1)$, there is a non-measurable subset $E \subset I$ such that $\mu^*(E) = r$. *Hint:* Use part (3) of Exercise B.2.11.

(3) Prove there is a subset $A \subset \mathbb{R}$ with the property that for every open set $U$ containing $A$ the outer measure of the set difference $\mu^*(U \setminus A)$ is infinite. (Cf. part (3) of Exercise B.4.3.)

# Bibliography


[C] Lennart Carleson, *On convergence and growth of partial sumas of Fourier series*, Acta Math. **116** (1966), 135–157.

[D] J. Dieudonné, *Foundations of Modern Analysis*, Academic Press, New York (1960).

[H] Paul Halmos, *Measure Theory*, D. Van Nostrand, Princeton, NJ (1950).

[J] Roger L. Jones, *New Proofs for the Maximal Ergodic Theorem and the Hardy-Littlewood Maximal Theorem*, Proc. of the Amer. Math. Soc., **87**, No. 4 (1983), 681-684.

[K] Yitzhak Katznelson and Yonatan Katznelson, *A (Terse) Introducton ot Linear Algebra*, Student Mathematical Library **44**, American Mathematical Society, Providence, RI (2008).

[L] Serge Lang, *Analysis II*, Addison Wesley, Reading, MA (1969).

[M] Ricardo Mañé, *Erdodic Theory and Differentiable Dynamics*, Ergebnisse der Mathematik und ihrer Grenzgebiete, 3. Folge, Bd. 8, Springer-Verlag, Berlin (1987).

[P] Mark Pinsky, *Introduction to Fourier Analysis and Wavelets*, Brooks Cole, Pacific Grove, CA (2002).

[Ro] H. L. Royden, *Real Analysis*, Third ed., Prentice Hall, Upper Saddle River, NJ (1988).

[Ru] Walter Rudin, *Real and Complex Analysis*, McGraw-Hill, New York, Third Edition (1987).

[T] S. J. Taylor, *Introduction to Measure and Integration*, Cambridge University Press, London (1966).

[Z] A. C. Zaanen, *An Introduction to the Theory of Integration*, North-Holland, New York (1961).

# Index

This book provides a student's first encounter with the concepts of measure theory and functional analysis. Its structure and content reflect the belief that difficult concepts should be introduced in their simplest and most concrete forms.

Despite the use of the word "terse" in the title, this text might also have been called *A (Gentle) Introduction to Lebesgue Integration*. It is terse in the sense that it treats only a subset of those concepts typically found in a substantial graduate-level analysis course. The book emphasizes the motivation of these concepts and attempts to treat them simply and concretely. In particular, little mention is made of general measures other than Lebesgue until the final chapter and attention is limited to $\mathbb{R}$ as opposed to $\mathbb{R}^n$.

After establishing the primary ideas and results, the text moves on to some applications. Chapter 6 discusses classical real and complex Fourier series for $L^2$ functions on the interval and shows that the Fourier series of an $L^2$ function converges in $L^2$ to that function. Chapter 7 introduces some concepts from measurable dynamics. The Birkhoff ergodic theorem is stated without proof and results on Fourier series from Chapter 6 are used to prove that an irrational rotation of the circle is ergodic and that the squaring map on the complex numbers of modulus 1 is ergodic.

This book is suitable for an advanced undergraduate course or for the start of a graduate course. The text presupposes that the student has had a standard undergraduate course in real analysis.

ISBN 978-0-8218-4862-3

9 780821 848623

STML/48

For additional information and updates on this book, visit

**www.ams.org/bookpages/stml-48**