Notes on Measure and Integration

John Franks
Department of Mathematics
Northwestern University

Preface

This text grew out of notes I have used in teaching a one quarter course on integration at the advanced undergraduate level. My intent is to introduce the Lebesgue integral in a quick, and hopefully painless, 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.

The actual construction of Lebesgue measure and proofs of its key properties are relegated to 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 σ-algebra. If a generalization of length has these properties then to make sense it should be defined on a σ-algebra.

The text introduces null sets (sets of measure zero) and shows that any generalization of length satisfying monotonicity must assign zero to them. We then *define* Lebesgue measurable sets to be sets in the σ-algebra generated by Borel 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 μ 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 σ-algebra generated by Borel sets and null sets.

The text presupposes a background which a student obtain from an undergraduate course in real analysis. Chapter 0 summarizes these prerequisites with many proofs and some references. Chapter 1 gives a brief treatment of the "regulated integral" (as found in Dieudonné [1]) and the Riemann integral in a way that permits drawing parallels with the presentation of the Lebesgue integral in subsequent chapters. Chapter 2 introduces Lebesgue measure in the way described above.

Chapter 3 discusses bounded Lebesgue measurable functions and their Lebesgue integral, while Chapter 4 considers unbounded functions and some of the standard convergence theorems. In Chapter 5 we consider the Hilbert space of L^2 functions on $[-1, 1]$ and show several elementary properties leading up to a definition of Fourier series.

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

Contents

Chapter 0

Background and Foundations

0.1 The Completeness of $\mathbb{R}$

This chapter gives a very terse summary of the properties of the real numbers which we will use 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. *We will denote the set of* real numbers *by* $\mathbb{R}$*, the* rational numbers *by* $\mathbb{Q}$*, the* integers *by* $\mathbb{Z}$ *and the* natural numbers *by* $\mathbb{N}$.

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 0.1.1. 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 0.1.2 (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 β is called the* least upper bound *or* supremum *of the set A if β 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 α is called the* greatest lower bound *or* infimum *of the set A if α is a lower bound and $\alpha \geq a$ for every lower bound a of A.*

Theorem 0.1.3. *If a non-empty set $A \subset \mathbb{R}$ has an upper bound, then it has a unique supremum β. If A has a lower bound, then it has a unique infimum α.*

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 β such that $x \leq \beta \leq y$ for every $x \in A$ and every $y \in B$. This implies that β is an upper bound of A and that $\beta \leq y$ for every upper bound y. Hence β is a supremum or least upper bound of A. It is unique, because any β' with the same properties must satisfy $\beta \leq \beta'$ (since β is a least upper bound) and $\beta' \leq \beta$ (since β' 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 $\inf A$.

Proposition 0.1.4. *If A has an upper bound and $\beta = \sup A$, then for any $\epsilon > 0$ there is an $x \in A$ with $\beta - \epsilon < x \leq \beta$. Moreover β 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 $x \in (\beta - \epsilon, \beta)$, then every $x \in A$ satisfies $x \leq \beta - \epsilon$. It follows that $\beta - \epsilon$ is an upper bound for A and is smaller than β contradicting the definition of β as the least upper bound. Hence there must be an $x \in A$ with $x \in (\beta - \epsilon, \beta)$.

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 β and hence β would not be an upper bound for A.

The proof for the *infimum* is similar. □

0.2 Sequences in $\mathbb{R}$

There are a number of equivalent formulations we could have chosen for the Completeness Axiom. For example, we could have take Theorem (0.1.3) 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 (0.2.2), Corollary (0.2.3), and Theorem (0.2.5).

We recall the definition of limit of a sequence.

Definition 0.2.1. *Suppose $\{x_n\}_{n=1}^{\infty}$ is a sequence in $\mathbb{R}$ and $L \in \mathbb{R}$. We say*

$$\lim_{n\to\infty} x_n = L$$

provided for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$|x_m - L| < \epsilon$$

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 0.2.2. *If $\{x_n\}_{n=1}^{\infty}$ is a bounded* monotone *sequence then $\lim_{n\to\infty} x_n$ exists.*

Proof. If $\{x_n\}_{n=1}^{\infty}$ is a bounded monotone increasing sequence, let $L = \sup\{x_n\}_{n=1}^{\infty}$. Given any $\epsilon > 0$ there is an N such that $L - \epsilon < x_N \leq L$ by Proposition (0.1.4). For any $n > N$ we have $x_N \leq x_n \leq L$ and hence $|L - x_n| < \epsilon$. Thus $\lim_{n\to\infty} x_n = L$.

If $\{x_n\}_{n=1}^{\infty}$ is a monotone decreasing sequence, then $\{-x_n\}_{n=1}^{\infty}$ is increasing and $\lim_{n\to\infty} x_n = -\lim_{n\to\infty} -x_n$. □

Corollary 0.2.3. *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. We will denote them by $\limsup_{n\to\infty} x_n$ and $\liminf_{n\to\infty} x_n$ respectively. The sequence $\{x_n\}_{n=1}^{\infty}$ has limit L, i.e., $\lim x_n = L$, if and only if

$$\liminf_{n\to\infty} x_n = \limsup_{n\to\infty} x_n = L.$$

Proof. If $y_m = \sup\{x_n\}_{n=m}^{\infty}$, then $\{y_m\}_{m=1}^{\infty}$ is a monotone decreasing sequence, so $\lim_{m\to\infty} y_m$ 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 $\lim_{n\to\infty} x_n$ exists and equals L. □

Definition 0.2.4 (Cauchy Sequence)**.** *A sequence* $\{x_n\}_{n=1}^{\infty}$ *is called a* Cauchy sequence *if for every* $\epsilon > 0$ *there is an* $N > 0$ *(depending on* ϵ*) such that* $|x_n - x_m| < \epsilon$ *for all* $n, m \geq N$*.*

Theorem 0.2.5 (Cauchy Sequences Have Limits)**.** *If* $\{x_n\}_{n=1}^{\infty}$ *is a Cauchy sequence, then* $\lim_{n\to\infty} x_n$ *exists.*

Proof. First we show that if $\{x_n\}_{n=1}^{\infty}$ is a Cauchy sequence, then it is bounded. For $\epsilon = 1$ 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 $\limsup_{n\to\infty} x_n$ exists.

Since the sequence is Cauchy, given $\epsilon > 0$ there is an N such that that $|x_n - x_m| < \epsilon/2$ for all $n, m \geq N$. Let

$$L = \limsup_{n\to\infty} x_n = \lim_{n\to\infty} \sup\{x_m\}_{m=n}^{\infty}.$$

Hence by Proposition (0.1.4) there is an $M \geq N$ such that $|x_M - L| < \epsilon/2$. It follows that for any $n > M$ we have $|x_n - L| \leq |x_n - x_M| + |x_M - L| < \epsilon/2 + \epsilon/2 = \epsilon$. So $\lim_{n\to\infty} x_n = L$. ☐

Definition 0.2.6 (Convergent and Absolutely Convergent)**.** *An infinite series* $\sum_{n=1}^{\infty} x_n$ *of real numbers is said to* converge *provided the sequence* $\{S_m\}_{m=1}^{\infty}$ *converges where* $S_m = \sum_{n=1}^{m} x_n$*. It is said to* converge absolutely *provided the series* $\sum_{n=1}^{\infty} |x_n|$ *converges.*

Theorem 0.2.7 (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 $\lim_{m\to\infty} S_m$ exists. We will do this by showing it is a Cauchy sequence. Since the series $\sum_{i=1}^{\infty} |x_i|$ converges, given $\epsilon > 0$, there is an $N > 0$ such that $\sum_{i=N}^{\infty} |x_i| < \epsilon$. Hence if $m > n \geq N$

$$|S_m - S_n| = \Big| \sum_{i=n+1}^{m} x_i \Big| \leq \sum_{i=n+1}^{m} |x_i| \leq \sum_{i=N}^{\infty} |x_i| < \epsilon.$$

Hence $\{S_n\}$ is a Cauchy sequence and converges. ☐

0.3 Set Theory and Countability

Proposition 0.3.1 (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*

$$A \cap \bigcup_{j \in J} B_j = \bigcup_{j \in J} (A \cap B_j) \text{ and } A \cup \bigcap_{j \in J} B_j = \bigcap_{j \in J} (A \cup B_j).$$

The proof which is straightforward is left to the reader.

Definition 0.3.2 (Set Difference, Complement)**.** *We define the* set difference *of sets A and B by*

$$A \setminus B = \{x \mid x \in A \text{ and } x \notin B\}.$$

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 0.3.3. *If for each j in some index set J there is a set $B_j \subset E$, then*

$$\bigcap_{j \in J} B_j^c = \Big(\bigcup_{j \in J} B_j\Big)^c \text{ and } \bigcup_{j \in J} B_j^c = \Big(\bigcap_{j \in J} B_j\Big)^c.$$

Again the elementary proof is left to the reader.

Proposition 0.3.4 (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 $B = \{n \in \mathbb{N} \mid n < a \text{ for all } a \in A\}$. If $1 \notin B$, then $1 \in A$ and it is the least element. Otherwise $1 \in B$ so $B \neq \emptyset$. Let b be the greatest element of the finite set B. The element $a_0 = b+1$ is in A and is its least element. □

Definition 0.3.5 (Injection, Surjection)**.** *Suppose A and B are sets and $\phi : A \to B$ is a function. Then*

(1) The function ϕ is called injective *(or* one-to-one*) if $\phi(x) = \phi(y)$ implies $x = y$.*

(2) The function ϕ is called surjective *(or* onto*) if for every $b \in B$ there exists $x \in A$ such that $\phi(x) = b$.*

(3) *The function* ϕ *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* $\phi^{-1}(\{c\})$.

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 0.3.6 (Countable). *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.

The following are standard properties of countable sets which we will need.

Proposition 0.3.7 (Countable Sets).

(1) *If* A *is countable, then any non-empty subset of* A *is countable.*

(2) *A set* A *is countable if and only if there is a surjective function* $f : \mathbb{N} \to A$.

Proof. Item (1) is trivial if A is finite. Hence in proving it we may assume there is a bijection from A to $\mathbb{N}$, and indeed, without loss of generality, we may assume A in fact equals $\mathbb{N}$.

To prove (1) suppose B is a non-empty subset of $A = \mathbb{N}$. 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 ϕ 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 ϕ is surjective.

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 $f : \mathbb{N} \to A$. □

Proposition 0.3.8 (Products and Unions of Countable Sets). *If* A *and* B *are countable, then their Cartesian product* $A \times B = \{(a, b) \mid a \in A,\ b \in B\}$ *is a countable set. If* A_n *is countable for each* $n \in \mathbb{N}$ *then* $\bigcup_{n=1}^{\infty} A_n$ *is countable.*

Proof. We first observe that from part (1) of the previous proposition a set A is countable if there is an injective function $\phi : A \to \mathbb{N}$. The function $\phi : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ given by $\phi(m, n) = 2^m 3^n$ is easily seen to be injective. This is because $2^m 3^n = 2^r 3^s$ only if $2^{m-r} = 3^{s-n}$. This is only possible if $m - r = s - n = 0$. Hence $\mathbb{N} \times \mathbb{N}$ is countable. To show $A \times B$ is countable when A and B are, we note that there are surjective functions $f : \mathbb{N} \to A$ and $g : \mathbb{N} \to B$ so

$$f \times g : \mathbb{N} \times \mathbb{N} \to A \times B$$

is surjective. Since $\mathbb{N} \times \mathbb{N}$ is countable it follows from part (2) of the previous proposition that $A \times B$ is countable.

To prove that a countable union of countable sets is countable note that if A_n is countable there is a surjection $\psi_n : \mathbb{N} \to A_n$. The function

$$\Psi : \mathbb{N} \times \mathbb{N} \to \bigcup_{n=1}^{\infty} A_n$$

given by

$$\Psi(n, m) = \psi_n(m)$$

is a surjection. Since $\mathbb{N} \times \mathbb{N}$ is countable it follows that $\bigcup_{n=1}^{\infty} A_n$ is countable by part (2) of the previous proposition. ☐

Corollary 0.3.9 ($\mathbb{Q}$ is countable)**.** *The rational numbers $\mathbb{Q}$ are countable.*

Proof. The set $\mathbb{Z}$ is countable (see exercises below) 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 so the set of rationals $\mathbb{Q}$ is countable. ☐

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 0.3.10. *Suppose A is a non-empty set and $f : A \to \mathcal{P}(A)$. 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. If $x \in B$, then, by the definition of B, we know $x \notin f(x)$ and $x \in B$ so $f(x) \neq B$. On the other hand if $x \notin B$, then by the definition of B we know $x \in f(x)$ and since $x \notin B$ we again conclude $f(x) \neq B$. Thus in either case $f(x) \neq B$, i.e. there is no x with $f(x) = B$, so f is not surjective. ☐

As an immediate consequence we have the existence of an uncountable set.

Corollary 0.3.11. *The set $\mathcal{P}(\mathbb{N})$ is uncountable.*

Proof. This is an immediate consequence of Proposition (0.3.10) and part (2) of Proposition (0.3.7). □

Corollary 0.3.12. *If $f : A \to B$ is surjective and B is uncountable, then A is uncountable.*

Proof. This is an immediate consequence of part (2) of Proposition (0.3.7), since if A were countable the set B would also have to be countable. □

Later we will give an easy proof using measure theory that the set of irrationals is not countable (see Corollary (A.2.7)). 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 0.3.13.

1. Prove Propositions (0.3.1) and (0.3.3).

2. (*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 can be referred to 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 (0.3.5)) 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.

3. Prove that any subset of $\mathbb{Z}$ is countable by finding an explicit bijection $f : \mathbb{Z} \to \mathbb{N}$.

4. (*Uncountabilitity of* $\mathbb{R}$)
Let $\mathcal{D}$ be the set of all infinite sequences $d_1 d_2 d_3 \dots d_n \dots$ 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 $f(A) = d_1 d_2 d_3 \dots d_n \dots$ where $d_n = 1$ if $n \in A$ and 0 otherwise.

   (b) Define $h : \mathcal{D} \to [0, 1]$ by letting $h(d_1 d_2 d_3 \dots d_n \dots)$ be the real number whose decimal expansion is $0.d_1 d_2 d_3 \dots d_n \dots$ Prove that h is injective.

   (c) Prove that the closed interval $[0, 1]$ is uncountable. *Hint:* Show there is a surjective function $\phi : [0, 1] \to \mathcal{D}$ defined by $\phi(x) = h^{-1}(x)$ if $x \in h(\mathcal{D})$ and $\phi(x) = 0$ otherwise.

   (d) Prove that if $a < b$, the closed interval $\{x \mid a \leq x \leq b\}$, the open interval $\{x \mid a < x < b\}$, the ray $\{x \mid a \leq x < \infty\}$, and $\mathbb{R}$ are all uncountable.

0.4 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 0.4.1 (Open, Closed, Dense)**.** *A subset $A \subset \mathbb{R}$ is called* open *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 open subset contains a point of A.*

Proposition 0.4.2 ($\mathbb{Q}$ is dense in $\mathbb{R}$)**.** *The rational numbers $\mathbb{Q}$ are a dense subset of $\mathbb{R}$.*

Proof. Let U be an 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 hence in U. □

Theorem 0.4.3. *An 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 $\epsilon > 0$, $[a_x - \epsilon, a_x + \epsilon] \subset U$ and hence $[a_x - \epsilon, x] \subset [a_x - \epsilon, a_x + \epsilon] \cup [a_x + \epsilon, x] \subset U$ 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 ϕ which assigns to each $r \in U \cap \mathbb{Q}$ the interval U_r is a surjective map onto this collection. By Proposition (0.3.7) this collection must be countable. □

Exercise 0.4.4.

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. A point x is called a *limit point* of a set S 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 its limit points.

0.5 Compact Subsets of $\mathbb{R}$

One of the most important concepts for analysis is the notion of compactness.

Definition 0.5.1. *A closed 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 0.5.2 (The Heine-Borel Theorem). *A subset X of $\mathbb{R}$ compact if and only if it is closed and bounded.*

Proof. To see that a compact set is bounded observe that $U_n = (-n, n)$ defines an open cover of any subset X of $\mathbb{R}$. If this cover has a finite subcover for a set X, 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 $m > 0$ such that $X \subset U_m$. It follows that $(y - 1/m, y + 1/m)$ is in the complement of X. Since y was 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 $X = [a, b]$ is a closed interval. Let $\mathcal{V}$ be an open cover of X and define

$$z = \sup\{x \in [a, b] \mid \text{The cover } \mathcal{V} \text{ of } [a, x] \text{ has a finite subcover}\}.$$

Our aim is to prove that $z = b$ which we do by showing that the assumption that $z < b$ leads to a contradiction. 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 $[a, z_1]$. Since $z < z_1$ this is contradiction arising from the assumption $z < b$.

For an arbitrary closed 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 bounded closed sets which we will use.

Theorem 0.5.3 (Nested Families of Compact Sets). *If $\{A_n\}_{n=1}^{\infty}$ is a nested family of closed bounded subsets of $\mathbb{R}$, i.e. $A_n \subset A_{n-1}$, then $\cap_{n=1}^{\infty} A_n$ is non-empty.*

Proof. Let $x_n = \inf A_n$. Then $\{x_n\}$ is a bounded monotonic sequence so the limit $z = \lim x_n$ exists by Theorem (0.2.2) Since A_n is closed $x_n \in A_n$ and hence $x_n \in A_m$ for all $m \leq n$. It follows that for any $m > 0$ we have $z \in A_m$, i.e. $z \in \cap_{n=1}^{\infty} A_n$. □

Exercise 0.5.4.

1. Prove that the set $\mathcal{D} = \{m/2^n \mid m \in \mathbb{Z},\ n \in \mathbb{N}\}$ is dense in $\mathbb{R}$.

2. Give an example of a nested family of non-empty open intervals $U_1 \supset U_2 \cdots \supset U_n \ldots$ such that $\cap U_n = \emptyset$.

0.6 Continuous and Differentiable Functions

Definition 0.6.1 (Continuous and Uniformly Continuous Functions)**.** *A function* $f : \mathbb{R} \to \mathbb{R}$ *is* continuous *if for every* x *and every* $\epsilon > 0$ *there is a* $\delta(x)$ *(depending on* x*) such that* $|f(y) - f(x)| < \epsilon$ *whenever* $|y - x| < \delta(x)$*. A function* $f : \mathbb{R} \to \mathbb{R}$ *is* uniformly continuous *if for every* $\epsilon > 0$ *there is a* δ *(independent of* x *and* y*) such that* $|f(y) - f(x)| < \epsilon$ *whenever* $|y - x| < \delta$*.*

Theorem 0.6.2. *If* f *is defined and continuous on a closed interval* $[a, b]$ *then it is uniformly continuous on that interval.*

Proof. Suppose $\epsilon > 0$ is given. For any $x \in [a, b]$ and any positive number δ 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 $y \in U(x, \delta(x))$ we have $|f(x) - f(y)| < \epsilon/2$. Therefore if y_1 and y_2 are both in $U(x, \delta(x))$ we note

$$|f(y_1) - f(y_2)| \leq |f(y_1) - f(x)| + |f(x) - f(y_2)| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

The collection $\{U(x, \delta(x)/2) \mid x \in [a, b]\}$ is an open cover of the compact set $[a, b]$ 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 $y_1, y_2 \in [a, b]$ and $|y_1 - y_2| < \delta$. Then y_1 is in $U(x_j, \delta(x_j)/2)$ 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 $|f(y_1) - f(y_2)| < \epsilon$. □

We will also make use of the following result from elementary calculus.

Theorem 0.6.3 (Mean Value Theorem)**.** *If f is 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 0.6.4. *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 0.6.5.

1. (*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 $C \subset \mathbb{R}$.

0.7 Real Vector Spaces

Definition 0.7.1 (Inner Product Space)**.** *A real vector space $\mathcal{V}$ is called an* inner product space *if there is 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**: $\langle v_1, v_2\rangle = \langle v_2, v\rangle$.
2. **Bi-linearity**: $\langle c_1 v_1 + c_2 v_2, w\rangle = c_1\langle v_1, w\rangle + c_2\langle v_2, w\rangle$.
3. **Positive Definiteness**: $\langle w, w\rangle \geq 0$ *with equality only if* $w = 0$.

Definition 0.7.2 (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 0.7.3 (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 $\|v\| = \|w\| = 1$. 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

$$\langle v, w\rangle^2 = \|\langle v, w\rangle w\|^2 \leq \|\langle v, w\rangle w\|^2 + \|v - \langle v, w\rangle w\|^2 = 1$$

with equality only if $\|v - \langle v, w\rangle w\| = 0$ or $v = \langle v, w\rangle w$. This implies the inequality $|\langle v, w\rangle| \leq 1 = \|v\|\ \|w\|$, when v and w are unit vectors. The 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 $v = \langle v, w\rangle w$, i.e. only if one of the vectors is a multiple of the other. □

Proposition 0.7.4 (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

$$\begin{aligned}
\|v+w\|^2 &= \langle v+w, v+w\rangle \\
&= \langle v,v\rangle + 2\langle v,w\rangle + \langle w,w\rangle \\
&= \|v\|^2 + 2\langle v,w\rangle + \|w\|^2 \\
&\le \|v\|^2 + 2|\langle v,w\rangle| + \|w\|^2 \\
&\le \|v\|^2 + 2\|v\|\ \|w\| + \|w\|^2 \text{ by Cauchy-Schwarz,} \\
&= (\|v\| + \|w\|)^2
\end{aligned}$$

To prove item (4), the parallelogram law, note $\|v-w\|^2 = \langle v-w, v-w\rangle = \|v\|^2 - 2\langle v,w\rangle + \|w\|^2$. Likewise $\|v+w\|^2 = \langle v+w, v+w\rangle = \|v\|^2 + 2\langle v,w\rangle + \|w\|^2$. Hence the sum $\|v-w\|^2 + \|v+w\|^2$ equals $2\|v\|^2 + 2\|w\|^2$. □

Chapter 1

The Regulated and Riemann Integrals

1.1 Introduction

We will consider several different approaches to defining the definite integral

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

of a function $f(x)$. These definitions will all assign the same value to the definite integral, but they differ in the size of the collection of functions for which they are defined. 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 $f(x)$ 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 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 also are closed under the operations of addition and scalar multiplication. We will call such a collection a *vector space* of functions. More precisely a non-empty set of real valued functions $\mathcal{V}$ defined on a fixed closed interval will be called 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 $r \in \mathbb{R}$, then $rf \in \mathcal{V}$.

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

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

II. Monotonicity: If 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

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

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 the integral $\int_a^b f(x)\ dx$ has been defined for all f in some vector space of functions $\mathcal{V}$ and for all subintervals $[a, b]$ of I. And suppose this integral satisfies properties I and II above. Then for any function $f \in \mathcal{V}$ for which $|f| \in \mathcal{V}$*

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

for all $a < b$ in I.

Proof. This follows from monotonicity and linearity. Since $f(x) \leq |f(x)|$ for all x we know $\int_a^b f(x)\ dx \leq \int_a^b |f(x)|\ dx$. 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 $|\int_a^b f(x)\ dx| \leq \int_a^b |f(x)|\ dx$. $\square$

1.3 Step Functions and the Regulated Integral

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 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).*

We will say that the points $x_0 = a < x_1 < x_2 < \cdots < x_{n-1} < x_n = b$ define an *interval partition* for the step function f. Note that the definition says 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 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 pretty 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 $c_i(x_i - x_{i-1})$. Of course 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*

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

Exercise 1.3.3.

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

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

Theorem 1.3.4. *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 $x_0 = a < x_1 < x_2 < \cdots < x_{n-1} < x_n = b$ 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

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

□

Recall the definition of a uniformly converging sequence of functions.

Definition 1.3.5 (Uniform Convergence)**.** *A sequence of functions $\{f_m\}$ is said to* converge uniformly *on $[a,b]$ to a function f if for every $\epsilon > 0$ there is an M (*independent of *x) such that for all $x \in [a,b]$*

$$|f(x) - f_m(x)| < \epsilon \text{ whenever } m \geq M.$$

Contrast this with the following.

Definition 1.3.6 (Pointwise Convergence)**.** *A sequence of functions $\{f_m\}$ is said to* converge pointwise *on $[a,b]$ to a function f if for each $\epsilon > 0$ and each $x \in [a,b]$ there is an M_x (*depending on *x) such that*

$$|f(x) - f_m(x)| < \epsilon \text{ whenever } m \geq M_x.$$

Definition 1.3.7 (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.*

Exercise 1.3.8.

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

2. Give an example of a sequence of step functions which converge uniformly to $f(x) = x$ on $[0,1]$. Give an example of a sequence of step functions which converge *pointwise* to 0 on $[0,1]$, but which do not converge uniformly.

Every regulated function can be uniformly approximated as closely as we wish by a step function. Since we know how to integrate step functions it is natural to take a sequence of better and better step function approximations to a regulated function $f(x)$ 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 exists and that it does not depend on the choice of approximating step functions.

Theorem 1.3.9. *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 $\epsilon > 0$ there is an M such that $|z_p - z_q| \le \epsilon$ whenever $p, q \ge M$.

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

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

Hence whenever, $p, q \ge M$

$$|f_p(x) - f_q(x)| < |f_p(x) - f(x)| + |f(x) - f_q(x)| < 2\frac{\epsilon}{2(b-a)} = \frac{\epsilon}{b-a}. \tag{1.3.2}$$

Therefore, whenever $p, q \ge M$

$$|z_p - z_q| = \Big| \int_a^b f_p(x) - f_q(x)\ dx \Big| \le \int_a^b |f_p(x) - f_q(x)|\ dx \le \int_a^b \frac{\epsilon}{b-a}\ dx = \epsilon,$$

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 (1.3.2). 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 $\epsilon > 0$ there is an M such that for all x

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

whenever $m \ge M$. It follows that

$$|f_m(x) - g_m(x)| \le |f_m(x) - f(x)| + |f(x) - g_m(x)| < 2\epsilon.$$

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

$$\Big| \int_a^b f_m(x) - g_m(x)\ dx \Big| \le \int_a^b |f_m(x) - g_m(x)|\ dx \le \int_a^b 2\epsilon\ dx = 2\epsilon(b-a),$$

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

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

This implies

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

□

This result enables us to define the regulated integral.

Definition 1.3.10 (The 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.

We next need to see that the regulated functions form a large class including all continuous functions.

Theorem 1.3.11 (Continuous functions are regulated)**.** *Every continuous function $f : [a, b] \to \mathbb{R}$ is a regulated function.*

Proof. By Theorem 0.6.2 a continuous function $f(x)$ defined on a closed interval $[a, b]$ is *uniformly continuous*. That is, given $\epsilon > 0$ there is a corresponding $\delta > 0$ such that $|f(x) - f(y)| < \epsilon$ whenever $|x - y| < \delta$. Let $\epsilon_n = 1/2^n$ and let δ_n be the corresponding δ 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

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

That is, on each half open interval $[x_{-1}i, x_i[$ 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

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

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 $x \in [a,b]$

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

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

Exercise 1.3.12.

1. Give an example of a continuous function on the *open* interval $(0,1)$ which is not regulated, i.e. which cannot be uniformly approximated by step functions.

2. Prove that the regulated integral, as given in (1.3.10), satisfies properties I-V of §1.2.

3. Prove that f is a regulated function on $I = [a,b]$ if and only if both the limits

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

   exist for every $c \in (a,b)$. (See section VII.6 of Dieudonné [1]).

1.4 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 slightly differently and show that the result above follows easily from that formulation.

Theorem 1.4.1. *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

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

To do this we note that

$$\begin{aligned} \Big| \frac{F(x_0 + h) - F(x_0)}{h} - f(x_0) \Big| &= \Big| \frac{\int_{x_0}^{x_0+h} f(t)\ dt}{h} - f(x_0) \Big| \\ &= \Big| \frac{\int_{x_0}^{x_0+h} f(t)\ dt - f(x_0)h}{h} \Big| \\ &= \frac{\big| \int_{x_0}^{x_0+h} (f(t) - f(x_0))\ dt \big|}{|h|} \end{aligned} \tag{1.4.1}$$

Monotonicity tells us that when h is positive

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

and if h is negative

$$\Big| \int_{x_0}^{x_0+h} (f(t) - f(x_0))\ dt \Big| \leq \int_{x_0+h}^{x_0} |f(t) - f(x_0)|\ dt = -\int_{x_0}^{x_0+h} |f(t) - f(x_0)|\ dt.$$

In either case we see

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

Combining this with inequality (1.4.1) above we obtain

$$\Big|\frac{F(x_0+h)-F(x_0)}{h} - f(x_0)\Big| \leq \frac{\big|\int_{x_0}^{x_0+h} |f(t)-f(x_0)|\ dt\big|}{|h|}. \tag{1.4.2}$$

But the continuity of f implies that given x_0 and any $\epsilon > 0$ there is a $\delta > 0$ such that whenever $|t - x_0| < \delta$ we have $|f(t) - f(x_0)| < \epsilon$. Thus if $|h| < \delta$ then $|f(t) - f(x_0)| < \epsilon$ for all t between x_0 and $x_0 + h$. It follows that $\big|\int_{x_0}^{x_0+h} |f(t) - f(x_0)|\ dt\big| < \epsilon|h|$ and hence that

$$\frac{\big|\int_{x_0}^{x_0+h} |f(t)-f(x_0)|\ dt\big|}{|h|} < \epsilon.$$

Putting this together with equation (1.4.2) above we have that

$$\Big|\frac{F(x_0+h)-F(x_0)}{h} - f(x_0)\Big| < \epsilon$$

whenever $|h| < \delta$ which was exactly what we needed to show. □

Corollary 1.4.2. 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 (1.4.1) 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 (0.6.4)).

Then

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

□

Exercise 1.4.3.

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

1.5 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 comparing with step functions.

Suppose that $f(x)$ is a bounded function on the interval $\mathcal{I} = [a, b]$ 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 $f(x) \le u(x)$ for all $x \in I$, then monotonicity implies that if we can define $\int_a^b f(x)\ dx$ it must satisfy $\int_a^b f(x)\ dx \le \int_a^b u(x)\ dx$.

This is true for every step function u satisfying $f(x) \le u(x)$ 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

$$\int_a^b f(x)\ dx \le \inf \Big\{ \int_a^b u(x)\ dx \ \Big|\ u \in \mathcal{U}(f) \Big\}. \tag{1.5.1}$$

The *infimum* exists because all 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

$$\sup \Big\{ \int_a^b v(x)\ dx \ \Big|\ v \in \mathcal{L}(f) \ \Big\} \le \int_a^b f(x)\ dx. \tag{1.5.2}$$

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

Putting inequalities (1.5.1) and (1.5.2) together, we see 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

$$\sup \Big\{ \int_a^b v(x)\ dx \ \Big|\ v \in \mathcal{L}(f) \Big\} \le \int_a^b f(x)\ dx \le \inf \Big\{ \int_a^b u(x)\ dx \ \Big|\ u \in \mathcal{U}(f) \Big\} \tag{1.5.3}$$

for every $f \in \mathcal{V}$. Even if we can't define an integral for f, however, we still have the inequalities of the ends.

Proposition 1.5.1. *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 $f(x) \leq u(x)$ 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. Then*

$$\sup\Big\{ \int_a^b v(x)\ dx \ \Big|\ v \in \mathcal{L}(f)\Big\} \leq \inf\Big\{ \int_a^b u(x)\ dx \ \Big|\ u \in \mathcal{U}(f)\Big\}.$$

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

$$V = \Big\{ \int_a^b v(x)\ dx \ \Big|\ v \in \mathcal{L}(f)\Big\} \text{ and } \quad U = \Big\{ \int_a^b u(x)\ dx \ \Big|\ u \in \mathcal{U}(f)\Big\}$$

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 ends of this inequality are not equal (see Exercise (1.5.4) below), but if it should happen that

$$\sup\Big\{ \int_a^b v(x)\ dx \ \Big|\ v \in \mathcal{L}(f)\Big\} = \inf\Big\{ \int_a^b u(x)\ dx \ \Big|\ u \in \mathcal{U}(f)\Big\}.$$

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 we can integrate. Henceforth we will use the more compact notation

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

and

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

Definition 1.5.2. The 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*

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

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

Theorem 1.5.3. *A bounded function $f : [a, b] \to \mathbb{R}$ is Riemann integrable if and only if, for every $\epsilon > 0$ 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*

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

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

$$\int_a^b v_0(x)\ dx \le \sup_{v \in \mathcal{L}(f)} \Big\{ \int_a^b v(x)\ dx \Big\} \le \inf_{u \in \mathcal{U}(f)} \Big\{ \int_a^b u(x)\ dx \Big\} \le \int_a^b u_0(x)\ dx.$$

This implies

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

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

Conversely if f is Riemann integrable, then by Proposition 0.1.4 there exists a step function $u_0 \in \mathcal{U}(f)$ such that

$$\int_a^b u_0(x)\ dx - \int_a^b f(x)\ dx = \int_a^b u_0(x)\ dx - \inf_{u \in \mathcal{U}(f)} \Big\{ \int_a^b u(x)\ dx \Big\} < \epsilon/2.$$

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

$$\int_a^b f(x)\ dx - \int_a^b v_0(x)\ dx < \epsilon/2.$$

Hence

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

and u_0 and v_0 are the desired functions. □

Exercise 1.5.4.

1. At the beginning of these notes 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

$$\sup_{v \in \mathcal{L}(f)} \Big\{ \int_0^1 v(x)\ dx \Big\} = 0 \text{ and } \inf_{u \in \mathcal{U}(f)} \Big\{ \int_0^1 u(x)\ dx \Big\} = 1.$$

   Hence f is not Riemann integrable.

There are several facts about the relation with the regulated integral we must establish. 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 1.5.5. Define the function $f : [0, 1] \to \mathbb{R}$ by

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

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

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

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

$$\inf_{u \in \mathcal{U}(f)} \Big\{ \int_0^1 u(x)\ dx \Big\} \le \inf_{m \in \mathbb{Z}^+} \Big\{ \int_0^1 u_m(x)\ dx \Big\} = \inf_{m \in \mathbb{Z}^+} \{\frac{1}{m}\} = 0.$$

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

$$0 \le \sup_{v \in \mathcal{L}(f)} \Big\{ \int_0^1 v(x)\ dx \Big\}.$$

Putting together the last two inequalities with Proposition (1.5.1) we obtain

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

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 $|f(x) - g(x)| \leq \epsilon$ for some $\epsilon > 0$. Then g is constant, say with value c_1 on the open interval $(0, x_1)$.

Now there are points $a_1, a_2 \in (0, x_1)$ with $f(a_1) = 0$ and $f(a_2) = 1$. Then $|c_1| = |c_1 - 0| = |g(a_1) - f(a_1)| \leq \epsilon$ and $|1 - c_1| = |f(a_2) - g(a_2)| \leq \epsilon$. 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 $\epsilon \geq 1/2$. That is, f cannot be uniformly approximated by any step function to within ϵ if $\epsilon < 1/2$. So f is not regulated. □

Theorem 1.5.6 (Regulated functions are 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 $\epsilon > 0$, it can be uniformly approximated within ϵ by a step function. In particular, if $\epsilon_n = 1/2^n$ there is a step function $g_n(x)$ such that $|f(x) - g_n(x)| < \epsilon_n$ 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 $u_n(x) - v_n(x) = g_n(x) + 1/2^n - (g_n(x) - 1/2^n) = 1/2^{n-1}$,

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

Hence we may apply Theorem(1.5.3) to conclude that f is Riemann integrable.

Also

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

Since for all n

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

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. $\square$

Theorem 1.5.7. *The set $\mathcal{R}$ of 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 $\epsilon > 0$ 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

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

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

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

This implies that

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

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

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

As $\epsilon > 0$ 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 $(f + g) \in \mathcal{R}$. $\square$

Exercise 1.5.8.

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

Chapter 2

Lebesgue Measure

2.1 Introduction

In the previous section we studied two definitions of integral that were based on two important facts: (1) There is only the one obvious way to define the integral of step 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 which f is constant, but with the sets *not necessarily intervals.* For example we will consider functions like

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

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 $3 \operatorname{len}(A) + 2 \operatorname{len}(B)$, but only if we can make sense of $\operatorname{len}(A)$ and $\operatorname{len}(B)$. 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 we did 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 it is not possible to achieve that.

There are several properties which we want any notion of "generalized length" to satisfy. 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 $\mu(A) = \operatorname{len}(A) = b - a$, 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 finite collections $\{A_n\}_{n=1}^{m}$ 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 $B \setminus A$ are disjoint and their union is B.

It should be fairly clear why we 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 only hold 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. If $\{x\}$ is the set consisting of a single point $x \in [0,1]$, then $\mu(\{x\}) = 0$ by property I and $[a,b]$ is an uncountable union of pairwise disjoint sets, namely each of the sets consisting of one point of $[a,b]$. Hence we would have $\mu([a,b]) = b - a$ is an *uncountable* sum of zeroes. This is one 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 μ 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 the properties of Lebesgue measure and how to use it.

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 sets whose "generalized length," or measure, is zero don't matter. 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 *provided for every $\epsilon > 0$ there is a collection of open intervals $\{U_n\}_{n=1}^{\infty}$ such that*

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

Perhaps surprisingly this definition makes no use of the measure μ. Indeed, we have not yet defined the measure μ of a set X for any choice of the set X! However, it is clear that if we can do so in a way that satisfies properties I-IV above and the Hence to denote X has measure zero, we will write $\mu(X) = 0$ even though we have not yet defined μ.

If X is a null set in $I = [0,1]$ we will say that its complement X^c has *full measure* in I.

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 μ 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.

It is not true that countable sets are the only sets which are null sets. We give an example in Exercise (2.6.1) below, namely, the Cantor middle third set, which is an uncountable null set.

2.3 Sigma algebras

As mentioned before there does not exist function μ satisfying properties I-IV and defined for every subset of $I = [0, 1]$. In this section we want to consider what is the best collection of subsets of I for which we can define a "generalized length" or *measure* μ. Suppose we have somehow defined μ 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 μ 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 Z^c, its complement in I, should also be in $\mathcal{A}$.

All 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* σ-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 σ-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 about countable intersections follows from the other two and Proposition (0.3.3) which says the intersection of a family of sets is the complement of the union of the complements of the sets. Also notice 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 σ-algebra of subsets of X (by definition), so is its complement, the empty set. A trivial example of a σ-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 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 σ-algebras is a σ-algebra. Thus for any collection $\mathcal{C}$ of subsets of $\mathbb{R}$ there is a smallest σ-algebra of subsets of $\mathbb{R}$ which contains all sets in $\mathcal{C}$, namely the intersection of all σ-algebras containing $\mathcal{C}$.

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

Exercise 2.3.3.

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

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

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

4. Let $\mathcal{A}$ be a σ-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 σ-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 $\mathcal{A}_i$ be the σ-algebra generated by $\mathcal{C}_i$. Prove that $\mathcal{B}_1, \mathcal{B}_2$, and $\mathcal{B}_3$ are all equal to the Borel σ-algebra $\mathcal{B}$.

2.4 Lebesgue Measure

The σ-algebra of primary interest to us is the one generated by Borel sets and null sets. Alternatively, as a consequence of part 5. of Exercise (2.3.3), it is the σ-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. *The σ-algebra of subsets of $\mathbb{R}$ generated by open intervals and null sets will be denoted $\mathcal{M}$. 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 use any other interval just as well. Notice that it is a consequence of part 4. of Exercise (2.3.3) that $\mathcal{M}(I)$ is a σ-algebra of subsets of I. It is by no means obvious that $\mathcal{M}$ is not the σ-algebra of all subsets of $\mathbb{R}$. However, in section (B) of the appendix 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 μ, 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 $A \subset I$, $c \in \mathbb{R}$ and $A + c \subset I$ where $A + c$ denotes the set $\{x + c \mid x \in A\}$. Then $\mu(A + c) = \mu(A)$*

III. Countable Additivity. *If $\{A_n\}_{n=1}^{\infty}$ is a countable collection of subsets of I, 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 $A, B \in \mathcal{M}(I)$ and $A \subset B$ then $\mu(A) \leq \mu(B)$*

V. Null Sets *A subset $A \subset I$ is a null set set if and only if $A \in \mathcal{M}(I)$ and $\mu(A) = 0$.*

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 most of this theorem to Appendix A, because it is somewhat technical and is a diversion from our main task of developing a theory of integration. However there are some properties of Lebesgue measure we can easily derive; so we do so now. For example, we will use properties I-III of Theorem (2.4.2) to prove property IV.

Proposition 2.4.3 (Monotonicity)**.** *If $A, B \in \mathcal{M}(I)$ and $A \subset B$, then $\mu(A) \leq \mu(B)$.*

Proof. Since $A \subset B$ we have $B = A \cup (B \setminus A)$. Also A and $B \setminus A$ are disjoint so by property III we know $\mu(A) + \mu(B \setminus A) = \mu(B)$. But $\mu(B \setminus A) \geq 0$ so $\mu(A) \leq \mu(A) + \mu(B \setminus A) = \mu(B)$. □

Proposition 2.4.4. *If $X \subset I$ is a null set, then $X \in \mathcal{M}(I)$ and $\mu(X) = 0$.*

Proof. If $X \subset I$ is a null set, then by the definition of $\mathcal{M}(I)$ we know $X \in \mathcal{M}(I)$. If $\epsilon > 0$ there is a collection of open intervals $\{U_n\}_{n=1}^\infty$ such that

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

Property I says $\operatorname{len}(U_n) = \mu(U_n)$, so

$$\sum_{n=1}^{\infty} \mu(U_n) = \sum_{n=1}^{\infty} \operatorname{len}(U_n) < \epsilon.$$

Since $X \subset \cup U_n$ properties II and III imply

$$\mu(X) \leq \mu\Big(\bigcup_{n=1}^{\infty} U_n\Big) \leq \sum_{n=1}^{\infty} \mu(U_n) < \epsilon,$$

This is true for any $\epsilon >$ so the only possible value for $\mu(X)$ is zero. □

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 2.4.5.

(1) The Lebesgue measure of I, $\mu(I)$, is 1 and hence $\mu(A^c) = 1 - \mu(A)$.

(2) If A and B are in $\mathcal{M}(I)$, then $A \setminus B$ is in $\mathcal{M}$ and $\mu(A \cup B) = \mu(A \setminus B) + \mu(B)$.

Proof. To see (1) observe that A and A^c are disjoint and $A \cup A^c = I$, so additivity implies $\mu(A) + \mu(A^c) = \mu(A \cup A^c) = \mu(I) = 1$.

For (2) note that $A \setminus B = A \cap B^c$ which is in $\mathcal{M}$. Also $A \setminus B$ and B are disjoint and their union is $A \cup B$. So once again additivity implies implies $\mu(A \setminus B) + \mu(B) = \mu(A \cup B)$. $\square$

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

Proposition 2.4.6. *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, $A_n = \cup_{i=1}^{n} F_i$ and $\cup_{i=1}^{\infty} A_i = \cup_{i=1}^{\infty} F_i$. 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

$$\mu\big(\bigcap_{n=1}^{\infty} B_n\big) = 1 - \mu\big(\bigcup_{i=1}^{\infty} E_i\big) = 1 - \lim_{n \to \infty} \mu(E_n) = \lim_{n \to \infty} (1 - \mu(E_n)) = \lim_{n \to \infty} \mu(B_n).$$

$\square$

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 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 $> p$ 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 the definition of outer measure and the fact that $\mu^*(A) = \mu(A)$, that for any $\epsilon > 0$ there is a countable open cover $\{U_n\}_{n=1}^{\infty}$ of A such that

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

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

$$\sum_{n=1}^{\infty} \operatorname{len}(U_n) \leq \mu(A) + (1-p)\mu(A)$$

$$\leq \mu(A) + (1-p)\sum_{n=1}^{\infty} \operatorname{len}(U_n)$$

so

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

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. □

There is a much stronger result than the theorem above which we now state, but do not prove. A proof can be found in Section 9.2 of [5].

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

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

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.*

2.6 Lebesgue Measurable Sets – Summary

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

1. The collection of Lebesgue measurable sets $\mathcal{M}$ is a *σ-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 $A \in \mathcal{M}(I)$, then there is a 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 $A \in \mathcal{M}$, then $\mu(A^c) = 1 - \mu(A)$.
   - If $A \in \mathcal{M}$ is a null set if and only if $\mu(A) = 0$.
   - *Countable Subadditivity*: If $A_n \in \mathcal{M}$ 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 $A_n \in \mathcal{M}$ for $n \in \mathbb{N}$ are pairwise disjoint, then
   $$\mu\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} \mu(A_n).$$

- *Increasing sequences*: If $A_n \in \mathcal{M}$ 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 $A_n \in \mathcal{M}$ 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.

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

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. (*The Cantor middle third set*) We first recursively define a nested sequence $\{J_n\}_{n=0}^{\infty}$ of closed subsets of I. Each J_n consists of a finite union of closed intervals. We define J_0 to be $I = [0,1]$ 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 $J_0 = [0,1]$, $J_1 = [0,1/3] \cup [2/3,1]$ and $J_2 = [0,1/9] \cup [2/9,1/3] \cup [2/3,7/9] \cup [8/9,1]$ etc.

   Let $C = \cap_{n=0}^{\infty} J_n$. It is called the *Cantor Middle Third set.*

   (a) 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 note $J_n = L_n \cup R_n$. 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 $1/3^n$.

   (b) Let $\mathcal{D}$ be the uncountable set 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 (0.3.13)) 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 ψ is surjective and hence by Corollary (0.3.12) the set C is uncountable. *Hint:* You will need to use Theorem (0.5.3).

   (c) 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 $\epsilon > 0$ a collection of open intervals $\{U_n\}_{n=1}^{\infty}$ such that

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

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 $\mathfrak{X}_A(x)$ *is defined by*

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

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* $\{A_i\}_{i=1}^n$ *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* $\int f \, d\mu = \sum_{i=1}^n r_i \mu(A_i)$.

The definition of 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 $\int_a^b 1\ dx = (b-a)$. The value of $\int f\ d\mu$ for a step 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 is 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 are 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 ≥ 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

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

□

Exercise 3.1.5.

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.

A function $f : [0,1] \to \mathbb{R} \cup \{\infty\} \cup \{-\infty\}$ will be 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 $[a, \infty]$.

Proposition 3.1.6. *If $f : [0,1] \to \mathbb{R}$ is an extended real valued function, then the following are equivalent:*

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

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

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

4. *For any* $a \in [-\infty, \infty]$ *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 by Theorem (A.3.8). 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.

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

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

which is measurable by Theorem (A.3.8). 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. □

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).*

Proposition 3.1.8. *If* $f(x)$ *is a function which has the value* 0 *except on a set of measure* 0*, then* $f(x)$ *is measurable.*

Proof. Suppose $f(x) = 0$ for all $x \notin A$ where $A \subset [0, 1]$ has measure 0. That is, if $A = f^{-1}([-\infty, 0[) \cup f^{-1}((0, \infty])$ then A is a null set. 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. □

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.

Proof. If $a \in [-\infty, \infty]$, 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 by Theorem (A.3.8). 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.*

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 just showed they are also measurable, so the bounded measurable functions are a vector subspace. ☐

Exercise 3.1.11.

1. Prove that if f is a measurable function, then so is f^2.

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

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 function bigger than f equals the *supremum* of the integrals of step function less than f. It is natural to alter both these definitions, replacing step function with simple function. 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*

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

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 $\epsilon_n = (b-a)/n$. We will partition the *range* $[a,b]$ of f by intervals follows: Let $c_i = a + i\epsilon_n$ so $a = c_0 < c_1 < \cdots < c_n = b$. Now define a measurable partition of $[0,1]$ by $A_i = f^{-1}([c_{i-1}, c_i))$ for $i < n$ and $A_n = f^{-1}([c_{n-1}, b])$. Then clearly $f_n(x) = \sum_{i=1}^n c_i \mathfrak{X}_{A_i}$ is a simple function. Moreover we note that for any $x \in [0,1]$ we have $|f(x) - f_n(x)| \leq \epsilon_n$. 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 $|f(x) - f_n(x)| \leq c_j - c_{j-1} = \epsilon_n$. 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 $\delta_n = \sup_{x\in[0,1]} |f(x) - f_n(x)|$, 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

$$\begin{aligned}
\inf_{u \in \mathcal{U}_\mu(f)} \Big\{ \int u \ d\mu \Big\} &\leq \liminf_{n\to\infty} \int u_n \ d\mu \\
&= \liminf_{n\to\infty} \int (f_n + \delta_n) \ d\mu \\
&= \liminf_{n\to\infty} \int f_n \ d\mu \\
&\leq \limsup_{n\to\infty} \int f_n \ d\mu \\
&= \limsup_{n\to\infty} \int (f_n - \delta_n) \ d\mu \\
&\leq \limsup_{n\to\infty} \int v_n \ d\mu \\
&\leq \sup_{v \in \mathcal{L}_\mu(f)} \Big\{ \int v \ d\mu \Big\}. \qquad (3.2.1)
\end{aligned}$$

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

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

Combining this with the inequality above we conclude that

$$\sup_{v \in \mathcal{L}_\mu(f)} \Big\{ \int v \ d\mu \Big\} = \inf_{u \in \mathcal{U}_\mu(f)} \Big\{ \int u \ d\mu \Big\}. \qquad (3.2.2)$$

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 $v_n(x) \leq f(x) \leq u_n(x)$ for all x and such that

$$\int u_n \ d\mu - \int v_n \ d\mu < 2^{-n}. \qquad (3.2.3)$$

By Theorem (3.1.9) the functions

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

are measurable. They are also bounded and satisfy $g_1(x) \leq f(x) \leq g_2(x)$. 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 $\int u_n \, d\mu - \int v_n \, d\mu = \int u_n - v_n \, d\mu \geq \int \frac{1}{m_0}\mathfrak{X}_{B_{m_0}} \, d\mu = \frac{1}{m_0}\mu(B_{m_0})$ for all n which contradicts equation (3.2.3) 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 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). □

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*

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

or equivalently (by Theorem (3.2.1)),

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

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 $\lim_{n\to\infty} \int g_n \, d\mu$ exists and is equal to $\int f \, d\mu$.*

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

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

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

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

Hence these inequalities must be equalities and $\lim_{n\to\infty} \int g_n \; d\mu = \int f \; d\mu$. □

Exercise 3.2.4.

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.

Theorem 3.2.5. *The Lebesgue integral, defined on the vector space of bounded Lebesgue measurable functions on* $[0, 1]$*, satisfies the following properties:*

I. Linearity: *If* f *and* g *are Lebesgue measurable 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* $f(x) \leq g(x)$ *for all* x*, then* $\int f \; d\mu \leq \int g \; d\mu$*.*

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

IV. Null Sets: *If f and g are bounded functions and $f(x) = g(x)$ except on a set of measure zero, then f is measurable if and only if g is measurable. If they are measurable, 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

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

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

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

If f and g are bounded functions which are equal except on a set E with $\mu(E) = 0$, then $h(x) = f(x) - g(x)$ is non-zero only on the set E. The function h is measurable by Proposition (3.1.8). Clearly, since $f = g + h$ the function f is measurable if and only if g is.

In case they are both measurable $|\int f\ d\mu - \int g\ d\mu| = |\int h\ d\mu| \leq \int |h|\ d\mu$. But the function h is bounded; say $|h(x)| \leq M$. Then $|h(x)| \leq M\mathfrak{X}_E(x)$ so 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 $\int f\ d\mu = \int g\ d\mu$. □

Definition 3.2.6. *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 3.2.7 (Additivity)**.** *If E and F are disjoint measurable subsets of $[0,1]$, then*

$$\int_{E\cup F} f\ d\mu = \int_E f\ d\mu + \int_F f\ d\mu.$$

Proof. If E and F are disjoint measurable subsets of $[0,1]$, then $\mathfrak{X}_{E\cup F} = \mathfrak{X}_E + \mathfrak{X}_F$ so

$$\int_{E\cup F} f\ d\mu = \int f\mathfrak{X}_{E\cup F}\ d\mu = \int f(\mathfrak{X}_E + \mathfrak{X}_F)\ d\mu = \int_E f\ d\mu + \int_F f\ d\mu.$$

□

Proposition 3.2.8 (Riemann integrable functions are 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

$$\sup_{v\in\mathcal{L}(f)} \Big\{ \int_0^1 v(t)\ dt \Big\} \leq \sup_{v\in\mathcal{L}_\mu(f)} \Big\{ \int v\ d\mu \Big\} \leq \inf_{u\in\mathcal{U}_\mu(f)} \Big\{ \int u\ d\mu \Big\} \leq \inf_{u\in\mathcal{U}(f)} \Big\{ \int_0^1 u(t)\ dt \Big\}.$$

The fact that f is Riemann integrable asserts the first and last of these values are equal. Hence they are all equal and f is measurable and the Riemann and Lebesgue integrals coincide. □

3.3 The Bounded Convergence Theorem

We want to investigate when the fact that a sequence of functions $\{f_n\}_{n=1}^\infty$ converges pointwise to a function f implies that their Lebesgue integrals converge to the integral of f. 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, because we prove a stronger result below. But first we consider an example which shows what can go wrong.

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

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

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 (The Bounded Convergence Theorem)**.** *Suppose $\{f_n\}_{n=1}^\infty$ is a sequence of measurable functions which converges pointwise to a function f and 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 $|f(x)| \leq M$ and by Theorem (3.1.9) that $f(x)$ is measurable.

We must show that

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

but

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

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

Given $\epsilon > 0$ define $E_n = \{x \mid |f_m(x) - f(x)| < \epsilon/2 \text{ for all } m \geq n\}$. Notice that if for some n the set E_n were all of $[0,1]$ we would be able to estimate $\int |f_m - f| \, d\mu \leq \int \epsilon/2 \, d\mu = \epsilon/2$ 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 $E_n \subset E_{n+1}$ by Proposition (2.4.6) we know $\lim_{n\to\infty} \mu(E_n) = \mu([0,1]) = 1$. Thus there is an n_0 such that $\mu(E_{n_0}) > 1 - \frac{\epsilon}{4M}$, so $\mu(E_{n_0}^c) < \frac{\epsilon}{4M}$.

Now for any $n > n_0$ we have

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

Thus we have shown $\lim_{n\to\infty} \int |f_n - f|\ d\mu = 0$. Putting this together with 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 (3.2.5) 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.

Theorem 3.3.4 (Better Bounded Convergence Theorem)**.** *Suppose $\{f_n\}_{n=1}^\infty$ is a sequence of bounded measurable functions and f 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 $n > 0$, $|f_n(x)| \le M$ for almost all $x \in [0,1]$. Then f is a measurable function, satisfying $|f(x)| \le M$ for almost all $x \in [0,1]$ 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 then $\mu(A) = 0$. Define the set $D_n = \{x \mid |f_n(x)| > M\}$, then then $\mu(D_n) = 0$ so if $E = A \cup \bigcup_{n=1}^{\infty} D_n$, then $\mu(E) = 0$. Let

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

Then $|g_n(x)| \leq M$ for *all* $x \in [0,1]$ and for any $x \notin E$ we have $\lim_{n\to\infty} g_n(x) = \lim_{n\to\infty} f_n(x) = f(x)$. Also for $x \in E$, $g_n(x) = 0$ so so for all $x \in [0,1]$ we have $\lim_{n\to\infty} g_n(x) = f(x)\mathfrak{X}_{E^c}(x)$.

Define the function g by $g(x) = f(x)\mathfrak{X}_{E^c}(x)$. For any $x \notin E$ we have $g(x) = \lim_{n\to\infty} g_n(x) = \lim_{n\to\infty} f_n(x) = f(x)$ so $g(x) = f(x)$ almost everywhere. We know from its definition that $|g(x)| \leq M$ since $|g_n(x)| \leq M$. And by Theorem (3.1.9) g is measurable. Since $f(x) - g(x)$ is zero almost everywhere it is measurable by Proposition (3.1.8). It follows that f is measurable and by Theorem (3.2.5)

$$\int f \, d\mu = \int g \, d\mu.$$

Since $f_n = g_n$ almost everywhere we also know that

$$\int f_n \, d\mu = \int g_n \, d\mu.$$

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 Theorem (3.3.2) □

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 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 this case.

4.1 Non-negative Functions

Definition 4.1.1 (Integrable Function)**.** *If $f : [0,1] \to \mathbb{R}$ 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 $\int f \, d\mu = \lim \int f_n \, d\mu \geq \lim n\mu(A) = +\infty$. □

Example 4.1.3. Let $f(x) = 1/\sqrt{x}$ for $x \in (0,1]$ and let $f(0) = +\infty$. Then $f : [0,1] \to \mathbb{R}$ 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 $f : [0,1] \to \mathbb{R}$ 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 $\mu(E) = 0$. □

Theorem 4.1.6 (Absolute Continuity)**.** *Suppose f is a non-negative integrable function. Then for any $\epsilon > 0$ there exists a $\delta > 0$ such that $\int_A f \, d\mu < \epsilon$ 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

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

Consequently

$$\int f_n \, d\mu = \int_{E_n} n \, d\mu + \int_{E_n^c} f \, d\mu.$$

Hence we have

$$\begin{aligned} \int f \, d\mu &= \int_{E_n} f \, d\mu + \int_{E_n^c} f \, d\mu \\ &= \int_{E_n} (f - n) \, d\mu + \int_{E_n} n \, d\mu + \int_{E_n^c} f \, d\mu \\ &= \int_{E_n} (f - n) \, d\mu + \int f_n \, d\mu. \end{aligned}$$

Thus $\int f \, d\mu - \int f_n \, d\mu = \int_{E_n} (f - n) \, d\mu$ and we conclude from integrability of f that

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

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

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

□

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 generalization of a result from Exercise (1.4.3).

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

Proof. Given $\epsilon > 0$ 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

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

by Theorem (4.1.6). We have in fact proven that F is uniformly continuous. □

Exercise 4.1.8.

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

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

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, however, because of Theorem (4.1.6).

Theorem 4.2.1 (Lebesgue Convergence for Non-negative functions)**.** *Suppose f_n is a sequence of non-negative measurable functions and g is a non-negative 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. 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

$$\lim_{n\to\infty} \left| \int h_n \, d\mu - \int h \, d\mu \right| = \lim_{n\to\infty} \left| \int (h_n - h) \, d\mu \right| \leq \lim_{n\to\infty} \int |h_n - h| \, d\mu. \quad (4.2.1)$$

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

Given $\epsilon > 0$ define $E_n = \{x \mid |h_m(x) - h(x)| < \epsilon/2 \text{ for all } m \geq n\}$. We know by Theorem (4.1.6) that there is a $\delta > 0$ such that $\int_A g \, d\mu < \epsilon/4$ 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 (2.4.6) we know $\lim_{n\to\infty} \mu(E_n) = \mu([0, 1]) = 1$. Thus there is an n_0 such that $\mu(E_{n_0}) > 1 - \delta$, so $\mu(E^c_{n_0}) < \delta$.

Now $|h_n(x) - h(x)| \leq |h_n(x)| + |h(x)| \leq 2g(x)$ so for any $n > n_0$ we have

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

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 (Monotone Convergence Theorem)**.** *Suppose g_n is an increasing sequence of non-negative measurable functions. 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 $f(x) \geq g_n(x)$ for almost all x allows us to apply the previous theorem to conclude the desired result.

Hence we need only show that if $\int f\ d\mu = +\infty$, then $\lim_{n\to\infty} \int g_n\ d\mu = +\infty$. But if $\int f\ d\mu = +\infty$, then for any $N > 0$ there exists n_0 such that $\int \min\{f(x), n_0\}\ d\mu > N$. And we know that

$$\lim_{n\to\infty} \min\{g_n(x), n_0\} = \min\{f(x), n_0\}$$

for almost all x. Since these are bounded measurable functions,

$$\lim_{n\to\infty} \int g_n\ d\mu \geq \lim_{n\to\infty} \int \min\{g_n, n_0\}\ d\mu = \int \min\{f, n_0\}\ d\mu > N,$$

where the equality comes from the bounded convergence Theorem (3.3.4). Since N is arbitrary we conclude that

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

□

Corollary 4.2.3 (Integral of infinite series)**.** *Suppose u_n is a non-negative measurable function and f is an 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 the previous theorem. □

4.3 Other Measures

There are other measures besides Lebesgue and indeed measures on other spaces besides $[0,1]$ or $\mathbb{R}$. We will limit our attention to measures defined on $I = [0,1]$.

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

Examples 4.3.1. The following are examples of σ-algebras on $I = [0,1]$:

1. *The trivial σ-algebra.* $\mathcal{A} = \{\emptyset, I\}$.
2. $\mathcal{A} = \{A \subset I \mid A \text{ is countable, or } A^c \text{ is countable}\}$.
3. $\mathcal{A} = \mathcal{M}$ the Lebesgue measurable sets
4. $\mathcal{A}$ is Borel sets, the smallest σ-algebra containing the open intervals.

Definition 4.3.2 (Finite Measure)**.** *If $\mathcal{A}$ is a σ-algebra of subsets of I, then a function $\nu : \mathcal{A} \to \mathbb{R}$ is called a* finite measure *provided*

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

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

We will restrict our attention to measures defined on the σ-algebra of Lebesgue measurable sets. The integral of a measurable function with respect to a measure ν is defined analogously to Lebesgue measure.

Definition 4.3.3. *Let* ν *be a finite measure defined on the* σ*-algebra* $\mathcal{M}(I)$*. If* $f(x) = \sum_{i=1}^{n} r_i \mathfrak{X}_{A_i}$ *is a simple function then its* integral with respect to ν *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* ν *by*

$$\int g \, d\nu = \inf_{u \in \mathcal{U}_\mu(g)} \Big\{ \int u \, d\nu \Big\}.$$

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

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

Definition 4.3.4 (Absolutely Continuous Measure)**.** *If* ν *is a measure defined on* $\mathcal{M}(I)$*, the Lebesgue measurable subsets of* I*,, then we say* ν *is* absolutely continuous *with respect to Lebesgue measure* μ *if* $\mu(A) = 0$ *implies* $\nu(A) = 0$.

The following result motivates the name "absolute continuity."

Theorem 4.3.5. *If* ν *is a measure defined on* $\mathcal{M}(I)$ *which is absolutely continuous with respect to Lebesgue measure, then for any* $\epsilon > 0$ *there is a* $\delta > 0$ *such that* $\nu(A) < \epsilon$ *whenever* $\mu(A) = \delta$.

Proof. We assume there is a counter-example and show this leads to a contradiction. If the measure ν does not satisfy the conclusion of the theorem, then there is an $\epsilon > 0$ for which it fails, i.e. there is no $\delta > 0$ which works for this ϵ. In particular, for any positive integer m there is a set B_m such that $\nu(B_m) \geq \epsilon$ 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. $A_n \supset A_{n+1}$. It follows from Proposition (2.4.6) that

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

The proof of Proposition (2.4.6) made use only of the countable additivity of the measure. Hence it is also valid for ν, i.e.

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

On the other hand $\nu(A_n) \geq \nu(B_{n+1}) \geq \epsilon$, so

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

This together with equation (4.3.1) contradicts the absolute continuity of ν with respect to μ. We have proven the contrapositive of the result we desire. □

Exercise 4.3.6. Given a point $x_0 \in [0,1]$ define the function $\delta_{x_0} : \mathcal{M} \to \mathbb{R}$ by $\delta_{x_0}(A) = 1$ if $x_0 \in A$ and $\delta_{x_0}(A) = 0$ if $x_0 \notin A$. Let $\nu(A) = \delta_{x_0}(A)$.

1. Prove that ν is a measure.
2. Prove that if f is a measurable function $\int f \, d\nu = f(x_0)$.
3. Prove that ν is not absolutely continuous with respect to Lebesgue measure μ.

The measure ν is called the *Dirac δ-measure.*

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 ν_f is a measure with σ-algebra $\mathcal{M}(I)$ which is absolutely continuous with respect to Lebesgue measure μ.

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 (4.2.3)

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

and so

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

Thus ν 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 ν is absolutely continuous with respect to μ. □

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 5 of Royden's book [4]

Theorem 4.3.8 (Radon-Nikodym)**.** *If ν is a measure with σ-algebra $\mathcal{M}(I)$ which is absolutely continuous with respect to Lebesgue measure μ, then there is a non-negative integrable function f on $[0, 1]$ such that define*

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

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 ν with respect to μ. In fact the Radon-Nikodym Theorem is more general than we have stated, since it applies to any two finite measures ν and μ defined on a σ-algebra $\mathcal{A}$ with ν absolutely continuous with respect to μ.

4.4 General Measurable Functions

In this section we consider extended measurable functions which may be unbounded both above and below. we will define

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

These are both non-negative measurable functions.

Definition 4.4.1. *If $f : [0,1] \to \mathbb{R}$ 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. *The measurable function $f : [0,1] \to \mathbb{R}$ is integrable if and only if the the function $|f|$ is integrable.*

Proof. Notice that $|f(x)| = f^+(x) + f^-(x)$. Thus if $|f|$ is integrable, since $|f(x)| \geq f^+(x)$ and $|f(x)| \geq f^-(x)$ 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 and g is an non-negative 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 $f_n^+(x) = \max\{f_n(x), 0\}$ and $f_n^-(x) = -\min\{f_n(x), 0\}$ satisfy

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

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)

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

Thus 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 $\epsilon > 0$ any integrable function can be approximated within ϵ by a step function if we are allowed to exclude a set of measure ϵ.

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

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

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

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

and hence

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

Also we may choose $N > 0$ such that

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

Let $V_N = \cup_{i=1}^N U_i$. 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 $\mu(A) < \epsilon$. 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 ϵ).

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

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

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. 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 $\lim \mu(A_n) = 0$. Hence there is an $N > 0$ such that $\mu(A_N) < \epsilon/2$.

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

$$|f(x) - g(x)| \leq |f(x) - f_N(x)| + |f_N(x) - g(x)| = |f_N(x) - g(x)| < \epsilon.$$

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 $|f(x)| \leq M$ for all x and g satisfies the conclusion of our theorem, then we define

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

The function g_1 is a step function with $|g_1(x)| \leq M$ and $g_1(x) = g(x)$ except when $|g(x)| > M$. Note if $g(x) > M$ and $x \notin A$ then $f(x) \leq M = g_1(x) < g(x)$ so $|g_1(x) - f(x)| < \epsilon$. The case $g(x) < -M$ is similar.

□

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

I. Linearity: *If f and g are Lebesgue measurable 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 $f(x) \leq g(x)$ for all x, then $\int f \; d\mu \leq \int g \; d\mu$.*

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

IV. Null Sets: *If f and g are bounded functions and $f(x) = g(x)$ except on a set of measure zero, then f is measurable if and only if g is measurable. If they are measurable, then $\int f \; d\mu = \int g \; d\mu$.*

The proof is left as an exercise.

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. Prove that if $f : [0, 1] \to \mathbb{R}$ is an integrable function, then given $\epsilon > 0$ there exists a continuous function $g : [0, 1] \to \mathbb{R}$ and a set A such that $\mu(A) < \epsilon$, $|f(x) - g(x)| < \epsilon$ for all $x \notin A$, and $g(0) = g(1)$.

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

4. Prove Theorem (4.4.6). Proposition (4.4.2) proves the null set property. Prove the remaining parts of this theorem, namely linearity, monotonicity, and the absolute value property. (You may use Theorem (3.2.5)).

Chapter 5

The Hilbert Space $L^2[-1,1]$

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 be a finite set with n elements, say, $X = \{1, 2, 3, \dots, n\}$ and we define a measure ν on X which is called the "counting measure".

More precisely, we take as σ-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, there is a vector space isomorphism of $L^2(X)$ and $\mathbb{R}^n$ given by $f \longleftrightarrow (x_1, x_2, \dots, x_n)$ where $x_i = f(i)$.

Under this isomorphism it is important to note what the inner product (or "dot" product $< x, y >= \sum_{i=1}^n x_i y_i$ becomes. 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

$$< x, y >= \sum_{i=1}^{n} x_i y_i = \sum_{i=1}^{n} f(i)g(i) = \int fg \ d\nu.$$

Also the norm (or length) of a vector is given by

$$\|x\|^2 =< x, x >= \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} f(i)^2 = \int f^2 \ d\nu.$$

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 this chapter it will be convenient (for notational purposes) to consider functions on the interval $[-1, 1]$ 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. *A measurable function $f : [-1, 1] \to \mathbb{R}$ is called* square integrable *if $f(x)^2$ is integrable. We denote the set of all square integrable functions by $L^2[-1, 1]$. We define the* norm *of $f \in L^2[-1, 1]$ by*

$$\|f\| = \Big(\int f^2 \ d\mu \Big)^{\frac{1}{2}}.$$

Proposition 5.1.2. *The norm $\| \ \|$ on $L^2[-1, 1]$ satisfies $\|af\| = |a|\|f\|$ for all $a \in \mathbb{R}$ and all $f \in L^2[-1, 1]$. Moreover for all f, $\|f\| \geq 0$ with equality only if $f = 0$ almost everywhere.*

Proof. We see

$$\|af\| = \Big(\int a^2 f^2 \ d\mu \Big)^{\frac{1}{2}} = \sqrt{a^2} \Big(\int f^2 \ d\mu \Big)^{\frac{1}{2}} = |a|\|f\|.$$

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 $f, g \in L^2[-1,1]$, then fg is integrable and*

$$2\int |fg|\ d\mu \le \|f\|^2 + \|g\|^2.$$

Equality holds if and only if $|f| = |g|$ almost everywhere.

Proof. Since

$$0 \le (|f(x)| - |g(x)|)^2 = f(x)^2 - 2|f(x)g(x)| + g(x)^2$$

we have $2|f(x)g(x)| \le f(x)^2 + g(x)^2$. Hence by Proposition (4.1.4) we conclude that $|fg|$ is integrable and that

$$2\int |fg|\ d\mu \le \|f\|^2 + \|g\|^2.$$

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. *$L^2[-1,1]$ is a vector space.*

Proof. We must show that if $f, g \in L^2[-1,1]$ and $c \in \mathbb{R}$, then $cf \in L^2[-1,1]$ and $(f+g) \in L^2[-1,1]$. The first of these is clear since f^2 integrable implies that c^2f^2 is integrable.

To check the second we observe that

$$(f+g)^2 = f^2 + 2fg + g^2 \le f^2 + 2|fg| + g^2.$$

Since f^2, g^2 and $|fg|$ are all integrable, it follows from Proposition (4.1.4) that $(f+g)^2$ is also. Hence $(f+g) \in L^2[-1,1]$. □

Theorem 5.1.5 (Hölder Inequality)**.** *If $f, g \in L^2[-1,1]$, then*

$$\int |fg|\ d\mu \le \|f\|\ \|g\|.$$

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. *If $f, g \in L^2[-1,1]$, then*

$$\Big|\int fg\ d\mu\Big| \leq \|f\|\ \|g\|.$$

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

$$\Big|\int fg\ d\mu\Big| \leq \int |fg|\ d\mu \leq \|f\|\ \|g\|.$$

Equality holds when both of these inequalities are equalities. If 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 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 vector space $L^2[-1, 1]$.

Theorem 5.1.7 (Minkowski's Inequality)**.** *If $f, g \in L^2[-1, 1]$, 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 $L^2[-1, 1]$)**.** *If $f, g \in L^2[-1, 1]$, then we define their* inner product *by*

$$\langle f, g \rangle = \int fg \, d\mu.$$

Theorem 5.1.9 (The Inner Product on $L^2[-1, 1]$)**.** *For any $f_1, f_2, g \in L^2[-1, 1]$ and any $c_1, c_2 \in \mathbb{R}$ the inner product on $L^2[-1, 1]$ satisfies the following properties:*

1. Commutativity*:* $\langle f, g \rangle = \langle g, f \rangle$.
2. Bi-linearity*:* $\langle c_1 f_1 + c_2 f_2, g \rangle = c_1 \langle f_1, g \rangle + c_2 \langle f_2, g \rangle$.
3. Positive Definiteness*:* $\langle g, g \rangle = \|g\|^2 \geq 0$ *with equality if and only if* $g = 0$ *almost everywhere.*

Proof. Clearly $\langle f,g\rangle = \int fg\ d\mu = \int gf\ d\mu = \langle g,f\rangle$. Bi-linearity 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. ☐

Notice that we have *almost* proved that $L^2[-1,1]$ is an inner product space. The one point where the definition is not quite satisfied is that $\|f\| = 0$ implies $f = 0$ *almost everywhere* rather than everywhere. The pedantic way to overcome this problem is to define $L^2[-1,1]$ 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 $L^2[-1,1]$ 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.

5.2 Convergence in $L^2[-1,1]$

We have discussed uniform convergence and pointwise convergence and now we wish to discuss convergence in the $L^2[-1,1]$ norm $\|\ \|$. The vector space $L^2[-1,1]$ is, of course, a metric space with distance function given by $\text{dist}(f,g) = \|f-g\|$. Note that $\text{dist}(f,g) = 0$ if and only if $f = g$ almost everywhere, so again if we wish to be pedantic this metric space is really the equivalence classes of functions which are equal almost everywhere.

Definition 5.2.1. *If $\{f_n\}_{n=1}^\infty$ is a sequence in $L^2[-1,1]$, then it is said to converge to in measure of order 2 or to converge in $L^2[-1,1]$ if there is a function $f \in L^2[-1,1]$ such that*

$$\lim_{n\to\infty} \|f - f_n\| = 0.$$

Lemma 5.2.2 (Density of Bounded Functions)**.** *If we define*

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

then

$$\lim_{n\to\infty} \|f - f_n\| = 0.$$

Proof. We will show that for any $\epsilon > 0$ there is an n such that $\|f - f_n\|^2 < \epsilon$. First we note that $|f_n(x)| \leq |f(x)|$ so

$$|f(x) - f_n(x)|^2 \leq |f(x)|^2 + 2|f(x)|\ |f_n(x)| + |f(x)|^2 \leq 4|f(x)|^2.$$

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 $\int_A |f|^2\ d\mu < \epsilon/4$ whenever $\mu(A) < \delta$. Thus we have

$$\|f - f_n\|^2 = \int |f - f_n|^2\ d\mu = \int_{E_n} |f - f_n|^2\ d\mu \leq \int_{E_n} 4|f|^2\ d\mu < 4\frac{\epsilon}{4} = \epsilon$$

whenever n is sufficiently large that $\mu(E_n) \leq C/n^2 < \delta$. □

Proposition 5.2.3 (Density of Step Functions and Continuous Functions)**.** *The step functions are dense in $L^2[-1,1]$. That is, for any $\epsilon > 0$ and any $f \in L^2[-1,1]$ there is a step function $g : [-1,1] \to \mathbb{R}$ such that $\|f - g\| < \epsilon$. Likewise, there is a continuous function $h : [-1,1] \to \mathbb{R}$ such that $\|f - h\| < \epsilon$. The function h may be chosen so $h(-1) = h(1)$.*

Proof. By the preceding result we may choose n so that $\|f - f_n\| < \frac{\epsilon}{2}$. Note that $|f_n(x)| \leq n$ for all x. Suppose now that δ 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

$$\begin{aligned}
\|f_n - g\|^2 &= \int |f_n - g|^2\ d\mu \\
&= \int_A |f_n - g|^2\ d\mu + \int_{A^c} |f_n - g|^2\ d\mu \\
&\leq \int_A 4n^2\ d\mu + \int_{A^c} \delta^2\ d\mu \\
&\leq 4n^2 \mu(A) + \delta^2 \mu(A^c) \leq 4n^2\delta + 2\delta^2.
\end{aligned}$$

Clearly if we choose δ sufficiently small, then

$$\|f_n - g\| \leq \sqrt{4n^2\delta + 2\delta^2} < \frac{\epsilon}{2}.$$

It follows that $\|f-g\| \leq \|f-f_n\| + \|f_n - g\| < \epsilon$.

The proof for continuous functions is the same, except Exercise (4.4.7) is used in place of Theorem (4.4.5). The details are left as an exercise. ☐

Definition 5.2.4. *An inner product space* $(\mathcal{V}, \langle\ ,\ \rangle)$ *which is* complete*, i.e. in which Cauchy sequences converge, is called a* Hilbert space*.*

For example, $\mathbb{R}^n$ with the usual dot product is a Hilbert space.

We want to prove that $L^2[-1,1]$ is a Hilbert space.

Theorem 5.2.5. $L^2[-1,1]$ *is a Hilbert space.*

Proof. We have already shown that $L^2[-1,1]$ is an inner product space. All that remains is to prove that the norm $\|\ \|$ is complete, i.e. that Cauchy sequences converge.

Let $\{f_n\}_{n=1}^\infty$ be a Cauchy sequence. Then we may choose numbers n_i such that $\|f_m - f_n\| < 1/2^i$ whenever $m, n \geq n_i$. Hence if we define $g_0 = 0$ and $g_i = f_{n_i}$ for $i > 0$, then $\|g_{i+1} - g_i\| < 1/2^i$ 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

$$\|h_n\| \leq \sum_{i=0}^{n-1} \|g_{i+1} - g_i\| < S.$$

Hence $\int h_n^2\, d\mu = \|h_n\|^2 < S^2$. 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 (4.2.2) that $\displaystyle\int h^2\, d\mu = \lim_{n\to\infty} \int h_n^2\, d\mu < S^2$ 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 by Theorem (0.2.6). 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

$$|g(x)| = \lim_{n\to\infty} |g_n(x)| \leq \lim_{n\to\infty} \sum_{i=0}^{n-1} |g_{i+1}(x) - g_i(x)| = \lim_{n\to\infty} h_n(x) = h(x)$$

for almost all x so $|g(x)|^2 \leq h(x)^2$ and hence $|g(x)|^2$ is integrable and $g \in L^2[-1,1]$.

We also observe that

$$|g(x) - g_n(x)|^2 \leq (|g(x)| + |g_n(x)|)^2 \leq (2h(x))^2.$$

Since $\lim\limits_{n\to\infty} |g(x) - g_n(x)|^2 = 0$ for almost all x the Lebesgue Convergence Theorem (4.4.4) tells us $\lim\limits_{n\to\infty} \int |g(x) - g_n(x)|^2 \, d\mu = 0$. This implies $\lim\limits_{n\to\infty} \|g - g_n\| = 0$.

Hence given $\epsilon > 0$ there is an i such that $\|g - g_i\| < \epsilon/2$ and $1/2^i < \epsilon/2$. Recalling that $g_i = f_{n_i}$ we see that whenever $m \geq n_i$ we have $\|g - f_m\| \leq \|g - g_i\| + \|g_i - f_m\| < \epsilon/2 + \epsilon/2 = \epsilon$. Hence $\lim\limits_{m\to\infty} \|g - f_m\| = 0$. □

5.3 Hilbert Space

In any Hilbert space we can, of course, talk about convergent sequences and series. The meaning is precisely what you would expect. In particular, if $\mathcal{H}$ is a Hilbert space and $\{x_n\}$ is a sequence, then

$$\lim_{n\to\infty} x_n = x$$

means that for any $\epsilon > 0$ there is an $N > 0$ such that $\|x - x_n\| < \epsilon$ whenever $n \geq N$. This is exactly the usual definition in $\mathbb{R}$ except we use the norm $\| \ \|$ in place of absolute value. Also if $\{u_n\}$ 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 $\sum_{m=1}^{\infty} \|u_m\|$ converges.

Proposition 5.3.1. *If a series in a Hilbert space converges absolutely then it converges.*

Proof. Given $\epsilon > 0$ there is an $N > 0$ such that whenever $n > m \geq N$,

$$\sum_{i=m}^{n} \|u_m\| \leq \sum_{i=m}^{\infty} \|u_m\| < \epsilon.$$

Let $s_n = \sum_{i=1}^{n} u_i$, then $\|s_n - s_m\| \leq \sum_{i=m}^{n} \|u_m\| < \epsilon$. 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. *If $\mathcal{H}$ is a Hilbert space, a* bounded linear functional *on $\mathcal{H}$ is a function $L : \mathcal{H} \to \mathbb{R}$ such that for all $v, w \in \mathcal{H}$ and $c_1, c_2 \in \mathbb{R}$, $L(c_1 v + c_2 w) = c_1 L(u) + c_2 L(w)$ and such that there is a constant M satisfying $|L(v)| \leq M\|v\|$ for all $v \in \mathcal{H}$.*

The following result was proved in Proposition (0.7.3). In the case of the Hilbert space $L^2[-1,1]$ it is just the corollary to Hölder's inequality, 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 functions from $\mathcal{H}$ to $\mathbb{R}$.

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 a limit in $\mathcal{V}$. To see this suppose $\lim x_n = x$ and $x_n \in \mathcal{V}$. Then $|L(x) - L(x_n)| = |L(x - x_n)| \leq M\|x - x_n\|$ for some M. Hence since $L(x_n) = 1$ for all n, we have $|L(x) - 1| \leq \lim M\|x - x_n\| = 0$. 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 $\|(x_n + x_m)/2\| \geq d$ or $\|x_n + x_m\| \geq 2d$. By the parallelogram law (Proposition (0.7.4)

$$\|x_n - x_m\|^2 + \|x_n + x_m\|^2 = 2\|x_n\|^2 + 2\|x_m\|^2.$$

Hence

$$\|x_n - x_m\|^2 = 2\|x_n\|^2 + 2\|x_m\|^2 - \|x_n + x_m\|^2 \leq 2\|x_n\|^2 + 2\|x_m\|^2 - 4d^2.$$

As m and n tend to infinity the right side of this equation 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 limit limit of this sequence. Since $\|x\| \leq \|x - x_n\| + \|x_n\|$ for all n, we have

$$\|x\| \leq \lim_{n\to\infty} \|x - x_n\| + \lim_{n\to\infty} \|x_n\| = d.$$

But $x \in \mathcal{V}$ implies $\|x\| \geq d$ 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 $\|x + y\| \geq 2d$. Hence using the parallelogram law again

$$\|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2 - \|x + y\|^2 \leq 4d^2 - 4d^2 = 0.$$

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 $\|x + tv\|^2 \geq \|x\|^2$. Hence

$$\|x\|^2 + 2t < x, v > +t^2\|v\|^2 \geq \|x\|^2, \text{ so}$$

$2t < x, v > +t^2\|v\|^2 \geq 0$ for all $t \in \mathbb{R}$. This is possible only if $< x, v >= 0$. ☐

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 $x = y$.

☐

5.4 Fourier Series

Definition 5.4.1. *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. *The family of functions* $\mathcal{F} = \{\frac{1}{\sqrt{2}}, \cos(n\pi x), \sin(n\pi x)\}_{n=1}^{\infty}$ *is an orthonormal family in* $L^2[-1,1]$.

For a proof see Chapter 1 of [3].

Theorem 5.4.3. *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

$$u = \sum_{n=0}^{N} a_n u_n, \text{ and } v = \sum_{n=0}^{N} c_n u_n.$$

Notice that by Theorem (5.3.2) $\langle u, u \rangle = \sum_{n=0}^{N} a_n^2$ 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 only if $c_n = a_n = \langle w, u_n \rangle$. That is, for all choices of v, the minimum value of $\|w - v\|^2$ occurs precisely when $v = u$. □

Definition 5.4.4. *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 (5.4.3) 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 5.4.5 (Fourier Series)**.** *The* Fourier coefficients *of w with respect to an orthonormal family $\{u_n\}_{n=0}^{\infty}$ are the numbers $\langle w, u_n \rangle$. The infinite series*

$$\sum_{n=0}^{\infty} \langle w, u_n \rangle u_n$$

is called the Fourier series*.*

Example 5.4.6 (Classical Fourier Series)**.** We will show later that the orthnormal family of functions $\mathcal{F} = \{\frac{1}{\sqrt{2}}, \cos(n\pi x), \sin(n\pi x), \}_{n=1}^{\infty}$ is complete. If $f \in L^2[-1,1]$, then the Fourier coefficients are

$$A_0 = \frac{1}{\sqrt{2}} \int f \; d\mu$$
$$A_n = \int f \cos(n\pi x) \; d\mu \text{ for } n > 0$$
$$B_n = \int f \sin(n\pi x) \; d\mu \text{ for } n > 0 \; ,$$

and the Fourier series is

$$\frac{1}{\sqrt{2}} A_0 + \sum_{n=1}^{\infty} A_n \cos(n\pi x) + \sum_{n=1}^{\infty} B_n \sin(n\pi x)$$

Theorem 5.4.7 (Bessel's Inequality)**.** *If $\{u_i\}_{i=0}^{\infty}$ is an orthonormal family of vectors in a Hilbert space $\mathcal{H}$ and $w \in \mathcal{H}$, then the series*

$$\sum_{i=0}^{\infty} \langle w, u_i \rangle^2 \leq \|w\|^2.$$

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 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 Theorem (5.3.2) 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. □

Corollary 5.4.8. *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.*

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 $\epsilon > 0$ there is an $N > 0$ such that $\|s_n - s_m\|^2 < \epsilon^2$ whenever $n, m \geq N$. In other words the sequence $\{s_n\}$ is Cauchy. □

If Bessel's inequality is actually an equality, then the Fourier series for w must converge to w in $\mathcal{H}$.

Theorem 5.4.9 (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, then, $\lim \|w - s_n\| = 0$ 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 $\sum_{n=0}^{\infty} \langle w, u_n \rangle^2 = \|w\|^2$. □

Recall that an *algebra of functions* is a vector space $\mathcal{A}$ of real valued functions with the additional property that if $f, g \in \mathcal{A}$, then $fg \in \mathcal{A}$. If $X = [a, b]$ is a closed interval in $\mathbb{R}$ we will denote by $C(X)$ the algebra of all continuous functions from X to $\mathbb{R}$ and by $C_{end}(X) = \{f \mid f(a) = f(b)\}$, the subalgebra of functions which agree at the endpoints. The following theorem is a special case of a much more general theorem called the Stone-Weierstrass Theorem.

Theorem 5.4.10. *Suppose that* $X = [-1, 1]$ *and* $\mathcal{A} \subset C_{end}(X)$ *is an algebra satisfying*

1. *The constant function* 1 *is in* $\mathcal{A}$, *and*
2. $\mathcal{A}$ *separates points (except endpoints): for any distinct* $x, y \in X$ *with* $\{x, y\} \neq \{-1, 1\}$ *there is* $f \in \mathcal{A}$ *such that* $f(x) \neq f(y)$.

Then $\mathcal{A}$ *is dense in* $C(X)$, *i.e. given any* $\epsilon > 0$ *and any* $g \in C(X)$ *there is* $f \in \mathcal{A}$ *such that* $|f(x) - g(x)| < \epsilon$ *for all* $x \in X$.

A proof can be found in 5.8.2 of [2] or in [4]. This result is usually stated in greater generality than we do here. For example the set X need only be a compact metric space, but since we have not defined these concepts we state only the special case above.

Corollary 5.4.11. *If* $\epsilon > 0$ *and* $g : [-1, 1] \to \mathbb{R}$ *is a continuous function satisfying* $g(-1) = g(1)$, *then there are* $a_n, b_n \in \mathbb{R}$ *such that* $|g(x) - p(x)| < \epsilon$, *for all* x, *where*

$$p(x) = a_0 + \sum_{n=1}^{N} a_n \cos(n\pi x) + \sum_{n=1}^{N} b_n \sin(n\pi x).$$

Proof. Let X be the unit circle in the plane $\mathbb{R}^2$, i.e. $X = \{(\cos(\pi x), \sin(\pi x)) \mid x \in [-1,1]\}$. So if $\phi : [-1,1] \to \mathbb{R}^2$ is given by $\phi(x) = (\cos(\pi x), \sin(\pi x))$, then $X = \phi([-1,1])$. For any function $f : [-1,1] \to \mathbb{R}$, with $f(-1) = f(1)$ we define $\hat{f} : X \to \mathbb{R}$ to be the continuous function such that $\hat{f}(\phi(x)) = f(x)$. We need the fact that $f(-1) = f(1)$ because $\phi(-1) = \phi(1)$. Conversely given any function $\hat{h} \in C(X)$ we can define $h : [-1,1] \to \mathbb{R}$ by $h(x) = \hat{h}(\phi(x))$ and we will have $h(-1) = h(1)$.

Let $\mathcal{A}$ be the collection of all functions on $[-1,1]$ of the form

$$q(x) = a_0 + \sum_{n=1}^{N} a_n \cos(n\pi x) + \sum_{n=1}^{N} b_n \sin(n\pi x).$$

for some choices of N, a_n, and b_n. Then $\mathcal{A}$ is a vector space and contains the constant function 1. It is an algebra as a consequence of the trigonometric identities

$$\sin(x)\cos(y) = \frac{1}{2}\big(\sin(x+y) + \sin(x-y)\big)$$
$$\cos(x)\cos(y) = \frac{1}{2}\big(\cos(x+y) + \cos(x-y)\big)$$
$$\sin(x)\sin(y) = \frac{1}{2}\big(\cos(x+y) - \cos(x-y)\big)$$

It is also the case that $\mathcal{A}$ separates points with the exception of the one pair of points $x = -1,\ y = 1$. To see this note that if x and y are not this pair and if one is positive and one negative, then $\sin(\pi x) \neq \sin(\pi y)$. On the other hand if both are ≥ 0 or both ≤ 0, then $\cos(\pi x) \neq \cos(\pi y)$.

It follows that if $\hat{\mathcal{A}} = \{\hat{q} \mid q \in \mathcal{A}\}$, then $\hat{\mathcal{A}}$ is an algebra which separates points of X and contains the constaint function 1. Note that the points $x = -1,\ y = 1$ correspond to a single point of X, namely $(-1,0) = \phi(-1) = \phi(1)$. So they cause no problem. Thus $\hat{\mathcal{A}}$ satisfies the hypothesis of the Stone-Weierstrass theorem.

Hence given $\epsilon > 0$ and $g : [-1,1] \to \mathbb{R}$, a continuous function satisfying $g(-1) = g(1)$, we consider $\hat{g}$. By Stone-Weierstrass there is a $\hat{p} \in \hat{\mathcal{A}}$ such that the value of $\hat{p}$ differs from the value of $\hat{g}$ by less than ϵ for all points of X. Thus if $p(x) = \hat{p}(\phi(x))$ we have $|g(x) - p(x)| = |\hat{g}(\phi(x)) - \hat{p}(\phi(x))| < \epsilon$ and $p \in \mathcal{A}$. □

Theorem 5.4.12. *If $f \in L^2[-1,1]$, then the Fourier series for f with respect to the orthonormal family $\mathcal{F}$ converges to f in $L^2[-1,1]$. In particular the orthonormal family $\mathcal{F}$ is complete.*

Proof. Given $\epsilon > 0$, we know by Proposition (5.2.3) there is a continuous function $g \in L^2[-1,1]$ such that $g(-1) = g(1)$ and $\|f - g\| < \epsilon/2$.

By the corollary to the Stone-Weierstrass theorem there is a function

$$p(x) = a_0 + \sum_{n=1}^{N} a_n \cos(n\pi x) + \sum_{n=1}^{N} b_n \sin(n\pi x).$$

with $|g(x) - p(x)| < \epsilon/4$ for all x. So

$$\|g - p\|^2 = \frac{1}{\pi}\int (g-p)^2 \, d\mu \le \frac{1}{\pi}\int \frac{\epsilon^2}{16} \, d\mu = \frac{\epsilon^2}{8}.$$

Hence, $\|f - p\| \le \|f - g\| + \|g - p\| < \epsilon/2 + \epsilon/\sqrt{8} < \epsilon$.

Let

$$S_N(x) = \frac{1}{\sqrt{2}} A_0 + \sum_{n=1}^{N} A_n \cos(n\pi x) + \sum_{n=1}^{N} B_n \sin(n\pi x)$$

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 (5.4.3) for every $m \ge N$, $\|f - S_m\| \le \|f - p\|$ so $\|f - S_m\| < \epsilon$. This proves $\lim \|f - S_m\| = 0$. □

Exercise 5.4.13. Suppose $X = [-1,1]$.

1. Prove that $C_{end}(X)$ is a subalgebra of $C(X)$, i.e. it is a vector subspace closed under multiplication.

2. Let $\mathcal{A}_p$ be the polynomials

Appendix A

Lebesgue Measure

A.1 Introduction

We want to 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 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 $\mu(A) = \text{len}(A) = b - a$, 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 finite collections $\{A_n\}_{n=1}^{m}$ 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 $B \setminus A$ are disjoint and their union is B.

It turns out that it is not possible to find a μ 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.

A.2 Outer Measure

We first describe the notion of "outer measure" which comes close to what we want. It is defined for *all* bounded sets of the reals and satisfies properties I and II 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 (called measurable) on which the additivity condition holds. Our task is to develop the definition of *measurable set*, to define the notion of *Lebesgue measure* for such a set and, then to 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 μ satisfying the properties I-IV above we would expect 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 $\mu(A)$ is 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 be a very useful definition.

Definition A.2.1 (Lebesgue Outer Measure)**.** *Suppose $A \subset \mathbb{R}$ is a bounded set 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.

Notice that this definition together with the definition of a null set, Definition (2.2.1), says 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 A.2.2. *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 $U_1 = (a - \epsilon, b + \epsilon)$ so $\mu^*([a,b]) \leq \text{len}(U_1) = b - a + 2\epsilon$. Since 2ϵ is arbitrary we conclude that $\mu^*([a,b]) \leq b - a$.

On the other hand by 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 $\sum \text{len}(U_i) \geq b - a$ 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 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

$$\sum_{i=1}^{n-1} \text{len}(V_i) \geq b - a.$$

But $\text{len}(U_{n-1}) + \text{len}(U_n) > \text{len}(V_{n-1})$ and $\text{len}(U_i) = \text{len}(V_{i-1})$ for $i > 2$. Hence

$$\sum_{i=1}^{n} \text{len}(U_i) > \sum_{i=1}^{n-1} \text{len}(V_i) \geq b - a.$$

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 $[a + \epsilon, b - \epsilon]$ so, as we just showed,

$$\sum_{i=1}^{\infty} \text{len}(U_i) \geq b - a - 2\epsilon.$$

As ϵ is arbitrary $\sum \text{len}(U_i) \geq b - a$ and hence $\mu^*((a,b)) \geq b - a$ which completes our proof. ∎

Two special cases are worthy of note:

Corollary A.2.3. *The outer measure of a set consisting of 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 A.2.4. *If A and B are bounded 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 $\mu^*(A) \leq \mu^*(B)$. □

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 bounded sets. 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 A.2.5 (Countable Subadditivity)**.** *If $\{A_n\}_{n=1}^{\infty}$ is a countable collection of bounded subsets of $\mathbb{R}$, then*

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

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

$$\sum_{i=1}^{\infty} \operatorname{len}(U_i^n) \leq \mu^*(A_n) + 2^{-n}\epsilon.$$

But the union of all these covers $\{U_i^n\}$ is a countable cover of $\bigcup_{n=1}^{\infty} A_n$. So

$$\mu^*(\bigcup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} \sum_{i=1}^{\infty} \operatorname{len}(U_i^n) \leq \sum_{n=1}^{\infty} \mu^*(A_n) + \sum_{n=1}^{\infty} 2^{-n}\epsilon = \sum_{n=1}^{\infty} \mu^*(A_n) + \epsilon.$$

Since this is true for every ϵ the result follows. The result for a finite collection $\{A_n\}_{n=1}^{m}$ follows from this by letting $A_i = \emptyset$ for $i > m$. □

Corollary A.2.6. *If A is countable, then $\mu^*(A) = 0$*

Proof. Suppose $A = \bigcup_{i=1}^{\infty}\{x_i\}$. We saw in Corollary (A.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 $\mu^*(A) = 0$. □

Since countable sets have outer measure 0 and $\mu^*([a, b]) = b - a$ we also immediately obtain the following non-trivial result (cf. part 4. of Exercise (0.3.13)).

Corollary A.2.7. *If $a < b$, then $[a, b]$ is not countable.*

Outer Lebesgue measure satisfies property II of those we enumerated at the beginning, namely it is translation invariant.

Theorem A.2.8. *If $c \in \mathbb{R}$ and A is a bounded 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.

Exercise A.2.9.

1. Prove Theorem (A.2.8).
2. Prove that given $\epsilon > 0$ 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 $\sum_{n=1}^{\infty} \operatorname{len}(U_n) = \epsilon$.
3. Give an example of a subset A of I 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 $\sum_{i=1}^{n} \operatorname{len}(U_i) \geq 1$.

A.3 Lebesgue Measurable Sets

In Definition (2.4.1) we defined the σ-algebra $\mathcal{M}$ to be the σ-algebra of subsets of $\mathbb{R}$ generated by open intervals and null sets (it is also the σ-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 σ-algebra. However now, in order to prove the existence of Lebesgue measure, we want to use a different, but equivalent definition.

Our program is roughly as follows:

- We will define a collection $\mathcal{M}_0$ of subsets of I The criterion used to define $\mathcal{M}_0$ is often given as the definition of Lebesgue measurable sets.
- We will define the Lebesgue measure $\mu(A)$ of a set A in $\mathcal{M}_0$ to be the outer measure of A.
- We will show that the collection $\mathcal{M}_0$ is a σ-algebra of subsets of I and in fact precisely the σ-algebra $\mathcal{M}(I)$ so we have defined $\mu(A)$ for all $A \in \mathcal{M}(I)$.
- We will prove that μ defined in this way satisfies the properties promised in Chapter 2, namely properties I-V of Theorem (2.4.2). Several of these properties follow from the corresponding properties for outer measure μ^*, which we proved in Section (A.2).

Henceforth for definiteness we will consider subsets of the unit interval $I = [0, 1]$. We could, of course, use any other closed interval or even, with some extra work, the whole real line. Lebesgue outer measure as in Definition (A.2.1) has most of the properties we want. 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 sets whose union is an interval.

Fortunately, the sets for which it fails are rather exotic and not too frequently encountered. Our strategy is to restrict our attention to only certain subsets of I which we will call "measurable" and to show that on these sets μ^* has all the properties we want.

If $A \subset I$ we will denote the *complement* of A by A^c, that is,

$$A^c = I \setminus A = \{x \in I \mid x \notin A\}.$$

Definition A.3.1 (Alternate Definition of Lebesgue Measurable)**.** *Let $\mathcal{M}_0$ denote the collection of all subsets of I defined as follows: A subset A of I is in $\mathcal{M}_0$ provided for* any *subset $X \subset I$*

$$\mu^*(A \cap X) + \mu^*(A^c \cap X) = \mu^*(X).$$

For any set $A \in \mathcal{M}_0$ we define $\mu(A)$ to be $\mu^*(A)$.

The goal of the remainder of this section is to prove that in fact $\mathcal{M}_0$ is nothing other than the σ-algebra $\mathcal{M}(I)$ of Lebesgue measurable subsets of I and the function $\mu : \mathcal{M}_0 \to \mathbb{R}$ satisfies the properties for Lebesgue measure we claimed 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* subset of I 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 μ which we want and, then show $\mathcal{M}_0 = \mathcal{M}(I)$ so the two definitions coincide. Indeed, we will prove in Corollary (A.3.9) that the sets in $\mathcal{M}_0$ are precisely the sets in $\mathcal{M}(I)$ the σ-algebra generated by Borel subsets and null subsets of I.

As a first step we show that $\mathcal{M}_0 \subset \mathcal{M}(I)$.

Proposition A.3.2. *Every set $A \in \mathcal{M}_0$ can be written as*

$$A = B \setminus N$$

where B is the intersection of a countable nested family of open sets (and, in particular, is a Borel set) and $N = A^c \cap B$ is a null set. It follows that $\mathcal{M}_0 \subset \mathcal{M}(I)$.

Proof. Since $\mu(A) = \mu^*(A)$ it follows from the definition of outer measure that for any $\epsilon > 0$ there is a cover $\mathcal{V}_\epsilon$ of A by open intervals U_n such that

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

The monotonicity of outer measure, then implies that if $V_\epsilon = \cup_{n=1}^{\infty} U_n$ we have $\mu^*(A) \leq \mu^*(V_\epsilon) \leq \mu^*(A) + \epsilon$. If we let

$$W_k = \bigcap_{i=1}^{k} V_{\frac{1}{k}}$$

, then each $\{W_k\}$ is a nested family of open sets and $\mu^*(A) \leq \mu^*(W_k) \leq \mu^*(A) + 1/k$. open.

Let $B = \cap_{k=1}^{\infty} W_k$. By monotonicity again we have

$$\mu^*(A) \leq \mu^*(B) \leq \mu^*(V_k) < \mu^*(A) + \frac{1}{k}.$$

Since this holds for all $k > 0$ we conclude $\mu^*(B) = \mu^*(A)$.

In the defining equation of $\mathcal{M}_0$ (see Definition (A.3.1) we take $X = B$ and conclude

$$\mu^*(A \cap B) + \mu^*(A^c \cap B) = \mu^*(B).$$

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 $\mu^*(A^c \cap B) = 0$. Therefore if $N = A^c \cap B$, then N is a null set. Finally $A = B \setminus (B \cap A^c)$ so $A = B \setminus N$. $\square$

The definition of $\mathcal{M}_0$ is relatively simple, but to show it has the properties we want requires some work. If we were to replace the = sign in this definition with $\geq$ we would obtain a statement which is true for all subsets of I. So to prove a set is in $\mathcal{M}_0$ we need only check the reverse inequality. More precisely,

Proposition A.3.3. *Suppose $A \subset I$, then*

(1) The set A is in $\mathcal{M}_0$ provided for any *subset $X \subset I$*

$$\mu^*(A \cap X) + \mu^*(A^c \cap X) \leq \mu^*(X).$$

(2) The set A is in $\mathcal{M}_0$ if and only if A^c is in $\mathcal{M}_0$. In this case $\mu(A^c) = 1 - \mu(A)$.

Proof. For part (1) observe $X = (A \cap X) \cup (A^c \cap X)$ so the subadditivity property of outer measure in Theorem (A.2.5) 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 I. The fact that $(A^c)^c = A$ implies immediately from Definition (A.3.1) that A is in $\mathcal{M}_0$ if and only if A^c is. Also taking $X = I$ in this definition we conclude

$$\mu^*(A \cap I) + \mu^*(A^c \cap I) = \mu^*(I) = 1.$$

So $\mu(A) + \mu(A^c) = 1$. □

Proposition A.3.4. *A set $A \subset I$ is a null set if and only if $A \in \mathcal{M}_0$ and $\mu(A) = 0$.*

Proof. 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 (A.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 (A.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) Proposition (A.3.3. □

Proposition A.3.5. *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 if two sets, A and B, then there union is in $\mathcal{M}_0$ requires some work. Suppose $X \subset I$. And since $(A \cup B) \cap X = (B \cap X) \cup (A \cap B^c \cap X)$, the subadditivity of Theorem (A.2.5) tells us

$$\mu^*((A \cup B) \cap X) \leq \mu^*(B \cap X) + \mu^*(A \cap B^c \cap X). \tag{A.3.1}$$

Also the definition of $\mathcal{M}_0$ tells us

$$\mu^*(B^c \cap X) = \mu^*(A \cap B^c \cap X) + \mu^*(A^c \cap B^c \cap X). \tag{A.3.2}$$

Notice that $(A \cup B)^c = A^c \cap B^c$. So we get

$$\begin{aligned}
&\mu^*((A \cup B) \cap X) + \mu^*((A \cup B)^c \cap X) \\
&\qquad = \mu^*((A \cup B) \cap X) + \mu^*(A^c \cap B^c \cap X) \\
&\qquad \leq \mu^*(B \cap X) + \mu^*(A \cap B^c \cap X) + \mu^*(A^c \cap B^c \cap X) && \text{by equation (A.3.1)}, \\
&\qquad = \mu^*(B \cap X) + \mu^*(B^c \cap X) && \text{by equation (A.3.2)}, \\
&\qquad = \mu^*(X).
\end{aligned}$$

According to part (1) of Proposition (A.3.3) this implies 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 A.3.6. *Any subinterval of I, open, closed or half open,in $\mathcal{M}_0$.*

Proof. First consider $[0, a]$ with complement $(a, 1]$. If X is an arbitrary subset of I we must show $\mu^*([0, a] \cap X) + \mu^*((a, 1] \cap X) = \mu^*(X)$. Let $X^- = [0, a] \cap X$ and $X^+ = (a, 1] \cap X$. Given $\epsilon > 0$, 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

$$\sum_{n=1}^{\infty} \operatorname{len}(U_n) \leq \mu^*(X) + \epsilon. \tag{A.3.3}$$

Let $U_n^- = U_n \cap [0, a]$ and $U_n^+ = U_n \cap (a, 1]$. Then $X^- \subset \bigcup_{n=1}^\infty U_n^-$ and $X^+ \subset \bigcup_{n=1}^\infty U_n^+$. Subadditivity of outer measure implies

$$\mu^*(X^-) \leq \mu^*(\bigcup_{n=1}^{\infty} U_n^-) \leq \sum_{n=1}^{\infty} \operatorname{len}(U_n^-)$$

and

$$\mu^*(X^+) \leq \mu^*(\bigcup_{n=1}^{\infty} U_n^+) \leq \sum_{n=1}^{\infty} \operatorname{len}(U_n^+).$$

Adding these inequalities and using equation (A.3.3) we get

$$\begin{aligned} \mu^*(X^-) + \mu^*(X^+) &\leq \sum_{n=1}^{\infty} \operatorname{len}(U_n^-) + \operatorname{len}(U_n^+) \\ &= \sum_{n=1}^{\infty} \operatorname{len}(U_n) \\ &\leq \mu^*(X) + \epsilon. \end{aligned}$$

Since ϵ is arbitrary we conclude that $\mu^*(X^-)+\mu^*(X^+) \leq \mu^*(X)$ which by Proposition (A.3.3) implies that $[0, a]$ is in $\mathcal{M}_0$ for any $0 \leq a \leq 1$. A similar argument implies that $[a, 1]$ is in $\mathcal{M}_0$. Taking complements, unions and intersections it is clear that any interval, open closed or half open, is in $\mathcal{M}_0$. □

Lemma A.3.7. *Suppose A and B are disjoint sets in $\mathcal{M}_0$ and $X \subset I$ is arbitrary. Then*

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

The analogous result for a finite union of disjoint measurable sets is also valid.

Proof. It is always true that

$$A \cap (A \cup B) \cap X = A \cap X.$$

Since A and B are disjoint

$$A^c \cap (A \cup B) \cap X = B \cap X.$$

Hence the fact that A is in $\mathcal{M}_0$ tells us

$$\begin{aligned} \mu^*((A \cup B) \cap X) &= \mu^*(A \cap (A \cup B) \cap X) + \mu^*(A^c \cap (A \cup B) \cap X) \\ &= \mu^*(A \cap X) + \mu^*(B \cap X). \end{aligned}$$

The result for a finite collection $A_1, A_2, \ldots, A_n$ follows immediately by induction on n. □

Theorem A.3.8. *The collection $\mathcal{M}_0$ of subsets of I is closed under countable unions and countable intersections. Hence $\mathcal{M}_0$ is a σ-algebra.*

Proof. We have already shown 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 measurable. 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 I, then since F_n is in $\mathcal{M}_0$

$$\mu^*(X) = \mu^*(F_n \cap X) + \mu^*(F_n^c \cap X) \geq \mu^*(F_n \cap X) + \mu^*(F^c \cap X)$$

since $F^c \subset F_n^c$. By Lemma (A.3.7)

$$\mu^*(F_n \cap X) = \sum_{i=1}^{n} \mu^*(B_i \cap X).$$

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 μ^* 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 (A.3.3).

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 A.3.9. *The σ-algebra $\mathcal{M}_0$ of subsets of I equals $\mathcal{M}(I)$ the σ-algebra of subsets of I generated by Borel sets and null sets.*

Proof. The σ-algebra $\mathcal{M}_0$ contains open intervals and closed intervals in I by Proposition (A.3.6) and hence contains the σ-algebra they generate, the Borel subsets of I. Also $\mathcal{M}_0$ contains null sets by Proposition (A.3.4). Therefore $\mathcal{M}_0$ contains $\mathcal{M}(I)$ the σ-algebra $\mathcal{M}_0$ generated by Borel sets and null sets.

On the other hand by Proposition (A.3.2) $\mathcal{M}_0 \subset \mathcal{M}(I)$. Hence $\mathcal{M}_0 = \mathcal{M}(I)$. □

Since we now know the sets in $\mathcal{M}_0$, i.e. the sets which satisfy Definition (A.3.1), coincide with the sets in $\mathcal{M}[I]$, we will refer to them sets 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 A.3.10 (Countable Additivity)**.** *If $\{A_n\}_{n=1}^{\infty}$ is a countable collection of measurable subsets of I, then*

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

If the sets are pairwise disjoint, then

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

The same equality and inequality are valid for a finite collection of measurable subsets $\{A_n\}_{n=1}^{m}$.

Proof. The first inequality is simply a special case of the subadditivity from Theorem (A.2.5). If the sets A_i are pairwise disjoint, then by Lemma (A.3.7) 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.

Theorem A.3.11 (Existence of Lebesgue Measure)**.** *There exists a unique function μ, 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 $A \subset I$, $c \in \mathbb{R}$ and $A + c \subset I$ where $A + c$ denotes the set $\{x + c \mid x \in A\}$. Then $\mu(A + c) = \mu(A)$*

III. Countable Additivity. *If $\{A_n\}_{n=1}^{\infty}$ is a countable collection of subsets of I, 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 $A, B \in \mathcal{M}(I)$ and $A \subset B$ then $\mu(A) \leq \mu(B)$*

V. Null Sets *A subset $A \subset I$ is a null set set if and only if $A \in \mathcal{M}(I)$ and $\mu(A) = 0$.*

Proof. The Lebesgue measure $\mu(A)$ of any set $A \in \mathcal{M}(I)$ is defined to be its outer measure $\mu^*(A)$. Hence properties I, II, and IV for μ follow from the corresponding properties of μ^*. These were established in Propostion (A.2.2), Theorem (A.2.8), and Proposition (A.2.4) respectively.

Property III, countable additivity, was proved in Theorem (A.3.10). And Property V is a consequence of Proposition (A.3.4).

We are left with the task of showing that μ is unique. Suppose μ_1 and μ_2 are two functions defined on $\mathcal{M}(I)$ and satisfying properties I-V. They must agree on any open interval by property I. By Theorem (0.4.3) any open set is a countable union of pairwise disjoint open intervals, so countable additivity implies μ_1 and μ_2 agree on open sets.

Suppose that B is the intersection of a countable nested family of open sets $U_1 \supset U_2 \supset \cdots \supset U_n \ldots$. Then Proposition (2.4.6) implies

$$\mu_1(B) = \mu_1\big(\bigcap_{n=1}^{\infty} U_n\big) = \lim_{n\to\infty} \mu_1(U_n) = \lim_{n\to\infty} \mu_2(U_n) = \mu_2\big(\bigcap_{n=1}^{\infty} U_n\big) = \mu_2(B).$$

Also if $N \in \mathcal{M}(I)$ is a null set, then $\mu_1(N) = 0 = \mu_2(N)$ by property V.

Finally if A is an arbitrary set in $\mathcal{M}(I)$ by Proposition (A.3.2) $A = B \setminus N$ where B is the intersection of a countable nested family of open sets and $N = A^c \cap B$ is a null set. Since B is the disjoint union of A and N It follows that

$$\mu_1(A) = \mu_1(B) - \mu_1(N) = \mu_2(B) - \mu_2(N) = \mu_2(A).$$

$\square$

Appendix B

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 Section (0.3)) and is highly non-constructive.

Lemma B.1.1. *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 $\epsilon > 0$ the set Δ contains the interval* $(-\epsilon, \epsilon)$.

Proof. By Theorem (2.5.1) there is an open interval U such that $\mu(A \cap U) > \frac{3}{4}\operatorname{len}(U)$. Let $\epsilon = \operatorname{len}(U)/2$, so $\operatorname{len}(U) = 2\epsilon$. Suppose $y \in (-\epsilon, \epsilon)$ and let $U+y = \{x+y \mid x \in U\}$,, then $U \cup (U+y)$ is an interval of length at most 3ϵ.

Now let $B = A \cap U$ and $B' = B + y$. Then $\mu(B') = \mu(B) > \frac{3}{4}\operatorname{len}(U) = \frac{3}{2}\epsilon$ so $\mu(B') + \mu(B) > 3\epsilon$. On the other hand the fact that $B \cup B' \subset U \cup (U+y)$ implies $\mu(B \cup B') \leq 3\epsilon$. It follows that B and B' cannot be disjoint since otherwise we would contradict additivity.

If $x_1 \in B \cap B'$, then $x_1 = x_2 + y$ for some $x_2 \in B$. Hence $y = x_1 - x_2 \in \Delta$. We have shown that any $y \in (-\epsilon, \epsilon)$ is in Δ. □

Theorem B.1.2 (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 "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 $z \in (\mathbb{Q} + x_1) \cap (\mathbb{Q} + x_2)$ 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 $\mathbb{Q} + x_2 = \{x_2 + r \mid r \in \mathbb{Q}\} = \{x_2 + r + (x_1 - x_2) \mid r \in \mathbb{Q}\} = \{x_1 + r \mid r \in \mathbb{Q}\} = \mathbb{Q} + x_1$.

Using the Axiom of Choice we construct a set E which 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. Now let $\{r_n\}_{n=1}^\infty$ be an enumeration of the rational numbers. We want to show that $\mathbb{R} = \bigcup_{n=1}^\infty E + r_n$. To see this let $x \in \mathbb{R}$ be arbitrary and let $\{x_0\} = E \cap (\mathbb{Q} + x)$. Then $x_0 = x + r$ for some $r \in \mathbb{Q}$ or $x = x_0 + r_0$ where $r_0 = -r$. Hence $x \in E - r$ so $x \in \bigcup_{n=1}^\infty E + r_n$. We have shown $\mathbb{R} = \bigcup_{n=1}^\infty E + r_n$.

We now make the assumption that E is measurable and show this leads to a contradiction, We first note that if we define $\Delta = \{x_1 - x_2 \mid x_1, x_2 \in E\}$, then Δ contains no rational points except 0. This is because $x_1 = x_2 + r$ for rational r would imply that $E \cap (\mathbb{Q} + x_2) \supset \{x_1, x_2\}$ and this intersection contains only one point. Since Δ contains at most one rational point it cannot contain an open interval so by Lemma (B.1.1) we must conclude that $\mu(E) = 0$.

But if we define $V_n = (E + r_n) \cap [0, 1]$, then $\mu(V_n) \leq \mu(E + r_n) = \mu(E) = 0$ so $\mu(V_n) = 0$. The fact that $\mathbb{R} = \bigcup_{n=1}^\infty E + r_n$ implies

$$[0,1] = \bigcup_{n=1}^{\infty} \big((E + r_n) \cap [0,1]\big) = \bigcup_{n=1}^{\infty} V_n.$$

Subadditivity, then would imply $\mu([0, 1]) \leq \sum_{n=1}^\infty \mu(V_n) = 0$ which clearly contradicts our assumption that E is measurable. □

Bibliography

[1] J. Dieudonné *Foundations of Modern Analysis,* Academic Press, New York. (1960).

[2] J. Marsden and M. Hoffman *Elementary Classical Analysis,* W.H. Freeman, (1993).

[3] M. Pinsky *Partial Differential Equations and Boundary Value Problems with Applications,* McGraw Hill, (1998).

[4] H. L. Royden *Real Analysis,* Macmillan, New York. (1963).

[5] S. J. Taylor *Introduction to Measure and Integration,* Cambridge University Press, London. (1966).

[6] A. C. Zaanen *An Introduction to the Theory of Integration,* North-Holland, New York. (1961).