MATH 4012
Real Analysis IIIB
Lebesgue Measure Theory

Paul Skoufranis

April 5, 2022

Preface:

These are a preliminary edition of these lecture notes for MATH 4012 (Real Analysis IIIB). In particular, I have yet to teach this course at York University and these notes are based on a version of this course that I taught at a different institution. Thus potentially substantial changes to these notes will be made the first time I have the opportunity to teach this course at York University.

In addition, there may be several typographical errors, missing exposition on necessary background, and more advance topics for which there will not be time in class to cover. Future iterations of these notes will hopefully be fairly self-contained provided one has the necessary background. If you come across any typos, errors, omissions, or unclear explanations, please feel free to contact me so that I may continually improve these notes.

Contents

Chapter 1

Cardinality

One important question in analysis is, “Given a set, how large is it?”. One idea to solve this problem would be to ‘count’ the number of elements. For finite sets, this enables us to determine whether two sets have the same number of elements or whether one set has more elements than the other. The problem is, “How do we count the number of elements in an infinite set?”

1.1 Equivalence Relations and Partial Orders

In order to determine when two sets have the same size and when one set is larger than another, we need generalize the notions of equality and of ordering. Both of these notions are a type of relation:

Definition 1.1.1. Given two non-empty sets X and Y, a *relation* is a subset of the product $X \times Y$. Given a relation R, we write xRy if $(x, y) \in R$.

Given a non-empty set X, by a relation on X we will mean a relation on $X \times X$.

Using a specific type of relation, we can generalize the notion of equality.

Definition 1.1.2. Let X be a set. A relation $\sim$ on the elements of X is said to be an *equivalence relation* if:

1. (reflexive) $x \sim x$ for all $x \in X$,

2. (symmetric) if $x \sim y$, then $y \sim x$ for all $x, y \in X$, and

3. (transitive) if $x \sim y$ and $y \sim z$, then $x \sim z$ for all $x, y, z \in X$.

Given an $x \in X$, the set $\{y \in X \mid y \sim x\}$ is called the *equivalence class* of x and is denoted $[x]$.

Notice that $[x] \cap [y] \neq \emptyset$ if and only if $x \sim y$. Thus by taking an index set consisting of one element from each equivalence class, the set X can be written as the disjoint union of its equivalence classes.

Example 1.1.3. Let V be a vector space and let W be a subspace of V. It is elementary to check that if we define $\vec{x} \sim \vec{y}$ if and only if $\vec{x} - \vec{y} \in W$, then $\sim$ is an equivalence relation on V. Note that the equivalence classes of V then become a vector space, denoted V/W, with the operations $[\vec{x}] + [\vec{y}] = [\vec{x} + \vec{y}]$ and $\alpha[\vec{x}] = [\alpha\vec{x}]$. Note the necessity of checking that these operations are well-defined; that is, for addition to make sense, one must show that if $\vec{x}_1 \sim \vec{x}_2$ and $\vec{y}_1 \sim \vec{y}_2$ then $\vec{x}_1 + \vec{y}_1 \sim \vec{x}_2 + \vec{y}_2$.

Similarity, specific types of relations produce orderings on elements of a set.

Definition 1.1.4. Let X be a set. A relation $\preceq$ on the elements of X is called a *partial ordering* if:

1. (reflexivity) $a \preceq a$ for all $a \in X$,
2. (antisymmetry) if $a \preceq b$ and $b \preceq a$, then $a = b$ for all $a, b \in X$, and
3. (transitivity) if $a, b, c \in X$ are such that $a \preceq b$ and $b \preceq c$, then $a \preceq c$.

Clearly $\leq$ is a partial ordering on $\mathbb{R}$. Here is another example:

Example 1.1.5. Given a set X, the relation $\preceq$ on $\mathcal{P}(X)$ defined by

$$Z \preceq Y \qquad \text{if and only if} \qquad Z \subseteq Y$$

is an equivalence relation on $\mathcal{P}(X)$.

The partial ordering in the previous example is not as nice as our ordering on $\mathbb{R}$. To see this, consider the sets $Z = \{1\}$ and $Y = \{2\}$. Then $Z \not\preceq Y$ and $Y \not\preceq Z$; that is, we cannot use the partial ordering to compare X and Z. However, if $x, y \in \mathbb{R}$, then either $x \leq y$ or $y \leq x$. Consequently, a partial ordering is nicer if it has the following property:

Definition 1.1.6. Let X be a set. A partial ordering $\preceq$ on X is called a *total ordering* if for all $x, y \in X$, either $x \preceq y$ or $y \preceq x$ (or both).

1.2 Definition of Cardinality

Let us return to the question of how to count the number of elements in a set and try to determine reasonable equivalence relations and partial orderings to compare the size of sets. One way to compare the number of elements in a set is to use functions. For example, one way to see that $\{1, 2, 3\}$ and $\{5, \pi, 42\}$ have the same number of elements is that we can pair up the elements via $\{(1,5),(3,\pi),(2,42)\}$ for example. However, we can see that $\{1,2,3\}$ and $\{5,\pi,42,29\}$ do not have the same number of elements since there is no such pairing.

Saying that there is such a pairing is precisely saying that there exists a bijection from one set to the other. Consequently, we define a relation $\sim$ on the 'collection' of all sets by $X \sim Y$ if and only if there exists a bijection $f : X \to Y$. Notice that $\sim$ 'is' an equivalence relation. Indeed, to see that $\sim$ satisfies the properties in Definition 1.1.2, first notice that $X \sim X$ as the function $f : X \to X$ defined by $f(x) = x$ for all $x \in X$ is a bijection. Next, if $f : X \to Y$ is a bijection, then $f^{-1} : Y \to X$ is a bijection so $X \sim Y$ implies $Y \sim X$. Finally, if $X \sim Y$ and $Y \sim Z$, then there exists bijections $f : X \to Y$ and $g : Y \to Z$. If we define $h : X \to Z$ to be the *composition* of f and g, denoted $g \circ f$, which is the function defined by $h(x) = g(f(x))$, it is not difficult to see that h is a bijection (either check h is injective and surjective directly, or check that $h^{-1} = f^{-1} \circ g^{-1}$) so $X \sim Z$.

Consequently, given a set X, we will use $|X|$ to denote the equivalence class of X under the above equivalence relation. Oppose to always referring to this equivalence relation, we make the following definition.

Definition 1.2.1. Given two sets X and Y, it is said that X and Y have the same *cardinality* (or are *equinumerous*), denoted $|X| = |Y|$, if there exists a bijection $f : X \to Y$.

Example 1.2.2. Notice that the sets $X = \{3, 7, \pi, 2\}$ and $Y = \{1, 2, 3, 4\}$ have the same cardinality via the function $f : Y \to X$ defined by $f(1) = 3$, $f(2) = \pi$, $f(3) = 2$, and $f(4) = 7$.

Example 1.2.3. We claim that $|\mathbb{N}| = |\mathbb{Z}|$ (which may seem odd as $\mathbb{N} \subseteq \mathbb{Z}$). To see this, define $f : \mathbb{N} \to \mathbb{Z}$ by

$$f(n) = \begin{cases} 0 & \text{if } n = 1 \\ \frac{n}{2} & \text{if } n \text{ is even} \\ -\frac{n-1}{2} & \text{if } n \text{ is odd and } n \geq 3 \end{cases}.$$

(For example $f(1) = 0$, $f(2) = 1$, $f(3) = -1$, $f(4) = 2$, $f(5) = -2$, etc.) It is not difficult to verify that f is a bijection.

Using bijections gives us a method for determining when two sets have the same size. However, we do not have any techniques for determining if two sets have the same cardinality other than explicitly writting a bijection (e.g. do $\mathbb{N}$, $\mathbb{Q}$, and $\mathbb{R}$ all have the same cardinality?). Thus it is useful to ask, how can we determine when one set has fewer elements than another?

We have already seen that $\{1,2,3\}$ and $\{5,\pi,42,29\}$ do not have the same number of elements. We know that $\{1,2,3\}$ has fewer elements than $\{5,\pi,42,29\}$. One way to see this is that we can define a function from

$\{1,2,3\}$ to $\{5,\pi,42,29\}$ that is optimal as possible; that is, we try to form a bijective pairing, but we only obtain an injective function as we cannot hit all of the elements of the later set. Consequently:

Definition 1.2.4. Given two sets X and Y, it is said that X has *cardinality less than* Y, denoted $|X| \leq |Y|$, if there exists an injective function $f : X \to Y$.

Note the above is a 'relation' on the equivalence classes used in Definition 1.2.1. Furthermore, it is not difficult to see that $|X| \leq |X|$ and if $|X| \leq |Y|$ and $|Y| \leq |Z|$ then $|X| \leq |Z|$ (as the composition of injections is an injection). However, it is not clear whether or not the relation in Definition 1.2.4 is antisymmetric, which must be demonstrated in order to show that this is a well-defined partial ordering. Let us postpone this question for now for the purpose of some examples.

Example 1.2.5. Let $n, m \in \mathbb{N}$ be such that $n < m$. Then $\{1, \ldots, n\}$ has cardinality less than $\{1, \ldots, m\}$ as $f : \{1, \ldots, n\} \to \{1, \ldots, m\}$ defined by $f(k) = k$ is injective.

Example 1.2.6. Since the function $f : \mathbb{N} \to \mathbb{Q}$ defined by $f(n) = n$ is injective, we see that $|\mathbb{N}| \leq |\mathbb{Q}|$. More generally, if $X \subseteq Y$, then $|X| \leq |Y|$. Thus $|\mathbb{Q}| \leq |\mathbb{R}|$.

Observe that when determining that $\{1,2,3\}$ has fewer elements than $\{5,\pi,42,29\}$, we could have thought of things in a different light. In particular, we could define a function from $\{5,\pi,42,29\}$ to $\{1,2,3\}$ that was onto. This should imply that $\{5,\pi,42,29\}$ has more elements than $\{1,2,3\}$, which is the case by the next result.

Proposition 1.2.7. *Let X and Y be non-empty sets. If $f : X \to Y$ is surjective, then $|Y| \leq |X|$.*

Proof. For each $y \in Y$, let

$$A_y = f^{-1}(\{y\}).$$

Since f is surjective, $A_y \neq \emptyset$ for all $y \in Y$. Hence, by the Axiom of Choice, there exists a function $g \in \prod_{y \in Y} A_y$; that is, $g : Y \to \bigcup_{y \in Y} A_y \subseteq X$ is such that $g(y) \in A_y$ for all $y \in Y$.

We claim that g is injective. To see this, suppose $y_1, y_2 \in Y$ are such that $g(y_1) = g(y_2)$. Let $x = g(y_1) = g(y_2) \in X$. By the properties of g, it must be the case that $x \in A_{y_1}$ and $x \in A_{y_2}$. Since $x \in A_{y_1}$, we must have $f(x) = y_1$ by the definition of A_{y_1}. Similarly, since $x \in A_{y_2}$, we must have $f(x) = y_2$. Therefore $y_1 = y_2$ as desired. ∎

1.3 Finite and Infinite Sets

Before we attempt to determine whether the relation in Definition 1.2.4 is a partial ordering, let us first formalize the notions of finite and infinite sets.

Definition 1.3.1. A non-empty set X is said to be *finite* if there exists an $n \in \mathbb{N}$ such that $|X| = |\{1, \ldots, n\}|$. In this case, we write $|X| = n$.

A non-empty set X is said to be *infinite* if X is not finite.

We intuitively know which sets are finite and which are infinite. However, there is a nicer characterization of infinite sets. To develop this characterization, we begin with the following.

Lemma 1.3.2. *If X is an infinite set, there exists an injection $f : \mathbb{N} \to X$.*

Proof. Since X is non-empty, $\mathcal{P}(X) \neq \{\emptyset\}$. By the Axiom of Choice there exists a function $f \in \prod_{A \in \mathcal{P}(X) \setminus \{\emptyset\}} A$ such that $f(A) \in A$ for all $A \in \mathcal{P}(X) \setminus \{\emptyset\}$.

Let $a_1 = f(X)$. As $|X| \neq 1$, $X \setminus \{a_1\}$ is non-empty. Hence define $a_2 = f(X \setminus \{a_1\})$. By construction $a_2 \in X \setminus \{a_1\}$ so $a_2 \neq a_1$. Similarly, as $|X| \neq 2$, we may define $a_3 = f(X \setminus \{a_1, a_2\})$ so that $a_3 \notin \{a_1, a_2\}$. Repeating this process, we obtain a sequence $\{a_n\}_{n \geq 1}$ of distinct elements of X (this is legal by taking the union of the finite sets we obtain). Therefore the function $g : \mathbb{N} \to X$ defined by $g(n) = a_n$ is an injection. ∎

Using the above, we can prove the following.

Proposition 1.3.3. *If X is an infinite set, then there exists a $Y \subseteq X$ such that $Y \neq X$ yet $|Y| = |X|$.*

Proof. By Lemma 1.3.2 there exists an injection $f : \mathbb{N} \to X$. For each $n \in \mathbb{N}$ let $a_n = f(n)$. Furthermore, let $Y = X \setminus \{a_1\}$. Clearly $Y \subseteq X$ and $Y \neq X$. To see that $|Y| = |X|$, define the function $g : X \to Y$ by

$$g(x) = \begin{cases} x & \text{if } x \notin f(\mathbb{N}) \\ a_{n+1} & \text{if } x = a_n \end{cases}.$$

It is clear that g is a bijection and thus $|Y| = |X|$ by definition. ∎

Since it is clear that any finite set is not equinumerous to a proper subset, we obtain the following.

Corollary 1.3.4. *A non-empty set X is infinite if and only if X is equinumerous to a proper subset.*

1.4 Cantor-Schröder-Bernstein Theorem

To show that $\leq$ from Definition 1.2.4 is a partial ordering, we must show that $\leq$ is antisymmetric. To begin, let us first consider the following. In Example 1.2.6, it was shown that $|\mathbb{N}| \leq |\mathbb{Q}|$. However, notice if

$$P = \left\{ \frac{m}{n} \,\middle|\, m \geq 0, n > 0, m \text{ and } n \text{ have no common divisors} \right\}$$
$$N = \left\{ \frac{m}{n} \,\middle|\, m < 0, n > 0, m \text{ and } n \text{ have no common divisors} \right\},$$

then $P \cap N = \emptyset$ and $P \cup N = \mathbb{Q}$. Furthermore, we may define $f : \mathbb{Q} \to \mathbb{N}$ by

$$f(q) = \begin{cases} 1 & \text{if } m = 0 \\ 2^m 3^n & \text{if } m > 0 \text{ and } n > 0 \\ 5^{-m} 7^n & \text{if } m < 0 \text{ and } n > 0 \end{cases}$$

where $q = \frac{m}{n}$ is the unique way to write q as an element of P or N. Using the uniqueness of prime factorization (something not covered in this course), we see f is an injective function. Hence $|\mathbb{Q}| \leq |\mathbb{N}|$!

As $|\mathbb{N}| \leq |\mathbb{Q}|$ and $|\mathbb{Q}| \leq |\mathbb{N}|$, is $|\mathbb{Q}| = |\mathbb{N}|$? It is seems difficult to construct a bijective function $f : \mathbb{N} \to \mathbb{Q}$, so what hope do we have?

To answer this question, we have the following result (alternatively, we could construct such a function, but it is not nice to define). Notice that if X and Y are sets such that there exists injective functions $f : X \to Y$ and $g : Y \to X$, then we may invoke the following theorem with $A = g(Y)$ and $B = f(X)$ to obtain that $|X| = |Y|$. Thus the following theorem demonstrates that $\leq$ is indeed a partial ordering and eases the verification that two sets have the same cardinality (as one need only find two injections instead of one bijection, with the former far easier to construct).

Theorem 1.4.1 (Cantor-Schröder–Bernstein Theorem). *Let X and Y be non-empty sets. Suppose $A \subseteq X$ and $B \subseteq Y$ are such that there exists bijective functions $f : X \to B$ and $g : Y \to A$. Then $|X| = |Y|$.*

Proof. Let $A_0 = X$ and $A_1 = A$. Define $h = g \circ f : A_0 \to A_0$ by $h(x) = g(f(x))$. Notice h is injective as f and g are injective.

Let $A_2 = h(A_0)$. Notice

$$A_2 = h(A_0) = g(f(A_0)) = g(B) \subseteq g(Y) = A_1.$$

Hence $A_2 \subseteq A_1 \subseteq A_0$. Next let $A_3 = h(A_1)$. Then

$$A_3 = h(A_1) \subseteq h(A_0) = A_2.$$

Consequently, if for each $n \in \mathbb{N}$ we recursively define $A_n = h(A_{n-2})$, then, by recursion (formally, we should apply the Principle of Mathematical Induction),

$$A_n = h(A_{n-2}) \subseteq h(A_{n-3}) = A_{n-1}$$

for all $n \in \mathbb{N}$. Hence we have constructed a sequence $A_0 \supseteq A_1 \supseteq A_2 \supseteq \cdots$ with $A_n = h(A_{n-2})$ for all $n \geq 2$.

We claim that $|A| = |X|$. To see this, notice that

$$X = A_0 = (A_0 \setminus A_1) \cup (A_1 \setminus A_2) \cup (A_2 \setminus A_3) \cup (A_3 \setminus A_4) \cup \cdots \cup \left(\bigcap_{n=1}^{\infty} A_n \right)$$

$$A = A_1 = (A_1 \setminus A_2) \cup (A_2 \setminus A_3) \cup (A_3 \setminus A_4) \cup (A_4 \setminus A_5) \cup \cdots \cup \left(\bigcap_{n=1}^{\infty} A_n \right).$$

Furthermore, notice that any two distinct sets chosen from either union have empty intersection as $A_0 \supseteq A_1 \supseteq A_2 \supseteq \cdots$.

Since h is injective

$$h(A_{2n} \setminus A_{2n+1}) = h(A_{2n}) \setminus h(A_{2n+1}) = A_{2n+2} \setminus A_{2n+3}$$

for all $n \in \mathbb{N} \cup \{0\}$. Therefore, as the sets in the union description of X are disjoint, we may define $h_0 : A_0 \to A_1$ via

$$h_0(x) = \begin{cases} x & \text{if } x \in \bigcap_{n=1}^{\infty} A_n \\ x & \text{if } x \in A_{2n-1} \setminus A_{2n} \text{ for some } n \in \mathbb{N} \\ h(x) & \text{if } x \in A_{2n} \setminus A_{2n+1} \text{ for some } n \in \mathbb{N} \end{cases}$$

Since

- h_0 maps $A_{2n} \setminus A_{2n+1}$ to $A_{2n+2} \setminus A_{2n+3}$ bijectively for all $n \in \mathbb{N}$,
- h_0 maps $A_{2n-1} \setminus A_{2n}$ to $A_{2n-1} \setminus A_{2n}$ bijectively for all $n \in \mathbb{N}$, and
- h_0 maps $\bigcap_{n=1}^{\infty} A_n$ to $\bigcap_{n=1}^{\infty} A_n$ bijectively,

we obtain that h_0 is a bijection (any two distinct sets chosen from either union have empty intersection). Hence $|A| = |X|$ as claimed.

However $|A| = |Y|$ as $g : Y \to A$ is a bijection. Hence $|Y| = |X|$ as having equal cardinality is an equivalence relation. ∎

Since we have shown $|\mathbb{N}| \leq |\mathbb{Q}|$ and $|\mathbb{Q}| \leq |\mathbb{N}|$, we have by the Cantor-Schröder–Bernstein Theorem (Theorem 1.4.1) that $|\mathbb{N}| = |\mathbb{Q}|$; that is $\mathbb{N}$ and $\mathbb{Q}$ have the same number of elements! Thus, is it possible that $|\mathbb{Q}| = |\mathbb{R}|$?

1.5 Countable Sets

One nice corollary about $|\mathbb{N}| = |\mathbb{Q}|$ (which will help us in determining whether $|\mathbb{Q}| = |\mathbb{R}|$) is that we can make a list of all rational numbers; that is, as there is a bijective function $f : \mathbb{N} \to \mathbb{Q}$, we can form the sequence of all rational numbers $(f(n))_{n\geq 1}$. Consequently, sets that are equinumerous to the natural numbers are particularity nice sets as we can index such sets by $\mathbb{N}$. This leads us to the study of such sets.

Definition 1.5.1. A non-empty set X is said to be

- *countable* if X is finite or $|X| = |\mathbb{N}|$,
- *countably infinite* if $|X| = |\mathbb{N}|$,
- *uncountable* if X is not countable.

A natural question is, "Under what operations is the countability of sets preserved?" The following demonstrates that subsets (and thus intersections) of countable sets are countable.

Lemma 1.5.2. *If X is a countable set, then any subset of X must also be countable.*

Proof. Let X be countable and let $Y \subset X$. If Y is finite, then clearly Y is countable. Otherwise Y is infinite. Hence $|Y| \geq |\mathbb{N}|$ by Lemma 1.3.2. Since Y is infinite, X is infinite. Thus, as X is countable, there exists a bijection $f : X \to \mathbb{N}$. Hence restricting f to Y produces an injection from Y to $\mathbb{N}$. Thus $|Y| \leq |\mathbb{N}|$ so $|Y| = |\mathbb{N}|$ and thus Y is countable. ∎

The following, which simply stated says the countable union of countable sets is countable, is an nice example of why it is useful to be able to write countable sets as a sequence.

Theorem 1.5.3. *For each $n \in \mathbb{N}$, let X_n be a countable set. Then $X = \bigcup_{n=1}^{\infty} X_n$ is countable.*

Proof. We first desire to restrict to the case that our countable sets are disjoint.

Let $B_1 = X_1$ and for each $k \geq 2$ let

$$B_k = X_k \setminus \left(\bigcup_{j=1}^{k-1} X_j \right).$$

Clearly $B_k \cap B_j = \emptyset$ for all $j \neq k$ and $X = \bigcup_{n=1}^{\infty} B_n$. Since $B_n \subseteq X_n$ for all n, each B_n is countable by Lemma 1.5.2. Consequently, for each $n \in \mathbb{N}$, we may write

$$B_n = (b_{n,1}, b_{n,2}, b_{n,3}, \ldots,).$$

We desire to define a function $f : X \to \mathbb{N}$ by

$$f(b_{n,m}) = 2^n 3^m.$$

Note such a function is well-defined since $B_k \cap B_j = \emptyset$ for all $j \neq k$. Since f is injective by the uniqueness of the prime decomposition of natural numbers, we obtain that $|X| \leq |\mathbb{N}|$. Hence X is countable. ∎

Corollary 1.5.4. *If X and Y are countable sets, $X \bigcup Y$ is a countable set.*

Proof. Apply Theorem 1.5.3 where $X_1 = X$, $X_2 = Y$, and $X_n = \emptyset$ for all $n \geq 3$. ∎

We briefly mention a few examples of countable sets.

Example 1.5.5. The set $\mathbb{N} \times \mathbb{N}$ is countable. To show that $\mathbb{N} \times \mathbb{N}$ is countable, it suffices by Lemma 1.5.2 to show that there exists an injective function $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$. Define $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ by

$$f(n, m) = 2^n 3^m$$

for all $n, m \in \mathbb{N}$. Since f is injective due to the uniqueness of the prime decomposition, the claim is complete.

Example 1.5.6. A real number α is said to be *algebraic* if there exists a non-zero polynomial $p(x)$ with integer coefficients such that $p(\alpha) = 0$. It turns out that the set of algebraic numbers is countable (and thus, as we will shortly see that $\mathbb{R}$ is uncountable, most numbers in $\mathbb{R}$ are not algebraic).

To begin, for each $n \in \mathbb{N} \cup \{0\}$, consider the set

$$A_n = \{(a_n, a_{n-1}, \ldots, a_1, a_0) \mid a_k \in \mathbb{Z}\}.$$

Notice that $A_0 = \mathbb{Z}$ so A_0 is countable. Furthermore, for each $n \in \mathbb{N}$ we may view A_n as a countable union of copies of A_{n-1}; that is,

$$\bigcup_{k \in \mathbb{Z}} A_{n-1} \sim A_n$$

where for all $(a_{n-1}, \ldots, a_0) \in A_{n-1}$ the k^{th} copy of $(a_{n-1}, \ldots, a_0)$ maps to $(k, a_{n-1}, \ldots, a_0)$. Hence A_n is countable for all $n \in \mathbb{N} \cup \{0\}$.

For each $n \in \mathbb{N} \cup \{0\}$ and for each $(a_n, a_{n-1}, \ldots, a_1, a_0) \in A_n \setminus \{(0, \ldots, 0)\}$, let

$$B_{(a_n, a_{n-1}, \ldots, a_1, a_0)} = \{\alpha \in \mathbb{R} \mid a_n \alpha^n + a_{n-1}\alpha^{n-1} + \cdots + a_1 \alpha + a_0 = 0\}.$$

As a non-zero polynomial of degree n has at most n roots (by, for example, the division algorithm), each $B_{(a_n, a_{n-1}, \ldots, a_1, a_0)}$ has at most n elements and thus is countable. Hence, if

$$C_n = \left\{\alpha \in \mathbb{R} \,\middle|\, \substack{a_n \alpha^n + a_{n-1}\alpha^{n-1} + \cdots + a_1 \alpha + a_0 = 0 \\ \text{for some } (a_n, a_{n-1}, \ldots, a_1, a_0) \in A_n \setminus \{(0, \ldots, 0)\}}\right\}$$

then C_n is a union over $A_n \setminus \{(0, \ldots, 0)\}$ of finite sets and thus is countable as $A_n \setminus \{(0, \ldots, 0)\}$ is countable.

Finally, let

$$\Psi = \{\alpha \in \mathbb{R} \mid \alpha \text{ is algebraic}\}.$$

Since $\Psi = \bigcup_{n \in \mathbb{N}} C_n$, Ψ is a countable union of countable sets and thus is countable.

The question of whether $\mathbb{Q}$ and $\mathbb{R}$ are equinumerous is equivalent to the question of whether $\mathbb{R}$ is countable or not. To show that $\mathbb{R}$ is not countable, we begin with the following.

Theorem 1.5.7. *The open interval* $(0,1)$ *is uncountable.*

Proof. The following proof is known as Cantor's diagonalization argument and has a wide variety of uses. Suppose that $(0,1)$ is countable. Then we may write $(0,1) = \{x_n \mid n \in \mathbb{N}\}$ and there exists numbers $\{a_{i,j} \mid i,j \in \mathbb{N}\} \subseteq \{0,1,\ldots,9\}$ such that

$$x_j = \sum_{k=1}^{\infty} \frac{a_{k,j}}{10^k}$$

for all $j \in \mathbb{N}$. Note that the sequence $(a_{k,j})_{k\geq 1}$ in the above expression for x_j represents the decimal expansion of x_j; that is

$$x_j = 0.a_{1,j}a_{2,j}a_{3,j}a_{4,j}a_{5,j}\cdots.$$

Consequently, this representation need not be unique due to the possibility of repeating 9s (and this is the only possibility).

For each $k \in \mathbb{N}$, define

$$y_k = \begin{cases} 3 & \text{if } a_{k,k} = 7 \\ 7 & \text{otherwise} \end{cases}$$

and let $y = \sum_{k=1}^{\infty} \frac{y_k}{10^k}$. It is not difficult to see that $y \in (0,1)$. Furthermore $y \neq x_n$ for all $n \in \mathbb{N}$ (as y and x_n will disagree in the n^{th} decimal place and this is not because of repeating 9s). Therefore, since $(0,1) = \{x_n \mid n \in \mathbb{N}\}$, we must have that $y \notin (0,1)$, which contradicts the fact that $y \in (0,1)$. ∎

Proposition 1.5.8. *A set containing an uncountable subset is uncountable.*

Proof. Let X be a set such that there exists an uncountable subset Y of X. Suppose X was countable. Then Y would be countable by Lemma 1.5.2, which contradicts the fact that Y is uncountable. Hence X must be uncountable. ∎

Combining Theorem 1.5.7 and Proposition 1.5.8, $\mathbb{R}$ is uncountable. In fact $|\mathbb{R}| = |(0,1)|$ as the function $f : (0,1) \to \mathbb{R}$ defined by $f(x) = \tan\left(\pi x - \frac{\pi}{2}\right)$ is a bijection. Furthermore we have the following.

Corollary 1.5.9. *The irrational numbers* $\mathbb{R} \setminus \mathbb{Q}$ *is an uncountable set.*

Proof. Suppose $\mathbb{R} \setminus \mathbb{Q}$ is a countable set. Since $\mathbb{Q}$ is countable and $\mathbb{R} = \mathbb{Q} \cup (\mathbb{R} \setminus \mathbb{Q})$, it would need to be the case that $\mathbb{R}$ is countable by Theorem 1.5.3. Since $\mathbb{R}$ is uncountable by Proposition 1.5.8, we have obtained a contradiction so $\mathbb{R} \setminus \mathbb{Q}$ is an uncountable set. ∎

One additional set that is important in analysis and measure theory is the following.

Definition 1.5.10. Let $P_0 = [0, 1]$. Construct P_1 from P_0 by removing the open interval of length $\frac{1}{3}$ from the middle of P_0 (i.e. $P_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$). Then construct P_2 from P_1 by removing the open intervals of length $\frac{1}{3^2}$ from the middle of each closed subinterval of P_1. Subsequently, having constructed P_n, construct P_{n+1} by removing the open intervals of length $\frac{1}{3^{n+1}}$ from the middle of each of the 2^n closed subintervals of P_n. Specifically, P_n is the union of the 2^n closed intervals of the form

$$\left[\sum_{k=1}^{n} \frac{a_k}{3^k}, \frac{1}{3^n} + \sum_{k=1}^{n} \frac{a_k}{3^k}\right]$$

where $a_1, \ldots, a_n \in \{0, 2\}$.

The set

$$\mathcal{C} = \bigcap_{n \geq 1} P_n$$

is known as the *Cantor set*.

Remark 1.5.11. The Cantor set has many interesting properties. In particular, the Cantor set is a compact set with no interior.

For an alternate description of the Cantor set, we prove the following.

Lemma 1.5.12. *Let $x \in \mathbb{R}$. Then $x \in \mathcal{C}$ if and only if there is a sequence $(a_n)_{n\geq 1}$ with $a_n \in \{0, 2\}$ for all $n \in \mathbb{N}$ such that $x = \lim_{n\to\infty} \sum_{k=1}^{n} \frac{a_k}{3^k}$ (i.e. $x \in [0, 1]$ and x has a ternary expansion using only 0s and 2s).*

Proof. Exercise. ∎

Using this alternate definition, it is possible to prove the following.

Theorem 1.5.13. *The Cantor set is uncountable.*

Proof. Exercise. ∎

Since $\mathbb{R}$ is uncountable, $|\mathbb{N}| < |\mathbb{R}|$ so there does not exist a list of real numbers. However, is $\mathbb{R}$ the 'smallest' set larger than $\mathbb{N}$? In particular:

Axiom 1.5.14 (The Continuum Hypothesis). *If $X \subseteq \mathbb{R}$ is uncountable, must it be the case that $|X| = |\mathbb{R}|$?*

The Continuum Hypothesis was originally postulated by Cantor whom spent many years (at the cost of his own health and possibly sanity) trying to prove the hypothesis. Consequently, we will not try. In fact, the reason for Cantor's difficulty is that there is no proof. However, nor is there any counter example. Like with the Axiom of Choice, the Continuum Hypothesis is independent of (Zermelo–Fraenkel) set theory, even if the Axiom of Choice is included. Most results in analysis do not require an assertion to whether the Continuum Hypothesis is true of false. Thus we move on.

1.6 Comparability of Cardinals

Using the Cantor-Schröder–Bernstein Theorem (Theorem 1.4.1), we saw that cardinality gives a partial ordering on the size of sets. However, is it a total ordering (Definition 1.1.6)? That is, if X and Y are non-empty sets, must it be the case that $|X| \leq |Y|$ or $|Y| \leq |X|$?

The above is a desirable property since it makes the ordering nicer. However, when given two sets, it is not clear whether there always exist an injection from one set to the other. The goal of this subsection is to develop the necessary tools in order to answer this problem in the subsequent subsection. The tools we require are related to partial ordering, so the following definition is made.

Definition 1.6.1. A *partially ordered set* (or *poset*) is a pair $(X, \preceq)$ where X is a non-empty set and $\preceq$ is a partial ordering on X.

For examples of posets, we refer the reader back to Section 1.1. Our main focus is a 'result' about totally ordered subsets of partially ordered sets:

Definition 1.6.2. Let $(X, \preceq)$ be a partially ordered set. A non-empty subset $Y \subseteq X$ is said to be a *chain* if Y is totally ordered with respect to $\preceq$; that is, if $a, b \in Y$, then either $a \preceq b$ or $b \preceq a$.

Clearly any non-empty subset of a totally ordered set is a chain. Here is a less obvious example.

Example 1.6.3. Recall from Example 1.1.5 that the power set $\mathcal{P}(\mathbb{R})$ of $\mathbb{R}$ has a partial ordering $\preceq$ where

$$A \preceq B \quad \Longleftrightarrow \quad A \subseteq B.$$

If $Y = \{A_n\}_{n=1}^{\infty} \subseteq \mathcal{P}(\mathbb{R})$ are such that $A_n \subseteq A_{n+1}$ for all $n \in \mathbb{N}$, then Y is a chain.

Like with the real numbers, upper bounds play an important role with respect to chains.

Definition 1.6.4. Let $(X, \preceq)$ be a partially ordered set. A non-empty subset $Y \subseteq X$ is said to be a *bounded above* if there exists a $z \in X$ such that $y \leq z$ for all $y \in Y$. Such an element z is said to be an *upper bound* for Y.

Example 1.6.5. Recall from Example 1.6.3 that if $Y = \{A_n\}_{n=1}^{\infty} \subseteq \mathcal{P}(\mathbb{R})$ are such that $A_n \subseteq A_{n+1}$ for all $n \in \mathbb{N}$, then Y is a chain with respect to the partial ordering defined by inclusion. If

$$A = \bigcup_{n=1}^{\infty} A_n$$

then clearly $A \in \mathcal{P}(\mathbb{R})$ and $A_n \subseteq A$ for all $n \in \mathbb{N}$. Hence A is an upper bound for Y.

Recall there are optimal upper bounds of subsets of $\mathbb{R}$ called least upper bounds which need not be in the subset. We desire a slightly different object when it comes to partially ordered sets as the lack of a total ordering means there may not be a unique 'optimal' upper bound.

Definition 1.6.6. Let X be a non-empty set and let $\preceq$ be a partial ordering on X. An element $x \in X$ is said to be *maximal* if there does not exist a $y \in X \setminus \{x\}$ such that $x \preceq y$; that is, there is no element of X that is larger than x with respect to $\preceq$.

Notice that $\mathbb{R}$ together with its usual ordering $\leq$ does not have a maximal element (by, for example, the Archimedean Property). However, many partially ordered sets do have maximal elements. For example $([0, 1], \leq)$ has 1 as a maximal element (although $((0, 1), \leq)$ does not).

For an example involving a partial ordering that is not a total ordering, suppose $X = \{x, y, z, w\}$ and $\preceq$ is defined such that $a \preceq a$ for all $a \in X$, $a \preceq b$ for all $a \in \{x, y\}$ and $b \in \{z, w\}$, and $a \npreceq b$ for all other pairs $(a, b) \in X \times X$. It is not difficult to see that z and w are maximal elements and x and y are not maximal elements. Thus it is possible, when dealing with a partial ordering that is not a total ordering, to have multiple maximal elements.

The result we require for the next subsection may now be stated using the above notions.

Axiom 1.6.7 (Zorn's Lemma). *Let $(X, \preceq)$ be a non-empty partially ordered set. If every chain in X has an upper bound, then X has a maximal element.*

We will not prove Zorn's Lemma. To do so, we would need to use the Axiom of Choice. In fact, Zorn's Lemma and the Axiom of Choice are logically equivalent; that is, assuming the axioms of (Zermelo–Fraenkel) set theory, one may use the Axiom of Choice to prove Zorn's Lemma, and one may use Zorn's Lemma to prove the Axiom of Choice.

Before using Zorn's Lemma to demonstrate that the ordering on cardinals is a total ordering, we analyze a simpler example.

Example 1.6.8. Let V be a (non-zero) vector space. We claim that V has a basis; that is, a linearly independent spanning set. To see this, let $\mathcal{L}$ denote the collection of all linearly independent subsets of V (which is clearly non-empty) and define a partial ordering on $\mathcal{L}$ by $A \preceq B$ if and only if $A \subseteq B$ (clearly this is a partial ordering on $\mathcal{L}$).

To invoke Zorn's Lemma, we need to demonstrate that every chain in $\mathcal{L}$ has an upper bound. Let $\{A_\alpha\}_{\alpha \in I}$ be a chain in $\mathcal{L}$ and let

$$A = \bigcup_{\alpha \in I} A_\alpha.$$

We claim that $A \in \mathcal{L}$. To see this, suppose $\vec{v}_1, \ldots, \vec{v}_n \in A$ and $a_1\vec{v}_1 + \cdots a_n\vec{v}_n = 0$ for some scalars a_k. By the definition of A and the fact that

$\{A_\alpha\}_{\alpha \in I}$ is a chain, there exists an $i \in I$ such that $\vec{v}_1, \ldots, \vec{v}_n \in A_i$ (that is, each $\vec{v}_k$ is in some A_α and as the A_α are totally ordered, take the largest). Hence, as A_i is a linearly independent set, $a_1\vec{v}_1 + \cdots a_n\vec{v}_n = 0$ implies $a_1 = \cdots = a_n = 0$. Hence $A \in \mathcal{L}$. As A is clearly an upper bound for $\{A_\alpha\}_{\alpha \in I}$, ever chain in $\mathcal{L}$ has an upper bound.

By Zorn's Lemma there exists a maximal element $B \in \mathcal{L}$. We claim that B is a basis for V. To see this, suppose to the contrary that $\operatorname{span}(B) \neq V$. Thus there exists a non-zero vector $\vec{v} \in V \setminus \operatorname{span}(B)$. This implies that $B \cup \{\vec{v}\}$ is linearly independent. However, as $B \preceq B \cup \{\vec{v}\}$ and $B \neq B \cup \{\vec{v}\}$, we have a contradiction to the fact that B is a maximal element in $\mathcal{L}$. Hence it must have been the case that $\operatorname{span}(B) = V$ and thus B is a basis for V.

Onto demonstrating the ordering on cardinals is a total ordering.

Theorem 1.6.9. *Let X and Y be non-empty sets. Then either $|X| \leq |Y|$ or $|Y| \leq |X|$.*

Proof. Let

$$\mathcal{F} = \{(A, B, f) \mid A \subseteq X, B \subseteq Y, f : A \to B \text{ is a bijection}\}.$$

Notice that $\mathcal{F}$ is non-empty since, by assumption, there exists an $x \in X$ and a $y \in Y$ so we may select $A = \{x\}$, $B = \{y\}$, and $f : A \to B$ defined by $f(x) = y$.

Given $(A_1, B_1, f_1), (A_2, B_2, f_2) \in \mathcal{F}$, define $(A_1, B_1, f_1) \preceq (A_2, B_2, f_2)$ if and only if

$$A_1 \subseteq A_2, \quad B_1 \subseteq B_2, \qquad \text{and} \qquad f_2(x) = f_1(x) \text{ for all } x \in A_1.$$

It is not difficult to verify that $\preceq$ is a partial ordering on $\mathcal{F}$.

We desire to invoke Zorn's Lemma (Axiom 1.6.7) in order to obtain a maximal element of $\mathcal{F}$. To invoke Zorn's Lemma, it must be demonstrated that every chain in $(\mathcal{F}, \preceq)$ has an upper bound. Let $\mathcal{C} = \{(A_\alpha, B_\alpha, f_\alpha) \mid \alpha \in I\}$ be an arbitrary chain in $(\mathcal{F}, \preceq)$. Let

$$A = \bigcup_{\alpha \in I} A_\alpha \qquad \text{and} \qquad B = \bigcup_{\alpha \in I} B_\alpha.$$

We desire to define $f : A \to B$ such that $f(x) = f_\alpha(x)$ whenever $x \in A_\alpha$. The question is, "Will such an f be well-defined as each x could be in multiple A_α?" To see that f is well-defined, suppose $x \in A_i$ and $x \in A_j$ for some $i, j \in I$. Since $\mathcal{C}$ is a chain, either $(A_i, B_i, f_i) \preceq (A_j, B_j, f_j)$ or $(A_j, B_j, f_j) \preceq (A_i, B_i, f_i)$. If $(A_i, B_i, f_i) \preceq (A_j, B_j, f_j)$, then $A_i \subseteq A_j$ and $\preceq$ implies that $f_j(x) = f_i(x)$. As the case that $(A_j, B_j, f_j) \preceq (A_i, B_i, f_i)$ is the same (reversing i and j), we obtain that f is well-defined.

In order for (A, B, f) to be an upper bound for $\mathcal{C}$, we must first demonstrate that $(A, B, f) \in \mathcal{F}$. Clearly $A \subseteq X$, $B \subseteq Y$, and $f : A \to B$ is a function. It remains to check that f is a bijection.

To see that f is injective, suppose $x_1, x_2 \in A$ are such that $f(x_1) = f(x_2)$. Since $A = \bigcup_{\alpha\in I} A_\alpha$, there exists $i, j \in I$ such that $x_i \in A_i$ and $x_j \in A_j$. Since $\mathcal{C}$ is a chain, we must have either $(A_i, B_i, f_i) \preceq (A_j, B_j, f_j)$ or $(A_j, B_j, f_j) \preceq (A_i, B_i, f_i)$. In the former case, we obtain that $f_j(x_1) = f(x_1) = f(x_2) = f_j(x_2)$. Therefore, since f_j is injective, it must be the case that $x_1 = x_2$. As the case that $(A_j, B_j, f_j) \preceq (A_i, B_i, f_i)$ is the same (reversing i and j), we obtain that f is injective.

To see that f is surjective, let $y \in B$ be arbitrary. Since $B = \bigcup_{\alpha\in I} B_\alpha$, there exists an $i \in I$ such that $y \in B_i$. Since f_i is surjective, there exists an $x \in A_i$ such that $f_i(x) = y$. Hence $x \in A$ and $f(x) = f_i(x) = y$. Therefore, as y was arbitrary, f is surjective. Hence f is a bijection and $(A, B, f) \in \mathcal{F}$.

As $(A, B, f) \in \mathcal{F}$, it is easy to see that (A, B, f) is an upper bound for $\mathcal{C}$ by the definition of (A, B, f) and the partial ordering $\preceq$. Hence, as $\mathcal{C}$ was an arbitrary chain, every chain in $\mathcal{F}$ has an upper bound. Thus Zorn's Lemma implies that $(\mathcal{F}, \preceq)$ has a maximal element.

Let $(A_0, B_0, f_0) \in \mathcal{F}$ be a maximal element. We claim that either $A_0 = X$ or $B_0 = Y$. To see this, suppose otherwise that $A_0 \neq X$ and $B_0 \neq Y$. Therefore, there exist $x_0 \in X \setminus A_0$ and $y_0 \in Y \setminus B_0$. Let $A' = A_0 \cup \{x_0\}$, $B' = B_0 \cup \{y_0\}$, and $g : A' \to B'$ be defined by $g(x_0) = y_0$ and $g(x) = f_0(x)$ for all $x \in A_0$. Clearly g is a well-defined bijection by construction so $(A', B', g) \in \mathcal{F}$. However, it is elementary to see that $(A_0, B_0, f_0) \preceq (A', B', g)$ and $(A_0, B_0, f_0) \neq (A', B', g)$. As this contradicts the fact that $(A_0, B_0, f_0) \in \mathcal{F}$ is a maximal element, we have obtained a contradiction. Hence either $A_0 = X$ or $B_0 = Y$.

If $A_0 = X$, then $f_0 : X \to B \subseteq Y$ is injective so $|X| \leq |Y|$ by definition. Otherwise, if $B_0 = Y$, then $f_0 : A_0 \to Y$ is surjective. Thus $|Y| \leq |A_0| \leq |X|$ by Proposition 1.2.7. ∎

1.7 Cardinal Arithmetic

One natural question to ask is, "If X and Y are disjoint sets and we know $|X|$ and $|Y|$, can we determine $|X \cup Y|$?" Of course if X and Y are finite sets, then $|X \cup Y| = |X| + |Y|$. Thus determining the cardinality of $X \cup Y$ from the cardinality of X and Y really is a form of cardinal arithmetic.

As we already know the answer when both sets are finite, we will focus on the case where at least one set is infinite. Furthermore, since we know if $|X| = |Y| = |\mathbb{N}|$ then $|X \cup Y| = |\mathbb{N}|$ by Theorem 1.5.3, we need not study this case.

We begin with the case that one set is finite. To show that adding a finite set to an infinite set does not change the cardinality, we prove the following.

Theorem 1.7.1. *Let X be an infinite set and let Y be a finite subset of X. Then $|X \setminus Y| = |X|$.*

Proof. Suppose X is an infinite set and Y is a finite subset of X. Then $Z = X \setminus Y$ is an infinite set as a finite union of finite sets is finite. Since Z is infinite, there exists an infinite countable set $W \subseteq Z$ by Lemma 1.3.2. Write $W = \{a_n\}_{n\in\mathbb{N}}$ and $Y = \{y_1, \ldots, y_m\}$ for some $m \in \mathbb{N}$. Define $f : Z \to X$ by

$$f(z) = \begin{cases} z & \text{if } z \notin W \\ y_n & \text{if } z = a_n \text{ for some } n \leq m \\ a_{n-m} & \text{if } z = a_n \text{ for some } n > m \end{cases}.$$

It is elementary to see that f is a well-defined bijection. Hence $|X| = |Z| = |X \setminus Y|$ ∎

To deal with the case that both sets are infinite, we will develop the following idea: "If X is an infinite set, then X can be divided into two disjoint subsets of the same cardinality". Seeing this idea is true in the case that X is countably infinite is rather trivial.

Lemma 1.7.2. *Let X be a countably infinite set. There exists two disjoint infinite countable sets Y and Z such that $Y \cup Z = X$.*

Proof. Let X be a countably infinite set. Hence there exists a bijection $f : \mathbb{N} \to X$. Let

$$Y = \{f(2n) \mid n \in \mathbb{N}\} \qquad \text{and} \qquad Z = \{f(2n-1) \mid n \in \mathbb{N}\}.$$

Since f is a bijection, it is elementary to verify that Y and Z have the desired properties. ∎

The extension of Lemma 1.7.2 to uncountable sets is more involved.

Lemma 1.7.3. *Let X be an infinite set. There exists two disjoint sets Y and Z such that $Y \cup Z = X$ and $|X| = |Y| = |Z|$.*

Proof. If X is countable, the result follows from Lemma 1.7.2. Thus suppose X is an uncountable set. Define

$$\mathcal{F} = \left\{ (W, A, B, f, g) \,\middle|\, \substack{A,B,W \subseteq X, f:W\to A \text{ and } g:W\to B \text{ bijections,} \\ A\cap B=\emptyset, W=A\cup B} \right\}.$$

For two elements $(W_1, A_1, B_1, f_1, g_1), (W_2, A_2, B_2, f_2, g_2) \in \mathcal{F}$, define

$$(W_1, A_1, B_1, f_1, g_1) \preceq (W_2, A_2, B_2, f_2, g_2)$$

if $W_1 \subseteq W_2$, $A_1 \subseteq A_2$, $B_1 \subseteq B_2$, and $f_2(w) = f_1(w)$ and $g_2(w) = g_1(w)$ for all $w \in W_1$. It is not difficult to verify that $\preceq$ is a partial ordering.

We desire to invoke Zorn's Lemma. To do this, first we must verify that $\mathcal{F}$ is non-empty. Since X is uncountable, by Lemma 1.3.2 there exists a $W \subseteq X$ such that W is infinite and countable. By Lemma 1.7.2 there exists

$A, B \subseteq W$ such that $A \cap B = \emptyset$, $W = A \cup B$, and $|A| = |B| = |W|$. As the later implies the existence of bijections $f : W \to A$ and $g : W \to B$, we obtain that $\mathcal{F}$ is non-empty.

Next let $\mathcal{C} = \{(W_\alpha, A_\alpha, B_\alpha, f_\alpha, g_\alpha) \mid \alpha \in I\}$ be an arbitrary chain in $\mathcal{F}$. Let

$$W = \bigcup_{\alpha \in I} W_\alpha, \qquad A = \bigcup_{\alpha \in I} A_\alpha, \qquad B = \bigcup_{\alpha \in I} B_\alpha,$$

and define $f : W \to A$ and $g : W \to B$ by $f(w) = f_\alpha(w)$ and $g(w) = g_\alpha(w)$ for all $w \in W_\alpha$. By the proof of Theorem 1.6.9, f and g are well-defined bijections. Furthermore, we claim that $A \cap B = \emptyset$. To see this, suppose to the contrary that $x \in A \cap B$. Hence there exists $\alpha, \beta \in I$ such that $x \in A_\alpha$ and $x \in B_\beta$. Since $\mathcal{C}$ is a chain, either $\alpha \leq \beta$ or $\beta \leq \alpha$. Hence if $\iota = \max\{\alpha, \beta\}$ we obtain that $x \in A_\iota \cap B_\iota$ as $\mathcal{C}$ is a chain. As this contradicts the definition of $\mathcal{F}$, we obtain that $A \cap B = \emptyset$. As it is clear that $W = A \cup B$, we see that $(W, A, B, f, g) \in \mathcal{F}$. As $(W_\alpha, A_\alpha, B_\alpha, f_\alpha, g_\alpha) \preceq (W, A, B, f, g)$ for all $\alpha \in I$, (W, A, B, f, g) is an upper bound for $\mathcal{C}$. Therefore, as $\mathcal{C}$ was arbitrary, every chain in $\mathcal{F}$ has an upper bound.

By Zorn's Lemma $\mathcal{F}$ has a maximal element. Let $(W_0, A_0, B_0, f_0, g_0)$ be a maximal element of $\mathcal{F}$. We claim that $X \setminus W_0$ is finite. To see this, suppose to the contrary that $X \setminus W_0$ is infinite. Thus there exists a countable subset $Z \subseteq X \setminus W_0$. By Lemma 1.7.2 there exists countable subsets A' and B' such that $A' \cap B' = \emptyset$ and $A' \cup B' = Z$. Thus there exist bijections $f' : Z \to A'$ and $g' : Z \to B'$.

Let $W = W_0 \cup Z$, $A = A_0 \cup A'$, and $B = B_0 \cup B'$. Define $f : W \to A$ and $g : W \to B$ by

$$f(w) = \begin{cases} f_0(w) & \text{if } w \in W_0 \\ f'(w) & \text{if } w \in Z \end{cases} \qquad \text{and} \qquad g(w) = \begin{cases} g_0(w) & \text{if } w \in W_0 \\ g'(w) & \text{if } w \in Z \end{cases}.$$

Since $W_0 \cap Z = A_0 \cap A' = B_0 \cap B' = \emptyset$, f and g are well-defined bijections. Clearly $(W, A, B, f, g) \in \mathcal{F}$ and $(W_0, A_0, B_0, f_0, g_0) \preceq (W, A, B, f, g)$, which contradicts the fact that $(W_0, A_0, B_0, f_0, g_0)$ was a maximal element. Hence $X \setminus W_0$ is finite.

By the above, we have that $A_0 \cap B_0 = \emptyset$, $W_0 = A_0 \cup B_0$, $|W_0| = |A_0| = |B_0|$, and $C = X \setminus W_0$ is finite. Therefore, if we let $Y = A_0 \cup C$ and $Z = B_0$, then $|X| = |W_0| = |Z| = |A_0| = |Y|$ by Theorem 1.7.1, $Y \cap Z = \emptyset$, and $X = Y \cup Z$ as desired. ∎

Finally, we obtain the following demonstrating that the cardinality of the union of two infinite sets is the larger of the cardinalities of the individual sets.

Theorem 1.7.4. *Let X and Y be non-empty sets with X infinite. If $|Y| \leq |X|$ then $|X \cup Y| = |X|$.*

Proof. Let X be an infinite set and let Y be a set such that $|Y| \leq |X|$. Let $Z = Y \setminus X$ so that $X \cap Z = \emptyset$ and $X \cup Z = X \cup Y$. Hence it suffices to show that $|X \cup Z| = |X|$. Since $X \subseteq X \cup Z$, we clearly have $|X| \leq |X \cup Z|$. For the other inequality, notice that $Z \subseteq Y$ so $|Z| \leq |Y| \leq |X|$. By Lemma 1.7.3 there exists two disjoint sets S and T such that $S \cup T = X$ and $|S| = |T| = |X|$. Since $|Z| \leq |S|$, there exists an injective function $f : Z \to S$. Similarly, since $|X| = |T|$, there exists a bijective function $g : X \to T$. Define $h : X \cup Z \to X$ by

$$h(q) = \begin{cases} f(q) & \text{if } q \in Z \\ g(q) & \text{if } q \in X \end{cases}.$$

Since $Z \cap X = \emptyset$, h is a well-defined function. Furthermore, since f and g are injective and since $S \cap T = \emptyset$, h is injective. Hence $|X \cup Z| \leq |X|$ so $|X| = |X \cup Z|$ as desired. ∎

As a corollary of the proof of Theorem 1.7.4, we note the following result which improves upon Theorem 1.5.3.

Corollary 1.7.5. *Let X be an infinite set. Let $\{X_n\}_{n\in\mathbb{N}}$ be a countable collection of infinite sets such that $|X_n| \leq |X|$ for all $n \in \mathbb{N}$. If $Y = \bigcup_{n=1}^{\infty} X_n$, then $|Y| \leq |X|$.*

Proof. By repeating the same argument as in Theorem 1.5.3, we may assume that the X_n are pairwise disjoint.

Since X is infinite, Lemma 1.7.3 implies there exists two subsets of X, denoted Y_1 and Z_1 such that $Y_1 \cup Z_1 = X$ and $|Y_1| = |Z_1| = |X|$. Since Y_1 is infinite, Lemma 1.7.3 implies there two subsets of Y_1, denoted Y_2 and Z_2 such that $Y_2 \cup Z_2 = Y_1$ and $|Y_2| = |Z_2| = |Y_1| = |X|$. By repeating this argument ad infinitum, there exists a collection $\{Z_n\}_{n\in\mathbb{N}}$ of pairwise disjoint subsets of X such that $|Z_n| = |X|$ for all $n \in \mathbb{N}$.

As $|X_n| \leq |X| = |Z_n|$ for all $n \in \mathbb{N}$, there exists an injective function $f_n : X_n \to Z_n$. Define $f : Y \to X$ by $f(x) = f_n(x)$ whenever $x \in X_n$. Notice that f is well-defined since $\{X_n\}_{n\in\mathbb{N}}$ are pairwise disjoint with union Y. Furthermore, since $\{Z_n\}_{n\in\mathbb{N}}$ are pairwise disjoint and since each f_n is injective, f is injective. Hence $|Y| \leq |X|$ as desired. ∎

To conclude this chapter on cardinality, we note that there are many other results pertaining to cardinality that we may study. For example, we can study how cardinality behaves under infinite unions, products, and exponentials. This would lead us to a rich notion of cardinal arithmetic. To be rigorous in this study would take substantial time and distract us from studying the main objects of focus in this course. Thus we mention the following two results.

Theorem 1.7.6 (Cantor's Theorem). *If X is an non-empty set, then $|X| \leq |\mathcal{P}(X)|$ but $|X| \neq |\mathcal{P}(X)|$.*

Proof. Exercise. ∎

Example 1.7.7. Let $X = \prod_{n=1}^{\infty}\{0,1\}$. The cardinality of X is denoted by $2^{|\mathbb{N}|}$ (as we are taking a $|\mathbb{N}|$ product of $\{0,1\}$ which has cardinality 2). We claim that $2^{|\mathbb{N}|} = |\mathbb{R}|$. To see this, first define $f : X \to [0,1]$ by

$$f((a_n)_{n\geq 1}) = \sum_{n=1}^{\infty} \frac{2a_n}{3^n}.$$

We claim that f is injective. To see this, we notice that $f((a_n)_{n\geq 1})$ is a ternary expansion of a number in $[0,1]$. Since the ternary expansion of a number in $[0,1]$ is unique up to repeating 2s (i.e. $\sum_{n=2}^{\infty} \frac{2}{3^n} = \frac{1}{3}$), and changing repeating 2s either changes a 1 to a 2 or a 0 to a 1, each number in $[0,1]$ that can be expressed using ternary numbers only involving 0s and 2s can be done so in a unique way. Hence f is injective so $|X| \leq |[0,1]| \leq |\mathbb{R}|$.

For the other direction, define $g : (0,1) \to X$ as follows: for each $x \in (0,1)$ write a binary expansion of x, say $x = \sum_{n=1}^{\infty} \frac{a_n}{2^n}$ where $a_n \in \{0,1\}$, and define $g(x) = (a_n)_{n\geq 1}$ (this is valid by the Axiom of Choice). Clearly g is well-defined. Furthermore, g is injective since if two numbers have the same binary expansion, they are the same number. Hence $|\mathbb{R}| = |(0,1)| \leq |X|$ so $2^{|\mathbb{N}|} = |\mathbb{R}|$ by Theorem 1.6.9 as desired.

Chapter 2

Measure Spaces

We have seen that the notion of cardinality of a set provides us with an idea about how large some sets are with respect to others. However, the notion of cardinality has some deficiencies. For example we have seen that the Cantor set has the same cardinality as the set $[0, 1]$ yet our intuition says that $[0, 1]$ has far more elements than the Cantor set as the Cantor set does not contain any intervals (i.e. we have shown the Cantor set is nowhere dense in Remark 1.5.11).

The goal for this section is to develop a notion of length for subsets of the real numbers. This will improve on our notion of cardinality in that we expect every countable subset of the real numbers to have zero length whereas larger sets should have positive length. We will endeavour to keep our theory as general as possible without becoming too distracted from the task at hand. In particular, this general theory is the basis for probability theory and describing the probability of a collection of events. Furthermore, this notion of length will enable us to generalize the notion of the Riemann integral in Chapter 4.

2.1 Measure Spaces

To begin to develop a notion of length for subsets of the real numbers, we must ask ourselves, "Do we expect to assign a length to every subset of the real numbers?" Of course this may be possible or it may not be possible. Thus, to keep our assumptions as weak as possible, we will define a structure that will encapsulate which sets can be measured. We may also think of the following structure through probability. The following structure will encapsulate which events may occur and we expect to be able to measure.

Definition 2.1.1. Let X be a non-empty set. A *σ-algebra* is a subset $\mathcal{A} \subseteq \mathcal{P}(X)$ such that

1. $\emptyset \in \mathcal{A}$ and $X \in \mathcal{A}$ (that is, we can measure the empty event and the full event),

2. if $A \in \mathcal{A}$ then $A^c = X \setminus A \in \mathcal{A}$ (that is, we can measure the complement of an event), and

3. if $\{A_n\}_{n=1}^\infty \subseteq \mathcal{A}$, then $\bigcup_{n=1}^\infty A_n \in \mathcal{A}$ (that is, we can measure the union of a countable collection of events).

The pair $(X, \mathcal{A})$ is called a *measurable space* and the elements of $\mathcal{A}$ are called *measurable sets*.

Remark 2.1.2. One may expect that we would like the intersection of a countable collection of measurable sets to be measurable. Indeed this is the case as if $(X, \mathcal{A})$ is a measurable space and $\{A_n\}_{n=1}^\infty \subseteq \mathcal{A}$, then

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

as complements and countable unions of elements of $\mathcal{A}$ are elements of $\mathcal{A}$. Furthermore, by using $\emptyset$ in unions and X is intersections, clearly a finite union or intersection of elements of $\mathcal{A}$ is an element of $\mathcal{A}$.

It is useful to have some examples of measurable spaces. Indeed we clearly have some trivial examples.

Example 2.1.3. Let X be a non-empty set. Then $(X, \mathcal{P}(X))$ is a measurable space and $(X, \{\emptyset, X\})$ is a measurable space.

Example 2.1.4. Let X be a non-empty set and let

$$\mathcal{A} = \{A \subseteq X \mid A \text{ is countable or } A^c \text{ is countable}\}.$$

Then $(X, \mathcal{A})$ is a measurable space.

Using the definition of a σ-algebra, it is trivial to verify the following result.

Lemma 2.1.5. *Let X be a non-empty set and let $\{\mathcal{A}_\alpha \mid \alpha \in I\}$ be a collection of σ-algebras of X. Then*

$$\bigcap_{\alpha \in I} \mathcal{A}_\alpha$$

is a σ-algebra of X.

Remark 2.1.6. Using the same idea as how one constructs the closure of a set, we can construct the smallest σ-algebra containing a collection of subsets. Indeed let X be a non-empty set and let $A \subseteq \mathcal{P}(X)$. Define

$$I = \{\mathcal{A} \mid \mathcal{A} \text{ is a } \sigma\text{-algerba of } X \text{ such that } A \subseteq \mathcal{A}\}.$$

Clearly $\mathcal{P}(X) \in I$ so I is non-empty. Hence Lemma 2.1.5 implies that

$$\sigma(A) = \bigcap_{\mathcal{A} \in I} \mathcal{A}$$

is a σ-algebra. Since clearly $A \in \sigma(A)$ by construction, $\sigma(A)$ is the smallest σ-algebra of X that contains A. As such, it is said that $\sigma(A)$ is the *σ-algebra generated by A*.

Definition 2.1.7. Let $(\mathcal{X}, d)$ be a metric space. The σ-algebra generated by the open subsets of $\mathcal{X}$ is called the *Borel σ-algebra* and is denoted $\mathfrak{B}(\mathcal{X})$. In particular, $\mathfrak{B}(\mathcal{X})$ is also the σ-algebra generated by the closed subsets of $\mathcal{X}$ as open and closed sets are complements of each other and as σ-algebras are closed under complements. Elements of $\mathfrak{B}(\mathcal{X})$ are called *Borel sets*.

Remark 2.1.8. In terms of the Borel subsets of $\mathbb{R}$, the sets

$$\begin{aligned}
&\{(a, b) \mid a, b \in \mathbb{R}, a < b\} \subseteq \mathcal{P}(\mathbb{R}) \\
&\{(a, b] \mid a, b \in \mathbb{R}, a < b\} \subseteq \mathcal{P}(\mathbb{R}) \\
&\{[a, b) \mid a, b \in \mathbb{R}, a < b\} \subseteq \mathcal{P}(\mathbb{R}) \\
&\{[a, b] \mid a, b \in \mathbb{R}, a < b\} \subseteq \mathcal{P}(\mathbb{R}) \\
&\{(-\infty, b) \mid b \in \mathbb{R}\} \subseteq \mathcal{P}(\mathbb{R}) \\
&\{(-\infty, b] \mid b \in \mathbb{R}\} \subseteq \mathcal{P}(\mathbb{R}) \\
&\{(a, \infty) \mid a \in \mathbb{R}\} \subseteq \mathcal{P}(\mathbb{R}) \\
&\{[a, \infty) \mid a \in \mathbb{R}\} \subseteq \mathcal{P}(\mathbb{R})
\end{aligned}$$

all can be shown to generate $\mathfrak{B}(\mathbb{R})$ via unions, intersections, and complements (i.e. show that $\mathfrak{B}(\mathbb{R})$ contains each of these sets and any σ-algebra containing one of these sets contains all open intervals and thus all open sets by the fact that every open set is a countable union of open intervals. In particular, the Borel subsets of $\mathbb{R}$ are sets we will definitely want to be able to assign length to.

It is possible to show that $|\mathfrak{B}(\mathbb{R})| = |\mathbb{R}|$. Unfortunately, the simplest proof uses *transfinite induction* to construct $\mathfrak{B}(\mathbb{R})$ via an uncountable union of sets obtained by taking countable unions and complements of a previous set, starting with the set of open subsets of $\mathbb{R}$. However, as by Cantor's Theorem (Theorem 1.7.6) we know that $|\mathbb{R}| < |\mathcal{P}(\mathbb{R})|$ so there are by far many more subsets of $\mathbb{R}$ than there are Borel subsets of $\mathbb{R}$. This raises the question, "Will we be able to assign lengths to only Borel sets?"

We now turn our attention the question of assigning lengths to our measurable sets with the desired properties or assigning probabilities to our events. For the rest of the course, if $a_n \in [0, \infty]$ for all $n \in \mathbb{N}$ and $a_k = \infty$ for some k, then $\sum_{n=1}^{\infty} a_n$ is defined to be ∞.

Definition 2.1.9. Let $(X, \mathcal{A})$ be a measurable space. A *(countably additive, positive) measure* on $(X, \mathcal{A})$ is a function $\mu : \mathcal{A} \to [0, \infty]$ such that

- $\mu(\emptyset) = 0$, and
- (countable additivity on disjoint subsets) if $\{A_n\}_{n=1}^\infty \subseteq \mathcal{A}$ are pairwise disjoint (i.e. $A_n \cap A_m = \emptyset$ if $n \neq m$), then

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

The triple $(X, \mathcal{A}, \mu)$ is called a *measure space* and given an element $A \in \mathcal{A}$, $\mu(A)$ is called the *μ-measure of A*.

Before we discuss a few examples, we note some trivial properties of our measures.

Remark 2.1.10. Notice if $(X, \mathcal{A}, \mu)$ is a measure space and $A_1, \ldots, A_n$ are pairwise disjoint subsets of $\mathcal{A}$, then

$$\mu\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n \mu(A_k)$$

by using the properties of a measure with $A_k = \emptyset$ for all $k > n$. Consequently, if $E, F \in \mathcal{A}$ and $E \subseteq F$, then $F \setminus E = F \cap E^c \in A$ is disjoint from E so

$$\mu(F) = \mu(E \cup (F \setminus E)) = \mu(E) + \mu(F \setminus E) \geq \mu(E).$$

In particular, if $\mathcal{A}$ is ordered by inclusion, then μ is monotone with respect to this inclusion. This implies if $\mu(X) < \infty$, then $\mu(A) < \infty$ for all $A \in \mathcal{A}$. Finally, notice in the above computation that if $\mu(E) < \infty$ then we may subtract $\mu(E)$ from both sides in order to obtain that $\mu(F \setminus E) = \mu(F) - \mu(E)$.

Finally, if $A, B \in \mathcal{A}$ are arbitrary and $\mu(A \cap B) < \infty$, then

$$\mu(A \cup B) = \mu(A \cup (B \setminus (B \cap A)) = \mu(A) + \mu(B \setminus (B \cap A)) = \mu(A) + \mu(B) - \mu(A \cap B)$$

which is a formula that may appear familiar in the context of probability.

Using the above, we desire to discuss some special types of measures.

Definition 2.1.11. A measure μ on a measurable space $(X, \mathcal{A})$ is said to be

- *finite* if $\mu(X) < \infty$.
- *σ-finite* if there exists a collection $\{A_n\}_{n=1}^\infty \subseteq \mathcal{A}$ such that $X = \bigcup_{n=1}^\infty A_n$ and $\mu(A_n) < \infty$ for all $n \in \mathbb{N}$.

Although finite measures are the simplest to handle, we will see later that σ-finite measures are easily handled in a similar way to finite measure in regards to analytic properties.

We may also define and obtain examples of measure spaces through probability now. In particular, every problem in probability theory takes place in the following context.

Definition 2.1.12. Let $(X, \mathcal{A}, \mu)$ be a measure space. It is said that $(X, \mathcal{A}, \mu)$ is a *probability space* and μ is a *probability measure* if $\mu(X) = 1$. In this case, elements of $\mathcal{A}$ are called *events* and given $A \in \mathcal{A}$, $\mu(A)$ denotes the probability that the event A occurs.

It is not difficult to see that a probability space is the correct notion in order to study probability theory. Indeed the probability of the entire space is one and whenever A and B are disjoint sets, which is the notion of independent events, then the probability of $A \cup B$ is the sum of the probability of A and the probability of B. Furthermore, Remark 2.1.10 is precisely the formula for the probability of $A \cup B$ when A and B are not disjoint; that is, the formula for the probability of the union of two not necessarily independent events.

For examples of measure, we need not go far.

Example 2.1.13. Let X be a non-empty set and let $x \in X$. The *point-mass measure at x* is the measure δ_x on $(X, \mathcal{P}(X))$ defined by

$$\delta_x(A) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}.$$

It is elementary to verify that δ_x is a measure.

Example 2.1.14. Let X be a non-empty set. The *counting measure on X* is the measure μ on $(X, \mathcal{P}(X))$ defined by

$$\mu(A) = \begin{cases} |A| & \text{if } A \text{ is finite} \\ \infty & \text{otherwise} \end{cases}.$$

It is elementary to verify that μ is a measure.

Example 2.1.15. A function $\mu : \mathcal{P}(\mathbb{N}) \to [0, \infty]$ is a measure on $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ if and only if there exists a sequence $(a_n)_{n\geq 1}$ of elements of $[0, \infty]$ such that

$$\mu(A) = \sum_{n \in A} a_n$$

for all $A \subseteq \mathbb{N}$. To see this, note that it is elementary to verify that if μ has the described form, then μ is a measure.

Conversely, suppose μ is a measure on $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$. As for each $n \in \mathbb{N}$ the set $\{n\}$ is measurable, for each $n \in \mathbb{N}$ we may define

$$a_n = \mu(\{n\}) \in [0, \infty].$$

We claim that

$$\mu(A) = \sum_{n \in A} a_n$$

for all $A \subseteq \mathbb{N}$. To see this, let $A \subseteq \mathbb{N}$ be arbitrary. Then, as A is countable and

$$A = \bigcup_{n \in A} \{n\},$$

we obtain by the properties of a measure that

$$\mu(A) = \mu\left(\bigcup_{n \in A} \{n\}\right) = \sum_{n \in A} \mu(\{n\}) = \sum_{n \in A} a_n$$

as desired.

It is also easy to construct new measures from old measures.

Example 2.1.16. Let $(X, \mathcal{A}, \mu)$ be a measure space, let $\{A_k\}_{k=1}^n \subseteq \mathcal{A}$ be disjoint, and let $\{a_k\}_{k=1}^n \in [0, \infty)$. Define $\nu : \mathcal{A} \to [0, \infty]$ by

$$\nu(A) = \sum_{k=1}^{n} a_k \mu(A_k \cap A)$$

for all $A \in \mathcal{A}$ where

$$a \times \infty = \begin{cases} 0 & \text{if } a = 0 \\ \infty & \text{otherwise} \end{cases}.$$

Then ν is a measure on $(X, \mathcal{A})$. To see this, we clearly note that $\nu(\emptyset) = 0$. Furthermore, if $\{B_m\}_{m=1}^\infty \subseteq \mathcal{A}$ are pairwise disjoint, then $\{A_k \cap B_m\}_{m=1}^\infty$ are pairwise disjoint for all k and thus, as μ is a measure,

$$\begin{aligned}
\nu\left(\bigcup_{m=1}^{\infty} B_m\right) &= \sum_{k=1}^{n} a_k \mu\left(A_k \cap \left(\bigcup_{m=1}^{\infty} B_m\right)\right) \\
&= \sum_{k=1}^{n} a_k \mu\left(\bigcup_{m=1}^{\infty} (A_k \cap B_m)\right) \\
&= \sum_{k=1}^{n} \sum_{m=1}^{\infty} a_k \mu\left(A_k \cap B_m\right) \\
&= \sum_{m=1}^{\infty} \sum_{k=1}^{n} a_k \mu\left(A_k \cap B_m\right) \quad \text{as all terms are non-negative} \\
&= \sum_{m=1}^{\infty} \nu(B_m).
\end{aligned}$$

Hence ν is a measure as desired.

We will see further examples of measures later, including the measure we wish to define on $\mathbb{R}$. Before moving on to constructing this measure, we desire some analytical properties of measures. In particular, in Definition 2.1.9 we defined the measure of a countable union of disjoint sets to be the sum of the measures.

Proposition 2.1.17 (Subadditivity of Measures). *Let $(X, \mathcal{A}, \mu)$ be a measure space and let $\{A_n\}_{n=1}^\infty \subseteq \mathcal{A}$. Then*

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

Proof. Define $E_1 = A_1$ and for each $n \in \mathbb{N}$ with $n \geq 2$, define

$$E_n = A_n \setminus \left(\bigcup_{k=1}^{n-1} A_k\right).$$

Clearly by the definite of a σ-algebra we have that $E_n \in \mathcal{A}$ for all $n \in \mathbb{N}$. Furthermore, it is clear that $E_n \cap E_m = \emptyset$ if $n \neq m$, $E_n \subseteq A_n$ for all n, and

$$\bigcup_{n=1}^\infty A_n = \bigcup_{n=1}^\infty E_n.$$

Hence by Definition 2.1.9 and the monotonicity of measures (Remark 2.1.10), we obtain that

$$\begin{aligned}\mu\left(\bigcup_{n=1}^\infty A_n\right) &= \mu\left(\bigcup_{n=1}^\infty E_n\right)\\ &= \sum_{n=1}^\infty \mu(E_n)\\ &\leq \sum_{n=1}^\infty \mu(A_n)\end{aligned}$$

as desired. ∎

For a more analytic result, we need only consider increase and decreasing sequences of measurable sets.

Theorem 2.1.18 (Monotone Convergence Theorem, Measures). *Let $(X, \mathcal{A}, \mu)$ be a measure space and let $\{A_n\}_{n=1}^\infty \subseteq \mathcal{A}$.*

(1) If $A_n \subseteq A_{n+1}$ for all $n \in \mathbb{N}$, then $\mu\left(\bigcup_{n=1}^\infty A_n\right) = \lim_{n\to\infty} \mu(A_n)$.

(2) If $A_{n+1} \subseteq A_n$ for all $n \in \mathbb{N}$ and $\mu(A_1) < \infty$, then $\mu\left(\bigcap_{n=1}^\infty A_n\right) = \lim_{n\to\infty} \mu(A_n)$.

Proof. To see (1), let $A_0 = \emptyset$ for notational simplicity. If for each $n \in \mathbb{N}$ we define

$$B_n = A_n \setminus A_{n-1},$$

then $\{B_n\}_{n=1}^\infty$ is a collection of pairwise disjoint elements of $\mathcal{A}$ such that $\bigcup_{k=1}^\infty B_k = \bigcup_{k=1}^\infty A_k$ and $\bigcup_{k=1}^n B_k = A_n$ for all n. Hence

$$\begin{aligned}\mu\left(\bigcup_{n=1}^\infty A_n\right) &= \sum_{k=1}^\infty \mu(B_k)\\ &= \lim_{n\to\infty} \sum_{k=1}^n \mu(B_k)\\ &= \lim_{n\to\infty} \mu\left(\bigcup_{k=1}^n B_k\right)\\ &= \lim_{n\to\infty} \mu(A_n)\end{aligned}$$

as desired.

For part (2), notice if $B_n = A_1 \setminus A_n$, then $\{B_n\}_{n=1}^\infty$ is a collection of elements of $\mathcal{A}$ with $B_n \subseteq B_{n+1}$ for all $n \in \mathbb{N}$. Hence, as

$$A_1 \setminus \left(\bigcap_{n=1}^\infty A_n\right) = \bigcup_{n=1}^\infty B_n$$

part (1) implies that

$$\mu\left(A_1 \setminus \left(\bigcap_{n=1}^\infty A_n\right)\right) = \lim_{n\to\infty} \mu(B_n) = \lim_{n\to\infty} \mu(A_1 \setminus A_n).$$

As $\mu(A_1) < \infty$, Remark 2.1.10 implies that $\mu(A_1 \setminus E) = \mu(A_1) - \mu(E)$ for all $E \in \mathcal{A}$ with $E \subseteq A_1$. Hence the result follows. ∎

2.2 The Caratheodory Method

With the above structures in place, we desire a measure λ on $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$ that emulates the length of a set. In particular, we desire such a measure to have some very natural properties, such as:

1. if I is an interval, then $\lambda(I)$ is the length of I.
2. if $A \in \mathcal{P}(\mathbb{R})$, $x \in \mathbb{R}$, and $x+A = \{x+a \mid a \in A\}$, then $\lambda(x+A) = \lambda(A)$; that is, λ is translation invariant.

However, it turns out that no such measure exists! This can be seen via the following example.

Example 2.2.1. Suppose λ is a measure on $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$ with the above two properties. We desire to construct some sets that will cause a contradiction.

Define an equivalence relation $\sim$ on $\mathbb{R}$ by $x \sim y$ if and only if $x - y \in \mathbb{Q}$. Clearly every equivalence class under $\sim$ has an element in $[0, 1)$ and by the Axiom of Choice there exists a subset A of $[0, 1)$ that contains precisely one element from each equivalence class.

Since $\mathbb{Q}$ is countable, we may enumerate $\mathbb{Q} \cap [0, 1)$ as $\mathbb{Q} \cap [0, 1) = \{r_n \mid n \in \mathbb{N}\}$. For each $n \in \mathbb{N}$, let

$$A_n = \{x \in [0, 1) \mid x \in r_n + A \text{ or } x + 1 \in r_n + A\}$$

(that is, A_n is $r_n + A$ modulo 1). Notice if $x \in [0, 1)$ then there exists a unique $y \in A$ such that $x - y \in \mathbb{Q}$. If $x - y \geq 0$ then $x - y = r_n$ for some n and thus $x \in A_n$. Otherwise if $x - y < 0$ then $(x + 1) - y = r_n$ for some n and thus $x \in A_n$. Therefore

$$[0, 1) = \bigcup_{n=1}^{\infty} A_n$$

and the uniqueness of y for x in the above argument implies $A_n \cap A_m = \emptyset$ if $n \neq m$. Thus $\{A_n\}_{n=1}^{\infty}$ is a collection of pairwise disjoint sets whose union is $[0, 1)$.

For each $n \in \mathbb{N}$, let

$$\begin{aligned} B_{n,1} &= (r_n + A) \cap [0, 1) \\ B_{n,2} &= -1 + ((r_n + A) \cap [1, 2)) . \end{aligned}$$

Clearly $A_n = B_{n,1} \cup B_{n,2}$ as $r_n + A \subseteq [0, 2)$ for all n. Furthermore, we claim that $B_{n,1} \cap B_{n,2} = \emptyset$. To see this, suppose to the contrary that $b \in B_{n,1} \cap B_{n,2}$. Hence there exists $x, y \in A$ such that $r_n + x \in [0, 1)$, $r_n + y \in [1, 2)$, and $b = r_n + x = -1 + r_n + y$. The first two conditions imply that $x \neq y$ whereas the later condition implies $x \sim y$. Therefore, as A contains exactly one element from each equivalence class, we have obtained a contradiction. Hence $B_{n,1} \cap B_{n,2} = \emptyset$.

To obtain our contradiction, suppose λ has the desire properties. Then

$$\begin{aligned}
1 &= \lambda([0,1)) \\
&= \lambda\left(\bigcup_{n=1}^{\infty} A_n\right) \\
&= \sum_{n=1}^{\infty} \lambda(A_n) \qquad \text{the } A_n \text{ are disjoint} \\
&= \sum_{n=1}^{\infty} \lambda(B_{n,1} \cup B_{n,2}) \\
&= \sum_{n=1}^{\infty} \lambda(B_{n,1}) + \lambda(B_{n,2}) \qquad B_{n,1} \cap B_{n,2} = \emptyset \\
&= \sum_{n=1}^{\infty} \lambda((r_n + A) \cap [0,1)) + \lambda\left(-1 + ((r_n + A) \cap [1,2))\right) \qquad \text{translation} \\
&= \sum_{n=1}^{\infty} \lambda(r_n + A) \qquad \text{disjointness} \\
&= \sum_{n=1}^{\infty} \lambda(A) \qquad \text{translation.}
\end{aligned}$$

However, this is impossible as no number in $[0, \infty]$ when summed an infinite number of times produces 1. Thus we have obtained a contradiction to the existence of such a λ defined on $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$.

The above example illustrates that $\mathcal{P}(\mathbb{R})$ is too large; that is, there are too many sets in $\mathcal{P}(\mathbb{R})$ to define such a measure in a consistent way. The set A in Example 2.2.1 is one of these sets.

To solve this problem, our answer is to reduce the number of sets we consider measurable. Of course, if we would like to do analysis, we need our open sets to be measurable and thus we require all Borel sets to be measurable. However, the problem still remains, "How do we construct our measure and determine which sets are measurable?"

To answer this problem, we will invoke a technique called the Caratheodory Method. The idea of this method is given a set X to define a function on the power set of X that is almost a measure, but has weaker properties. We will then define sets that behave 'nicely' and show these nice sets form a σ-algebra. Finally, we will demonstrate that restricting the function to these nice sets does indeed produce a measure space.

To begin, we define the 'function' that behaves almost like a measure.

Definition 2.2.2. Let X be a non-empty set. A function $\mu^* : \mathcal{P}(X) \to [0, \infty]$ is said to be an *outer measure* if

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

(2) $\mu^*(A) \leq \mu^*(B)$ whenever $A, B \in \mathcal{P}(X)$ are such that $A \subseteq B$, and

(3) (countable subadditivity on disjoint subsets) if $\{A_n\}_{n=1}^\infty \subseteq \mathcal{P}(X)$, then $\mu^*\left(\bigcup_{n=1}^\infty A_n\right) \leq \sum_{n=1}^\infty \mu^*(A_n)$.

Notice that every measure is an outer measure by the results of Section 2.1 whereas an outer measure need not be a measure as it is not necessary that equality occur in the third property of Definition 2.2.2 when the collection $\{A_n\}_{n=1}^\infty$ are pairwise disjoint.

The outer measure we are interested in studying is as follows and is motivated by the notion of open covers.

Definition 2.2.3. Given an interval $I \subseteq \mathbb{R}$, let $\ell(I)$ denote the length of I (where the length of an infinite interval is assigned ∞ and the length of the empty set is 0). The *Lebesgue outer measure* is the function $\lambda^* : \mathcal{P}(\mathbb{R}) \to [0, \infty]$ defined by

$$\lambda^*(A) = \inf\left\{ \sum_{n=1}^\infty \ell(I_n) \,\middle|\, \begin{array}{c}\{I_n \mid n \in \mathbb{N}\} \text{ are open intervals} \\ \text{such that } A \subseteq \bigcup_{n=1}^\infty I_n\end{array} \right\}$$

for all $A \subseteq \mathbb{R}$ (where $\inf\{\infty\} = \infty$).

Theorem 2.2.4. *The Lebesgue outer measure is an outer measure.*

Proof. Clearly $\lambda^* : \mathcal{P}(\mathbb{R}) \to [0, \infty]$, $\lambda^*(\emptyset) = 0$, and $\lambda^*(A) \leq \lambda^*(B)$ if $A \subseteq B$. Verify the final property from Definition 2.2.2 for λ^*, let $\{A_n\}_{n=1}^\infty \subseteq \mathcal{P}(X)$ and let $A = \bigcup_{n=1}^\infty A_n$. Fix $\epsilon > 0$. By the definition of λ^*, for each $n \in \mathbb{N}$ there exists a collection $\{I_{n,k} \mid k \in \mathbb{N}\}$ of open intervals such that $A_n \subseteq \bigcup_{k=1}^\infty I_{n,k}$ and

$$\sum_{k=1}^\infty \ell(I_{n,k}) \leq \lambda^*(A_n) + \frac{\epsilon}{2^n}.$$

Clearly $\{I_{n,k} \mid n, k \in \mathbb{N}\}$ is a collection of open intervals such that

$$A \subseteq \bigcup_{n,k=1}^\infty I_{n,k}.$$

Hence, by the definition of λ^*

$$\lambda^*(A) \leq \sum_{n,k=1}^\infty \ell(I_{n,k}) \leq \sum_{n=1}^\infty \lambda^*(A_n) + \frac{\epsilon}{2^n} = \epsilon + \sum_{n=1}^\infty \lambda^*(A_n).$$

Therefore, as $\epsilon > 0$ was arbitrary, we obtain that

$$\lambda^*(A) \leq \sum_{n=1}^\infty \lambda^*(A_n).$$

Hence λ^* is an outer measure. ∎

Now that we have what will be our outer measure (although we have not shown any properties of λ^* that appear like those we want which we are postponing as we have yet to demonstrate how we will get our measure) we desire to define which sets we will be restricting our attention to.

Definition 2.2.5. Let X be a non-empty set and let $\mu^* : \mathcal{P}(X) \to [0, \infty]$ be an outer measure on X. A subset $A \subseteq X$ is said to be *μ^*-measurable* or *outer measurable* if for every $B \in \mathcal{P}(X)$

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

Notice by the properties of an outer measure that if $A, B \in \mathcal{P}(X)$ then

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

Thus to show that A is outer measurable, it suffices to show that

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

for all $B \in \mathcal{P}(X)$. Furthermore, clearly it suffices to restrict our attention to B such that $\mu^*(B) < \infty$. The reason we are interested in outer measurable sets is that if $A \subseteq X$ has the property that

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

for some $B \in \mathcal{P}(X)$, it is likely we don't want to consider A to be measurable as it causes μ^* to fail to be additive on specific disjoint sets if B was also measurable.

The notion of outer measurable sets allows us to construct a measure space as follows.

Theorem 2.2.6. *Let X be a non-empty set and let $\mu^* : \mathcal{P}(X) \to [0, \infty]$ be an outer measure on X. The set $\mathcal{A}$ of all outer measurable sets is a σ-algebra. Furthermore $\mu^*|_{\mathcal{A}}$ is a measure on $(X, \mathcal{A})$.*

Proof. First notice for all $B \in \mathcal{P}(X)$ that

$$\mu^*(B) = \mu^*(B) + 0 = \mu^*(B \cap \emptyset^c) + \mu^*(B \cap \emptyset).$$

Hence $\emptyset \in \mathcal{A}$. Furthermore, clearly if $A \in \mathcal{A}$ then clearly $A^c \in \mathcal{A}$ due to the symmetry in the definition of an outer measurable set. Hence $X \in \mathcal{A}$.

In order to demonstrate a countable union of elements of $\mathcal{A}$ is in $\mathcal{A}$, let us first verify that $\mathcal{A}$ is closed under finite unions. To begin, let $A_1, A_2 \in \mathcal{A}$. To see that $A_1 \cup A_2 \in \mathcal{A}$, let $B \subseteq X$ be arbitrary. Since A_1 is outer measurable, we know that

$$\mu^*(B) = \mu^*(B \cap A_1) + \mu^*(B \cap A_1^c).$$

Furthermore, since A_2 is outer measurable, we see that

$$\mu^*(B \cap A_1^c) = \mu^*(B \cap A_1^c \cap A_2) + \mu^*(B \cap A_1^c \cap A_2^c).$$

Hence

$$\mu^*(B) = \mu^*(B \cap A_1) + \mu^*(B \cap A_1^c \cap A_2) + \mu^*(B \cap A_1^c \cap A_2^c).$$

However, since

$$B \cap (A_1 \cup A_2) = (B \cap A_1) \cup (B \cap (A_2 \cap A_1^c)),$$

subadditivity implies that

$$\begin{aligned} \mu^*(B) &= \mu^*(B \cap A_1) + \mu^*(B \cap A_1^c \cap A_2) + \mu^*(B \cap A_1^c \cap A_2^c) \\ &\geq \mu^*(B \cap (A_1 \cup A_2)) + \mu^*(B \cap A_1^c \cap A_2^c) \\ &= \mu^*(B \cap (A_1 \cup A_2)) + \mu^*(B \cap (A_1 \cup A_2)^c) \end{aligned}$$

Therefore, as $B \subseteq X$ was arbitrary, we obtain that $A_1 \cup A_2 \in \mathcal{A}$. Thus, as $\mathcal{A}$ is also closed under complements, we also obtain that $A_1 \cap A_2 \in \mathcal{A}$.

With the above complete, we may now proceed to show that $\mathcal{A}$ is closed under countable unions. Let $\{A_n\}_{n=1}^\infty \subseteq \mathcal{A}$. For each $n \in \mathbb{N}$, define E_n via $E_1 = A_1$ and for $n \geq 1$,

$$E_n = A_n \setminus \left(\bigcup_{k=1}^{n-1} A_k \right) = A_n \cap \left(\bigcup_{k=1}^{n-1} A_l \right)^c.$$

Clearly $\{E_n\}_{n=1}^\infty$ are pairwise disjoint such that $\bigcup_{n=1}^\infty E_n = \bigcup_{n=1}^\infty A_n$. Furthermore, $E_n \in \mathcal{A}$ for all $n \in \mathbb{N}$ by the above argument.

To see that $E = \bigcup_{n=1}^\infty E_n$ is an element of $\mathcal{A}$, let $B \subseteq X$ be arbitrary. For each $n \in \mathbb{N}$, let $F_n = \bigcup_{k=1}^n E_k$, which is an element of $\mathcal{A}$ by the above work. Therefore, as $F_n \subseteq E$ so $E^c \subseteq F_n^c$, we obtain by the fact that F_n is outer measurable and by the fact that μ^* is monotone that

$$\mu^*(B) = \mu^*(B \cap F_n) + \mu^*(B \cap F_n^c) \geq \mu^*(B \cap F_n) + \mu^*(B \cap E^c)$$

for all $n \in \mathbb{N}$. However, notice that $F_n = F_{n-1} \cup E_n$ and $F_{n-1} \cap E_n = \emptyset$ by construction. Therefore, since $E_n \in \mathcal{A}$,

$$\mu^*(B \cap F_n) = \mu^*(B \cap F_n \cap E_n) + \mu^*(B \cap F_n \cap E_n^c) = \mu^*(B \cap E_n) + \mu^*(B \cap F_{n-1}).$$

By repeating the above argument another $n - 2$ times, we obtain that

$$\mu^*(B \cap F_n) = \sum_{k=1}^n \mu^*(B \cap E_k)$$

for all $n \in \mathbb{N}$. Hence

$$\mu^*(B) \geq \mu^*(B \cap E^c) + \sum_{k=1}^{n} \mu^*(B \cap E_k)$$

for all $n \in \mathbb{N}$. By taking the supremum of the right-hand-side of the above expression, we obtain that

$$\mu^*(B) \geq \mu^*(B \cap E^c) + \sum_{k=1}^{\infty} \mu^*(B \cap E_k).$$

Therefore, by subadditivity, we obtain that

$$\mu^*(B) \geq \mu^*\left(B \cap \left(\bigcup_{k=1}^{\infty} E_k\right)\right) + \mu^*(B \cap E^c) = \mu^*(B \cap E) + \mu^*(B \cap E^c).$$

Therefore, as $B \subseteq X$ was arbitrary, we obtain that $E \in \mathcal{A}$ as desired. Hence $\mathcal{A}$ is a σ-algebra.

To see that μ^* is a measure when restricted to $\mathcal{A}$, first notice that $\mu^*(\emptyset) = 0$ by design. To check the other property of Definition 2.1.9, let $\{E_n\}_{n=1}^{\infty}$ be an arbitrary collection of pairwise disjoint elements of $\mathcal{A}$ and let $E = \bigcup_{n=1}^{\infty} E_n$. Using the above computation with E in place of B, we see that

$$\mu^*(E) \geq \mu^*(E \cap E^c) + \sum_{k=1}^{\infty} \mu^*(E \cap E_k) = \sum_{k=1}^{\infty} \mu^*(E_k).$$

However, since subadditivity of outer measures implies

$$\mu^*(E) \leq \sum_{k=1}^{\infty} \mu^*(E_k)$$

we obtain that

$$\mu^*(E) = \sum_{k=1}^{\infty} \mu^*(E_k).$$

Hence μ^* is a measure when restricted to $\mathcal{A}$ as desired. ∎

Before moving on to studying the properties of the Lebesgue outer measure and its restriction of Lebesgue outer measurable sets, we note measures obtained from Theorem 2.2.6 have one additional property.

Definition 2.2.7. A measure space $(X, \mathcal{A}, \mu)$ is said to be *complete* if whenever $A \in \mathcal{A}$ and $B \in \mathcal{P}(X)$ are such that $B \subseteq A$ and $\mu(A) = 0$, then $B \in \mathcal{A}$.

Proposition 2.2.8. *Let X be a non-empty set, let $\mu^* : \mathcal{P}(X) \to [0, \infty]$ be an outer measure on X, and let $\mathcal{A}$ be the σ-algebra of all outer measurable sets. If $A \in \mathcal{P}(X)$ and $\mu^*(A) = 0$, then $A \in \mathcal{A}$. Hence $(X, \mathcal{A}, \mu^*|_{\mathcal{A}})$ is complete by the monotonicity of μ^*.*

Proof. Let $A \in \mathcal{P}(X)$. To see that $A \in \mathcal{A}$, let $B \in \mathcal{P}(X)$ be arbitrary. Then

$$0 \leq \mu^*(B \cap A) \leq \mu^*(A) = 0$$

by monotonicity. Hence, by monotonicity,

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

Therefore, as $B \in \mathcal{P}(X)$ was arbitrary, $A \in \mathcal{A}$. ∎

2.3 The Lebesgue Measure

Let λ^* be the Lebesgue outer measure from Definition 2.2.3. By Theorem 2.2.6 the collection $\mathcal{M}(\mathbb{R})$ of λ^*-measurable sets is a σ-algebra and $\lambda^*|_{\mathcal{M}(\mathbb{R})}$ is a measure. As these objects will be the focus for the remainder of our course, we make the following definition.

Definition 2.3.1. The *Lebesgue measure* on $\mathbb{R}$ is the measure $\lambda = \lambda^*|_{\mathcal{M}(\mathbb{R})}$. The elements of $\mathcal{M}(\mathbb{R})$ are called *Lebesgue measurable sets*.

Remark 2.3.2. Note by Proposition 2.2.8 that λ is a complete measure.

Our goal for this section is to study λ. In particular, we must answer the following questions:

- What types of sets are contained in $\mathcal{M}(\mathbb{R})$?
- Does λ have the desired properties discussed at the beginning of Section 2.2?

To begin these discussions, we show that the Lebesgue outer measure of an interval is what one would expect.

Theorem 2.3.3. *Let $I \subseteq \mathbb{R}$ be an interval. Then $\lambda^*(I) = \ell(I)$.*

Proof. First suppose $I = [a, b]$. To see that $\lambda^*(I) \leq b - a$, let $\epsilon > 0$ be arbitrary. Then $I_1 = (a - \epsilon, b + \epsilon)$ is an open interval such that $I \subseteq I_1$. Hence, by the definition of the Lebesgue outer measure (using the empty set for all other open intervals in our countable collection which covers I), we obtain that

$$\lambda^*(I) \leq \ell(I_1) = b - a + 2\epsilon.$$

Therefore, as $\epsilon > 0$ was arbitrary, we obtain that $\lambda^*(I) \leq b - a$.

For the other inequality, let $\{I_n \mid n \in \mathbb{N}\}$ be an arbitrary collection of open intervals of $\mathbb{R}$ such that $I \subseteq \bigcup_{n=1}^{\infty} I_n$. Hence $\{I_n \mid n \in \mathbb{N}\}$ is an open subcover of I. Therefore, since I is compact, there must exists a finite subcover of I from $\{I_n \mid n \in \mathbb{N}\}$. By reindexing the intervals if necessary, we may assume that $I \subseteq \bigcup_{k=1}^{m} I_k$ for some $m \in \mathbb{N}$.

Since $a \in I$, there exists a $k \in \{1, \ldots, m\}$ such that $a \in I_k$. By reindexing the intervals if necessary, we may assume that $a \in I_1$. Write $I_1 = (a_1, b_1)$. Hence $a_1 < a < b_1$. If $b \in I_1$ terminate this argument here. Otherwise $b_1 \leq b$ so $b_1 \in I$. Hence, as $I \subseteq \bigcup_{k=1}^m I_k$, there exists a $k \in \{1, \ldots, m\}$ such that $b_1 \in I_k$. By reindexing the intervals if necessary, we may assume that $b_1 \in I_2$. Write $I_2 = (a_2, b_2)$. Hence $a_1 < a < b_1 < b_2$. If $b < b_2$, terminate this argument here. Otherwise, as there are a finite number (specifically m) of intervals we need to consider, we may continue this process a finite number of times to obtain an $m' \leq m$ and intervals $I_k = (a_k, b_k)$ for $k \leq m'$ such that $a_{k+1} < b_k < b_{k+1}$ for all k and

$$a_1 < a < b_1 < b_2 < \cdots < b_{m'-1} < b < b_{m'}.$$

Hence

$$\begin{aligned} \sum_{k=1}^{\infty} \ell(I_k) &\geq \sum_{k=1}^{m'} \ell(I_k) \\ &= \sum_{k=1}^{m'} b_k - a_k \\ &\geq (b_1 - a_1) + \sum_{k=2}^{m'} b_k - b_{k-1} \\ &\geq b_{m'} - a_1 > b - a. \end{aligned}$$

Therefore, as $\{I_n \mid n \in \mathbb{N}\}$ was arbitrary, we obtain that $\lambda^*(I) \geq b - a$. Hence $\lambda^*(I) = b - a$ as desired.

To complete the proof, let $I \subseteq \mathbb{R}$ be an arbitrary interval. If I is finite in length, then for each $\epsilon > 0$ there exists a closed interval J such that $J \subseteq I$ and $\ell(I) \leq \ell(J) + \epsilon$. Hence, as $\overline{I}$ is a closed interval of the same length as I, we obtain that

$$\ell(I) - \epsilon \leq \ell(J) = \lambda^*(J) \leq \lambda^*(I) \leq \lambda^*(\overline{I}) = \ell(\overline{I}) = \ell(I)$$

by the above case for closed intervals. Therefore, as $\epsilon > 0$ was arbitrary, we obtain that $\lambda^*(I) = \ell(I)$.

Otherwise, if I is an infinite interval, fix $M > 0$. As I is infinite there exists a closed interval J such that $J \subseteq I$ and $\ell(J) = M$. Hence

$$\lambda^*(I) \geq \lambda^*(J) = \ell(J) = M$$

by the above case for closed intervals. Therefore, as $M > 0$ was arbitrary, we obtain that $\lambda^*(I) = \infty = \ell(I)$ as desired. ∎

Using Theorem 2.3.3, we can demonstrate certain sets are Lebesgue measurable.

Theorem 2.3.4. *For each $a \in \mathbb{R}$, (a, ∞) and $(-\infty, a]$ are Lebesgue measurable.*

Proof. As $(-\infty, a] = (a, \infty)^c$ and $\mathcal{M}(\mathbb{R})$ is a σ-algebra, it suffices to show that $(a, \infty) \in \mathcal{M}(\mathbb{R})$. To see this, let $B \subseteq \mathbb{R}$ be arbitrary. Therefore $B_1 = B \cap (a, \infty)$ and $B_2 = B \cap (-\infty, a]$ are disjoint sets such that $B = B_1 \cup B_2$.

Let $\epsilon > 0$ be arbitrary. By the definition of the Lebesgue outer measure, there exists a collection $\{I_n \mid n \in \mathbb{N}\}$ of open intervals such that $B \subseteq \bigcup_{n=1}^{\infty} I_n$ and

$$\sum_{n=1}^{\infty} \ell(I_n) \leq \lambda^*(B) + \epsilon.$$

For each $n \in \mathbb{N}$, let $I'_n = I_n \cap (a, \infty)$ and $I''_n = I_n \cap (\infty, a]$. Clearly I'_n an I''_n are disjoint intervals such that $I_n = I'_n \cup I''_n$ and $\ell(I_n) = \ell(I'_n) + \ell(I''_n)$. Furthermore, clearly $\{I'_n \mid n \in \mathbb{N}\}$ and $\{I''_n \mid n \in \mathbb{N}\}$ are collections of open intervals such that $B_1 \subseteq \bigcup_{n=1}^{\infty} I'_n$ and $B_2 \subseteq \bigcup_{n=1}^{\infty} I''_n$. Hence

$$\begin{aligned}
\lambda^*(B \cap (a, \infty)) + \lambda^*(B \cap (a, \infty)^c) &= \lambda^*(B_1) + \lambda^*(B_2) \\
&\leq \sum_{n=1}^{\infty} \lambda^*(I'_n) + \sum_{n=1}^{\infty} \lambda^*(I''_n) \\
&= \sum_{n=1}^{\infty} \ell(I'_n) + \sum_{n=1}^{\infty} \ell(I''_n) \\
&= \sum_{n=1}^{\infty} \ell(I_n) \\
&\leq \lambda^*(B) + \epsilon.
\end{aligned}$$

Therefore, as $\epsilon > 0$ was arbitrary, we obtain that

$$\lambda^*(B \cap (a, \infty)) + \lambda^*(B \cap (a, \infty)^c) \leq \lambda^*(B).$$

Therefore, as $B \subseteq \mathbb{R}$ was arbitrary, (a, ∞) is Lebesgue measurable as desired. ∎

Corollary 2.3.5. *Every Borel subset of $\mathbb{R}$ is Lebesgue measurable.*

Proof. As $\mathcal{M}(\mathbb{R})$ is a σ-algebra and Theorem 2.3.4 implies $(a, \infty), (-\infty, b] \in \mathcal{M}(\mathbb{R})$ for all a, b, we obtain that $(a, b] \in \mathcal{M}(\mathbb{R})$ for all $a < b$. Hence, as $\mathfrak{B}(\mathbb{R})$ is the smallest σ-algebra containing the Borel sets, Remark 2.1.8 implies that $\mathfrak{B}(\mathbb{R}) \subseteq \mathcal{M}(\mathbb{R})$. ∎

Example 2.3.6. We claim that the Cantor set $\mathcal{C}$ has Lebesgue measure zero. To see this, first recall that C is a closed set, hence C is a Borel set, and thus Lebesgue measurable.

To see that $\lambda(C) = 0$, recall from definition 1.5.10 that

$$C = \bigcap_{n \geq 1} P_n$$

where $P_n \subseteq [0,1]$ is the union of 2^n closed intervals each of length $\frac{1}{3^{n+1}}$. Therefore, we obtain for each $n \in \mathbb{N}$ that

$$0 \leq \lambda(C) \leq \lambda(P_n) \leq \frac{2^n}{3^{n+1}}.$$

Hence, as $\lim_{n\to\infty} \frac{2^n}{3^{n+1}} = 0$, we obtain that $\lambda(C) = 0$ as desired.

In fact, many other subsets of $\mathbb{R}$ are Lebesgue measurable.

Proposition 2.3.7. *Let $A \subseteq \mathbb{R}$ be countable. Then $A \in \mathcal{M}(\mathbb{R})$ and $\lambda(A) = 0$.*

Proof. Let $A \subseteq \mathbb{R}$ be countable. First we will show that $\lambda^*(A) = 0$. This implies A is Lebesgue measurable and $\lambda(A) = 0$ as λ is complete.

To see that $\lambda^*(A) = 0$, let $\epsilon > 0$ be arbitrary. Since A is countable, we may write $A = \{a_n\}_{n=1}^\infty$. For each $n \in \mathbb{N}$, let

$$I_n = \left(a_n - \frac{\epsilon}{2^{n+1}}, a_n + \frac{\epsilon}{2^{n+1}}\right).$$

Clearly for all $n \in \mathbb{N}$ we have I_n is an open interval of length $\frac{\epsilon}{2^n}$ with $a_n \in I_n$. Hence we obtain that

$$A \subseteq \bigcup_{n\geq 1} I_n.$$

Therefore, by the definition of the Lebesgue outer measure, we obtain that

$$0 \leq \lambda^*(A) \leq \sum_{n=1}^\infty \ell(I_n) = \sum_{n=1}^\infty \frac{\epsilon}{2^n} = \epsilon.$$

Therefore, as $\epsilon > 0$ was arbitrary, we obtain that $\lambda^*(A) = 0$ as desired. ∎

Finally, we demonstrate that the Lebesgue measure satisfies the final desired property: it is translation invariant.

Proposition 2.3.8. *If $A \in \mathcal{M}(\mathbb{R})$ and $x \in \mathbb{R}$, then $x + A \in \mathcal{M}(\mathbb{R})$ and $\lambda(x + A) = \lambda(A)$.*

Proof. Fix $A \in \mathcal{M}(\mathbb{R})$ and $x \in \mathbb{R}$. Since the translation of an open interval is an open interval of the same length, it is elementary to see that if $B \subseteq \mathbb{R}$ then

$$\lambda^*(x + B) = \lambda^*(B).$$

Thus it suffices to show that $x + A$ is measurable.

To see that $x + A$ is Lebesgue measurable, let $B \subseteq \mathbb{R}$ be arbitrary. Then

$$\begin{aligned}
\lambda^*(B) &= \lambda^*(-x + B) \\
&= \lambda^*((-x + B) \cap A) + \lambda^*((-x + B) \cap A^c) \qquad A \in \mathcal{M}(\mathbb{R}) \\
&= \lambda^*(B \cap (x + A)) + \lambda^*(B \cap (x + A^c)) \\
&= \lambda^*(B \cap (x + A)) + \lambda^*(B \cap (x + A)^c).
\end{aligned}$$

Therefore, as $B \subseteq \mathbb{R}$ was arbitrary, $x + A \in \mathcal{M}(\mathbb{R})$. ∎

In addition, the Lebesgue measure has one more important property we will use later: it is inversion invariant.

Proposition 2.3.9. *If* $A \in \mathcal{M}(\mathbb{R})$ *and* $-A = \{-a \mid a \in A\}$, *then* $-A \in \mathcal{M}(\mathbb{R})$ *and* $\lambda(-A) = \lambda(A)$.

Proof. Fix $A \in \mathcal{M}(\mathbb{R})$. Since if I is an open interval then $-I$ is an open interval is an open interval of the same length, it is elementary to see that if $B \subseteq \mathbb{R}$ then

$$\lambda^*(-B) = \lambda^*(B).$$

Thus it suffices to show that $-A$ is measurable.

To see that $-A$ is Lebesgue measurable, let $B \subseteq \mathbb{R}$ be arbitrary. Then

$$\begin{aligned} \lambda^*(B) &= \lambda^*(-B) \\ &= \lambda^*((-B) \cap A) + \lambda^*((-B) \cap A^c) \qquad A \in \mathcal{M}(\mathbb{R}) \\ &= \lambda^*(B \cap (-A)) + \lambda^*(B \cap (-A^c)) \\ &= \lambda^*(B \cap (-A)) + \lambda^*(B \cap (-A)^c). \end{aligned}$$

Therefore, as $B \subseteq \mathbb{R}$ was arbitrary, $-A \in \mathcal{M}(\mathbb{R})$. ∎

Remark 2.3.10. Note Corollary 2.3.5 shows us that $\mathfrak{B}(\mathbb{R}) \subseteq \mathcal{M}(\mathbb{R}) \subseteq \mathcal{P}(\mathbb{R})$. However, we have seen (claimed really) that $|\mathfrak{B}(\mathbb{R})| = |\mathbb{R}|$ whereas Cantor's Theorem implies that $|\mathcal{P}(\mathbb{R})| > |\mathbb{R}|$. Thus it is natural to ask, what is the cardinality of $\mathcal{M}(\mathbb{R})$? After all, if not that many subsets of $\mathbb{R}$ are Lebesgue measurable, do we really have a suitably general measure?

Recall by Example 2.3.6 that the Cantor set $\mathcal{C}$ is Lebesgue measurable with $\lambda(\mathcal{C}) = 0$. Therefore, as $|\mathcal{C}| = |\mathbb{R}|$, we obtain that $|\mathcal{P}(\mathcal{C})| = |\mathcal{P}(\mathbb{R})|$ and thus $|\mathcal{M}(\mathbb{R})| = |\mathcal{P}(\mathbb{R})|$. Thus, in terms of cardinality, the set of Lebesgue measurable subsets of $\mathbb{R}$ is as large as possible.

However $\mathcal{M}(\mathbb{R}) \neq \mathcal{P}(\mathbb{R})$. Indeed as we have exhibited that λ is a measure on $\mathcal{M}(\mathbb{R})$ that is translation invariant and such that $\lambda(I) = \ell(I)$ for all intervals I, Example 2.2.1 produces an example of a subsets of $\mathbb{R}$ that is not Lebesgue measurable.

However, there are $|\mathcal{P}(\mathbb{R})|$ subsets of $\mathbb{R}$ that are not measurable. Indeed if $A \subseteq [0, 1)$ is the non-Lebesgue measurable set from Example 2.2.1, then $A' = 2+A$ is not Lebesgue measurable. Therefore $A' \cup C$ cannot be Lebesgue measurable as $(A' \cup C) \cap C^c = A'$ (since $A' \cap C = \emptyset$) would then be Lebesgue measurable. Similarly, if $S \subseteq C$ then $A' \cup S$ is not Lebesgue measurable. Therefore, as $A' \cap C = \emptyset$ and as there are $|\mathcal{P}(\mathbb{R})|$ subsets of C, we obtain that there are $|\mathcal{P}(\mathbb{R})|$ subsets of $\mathbb{R}$ that are not measurable.

To conclude this section, we list several approximation properties for the Lebesgue measure.

Proposition 2.3.11. *Let* $A \in \mathcal{M}(\mathbb{R})$. *Then* *(a)* $\lambda(A) = \inf\{\lambda(U) \mid U \subseteq \mathbb{R}$ *is an open set such that* $A \subseteq U\}$.

(b) $\lambda(A) = \sup\{\lambda(K) \mid K \subseteq \mathbb{R}$ *is a compact set such that* $K \subseteq A\}$.

Proof. Exercise. ∎

Proposition 2.3.12. *Let* $A \subseteq \mathbb{R}$. *The following are equivalent:*

1. $A \in \mathcal{M}(\mathbb{R})$.
2. *For all* $\epsilon > 0$ *there exists an open subset* $U \subseteq \mathbb{R}$ *such that* $A \subseteq U$ *and* $\lambda^*(U \setminus A) < \epsilon$.
3. *For all* $\epsilon > 0$ *there exists a closed subset* $F \subseteq \mathbb{R}$ *such that* $F \subseteq A$ *and* $\lambda^*(A \setminus F) < \epsilon$.
4. *There exists a* G_δ *set* $G \subseteq \mathbb{R}$ *such that* $A \subseteq G$ *and* $\lambda^*(G \setminus A) = 0$.
5. *There exists an* F_σ *set* $F \subseteq \mathbb{R}$ *such that* $F \subseteq A$ *and* $\lambda^*(A \setminus F) = 0$.

Proof. Exercise. ∎

Chapter 3

Measurable Functions

In mathematics, the natural progression of study of constructs is to first define and student the properties of objects, and then study the morphisms (i.e. maps) between them. For example, in analysis and topology, first we define the notion of a topological space and study the properties such as characterizations of open/closed sets. The appropriate morphisms between topological spaces are then the continuous functions as they preserve the notion of convergence.

With the construction of the Lebesgue measure, we desire to analyse and use functions that preserve measurability of sets. These so called measurable functions will be the focus of this chapter. After developing the basic properties of real and complex valued measurable functions, we will demonstrate that every measurable function can be 'approximated' by 'simple' functions. We will also demonstrate that convergence of measurable functions occurs 'uniformly almost everywhere' and that measurable functions on closed intervals are 'almost everywhere continuous'. The theory of measurable function is vital for a theory of integration as we will see in Chapter 4.

3.1 Measurable Functions

To begin, we define the notion of a measurable function. Note the flavour of this definition is very similar to the definition of continuous functions between topological spaces where it is said that a function is continuous if the inverse image of every open set is open.

Definition 3.1.1. Let $(X, \mathcal{A}_X)$ and $(Y, \mathcal{A}_Y)$ be measurable spaces. A function $f : X \to Y$ is said to be *measurable* if $f^{-1}(A) \in \mathcal{A}_X$ for all $A \in \mathcal{A}_Y$; that is, the inverse image of every measurable set in Y is measurable in X.

Of course, we have a collection of trivial examples.

Example 3.1.2. Let $(X, \mathcal{A}_X)$ and $(Y, \mathcal{A}_Y)$ be measurable spaces and let

$f : X \to Y$. If f is constant, then f is measurable as either $f^{-1}(A) = X$ or $f^{-1}(A) = \emptyset$ for all $A \in \mathcal{A}_Y$.

Alternatively, if $\mathcal{A}_X = \mathcal{P}(X)$, then f is automatically measurable.

Similarly, if $\mathcal{A}_Y = \{\emptyset, Y\}$, then f is automatically measurable as $f^{-1}(\emptyset) = \emptyset$ and $f^{-1}(Y) = X$.

For a more robust collection of examples, we look at the following.

Definition 3.1.3. Let X be a non-empty set and let $A \subseteq X$. The *characteristic function of* A (or *indicator function*) is the function $\chi_A : X \to \mathbb{R}$ defined by

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

for all $x \in X$.

In the sense of probability theory, the characteristic function of an event takes on the value one at a point where the event occurs and zero otherwise. Of course, for a characteristic function to make sense in probability, we would want the event to be in our probability space; that is, we would want the set to be measurable.

Example 3.1.4. Let $(X, \mathcal{A})$ be a measurable space and let $A \subseteq X$. The characteristic function χ_A is measurable as a function to $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$ if and only if $A \in \mathcal{A}$. Indeed, notice for all $B \subseteq \mathbb{R}$ that

$$\chi_A^{-1}(B) = \begin{cases} \emptyset & \text{if } 0, 1 \notin B \\ A & \text{if } 0 \notin B \text{ and } 1 \in B \\ A^c & \text{if } 1 \notin B \text{ and } 0 \in B \\ X & \text{if } 0, 1 \in B \end{cases}.$$

From this and the fact that all cases are possibly by choosing select $B \in \mathfrak{B}(\mathbb{R})$, clearly χ_A is measurable if and only if $A, A^c \in \mathcal{A}$ if and only if $A \in \mathcal{A}$.

Of course, we will mainly be interested in functions from a measure space into either the real or complex numbers. As such, we will use $\mathbb{K}$ to denote either the real or complex numbers.

However we have a notion of a measurable function for each σ-algebra on $\mathbb{K}$. One might think we could use the σ-algebra $\{\emptyset, \mathbb{K}\}$ to force every function to be measurable. However we will see later that this notion of measurability does not preserve enough topology to guarantee certain desirable facts about measurable functions (such as the pointwise limit of measurable functions is measurable). Since we desire any continuous function to be measurable, the σ-algebra on $\mathbb{K}$ should at least contain every open set, and thus must contain the Borel σ-algebra. Thus we define the following notion.

Definition 3.1.5. Let $(X, \mathcal{A})$ be a measurable space. A function $f : X \to \mathbb{K}$ is said to be *measurable* if f is measurable as a function from $(X, \mathcal{A})$ to $(\mathbb{K}, \mathfrak{B}(\mathbb{K}))$. The set of all measurable functions from $(X, \mathcal{A})$ to $(\mathbb{K}, \mathfrak{B}(\mathbb{K}))$ is denoted $\mathcal{M}(X, \mathbb{K})$.

Of course, we do not have much precise information about the Borel σ-algebra in the sense that we do not have an easy method for testing whether a set is Borel. In particular, how can we determine whether $f^{-1}(A) \in \mathcal{A}$ for all $A \in \mathfrak{B}(\mathbb{K})$ if we cannot describe the elements of $\mathfrak{B}(\mathbb{K})$? However, we do know several sets which generate $\mathfrak{B}(\mathbb{K})$. Hence the following result will easily enable us to check whether a function is in $\mathcal{M}(X, \mathbb{K})$.

Proposition 3.1.6. *Let $(X, \mathcal{A}_X)$ and $(Y, \mathcal{A}_Y)$ be measurable spaces and let $f : X \to Y$. If $A \subseteq \mathcal{A}_Y$ and $\mathcal{A}_Y = \sigma(A)$ (that is, $\mathcal{A}_Y$ is the smallest σ-algebra containing A), then f is measurable if and only if*

$$\{f^{-1}(B) \mid B \in A\} \subseteq \mathcal{A}_X.$$

Proof. If f is measurable, then clearly $\{f^{-1}(B) \mid B \in A\} \subseteq \mathcal{A}_X$ by definition.

Conversely, suppose $\{f^{-1}(B) \mid B \in A\} \subseteq \mathcal{A}_X$. To see that f is measurable, consider the set

$$\mathcal{A} = \{B \subseteq Y \mid f^{-1}(B) \in \mathcal{A}_X\}.$$

We claim that $\mathcal{A}$ is a σ-algebra. To see this, we notice that $f^{-1}(\emptyset) = \emptyset \in \mathcal{A}_X$ and $f^{-1}(Y) = X \in \mathcal{A}_X$ so clearly $\emptyset, X \in \mathcal{A}$. Next, if $B \subseteq Y$ then $f^{-1}(B) \in \mathcal{A}_X$ so $f^{-1}(B^c) = (f^{-1}(B))^c \in \mathcal{A}_X$ so $B^c \in \mathcal{A}$. Finally, let $\{B_n\}_{n=1}^{\infty} \subseteq \mathcal{A}$ be arbitrary. Hence $\{f^{-1}(B_n)\}_{n=1}^{\infty} \subseteq \mathcal{A}_X$. Since

$$f^{-1}\left(\bigcup_{n=1}^{\infty} B_n\right) = \bigcup_{n=1}^{\infty} f^{-1}(B_n) \in \mathcal{A}_X$$

we see that $\bigcup_{n=1}^{\infty} B_n \in \mathcal{A}$. Hence, as $\{B_n\}_{n=1}^{\infty}$ was arbitrary, $\mathcal{A}$ is a σ-algebra.

Since $A \subseteq \mathcal{A}$ by assumption and since $\mathcal{A}_Y = \sigma(A)$, we obtain that $\mathcal{A}_Y \subseteq \mathcal{A}$. Hence f is measurable by definition. ∎

Corollary 3.1.7. *Let $(X, \mathcal{A})$ be a measurable space, let Y be a metric space, and let $f : X \to Y$. Then f is measurable as a function from $(X, \mathcal{A})$ to $(Y, \mathcal{B}(Y))$ if and only if $f^{-1}(U) \in \mathcal{A}$ for all open subsets $U \subseteq Y$; that is, a function to a metric space equipped with the Borel σ-algebra is measurable if and only if the inverse image of every open set is measurable.*

As the inverse image of an open set under a continuous function is open, Corollary 3.1.7 trivially implies the following.

Corollary 3.1.8. *Let $(\mathcal{X}, d)$ be a metric space and let $\mathcal{A}$ be a σ-algebra of X containing $\mathfrak{B}(\mathcal{X})$. If $f : \mathcal{X} \to \mathbb{K}$ is continuous, then f is measurable.*

Corollary 3.1.9. *Let $(X, \mathcal{A})$ be a measurable space and let $f : X \to \mathbb{R}$. The following are equivalent:*

1. $f \in \mathcal{M}(X, \mathbb{R})$.
2. $\{x \in X \mid f(x) > a\} \in \mathcal{A}$ *for all* $a \in \mathbb{R}$.
3. $\{x \in X \mid f(x) \geq a\} \in \mathcal{A}$ *for all* $a \in \mathbb{R}$.
4. $\{x \in X \mid f(x) < a\} \in \mathcal{A}$ *for all* $a \in \mathbb{R}$.
5. $\{x \in X \mid f(x) \leq a\} \in \mathcal{A}$ *for all* $a \in \mathbb{R}$.
6. $\{x \in X \mid a < f(x) < b\} \in \mathcal{A}$ *for all* $a, b \in \mathbb{R}$.

Proof. The result follows easily from Proposition 3.1.6 as Remark 2.1.8 implies each of the sets used in the inverse images generate $\mathfrak{B}(\mathbb{R})$. ∎

Now that we have a good notion of real- and complex-valued measurable functions on a measurable space, it is useful to see which operations preserve measurability. To do so, we note the following important result.

Proposition 3.1.10. *Let $(X, \mathcal{A}_X)$ be a measurable space and let $(\mathcal{Y}, d_{\mathcal{Y}})$, and $(\mathcal{Z}, d_{\mathcal{Z}})$ be metric spaces. If $\mathcal{Y}$ and $\mathcal{Z}$ are equipped with their respective Borel σ-algebras, if $f : \mathcal{X} \to \mathcal{Y}$ is measurable, and if $g : \mathcal{Y} \to \mathcal{Z}$ is continuous, then $g \circ f : \mathcal{X} \to \mathcal{Z}$ is measurable.*

Proof. Let $U \subseteq \mathcal{Z}$ be an arbitrary open set. Since g is continuous, $g^{-1}(U) \subseteq \mathcal{Y}$ is open and thus measurable in $\mathcal{Y}$ as $\mathcal{Y}$ was equipped with the Borel σ-algebra. Hence, as f is measurable,

$$(g \circ f)^{-1}(U) = f^{-1}(g^{-1}(U)) \in \mathcal{A}_X.$$

Therefore, as $U \subseteq \mathcal{Z}$ was an arbitrary open set and as open sets generated the Borel σ-algebra on $\mathcal{Z}$, the result follows from Proposition 3.1.6. ∎

Using Corollary 3.1.7 and Proposition 3.1.10, we can easily use continuous operations to show specific operations on measurable functions preserve measurability. To get all of the operations we want, sometimes we need to double up.

Proposition 3.1.11. *Let $(X, \mathcal{A})$ be a measurable space and let $f, g \in \mathcal{M}(X, \mathbb{K})$. If $h : X \to \mathbb{K}^2$ is defined by*

$$h(x) = (f(x), g(x))$$

for all $x \in X$, then h is a measurable function from $(X, \mathcal{A})$ to $(\mathbb{K}^2, \mathfrak{B}(\mathbb{K}^2))$.

Proof. Let $f, g \in \mathcal{M}(X, \mathbb{K})$. Note that $\mathfrak{B}(\mathbb{K}^2)$ is generated by open balls with respect to the infinity norm (i.e. check that if U is open in $\mathbb{K}^2$ then U is the union of all balls of the form $B((z_1, z_2), r_{(z_1,z_2)})$ where z_1 and z_2 are rational (or complex rational) numbers and $r_{(z_1,z_2)}$ is the largest radius r such that $B((z_1, z_2), r) \subseteq U$). However each open ball in the infinity norm is of the form $I_1 \times I_2$ where $I_1, I_2 \subseteq \mathbb{K}$ are open sets with respect to $|\cdot|$. Hence if $I_1, I_2 \subseteq \mathbb{K}$ are open, then $f^{-1}(I_1), g^{-1}(I_2) \in \mathcal{A}$ as $f, g \in \mathcal{M}(X, \mathbb{K})$ and thus

$$h^{-1}(I_1 \times I_2) = f^{-1}(I_1) \cap g^{-1}(I_2) \in \mathcal{A}.$$

Therefore Proposition 3.1.6 implies that h is measurable. ∎

Corollary 3.1.12. *Let $(X, \mathcal{A})$ be a measurable space and let $f, g \in \mathcal{M}(X, \mathbb{K})$. Then*

1. $cf \in \mathcal{M}(X, \mathbb{K})$ *for all* $c \in \mathbb{K}$.
2. $f + g \in \mathcal{M}(X, \mathbb{K})$.
3. $fg \in \mathcal{M}(X, \mathbb{K})$.
4. $|f| \in \mathcal{M}(X, \mathbb{K})$.
5. $\frac{1}{f} \in \mathcal{M}(X, \mathbb{K})$ *if* $f(x) \neq 0$ *for all* $x \in X$.
6. $\overline{f} \in \mathcal{M}(X, \mathbb{K})$ *where* $\overline{f}(z) = \overline{f(z)}$.

Proof. As constant functions are measurable, (1) will follow from (3).

Let $f, g \in \mathcal{M}(X, \mathbb{K})$. By Proposition 3.1.11, the function $h : X \to \mathbb{K}^2$ defined by $h(x) = (f(x), g(x))$ is measurable. Since the functions $h_+, h_\times : \mathbb{K}^2 \to \mathbb{K}$ defined by $h_+(x, y) = x + y$ and $h_\times(x, y) = xy$ are continuous functions, Propositions 3.1.10 implies that $f + g = h_+ \circ h$ and $fg = h_\times \circ h$ are measurable. Hence (2) and (3) follow.

Since the functions $a, C : \mathbb{K} \to \mathbb{R}$ defined by $a(z) = |z|$ and $C(z) = \overline{z}$ are continuous, Proposition 3.1.10 implies that $|f| = a \circ f$ and $\overline{f} = C \circ f$ are measurable. Hence (4) and (6) follow.

Finally define the function $q : \mathbb{K} \setminus \{0\} \to \mathbb{K} \setminus \{0\}$ by $q(z) = \frac{1}{z}$. Clearly q is continuous with respect to the metric on $\mathbb{K} \setminus \{0\}$ induced by $|\cdot|$. Since $\frac{1}{f} = q \circ f$ is well-defined as $f(x) \neq 0$ for all $x \in X$, Proposition 3.1.10 implies that $\frac{1}{f}$ is measurable. Hence (5) follows. ∎

Remark 3.1.13. Using Corollary 3.1.12 we may reduce the study of complex-valued measurable functions to real-valued measurable functions. Indeed let $(X, \mathcal{A})$ be a measurable space and let $f : X \to \mathbb{C}$. Define $\mathrm{Re}(f), \mathrm{Im}(f) : X \to \mathbb{R}$ by

$$\mathrm{Re}(f)(x) = \frac{1}{2}\left(f(x) + \overline{f(x)}\right) \qquad \text{and} \qquad \mathrm{Im}(f)(x) = \frac{1}{2i}\left(f(x) - \overline{f(x)}\right)$$

for all $x \in X$. Hence $f(x) = \mathrm{Re}(f)(x) + i\mathrm{Im}(f)(x)$ for all $x \in X$. Hence f is measurable if and only if $\mathrm{Re}(f)$ and $\mathrm{Im}(f)$ are measurable by Corollary 3.1.12. The functions $\mathrm{Re}(f)$ and $\mathrm{Im}(f)$ are called the *real and imaginary parts of* f respectively.

Remark 3.1.14. In fact, the theory of measurable functions can be reduced to non-negative measurable functions. Indeed let $(X, \mathcal{A})$ be a measurable space and let $f : X \to \mathbb{R}$. Define $f_+, f_- : X \to [0, \infty)$ by

$$f_+(x) = \frac{1}{2}(|f(x)| + f(x)) = \begin{cases} f(x) & \text{if } f(x) \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

and

$$f_-(x) = \frac{1}{2}(|f(x)| - f(x)) = \begin{cases} -f(x) & \text{if } f(x) \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

for all $x \in X$. Hence $|f|(x) = f_+(x) + f_-(x)$ and $f(x) = f_+(x) - f_-(x)$ for all $x \in X$. Hence f is measurable if and only if f_+ and f_- are measurable by Corollary 3.1.12. The functions f_+ and f_- are called the *positive and negative parts of* f respectively.

Of course, when dealing with limits of functions, often the collection of functions diverges at specific points. Consequently, it is useful to extend the notion of measurable functions to allow infinite values.

Definition 3.1.15. Let $(X, \mathcal{A})$ be a measurable space. An extended real-valued function $f : X \to [-\infty, \infty]$ is said to be *measurable* if

$$f^{-1}(\{-\infty\}), f^{-1}(\{\infty\}) \in \mathcal{A}$$

and $f^{-1}(A) \in \mathcal{A}$ for all $A \in \mathfrak{B}(\mathbb{R})$.

Remark 3.1.16. It is not difficult to see that the characterization of measurable real-valued functions from Corollary 3.1.9 extends to extended real-valued functions. Indeed the second characterization of Corollary 3.1.9 will extend since

$$f^{-1}(\{\infty\}) = \bigcap_{n=1}^{\infty} f^{-1}((n, \infty]) \quad \text{and} \quad f^{-1}(\{-\infty\}) = \left(\bigcup_{n=1}^{\infty} f^{-1}((-n, \infty])\right)^c.$$

One reason for dealing with extended real-valued functions is it enables us to take supremums and infimums of functions without worrying about pointwise boundedness. Thus we arrive at the following result, which would be impossible to prove if we used $\mathcal{P}(\mathbb{R})$ as the σ-algebra on $\mathbb{R}$ when we defined measurable functions. In particular, this is the first instance where the Borel σ-algebra is essential as the proof will demonstrate.

Proposition 3.1.17. *Let $(X, \mathcal{A})$ be a measurable space. For each $n \in \mathbb{N}$, let $f_n : X \to [-\infty, \infty]$ be a measurable function. Then the functions*

$$\sup_{n\geq 1} f_n, \quad \inf_{n\geq 1} f_n, \quad \limsup_{n\to\infty} f_n, \quad \text{and} \quad \liminf_{n\to\infty} f_n$$

are measurable (where by sup, inf, lim sup, *and* lim inf *of functions, we mean the functions that are defined pointwise by taking the respective operation applied to the sequence of functions pointwise). Consequently, if $f : X \to [-\infty, \infty]$ is such that $f(x) = \lim_{n\to\infty} f_n(x)$ (that is, f_n converge to f pointwise), then f is measurable.*

Proof. For each $n \in \mathbb{N}$, let $f_n : X \to [-\infty, \infty]$ be a measurable function. To see that $\sup_{n\geq 1} f_n$ is measurable, notice for all $a \in \mathbb{R}$ that

$$\left(\sup_{n\geq 1} f_n\right)^{-1}((a,\infty]) = \bigcup_{n=1}^{\infty} f_n^{-1}((a,\infty]) \in \mathcal{A}.$$

Hence $\sup_{n\geq 1} f_n$ is measurable by Corollary 3.1.9 (this is where we require the Borel σ-algebra on $\mathbb{R}$ for otherwise how would we determine the behaviour of the inverse image of the supremum). Similarly, to see that $\inf_{n\geq 1} f_n$ is measurable, notice for all $a \in \mathbb{R}$ that

$$\left(\inf_{n\geq 1} f_n\right)^{-1}([a,\infty]) = \bigcap_{n=1}^{\infty} f_n^{-1}([a,\infty]) \in \mathcal{A}.$$

Hence $\inf_{n\geq 1} f_n$ is measurable by Corollary 3.1.9.

Next, for each $k \in \mathbb{N}$ let

$$g_k = \sup_{n\geq k} f_n \qquad \text{and} \qquad h_k = \inf_{n\geq k} f_n.$$

Since each g_k and h_k is measurable from above and since

$$\limsup_{n\to\infty} f_n = \inf_{k\geq 1} g_k \qquad \text{and} \qquad \liminf_{n\to\infty} f_n = \sup_{k\geq 1} h_k,$$

we obtain that $\limsup_{n\to\infty} f_n$ and $\liminf_{n\to\infty} f_n$ are measurable from above.

Finally, if $f : X \to [-\infty, \infty]$ is such that $f(x) = \lim_{n\to\infty} f_n(x)$ then $f = \limsup_{n\to\infty} f_n$ so f is measurable. ∎

Corollary 3.1.18. *Let $(X, \mathcal{A})$ be a measurable space. For each $n \in \mathbb{N}$, let $f_n : X \to \mathbb{C}$ be a measurable function. If $f : X \to \mathbb{C}$ is such that $f(x) = \lim_{n\to\infty} f_n(x)$ (that is, f_n converge to f pointwise), then f is measurable.*

Proof. Clearly for each $x \in X$ we have $f(x) = \lim_{n\to\infty} f_n(x)$ if and only if

$$\mathrm{Re}(f)(x) = \lim_{n\to\infty} \mathrm{Re}(f_n)(x) \qquad \text{and} \qquad \mathrm{Im}(f)(x) = \lim_{n\to\infty} \mathrm{Im}(f_n)(x).$$

Since $\mathrm{Re}(f_n)$ and $\mathrm{Im}(f_n)$ are measurable by Remark 3.1.13, we obtain that $\mathrm{Re}(f)$ and $\mathrm{Im}(f)$ are measurable by Proposition 3.1.17. Hence f is measurable by Remark 3.1.13. ∎

Of course, asking for pointwise convergence at every point in X is a lot to ask. However, we are dealing with measures which determine the size of a set. As sets with zero measure have 'no mass', it is natural to ask whether we can have pointwise convergence except on a set of zero measure and still have measurability? This leads us to the following notion.

Definition 3.1.19. Let $(X, \mathcal{A}_X, \mu)$ be a measure space and let P be a property that at each point in X is either true or false. It is said that P holds *μ-almost everywhere* (abbreviated μ-a.e. or simply a.e. if μ is clear) if there exists a set $A \subseteq \mathcal{A}_X$ such that $P(x)$ is true for all $x \in A$ and $\mu(A^c) = 0$.

Remark 3.1.20. For example, given a measure space $(X, \mathcal{A}_X, \mu)$, two functions $f, g : X \to \mathbb{K}$ are equal almost everywhere if there exists a set $A \subseteq \mathcal{A}_X$ such that $f(x) = g(x)$ for all $x \in A$ and $\mu(A^c) = 0$. Note this is not necessarily the same as saying

$$\mu(\{x \in X \mid f(x) \neq g(x)\}) = 0$$

since we do not know whether this set is measurable. However, if we know f and g are measurable, then $f - g$ is measurable so the set

$$\{x \in X \mid f(x) \neq g(x)\} = \{x \in X \mid (f - g)(x) \neq 0\}$$

is indeed measurable. Thus $f = g$ almost everywhere is equivalent to $\mu(\{x \in X \mid f(x) \neq g(x)\}) = 0$ when f and g are measurable.

Example 3.1.21. It is elementary to see that $\chi_{\mathbb{Q}} = 0$ almost everywhere with respect to the Lebesgue measure. Similarly, if A is any measurable set with zero μ-measure, then $\chi_A = 0$ μ-almost everywhere.

It is not difficult to see that measurable functions behave well if properties only hold almost everywhere.

Proposition 3.1.22. *Let $(X, \mathcal{A}_X, \mu)$ be a complete measure space, let $(Y, \mathcal{A}_Y)$ be a measurable space, and let $f, g : X \to Y$ be such that $f = g$ μ-almost everywhere. If f is measurable, then g is measurable.*

Proof. Let $f, g : X \to Y$ be such that f is measurable and $f = g$ almost everywhere. Hence there exists a set $A \subseteq \mathcal{A}_X$ such that $f(x) = g(x)$ for all $x \in A$ and $\mu(A^c) = 0$. Let $B \in \mathcal{A}_Y$ be arbitrary. Notice

$$g^{-1}(B) = \left(A \cap g^{-1}(B)\right) \cup \left(A^c \cap g^{-1}(B)\right) = \left(A \cap f^{-1}(B)\right) \cup \left(A^c \cap g^{-1}(B)\right)$$

as $f(x) = g(x)$ for all $x \in A$. Since $A^c \cap g^{-1}(B) \subseteq A^c$, since $A^c \in \mathcal{A}_X$ as $A \in \mathcal{A}_X$, since $\mu(A^c) = 0$, and since $(X, \mathcal{A}_X, \mu)$ is complete, we obtain that $A^c \cap g^{-1}(B) \in \mathcal{A}_X$ by definition. Furthermore, since f is measurable, $f^{-1}(B) \in \mathcal{A}_X$. Hence, we obtain that $A \cap f^{-1}(B) \in \mathcal{A}_X$. Hence $g^{-1}(B) \in \mathcal{A}_X$. Therefore, as $B \in \mathcal{A}_Y$ was arbitrary, g is measurable. ∎

The following illustrates our first use of how we can correct functions on measure zero sets.

Corollary 3.1.23. *Let $(X, \mathcal{A}_X, \mu)$ be a complete measure space. For each $n \in \mathbb{N}$, let $f_n : X \to \mathbb{K}$ be a measurable function. If $f : X \to \mathbb{K}$ is such that $f(x) = \lim_{n\to\infty} f_n(x)$ a.e. (that is, f_n converge to f pointwise except on a set of measure zero), then f is measurable.*

Proof. Since $f(x) = \lim_{n\to\infty} f_n(x)$ for a.e. $x \in X$, there exists a set $A \in \mathcal{A}$ such that $f(x) = \lim_{n\to\infty} f_n(x)$ for all $x \in A$ and $\mu(A^c) = 0$. Consider the sequence of functions $(f_n\chi_A)_{n\geq 1}$. Clearly $f_n\chi_A$ is measurable for all $n \in \mathbb{N}$ by Corollary 3.1.12 since f_n is measurable and χ_A is measurable as $A \in \mathcal{A}$. Therefore, since $f(x)\chi_A(x) = \lim_{n\to\infty} f_n(x)\chi_n(x)$ for all $x \in X$, $f\chi_A$ is measurable by Corollary 3.1.18. Therefore, since $\mu(A^c) = 0$ and $f(x)\chi_A(x) = f(x)$ for all $x \in A$, we see that $f = f\chi_A$ almost everywhere. Hence Proposition 3.1.22 implies that f is measurable. ∎

3.2 Simple Functions

As we desire to study measurable functions beyond the properties developed above and measurable functions may appear on the surface to be difficult to describe, it is useful to have a 'simple' collection of measurable functions that are easy to understand yet well-approximate all measurable functions. We find such a collection in the following definition.

Definition 3.2.1. Let $(X, \mathcal{A})$ be a measurable space. A function $\varphi : X \to [0, \infty)$ is said to be *simple* if there exists an $n \in \mathbb{N}$, non-empty, pairwise disjoint sets $\{A_k\}_{k=1}^n \subseteq \mathcal{A}$ such that $X = \bigcup_{k=1}^n A_k$, and $\{a_k\}_{k=1}^n \subseteq [0, \infty)$ distinct (i.e. $a_i \neq a_j$ whenever $i \neq j$) such that

$$\varphi = \sum_{k=1}^{n} a_k \chi_{A_k}.$$

Note clearly all simple functions are measurable by Example 3.1.4 and Corollary 3.1.12. In particular, students have already encountered specific types of simple functions in previous courses.

Example 3.2.2. Recall that $\varphi : [a, b] \to [0, \infty)$ is said to be a *step function* if $\varphi = \sum_{k=1}^n a_k\chi_{A_k}$ where $\{A_k\}_{k=1}^\infty$ are disjoint intervals whose union is $[a, b]$.

Remark 3.2.3. Suppose $(X, \mathcal{A})$ is a measurable space and $\varphi : X \to [0, \infty)$ is measurable with finite range. We claim that φ is a simple function. Indeed write $\varphi(X) = \{b_1, \ldots, b_m\}$. As φ is measurable, $A_k = \varphi^{-1}(\{b_k\}) \in \mathcal{A}$ for all $k \in \{1, \ldots, m\}$. It is then easy to see that $\varphi = \sum_{k=1}^m b_k\chi_{A_k}$ and $\{A_k\}_{k=1}^n \subseteq \mathcal{A}$ pairwise disjoint non-empty with $X = \bigcup_{k=1}^n A_k$.

Consequently, if $g : X \to [0, \infty]$ is such that $g = \sum_{k=1}^n a_k \chi_{A_k}$ where $\{A_k\}_{k=1}^n \subseteq \mathcal{A}$ and $\{a_k\}_{k=1}^n \subseteq [0, \infty)$, then g is a simple function. Consequently, the representation of a simple function given in Definition 3.2.1 is called the *canonical representation of a simple function*.

The reason for analyzing simple functions and why simple functions are so essential to this course is the following result. This result will most often be used to conclude a result for all measurable functions provided one can verify the result for simple function and take limits.

Theorem 3.2.4. *Let $(X, \mathcal{A})$ be a measurable space and let $f : X \to [0, \infty]$. Then f is measurable if and only if there exists a sequence $(\varphi_n)_{n\geq 1}$ of simple functions on X such that $\varphi_n \leq \varphi_{n+1}$ for all $n \in \mathbb{N}$ and $(\varphi_n)_{n\geq 1}$ converges to f pointwise (that is, $\lim_{n\to\infty} \varphi_n(x) = f(x)$ for all $x \in X$).*

Proof. Suppose there exists a sequence $(\varphi_n)_{n\geq 1}$ of simple functions such that $\varphi_n \leq \varphi_{n+1}$ for all $n \in \mathbb{N}$ and $(\varphi_n)_{n\geq 1}$ converges to f pointwise. Since each simple function is measurable, we obtain that f is measurable by Proposition 3.1.17.

Conversely suppose that f is measurable. For each $n \in \mathbb{N}$ and for each $k \in \{1, \ldots, n2^n\}$, consider the sets

$$A_{n,k} = f^{-1}\left(\left[\frac{k-1}{2^n}, \frac{k}{2^n}\right)\right) \qquad \text{and} \qquad B_n = \left(\bigcup_{k=1}^{n2^n} A_{n,k}\right)^c.$$

Clearly B_n and each $A_{n,k}$ is measurable as f is a measurable function, and $\{A_{n,k}\}_{k=1}^{n2^n}$ are pairwise disjoint. Furthermore, $x \in B_n$ if and only if $x \notin A_{n,k}$ for all $k \in \{1, \ldots, n2^n\}$ if and only if $f(x) \notin \left[\frac{k-1}{2^n}, \frac{k}{2^n}\right)$ for all $k \in \{1, \ldots, n2^n\}$ if and only if $f(x) \geq n$.

For each $n \in \mathbb{N}$ the function $\varphi_n : X \to [0, \infty)$ defined by

$$\varphi_n = n\chi_{B_n} + \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \chi_{A_{n,k}}$$

is a simple function. Clearly $\varphi_n \leq \varphi_{n+1}$ for all $n \in \mathbb{N}$ due to the refining nature of the construction (i.e. $A_{n,k}$ is refined into two $A_{n+1,k'}$ each of which has the property that $\frac{k'-1}{2^{n+1}} \geq \frac{k-1}{2^n}$ and part of B_n becomes 2^{n+1} $A_{n+1,k'}$ each of which has the property that $\frac{k'-1}{2^{n+1}} \geq n$).

To see that $(\varphi_n)_{n\geq 1}$ converges to f pointwise, fix $x \in X$. If $f(x) < \infty$ then for all $n \in \mathbb{N}$ such that $f(x) < n$ we see that $|f(x) - \varphi_n(x)| \leq \frac{1}{2^n}$ for all as $f(x) < n$ implies $x \in A_{n,k}$ for some k. Hence $\lim_{n\to\infty} \varphi_n(x) = f(x)$ when $f(x) < \infty$. Otherwise, if $f(x) = \infty$ then $\varphi_n(x) = n$ for all $n \in \mathbb{N}$ so $\lim_{n\to\infty} \varphi_n(x) = \infty = f(x)$. Hence the result follows. ∎

3.3 Egoroff's Theorem

Theorem 3.2.4 will be essential to us because, upto an increasing pointwise limit, every measurable function is almost a simple function. In this and the subsequence two sections, we will look at the three Littlewood principles which give us more control over the behaviour of measurable sets and functions. The following Littlewood principle (which is actually the third of Littlewood's principles) enables us to control the behaviour of convergence of measurable functions. In particular, up to a set of small measure, we have pointwise convergence of measurable functions implies uniform convergence of measurable functions.

Theorem 3.3.1 (Egoroff's Theorem). *Let $a, b \in \mathbb{R}$ be such that $a < b$. For each $n \in \mathbb{N}$ let $f_n : [a, b] \to \mathbb{C}$ be a measurable function. If $f : [a, b] \to \mathbb{C}$ is a measurable function such that $f(x) = \lim_{n\to\infty} f_n(x)$ for all $x \in [a, b]$, then for all $\delta > 0$ there exists a Lebesgue measurable set $B \subseteq [a, b]$ such that $\lambda(B) < \delta$ and $f = \lim_{n\to\infty} f_n$ uniformly on $[a, b] \setminus B$.*

Proof. Fix $\delta > 0$. For each $m, k \in \mathbb{N}$ let

$$B_{m,k} = \bigcup_{n=m}^{\infty} \left\{ x \in [a, b] \,\middle|\, |f_n(x) - f(x)| \geq \frac{1}{k} \right\}.$$

Therefore, as f and $(f_n)_{n\geq1}$ are measurable functions, we see that $B_{m,k}$ is measurable for all $m, k \in \mathbb{N}$. Notice that $B_{m+1,k} \subseteq B_{m,k}$ for all $m \in \mathbb{N}$ and, as $f(x) = \lim_{n\to\infty} f_n(x)$ for all $x \in [a, b]$, we see that

$$\bigcap_{m=1}^{\infty} B_{m,k} = \emptyset$$

for all $k \in \mathbb{N}$. Therefore, as $\lambda(\emptyset) = 0$ and $\lambda([a, b]) < \infty$, the Monotone Convergence Theorem (Theorem 2.1.18) implies that

$$\lim_{m\to\infty} \lambda(B_{m,k}) = 0$$

for all $k \in \mathbb{K}$. Hence for each $k \in \mathbb{K}$, there exists an $n_k \in \mathbb{N}$ such that $\lambda(B_{n_k,k}) < \frac{\delta}{2^k}$.

Let $B = \bigcup_{k=1}^{\infty} B_{n_k,k}$. Clearly B is measurable being the countable union of measurable sets. Furthermore, clearly

$$\lambda(B) \leq \sum_{k=1}^{\infty} \lambda(B_{n_k,k}) \leq \sum_{k=1}^{\infty} \frac{\delta}{2^k} = \delta.$$

Hence, to complete the proof, it suffices to show that $(f_n)_{n\geq1}$ converges uniformly to f on $[a, b] \setminus B$.

To see that $(f_n)_{n\geq1}$ converges uniformly to f on $[a, b] \setminus B$, let $\epsilon > 0$ be arbitrary. Let $k \in \mathbb{N}$ be such that $\frac{1}{k} < \epsilon$. Notice that if $x \in [a, b] \setminus B$ then

$x \notin B$ so $x \notin B_{n_k,k}$. Hence for all $x \in [a,b] \setminus B$ and for all $n \geq n_k$ we have that

$$|f_n(x) - f(x)| < \frac{1}{k} < \epsilon.$$

Therefore, as $\epsilon > 0$ was arbitrary, we obtain that $(f_n)_{n\geq 1}$ converges uniformly to f on $[a,b] \setminus B$ as desired. ∎

Remark 3.3.2. If in the statement of Egoroff's Theorem (Theorem 3.3.1) one only knew that $f(x) = \lim_{n\to\infty} f_n(x)$ almost everywhere, then the conclusions still hold. Indeed, suppose $\delta > 0$ and $f(x) = \lim_{n\to\infty} f_n(x)$ almost everywhere. Then there exists a Lebesgue measurable set A such that $\lambda(A^c) = 0$ and $f(x) = \lim_{n\to\infty} f_n(x)$ for all $x \in A$. Hence the sequence $(\chi_A f_n)_{n\geq 1}$ is a sequence of measurable functions that converges pointwise to the measurable function $\chi_A f$. By Egoroff's Theorem (Theorem 3.3.1) as stated, there exists a Lebesgue measurable set B such that $\lambda(B) < \delta$ and $f\chi_A = \lim_{n\to\infty} f_n\chi_A$ uniformly on $[a,b] \setminus B$. Hence, if $C = B \cup A^c$, then C is Lebesgue measurable, $\lambda(C) < \delta$, and $f = \lim_{n\to\infty} f_n$ uniformly on C as desired.

Example 3.3.3. The conclusions of Egoroff's Theorem (Theorem 3.3.1) fail if we do not restrict to a finite interval. Indeed consider the functions $f_n = \chi_{[n,\infty)}$. Clearly $(f_n)_{n\geq 1}$ converges pointwise to the constant function 0. However there does not exists a Lebesgue measurable set $B \subseteq \mathbb{R}$ such that $(f_n)_{n\geq 1}$ converges uniformly to 0 on B^c and $\lambda(B)$ is finite. To see this, suppose $(f_n)_{n\geq 1}$ converged uniformly to 0 on B^c for some Lebesgue measurable set B. Thus if $\epsilon = 1$ there exists an $N \in \mathbb{N}$ such that

$$|f_n(x)| < \epsilon = 1$$

for all $n \geq N$ and for all $x \in B^c$. Due to the description of f_n, the above implies $B^c \subseteq (-\infty, N)$ as $f_n(x) = 1$ when $x \geq n$. Therefore $[N, \infty) \subseteq B$ so $\lambda(B) = \infty$.

3.4 Littlewood's First Principle

Our next goal in this course is to proof Lusin's Theorem (Theorem 3.5.1), which is also know as Littlewood's second principle. One proof of Lusin's Theorem can be constructed using Littlewood's first principle. However, we will present a different proof of Lusin's Theorem that is shorter and bypasses Littlewood's first principle. Thus, for completeness and to introduce concepts required for the proof of Lusin's Theorem, we will prove Littlewood's first principle first. Consequently, we begin with the following notions.

Theorem 3.4.1 (Littlewood's First Principle). *Let $A \subseteq \mathbb{R}$ be a Lebesgue measurable set such that $\lambda(A) < \infty$. Then for all $\epsilon > 0$ there exists a finite* *number of open intervals $I_1, \ldots, I_n$ such that if $U = \bigcup_{k=1}^n I_k$ then*

$$\lambda((A \setminus U) \cup (U \setminus A)) < \epsilon.$$

Proof. Let $\epsilon > 0$. By Proposition 2.3.11 there exists an open set V such that $A \subseteq V$ and

$$\lambda(V) < \lambda(A) + \frac{\epsilon}{2}.$$

Since $\lambda(A) < \infty$, the above implies $\lambda(V) < \infty$ and

$$\lambda(V \setminus A) < \frac{\epsilon}{2}.$$

As every open subset of $\mathbb{R}$ is a countable union of open intervals, we can write $V = \bigcup_{k=1}^\infty I_k$ where each I_k is an open interval. By the Monotone Convergence Theorem for measures (Theorem 2.1.18), we know that

$$\lambda(V) = \lim_{n\to\infty} \lambda\left(\bigcup_{k=1}^n I_k\right).$$

Hence there exists an $N \in \mathbb{N}$ such that

$$\lambda(V) < \lambda\left(\bigcup_{k=1}^N I_k\right) + \frac{\epsilon}{2}.$$

Therefore, if $U = \bigcup_{k=1}^N I_k$, we see that $U \subseteq V$ so $\lambda(U) < \infty$, and thus the above equation gives us that $\lambda(V \setminus U) < \frac{\epsilon}{2}$. Hence

$$\lambda(A \setminus U) \leq \lambda(V \setminus U) < \frac{\epsilon}{2}$$

and

$$\lambda(U \setminus A) \leq \lambda(V \setminus A) < \frac{\epsilon}{2}.$$

Hence $\lambda((A \setminus U) \cup (U \setminus A)) < \epsilon$ as desired. ∎

3.5 Lusin's Theorem

With the proof of Littlewood's first principle complete, we turn to the last of the remaining Littlewood's principles in the hopes to further understand measurable functions. This principle roughly states that 'every measurable function is continuous except on a set of small measure'. Formally, we have the following.

Theorem 3.5.1 (Lusin's Theorem). *Let $a, b \in \mathbb{R}$ with $a < b$, and let $f : [a, b] \to \mathbb{C}$ be measurable. For all $\epsilon > 0$ there exists a closed subset $F \subseteq [a, b]$ such that $\lambda([a, b] \setminus F) < \epsilon$ and $f|_F$ is continuous.*

Consequently, for all $\epsilon > 0$ there exists a exists a continuous function $g : [a, b] \to \mathbb{C}$ such that

$$\sup(\{|g(x)| \mid x \in [a, b]\}) \leq \sup(\{|f(x)| \mid x \in [a, b]\})$$

and

$$\mu(\{x \in [a, b] \mid f(x) \neq g(x)\}) < \epsilon.$$

To see why the first part of Lusin's Theorem implies the second, we note the following that will also be of use in the proof. In addition, the following implies we need only demonstrate the first portion of the statement of Lusin's Theorem.

Theorem 3.5.2 (Tietze's Extension Theorem on $\mathbb{R}$). *Let $F \subseteq \mathbb{R}$ be closed and let $h : F \to \mathbb{C}$ be continuous. There exists a continuous function $g : \mathbb{R} \to \mathbb{C}$ such that $g(x) = h(x)$ for all $x \in F$ and*

$$\sup(\{|g(x)| \mid x \in \mathbb{R}\}) \leq \sup(\{|h(x)| \mid x \in F\}).$$

Proof. Since F^c is open, we may write F^c as a countable union of disjoint non-empty open intervals, say $\bigcup_{n=1}^{\infty}(a_n, b_n)$. Define $g : \mathbb{R} \to \mathbb{C}$ by

$$g(x) = \begin{cases} h(x) & \text{if } x \in F \\ h(a_n) & \text{if } x \in (a_n, b_n) \text{ and } b_n = \infty \\ h(b_n) & \text{if } x \in (a_n, b_n) \text{ and } a_n = -\infty \\ \frac{f(b_n)-f(a_n)}{b_n-a_n}(x - a_n) + h(a_n) & \text{otherwise} \end{cases}$$

for all $x \in \mathbb{R}$; that is g agrees with h on F and is linear on each (a_n, b_n). It is not difficult to see that g is continuous and

$$\sup(\{|g(x)| \mid x \in \mathbb{R}\}) \leq \sup(\{|h(x)| \mid x \in F\}).$$ ∎

To proceed with the proof of Lusin's Theorem (Theorem 3.5.1), we begin with the simplest case.

Lemma 3.5.3. *Lusin's Theorem (Theorem 3.5.1) holds under the additional assumption that the function f is simple.*

Proof. Let

$$f = \sum_{k=1}^{N} a_k \chi_{A_k}$$

be the canonical representations of the simple function f. Thus $\{A_k\}_{k=1}^{N}$ are pairwise disjoint measurable sets with union $[a, b]$ and $a_k \geq 0$ for all k.

Fix $\epsilon > 0$. By Proposition 2.3.11, for every k there exists a compact subset $F_k \subseteq A_k$ such that

$$\lambda(A_k) < \lambda(F_k) + \frac{\epsilon}{N}.$$

Clearly $\{F_k\}_{k=1}^N$ are pairwise disjoint as $\{A_k\}_{k=1}^N$ are pairwise disjoint .

Let $F = \bigcup_{k=1}^N F_k$. Then F is compact (and thus closed) being the finite union of compact (and thus closed) sets. Moreover, notice since λ is finite, $\{A_k\}_{k=1}^N$ are pairwise disjoint, and $\{F_k\}_{k=1}^N$ are pairwise disjoint that

$$\lambda([a,b] \setminus F) = \lambda([a,b]) - \lambda(F) = \sum_{k=1}^N \lambda(A_k) - \lambda(F_k) < \epsilon.$$

It remains to show that $f|_F$ is continuous. To see this, suppose $(x_n)_{n\geq 1}$ is a sequence of elements in F that converge to a point $x \in F$. Since F is the union of the pairwise disjoint closed sets $\{F_k\}_{k=1}^N$, it must be the case that there exists an k_0 such that $x \in F_{k_0}$ and $x_n \in F_{k_0}$ for all $n \geq M$ (for otherwise there would exist a sequence in some F_k where $k \neq k_0$ that converges to x, which would imply $x \in F_k$ as F_k is closed thereby contradicting the disjointness of F_k and F_{k_0}). Therefore, as $x_n \in F_{k_0}$ for all $n \geq M$, $f(x_n) = a_{k_0} = f(x)$ for all $n \geq M$. Hence $f|_F$ is continuous as desired. ∎

Using our knowledge of simple functions, we are in a position to prove Lusin's Theorem (Theorem 3.5.1).

Proof of Lusin's Theorem (Theorem 3.5.1). Let $f : [a,b] \to \mathbb{C}$ be an arbitrary measurable function and fix $\epsilon > 0$. By applying Theorem 3.2.4 to the positive and negative parts of the real and imaginary parts of f, we can construct a sequence $(f_n)_{n\geq 1}$ of functions that are linear combinations of simple functions that converge to f pointwise. By applying Lemma 3.5.3 to each of the four simple functions in the linear combination of f_n and by taking the intersection of four closed sets (each whose measure in $[a,b]$ is at least $(b-a) - \frac{\epsilon}{2^{n+3}}$), there exists a closed subset $F_n \subseteq [a,b]$ and a continuous function $g_n : [a,b] \to \mathbb{C}$ such that $f_n(x) = g_n(x)$ for all $x \in F_n$ and $\lambda([a,b] \setminus F_n) < \frac{\epsilon}{2^{n+1}}$.

As $(f_n)_{n\geq 1}$ converges pointwise to f, Egoroff's Theorem (Theorem 3.3.1) implies there exists a measurable set B such that $\lambda(B) < \frac{\epsilon}{4}$ and $(f_n)_{n\geq 1}$ converges uniformly to f on $[a,b] \setminus B$. By Proposition 2.3.11 there exists an open set U (by this we really mean an open subset $U \subseteq [a,b]$, but by the relative topology we can view U as an open subset of $\mathbb{R}$) such that $B \subseteq U$ and

$$\lambda(U) < \lambda(B) + \frac{\epsilon}{4} < \frac{\epsilon}{2}$$

Hence, if $F_0 = [a,b] \setminus U \subseteq [a,b] \setminus B$, then F_0 is a closed subset such that $(f_n)_{n\geq 1}$ converges uniformly to f on F_0 and

$$\lambda([a,b] \setminus F_0) \leq \lambda(U) < \frac{\epsilon}{2}.$$

Let $F = \bigcap_{k=0}^\infty F_k$. Then clearly F is a closed subset of $[a,b]$ such that

$$\mu([a,b] \setminus F) = \mu\left(\bigcup_{k=0}^\infty ([a,b] \setminus F_k)\right) \leq \sum_{k=0}^\infty \mu([a,b] \setminus F_k) \leq \sum_{k=0}^\infty \frac{\epsilon}{2^{k+1}} = \epsilon.$$

Since $F \subseteq F_0$, we see that $(f_n)_{n\geq 1}$ converge uniformly to f on F. Therefore, since $F \subseteq F_n$ for all n and thus $f_n(x) = g_n(x)$ for all $x \in F_n$, we see that the continuous functions $(g_n|_F)_{n\geq 1}$ converge uniformly to $f|_F$ on F. Hence $f|_F$ is continuous as desired. ∎

Although Lusin's Theorem (Theorem 3.5.1) appears to rely on the finiteness of the measure used, this is not necessarily required as the following result demonstrates.

Theorem 3.5.4 (Lusin's Theorem, Lebesgue measure on $\mathbb{R}$). *Let $f : [a,b] \to \mathbb{C}$ be Lebesgue measurable. For all $\epsilon > 0$ there exists a closed subset $F \subseteq \mathbb{R}$ such that $\lambda(F^c) < \epsilon$ and $f|_F$ is continuous.*

Consequently, for all $\epsilon > 0$ there exists a exists a continuous function $g : \mathbb{R} \to \mathbb{C}$ such that

$$\sup(\{|g(x)| \mid x \in [a,b]\}) \leq \sup(\{|f(x)| \mid x \in [a,b]\})$$

and

$$\lambda(\{x \in [a,b] \mid f(x) \neq g(x)\}) < \epsilon.$$

Proof. Exercise. ∎

Chapter 4

The Lebesgue Integral

The Riemann integral is the notion of integral taught in most entry level calculus courses. This is because of its simplicity and sufficiency for most applications in science and engineering. However, there are many deficiencies in the Riemann integral. Firstly, only functions that are bounded and 'mostly' continuous are Riemann integrable. Next, the Riemann integral does not behave well with respect to limits (see Example 4.1.11). Finally, it is only possible to take the Riemann integral of functions defined on intervals.

In this chapter, after a brief review of the Riemann integral, we will generalize the notion of an integral. In particular, for every measure space there is a notion of real-valued (or complex-valued) functions on the measure space. In particular, by applying these notions to the Lebesgue measure, we obtain a generalization to the Riemann integral known as the Lebesgue integral. Furthermore, we will see that all such integrals are well-behaved with respect to limit operations; something that is particularly useful in analysis.

4.1 Review of the Riemann Integral

In this section, we will briefly review the Riemann integral. Recall the Riemann integral of a function over an interval $[a, b]$ is defined based on dividing up the interval $[a, b]$ into very small regions:

Definition 4.1.1. A *partition* of a closed interval $[a, b]$ is a finite list of real numbers $\{t_k\}_{k=0}^n$ such that

$$a = t_0 < t_1 < t_2 < \cdots < t_{n-1} < t_n = b.$$

Once we have divided $[a, b]$ into regions, we can estimate the area beneath a function $f : [a, b] \to \mathbb{R}$ that is bounded. To do so, we obtain lower and upper estimates for this area.

Definition 4.1.2. Let $\mathcal{P} = \{t_k\}_{k=0}^n$ be a partition of $[a, b]$ and let $f : [a, b] \to \mathbb{R}$ be bounded. The *lower Riemann sum* of f associated to $\mathcal{P}$, denoted $L(f, \mathcal{P})$, is

$$L(f, \mathcal{P}) = \sum_{k=1}^{n} m_k(t_k - t_{k-1})$$

where, for all $k \in \{1, \ldots, n\}$,

$$m_k = \inf\{f(x) \mid x \in [t_{k-1}, t_k]\}.$$

The *upper Riemann sum* of f associated to $\mathcal{P}$, denoted $U(f, \mathcal{P})$, is

$$U(f, \mathcal{P}) = \sum_{k=1}^{n} M_k(t_k - t_{k-1})$$

where, for all $k \in \{1, \ldots, n\}$,

$$M_k = \sup\{f(x) \mid x \in [t_{k-1}, t_k]\}.$$

In order to obtain better and better estimates for the area under the curve, we should simply add more and more points into the partition.

Definition 4.1.3. Let $\mathcal{P}$ and $\mathcal{Q}$ be partitions of $[a, b]$. It is said that $\mathcal{Q}$ is a *refinement* of $\mathcal{P}$, denoted $\mathcal{P} \leq \mathcal{Q}$, if $\mathcal{P} \subseteq \mathcal{Q}$; that is $\mathcal{Q}$ has all of the points that $\mathcal{P}$ has, and possibly more.

Of course, adding more and more points into our partition should only improve the estimates our lower and upper Riemann sums give for the area under the curve.

Lemma 4.1.4. *Let $\mathcal{P}$ and $\mathcal{Q}$ be partitions of $[a, b]$ and let $f : [a, b] \to \mathbb{R}$ be bounded. If $\mathcal{Q}$ is a refinement of $\mathcal{P}$, then*

$$L(f, \mathcal{P}) \leq L(f, \mathcal{Q}) \leq U(f, \mathcal{Q}) \leq U(f, \mathcal{P}).$$

Of course, one natural expectation is that any lower estimate for the area under a curve must be smaller than any upper estimate for the area under a curve. Using Lemma 4.1.4 and the following notion will allow us to guarantee this to be true.

Definition 4.1.5. Given two partitions $\mathcal{P}$ and $\mathcal{Q}$ of $[a, b]$, the *common refinement* of $\mathcal{P}$ and $\mathcal{Q}$ is the partition $\mathcal{P} \cup \mathcal{Q}$ of $[a, b]$.

Clearly, given two partitions $\mathcal{P}$ and $\mathcal{Q}$, $\mathcal{P} \cup \mathcal{Q}$ is a refinement of $\mathcal{P}$ and $\mathcal{Q}$. Consequently, if $f : [a, b] \to \mathbb{R}$ is bounded, then Lemma 4.1.4 implies that

$$L(f, \mathcal{P}) \leq L(f, \mathcal{P} \cup \mathcal{Q}) \leq U(f, \mathcal{P} \cup \mathcal{Q}) \leq U(f, \mathcal{Q}).$$

Hence any lower bound for the area under a curve is smaller than any upper bound for the area under a curve.

The notion of a Riemann integrable function is now easy to define.

Definition 4.1.6. Let $f : [a, b] \to \mathbb{R}$ be bounded. It is said that f is *Riemann integrable* on $[a, b]$ if

$$\begin{aligned}&\sup\{L(f, \mathcal{P}) \mid \mathcal{P} \text{ a partition of } [a, b]\}\\ &= \inf\{U(f, \mathcal{P}) \mid \mathcal{P} \text{ a partition of } [a, b]\}.\end{aligned}$$

If f is Riemann integrable on $[a, b]$, the *Riemann integral of f from a to b*, denoted $\int_a^b f(x)\,dx$, is

$$\begin{aligned}\int_a^b f(x)\,dx &= \sup\{L(f, \mathcal{P}) \mid \mathcal{P} \text{ a partition of } [a, b]\}\\ &= \inf\{U(f, \mathcal{P}) \mid \mathcal{P} \text{ a partition of } [a, b]\}.\end{aligned}$$

Using the notion of common refinements together with Lemma 4.1.4, we see that the definition of a Riemann integrable function has the following equivalent characterization.

Theorem 4.1.7. *Let $f : [a, b] \to \mathbb{R}$ be bounded. Then f is Riemann integrable if and only if for every $\epsilon > 0$ there exists a partition $\mathcal{P}$ of $[a, b]$ such that*

$$0 \leq U(f, \mathcal{P}) - L(f, \mathcal{P}) < \epsilon.$$

It is not difficult to verify that many functions are Riemann integrable.

Theorem 4.1.8. *If $f : [a, b] \to \mathbb{R}$ is*

- *monotonic and bounded, or*
- *continuous,*

then f is Riemann integrable on $[a, b]$.

Theorem 4.1.9. *Let $f, g : [a, b] \to \mathbb{R}$ be bounded, Riemann integrable functions on $[a, b]$. Then:*

a) If $\alpha \in \mathbb{R}$, then αf is Riemann integrable on $[a, b]$ and

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

b) $f + g$ is Riemann integrable on $[a, b]$ and

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

c) If $a \leq c \leq b$, then f is Riemann integrable on $[a, c]$ and $[c, b]$ with

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

d) If $f(x) \leq g(x)$ for all $x \in [a,b]$, then

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

e) If $m \leq f(x) \leq M$ for all $x \in [a,b]$, then

$$m(b-a) \leq \int_a^b f(x)\,dx \leq M(b-a).$$

f) $|f| : [a,b] \to \mathbb{R}$ is Riemann integrable on $[a,b]$ and

$$\left|\int_a^b f(x)\,dx\right| \leq \int_a^b |f(x)|\,dx.$$

g) $fg : [a,b] \to \mathbb{R}$ is Riemann integrable on $[a,b]$ (although there is no nice formula for the integral of a product).

Unfortunately, not every bounded function is Riemann integrable.

Example 4.1.10. Let $f : [0,1] \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \in \mathbb{R} \setminus \mathbb{Q} \end{cases}$$

for all $x \in [0,1]$. We easily see that $L(f,\mathcal{P}) = 0$ and $U(f,\mathcal{P}) = 1$ for all partitions $\mathcal{P}$ of $[0,1]$. Hence f is not Riemann integrable

Using Example 4.1.10 we can see that the Riemann integral is quite deficient when it comes to pointwise (increasing) limits.

Example 4.1.11. To see this, recall that $\mathbb{Q} \cap [0,1]$ is countable so we may write $\mathbb{Q} \cap [0,1] = \{r_n \mid n \in \mathbb{N}\}$. For each $m \in \mathbb{N}$, let

$$f_m(x) = \begin{cases} 1 & \text{if } x = r_n \text{ for some } n \leq m \\ 0 & \text{otherwise} \end{cases}$$

for all $x \in [0,1]$. Notice each f_m is Riemann integrable with $\int_0^1 f_m(x)\,dx = 0$. However, if f is the function from Example 4.1.10, it is elementary to see that $\lim_{m\to\infty} f_m(x) = f(x)$ for all $x \in [0,1]$ yet f is not Riemannn integrable.

The above seems quite the odd deficit from the standpoint of the Lebesgue measure. The rational numbers have zero Lebesgue measure; that is, they have no length. Consequently, it seems natural that the area under the function f from Example 4.1.10 should be zero. In particular, any function that is one on a measurable set A and zero everywhere else should have integral being the measure of A. Thus we begin our search for such an integral.

4.2 The Integral of Positive Functions

With the above review of Riemann integrals complete, we turn our attention to defining the integral of measurable functions. When defining Riemann integration, one defines the integral via upper and lower Riemann sums of partitions for any function and later determines which functions were integrable. Our approach for the integral over a measure space will be different. We want all measurable functions to be integrable, and we will build-up our integral systematically. We will do this by first developing a notion of an integral for all non-negative measurable functions. This will enable us to construct an integral for other measurable functions using a linear combination of integrals for non-negative measurable functions.

To begin, if A is a measurable set, it would be natural to expect to be able to integrate the characteristic function of A, whose integral should just be $\mu(A)$. Of course, this enables us to integrate $\chi_{\mathbb{Q}}$ and obtain zero thereby avoiding one of the pitfalls of the Riemann integral we saw in Example 4.1.11. Furthermore, it will not be difficult to see how we should define the integral for the simplest of functions.

Definition 4.2.1. Let $(X, \mathcal{A}, \mu)$ be a measure space and let $\varphi : X \to [0, \infty)$ be a simple function with canonical representation $\varphi = \sum_{k=1}^{n} a_k \chi_{A_k}$. For every $A \in \mathcal{A}$, we define the *integral of φ over A against μ* to be

$$\int_A \varphi \, d\mu = \sum_{k=1}^{n} a_k \mu(A \cap A_k) \in [0, \infty]$$

where

$$a \times \infty = \begin{cases} 0 & \text{if } a = 0 \\ \infty & \text{otherwise} \end{cases}.$$

In fact, we have see the quantity in Definition 4.2.1 before.

Remark 4.2.2. Let $(X, \mathcal{A}, \mu)$ be a measure space and let $\varphi : X \to [0, \infty)$ be a simple function. If one defines $\nu : \mathcal{A} \to [0, \infty]$ by

$$\nu(A) = \int_A \varphi \, d\mu,$$

then ν is a measure as shown in Example 2.1.16.

Example 4.2.3. Consider $\chi_{\mathbb{Q}}$. Then $\chi_Q = 1\chi_{\mathbb{Q}} + 0\chi_{\mathbb{R}\setminus\mathbb{Q}}$ is the canonical representation of $\chi_{\mathbb{Q}}$ so

$$\int_{[0,1]} \chi_{\mathbb{Q}} \, d\lambda = 1\lambda(\mathbb{Q} \cap [0, 1]) + 0\lambda((\mathbb{R} \setminus \mathbb{Q}) \cap [0, 1]) = 1(0) + 0(1) = 0.$$

Remark 4.2.4. Let $(X, \mathcal{A}, \mu)$ be a measure space and let $g : X \to [0, \infty)$ be such that $g = \sum_{k=1}^{n} a_k \chi_{A_k}$ where $\{A_k\}_{k=1}^{n} \subseteq \mathcal{A}$ are pairwise disjoint possibly empty sets with union X, and $\{a_k\}_{k=1}^{n} \subseteq [0, \infty)$. By Remark 3.2.3 we know that g is a simple function. In particular, Remark 3.2.3 shows that if $g(X) = \{b_1, \ldots, b_m\}$ and $B_j = g^{-1}(\{b_j\})$ then $g = \sum_{j=1}^{m} b_j \chi_{B_j}$ is the canonical representation of g. In particular, if for each $j \in \{1, \ldots, m\}$ we define

$$K_j = \{k \in \{1, \ldots, n\} \mid a_k = b_j\}$$

then $\bigcup_{j=1}^{m} K_j = \{k \in \{1, \ldots, n\} \mid A_k \neq \emptyset\}$ and

$$B_j = \bigcup_{k \in K_j} A_k.$$

Hence if $A \in A$, then

$$\begin{aligned}
\sum_{k=1}^{n} a_k \mu(A_k \cap A) &= \sum_{\{k \in \{1, \ldots, n \mid A_k \neq \emptyset\}} a_k \mu(A_k \cap A) \\
&= \sum_{j=1}^{m} \sum_{k \in K_j} a_k \mu(A_k \cap A) \\
&= \sum_{j=1}^{m} \sum_{k \in K_j} b_j \mu(A_k \cap A) \\
&= \sum_{j=1}^{m} b_j \mu(B_j \cap A) \\
&= \int_A g \, d\mu.
\end{aligned}$$

Hence in Definition 4.2.1 it is not necessary for the $\{a_k\}_{k=1}^{n} \subseteq [0, \infty]$ to be distinct.

In order to have a good notion of an integral, we expect several properties that we saw for the Riemann integral. The following verifies these properties for the integral of simple functions. In particular, the following demonstrates that we can extend Remark 4.2.4 to the case where the $\{A_k\}_{k=1}^{n} \subseteq \mathcal{A}$ are not necessarily pairwise disjoint. Note for the remainder of these notes, given $f, g : X \to \mathbb{R}$, we will use $f \leq g$ to denote $f(x) \leq g(x)$ for all $x \in X$.

Theorem 4.2.5. *Let $(X, \mathcal{A}, \mu)$ be a measure space, let $A \in \mathcal{A}$, and let $\varphi, \psi : X \to [0, \infty)$ be simple functions. Then:*

(1) if $c \geq 0$ then $c\varphi$ is a simple function with $\int_A c\varphi \, d\mu = c \int_A \varphi \, d\mu$.

(2) $\varphi + \psi$ is a simple function with $\int_A \varphi + \psi \, d\mu = \int_A \varphi \, d\mu + \int_A \psi \, d\mu$.

(3) If $B \in \mathcal{A}$ and $B \subseteq A$, then $\int_B \varphi \, d\mu \leq \int_A \varphi \, d\mu$.

(4) $\varphi\chi_A$ is a simple function with $\int_X \chi_A\varphi \, d\mu = \int_A \varphi \, d\mu$.

(5) If $\varphi\chi_A \leq \psi\chi_A$, then $\int_A \varphi \, d\mu \leq \int_A \psi \, d\mu$.

Proof. Let

$$\varphi = \sum_{k=1}^{n} a_k \chi_{A_k} \quad \text{and} \quad \psi = \sum_{k=1}^{m} b_k \chi_{B_k}$$

be the canonical representations of φ and ψ respectively. Thus $\{A_k\}_{k=1}^n$ are pairwise disjoint sets with union X and $\{B_k\}_{k=1}^m$ are pairwise disjoint sets with union X.

To see (1), notice the result is trivial if $c = 0$. Otherwise, if $c > 0$ then

$$c\varphi = \sum_{k=1}^{n} c a_k \chi_{A_k}$$

so $c\varphi$ is a simple function and the above is the canonical representation of $c\varphi$. Hence, by definition,

$$\int_A c\varphi \, d\mu = \sum_{k=1}^{n} c a_k \mu(A \cap A_k) = c\left(\sum_{k=1}^{n} a_k \mu(A \cap A_k)\right) = c \int_A \varphi \, d\mu.$$

To see (2), for each $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m\}$, let $C_{i,j} = A_i \cap B_j$. Clearly

$$\{C_{i,j} \mid i \in \{1, \ldots, n\}, j \in \{1, \ldots, m\}\}$$

is a collection of pairwise disjoint measurable sets such that $\bigcup_{i=1}^n C_{i,j} = B_j$ for all $j \in \{1, \ldots, m\}$, $\bigcup_{j=1}^m C_{i,j} = A_i$ for all $i \in \{1, \ldots, n\}$, and

$$\varphi + \psi = \sum_{i=1}^{n} \sum_{j=1}^{m} (a_i + b_j) \chi_{C_{i,j}}.$$

Hence by Remark 4.2.4,

$$\begin{aligned}
\int_A \varphi + \psi \, d\mu &= \sum_{i=1}^{n} \sum_{j=1}^{m} (a_i + b_j) \mu(C_{i,j} \cap A) \\
&= \sum_{i=1}^{n} a_i \sum_{j=1}^{m} \mu(C_{i,j} \cap A) + \sum_{j=1}^{m} b_j \sum_{i=1}^{n} \mu(C_{i,j} \cap A) \\
&= \sum_{i=1}^{n} a_i \mu(A_i \cap A) + \sum_{j=1}^{m} b_j \mu(B_j \cap A) \quad \text{as the } C_{i,j} \text{ are disjoint} \\
&= \int_A \varphi \, d\mu + \int_A \psi \, d\mu.
\end{aligned}$$

Hence (2) follows.

To see (3), note if $B \subseteq A$ then $\mu(A_k \cap B) \leq \mu(A_k \cap A)$ for all $k \in \{1, \ldots, n\}$ by the monotonicity of measures. Hence

$$\int_B \varphi \, d\mu = \sum_{k=1}^n a_k \mu(B \cap A_k) \leq \sum_{k=1}^n a_k \mu(A \cap A_k) = \int_A \varphi \, d\mu$$

as desired.

To see (4), we notice that

$$\chi_A \varphi = \sum_{k=1}^n a_k \chi_{A_k} \chi_A = \sum_{k=1}^n a_k \chi_{A_k \cap A}$$

(as $\chi_{A_k}(x)\chi_A(x) = 1$ if and only if $x \in A_k$ and $x \in A$ if and only if $\chi_{A_k \cap A}(x) = 1$). Hence (4) easily follows.

To see (3), note that $\psi\chi_A - \varphi\chi_A$ is a simple function by Remark 3.2.3 and by (4). Hence by (2)

$$\int_X \psi\chi_A \, d\mu = \int_X \varphi\chi_A + (\psi\chi_A - \varphi\chi_A) \, d\mu = \int_X \varphi\chi_A \, d\mu + \int_X \psi\chi_A - \varphi\chi_A \, d\mu.$$

Therefore, as $\int_X \psi\chi_A - \varphi\chi_A \, d\mu \geq 0$, the result follows by (4). ∎

We record the following fact implied by Theorem 4.2.5.

Corollary 4.2.6. *Suppose $(X, \mathcal{A}, \mu)$ is a measure space and $\varphi : X \to [0, \infty]$ is such that $\varphi = \sum_{k=1}^n a_k \chi_{A_k}$ where $\{A_k\}_{k=1}^n \subseteq \mathcal{A}$ and $\{a_k\}_{k=1}^n \subseteq [0, \infty]$ (that is, $\{A_k\}_{k=1}^n$ are not necessarily disjoint with union X and $\{a_k\}_{k=1}^n$ need not be distinct). Then for all $A \in \mathcal{A}$,*

$$\int_A \varphi \, d\mu = \sum_{k=1}^n a_k \mu(A_k \cap A).$$

Hence the representation of a simple function does not affect the value of the integral.

Our next goal is to extend the notion of integral of simple functions to other functions. To do so, we will use some form of approximation. Recall the notion of the Riemann integral was obtained by approximating the area under the curve from above and below. If f is positive and measurable and we wanted to approximate the under f, we could approximate f with a simple function from below and integrate the simple function. Furthermore, Theorem 3.2.4 says we can make these approximations as close as we would like if we only require pointwise limits. As such, it is natural to define the integral of a non-negative measurable function as follows.

Definition 4.2.7. Let $(X, \mathcal{A}, \mu)$ be a measure space, let $A \in \mathcal{A}$, and let $f : X \to [0, \infty]$ be measurable. The *integral of f over A against μ* is defined to be

$$\int_A f\, d\mu = \sup\left\{ \int_A \varphi\, d\mu \,\middle|\, \varphi : X \to [0, \infty) \text{ simple}, \varphi \leq f \right\}.$$

In the case $(X, \mathcal{A}, \mu) = (\mathbb{R}, \mathcal{M}(\mathbb{R}), \lambda)$, the above integral is called the *Lebesgue integral of f over A*.

Remark 4.2.8. Of course the one subtlety in the above definition is that if ψ is a simple function, then we have defined $\int_A \psi\, d\mu$ in two ways! One way as a simple function and one way as a non-negative measurable function.

Let $\alpha = \int_A \psi\, d\mu$ when we evaluate the integral viewing ψ as a simple function and let $\beta = \int_A \psi\, d\mu$ when we evaluate the integral viewing ψ as a positive measurable function. By Definition 4.2.7, we see using $\varphi = \psi$ that $\alpha \leq \beta$. However, if $\varphi : X \to [0, \infty)$ is a simple function such that $\varphi \leq \psi$, we obtain by part (5) of Theorem 4.2.5 that

$$\int_A \varphi\, d\mu \leq \alpha.$$

Hence by taking the supremum in Definition 4.2.7 yields $\beta \leq \alpha$ so $\alpha = \beta$. Hence our two definitions for the integral of a simple function are equal.

Using Theorem 4.2.5, several properties of integrating simple functions transfer to integrating non-negative measurable functions.

Theorem 4.2.9. *Let $(X, \mathcal{A}, \mu)$ be a measure space, let $A \in \mathcal{A}$, and let $f, g : X \to [0, \infty]$ be measurable functions. Then:*

(1) if $c \geq 0$ then $\int_A cf\, d\mu = c \int_A f\, d\mu$.

(2) If $B \in \mathcal{A}$ and $B \subseteq A$, then $\int_B f\, d\mu \leq \int_A f\, d\mu$.

(3) $\int_X \chi_A f\, d\mu = \int_A f\, d\mu$.

(4) If $f\chi_A \leq g\chi_A$, then $\int_A f\, d\mu \leq \int_A g\, d\mu$.

(5) $\int_A f\, d\mu = 0$ if and only if $\mu(\{x \in X \mid f(x) > 0\} \cap A) = 0$.

(6) If $\mu(A) = 0$, then $\int_A f\, d\mu = 0$.

Proof. Clearly (1) holds if $c = 0$. Otherwise if $c > 0$, it is clear that if $\varphi : X \to [0, \infty)$ is a simple function and $\varphi \leq f$ then $c\varphi$ is a simple function and $c\varphi \leq cf$. Hence as Theorem 4.2.5 implies

$$c \int_A \varphi\, d\mu = \int_A c\varphi\, d\mu \leq \int_A cf\, d\mu.$$

Thus we obtain that $\int_A cf\,d\mu \geq c\int_A f\,d\mu$. Similarly, if $\varphi : X \to [0,\infty)$ is a simple function and $\varphi \leq cf$ then $\frac{1}{c}\varphi$ is a simple function and $c\frac{1}{c}\varphi \leq f$. Hence as Theorem 4.2.5 implies

$$\frac{1}{c}\int_A \varphi\,d\mu = \int_A \frac{1}{c}\varphi\,d\mu \leq \int_A f\,d\mu.$$

Thus we obtain that $\int_A cf\,d\mu = c\int_A f\,d\mu$ as desired.

Note (2) clearly follows from Theorem 4.2.5 and (4) clearly follows by Definition 4.2.7 once (3) is complete. Similarly, (6) follows easily from (5).

To see (3), notice by Theorem 4.2.5

$$\begin{aligned}\int_A f\,d\mu &= \sup\left\{\int_A \varphi\,d\mu \,\middle|\, \varphi : X \to [0,\infty) \text{ simple}, \varphi \leq f\right\}\\ &= \sup\left\{\int_X \chi_A\varphi\,d\mu \,\middle|\, \varphi : X \to [0,\infty) \text{ simple}, \varphi \leq f\right\}\\ &= \sup\left\{\int_X \psi\,d\mu \,\middle|\, \varphi : X \to [0,\infty) \text{ simple}, \psi \leq \chi_A f\right\}\\ &= \int_X f\chi_A\,d\mu\end{aligned}$$

as desired.

Finally, let $B = \{x \in X \mid f(x) > 0\} \cap A$. Suppose $\int_A f\,d\mu = 0$. Notice B is measurable as A is measurable and f is measurable. For each $n \in \mathbb{N}$ let

$$A_n = \left\{x \in X \mid f(x) > \frac{1}{n}\right\}.$$

Since f is measurable, A_n is measurable for all $n \in \mathbb{N}$. Hence $\frac{1}{n}\chi_{A_n}$ is a simple function for each $n \in \mathbb{N}$. Since $\frac{1}{n}\chi_{A_n} \leq f$, the definition of the integral implies that

$$\frac{1}{n}\mu(A_n \cap A) = \int_A \frac{1}{n}\chi_{A_n}\,d\mu \leq \int_A f\,d\mu = 0.$$

Hence $\mu(A_n \cap A) = 0$ for all $n \in \mathbb{N}$. Since

$$B = \bigcup_{n=1}^{\infty} A_n \cap A$$

as $f : X \to [0,\infty]$, we obtain by the subadditivity of measures (Proposition 2.1.17) that $\mu(B) = 0$.

Conversely, suppose $\mu(B) = 0$. Suppose $\varphi : X \to [0,\infty)$ is a simple function such that $\varphi \leq f$. Write $\varphi = \sum_{k=1}^{n} a_k\chi_{A_k}$ where $a_k > 0$ for all $k \in \{1,\ldots,n\}$. As $\varphi \leq f$, we see that

$$A_k \subseteq \{x \in X \mid f(x) > 0\}.$$

Hence the monotonicity of measures implies that

$$\mu(A_k \cap A) \leq \mu(B) = 0.$$

Hence

$$\int_A \varphi \, d\mu = \sum_{k=1}^{n} a_k \mu(A_k \cap A) = 0.$$

Hence, by the definition of the integral, $\int_A f \, d\mu = 0$. ∎

Of course, one omission in Theorem 4.2.9 is the additivity of integrals:

$$\int_X f + g \, d\mu = \int_X f \, d\mu + \int_X g \, d\mu.$$

Clearly if φ and ψ are simple functions with $\varphi \leq f$ and $\psi \leq g$, then $\varphi + \psi$ is a simple function with $\varphi + \psi \leq f + g$. Thus Theorem 4.2.5 clearly implies

$$\int_X f \, d\mu + \int_X g \, d\mu \leq \int_X f + g \, d\mu.$$

However, difficulty occurs with the reverse inequality as if φ were a simple function with $\varphi \leq f + g$, how can we find simple functions φ_1 and φ_2 such that $\varphi_1 \leq f$, $\varphi_2 \leq g$, and $\varphi_1 + \varphi_2 = \varphi$?

We will investigate this question in the next section. For now, we note some essential connections.

Remark 4.2.10. As Theorem 4.2.9 implies that $\int_X \chi_A f \, d\mu = \int_A f \, d\mu$, when developing the theory of integrals, it suffices to consider only integrals over all of X as the results for integrating over an arbitrary measurable subset A will then follow from multiplying the functions under consideration by χ_A.

With Theorem 4.2.9 we may compare the Lebesgue integral to the Riemann integral. We begin with the following.

Proposition 4.2.11. *A function $f : [a, b] \to [0, \infty)$ is Riemann integrable if and only if f is bounded and f is continuous almost everywhere.*

Proof. To begin, suppose f is Riemann integrable. Clearly this implies f is bounded by definition. To see that f is continuous almost everywhere, for each $n \in \mathbb{N}$ let

$$D_n(f) = \left\{ x \in \mathbb{R} \,\middle|\, \begin{array}{c} \text{for every } \delta > 0 \text{ there exists } y, z \text{ such that} \\ |x - y| < \delta, |x - z| < \delta, \text{ and } |f(y) - f(z)| \geq \frac{1}{n} \end{array} \right\}.$$

Recall the discontinuities of f are given by $\bigcup_{n=1}^{\infty} D_n(f)$. Hence, to show that f is continuous almost everywhere (i.e. the set of discontinuities of f have measure zero), it suffices to show that each $D_n(f)$ has Lebesgue measure zero.

Suppose to the contrary that there exists an $q \in \mathbb{N}$ such that $\lambda(D_q(f)) > 0$. Since f is Riemann integrable, there exists a partition $\mathcal{P} = \{t_k\}_{k=0}^n$ of $[a, b]$ such that if for all $k \in \{1, \dots, n\}$ we define

$$m_k = \inf\{f(x) \mid x \in [t_{k-1}, t_k]\} \quad \text{and} \quad M_k = \sup\{f(x) \mid x \in [t_{k-1}, t_k]\}$$

then

$$U(f, \mathcal{P}) - L(f, \mathcal{P}) = \sum_{k=1}^n (M_k - m_k)(t_k - t_{k-1}) < \frac{1}{q}\lambda(D_q(f)).$$

For each $k \in \{1, \dots, n\}$ let $I_k = [t_{k-1}, t_k]$. Notice if $D_q(f) \cap I_k \neq \emptyset$, then $M_k - m_k \geq \frac{1}{q}$ by the definition of $D_q(f)$. Hence as

$$D_q(f) \subseteq \bigcup_{\substack{k \in \{1, \dots, n\} \\ I_k \cap D_q(f) \neq \emptyset}} I_k$$

we obtain that

$$\begin{aligned}
\sum_{k=1}^n (M_k - m_k)(t_k - t_{k-1}) &\geq \sum_{\substack{k \in \{1, \dots, n\} \\ I_k \cap D_q(f) \neq \emptyset}} (M_k - m_k)(t_k - t_{k-1}) \\
&\geq \sum_{\substack{k \in \{1, \dots, n\} \\ I_k \cap D_q(f) \neq \emptyset}} \frac{1}{q}(t_k - t_{k-1}) \\
&\geq \frac{1}{q}\lambda\left(\bigcup_{\substack{k \in \{1, \dots, n\} \\ I_k \cap D_q(f) \neq \emptyset}} I_k \right) \\
&\geq \frac{1}{q}\lambda(D_q(f)).
\end{aligned}$$

As this contradicts our selection of $\mathcal{P}$, it must be the case that f is continuous almost everywhere.

Conversely, suppose that f is bounded and continuous almost everywhere. To show that f is Riemann integrable, we demonstrate that for all $\epsilon > 0$ there exists a partition $\mathcal{P}$ of $[a, b]$ such that

$$U(f, \mathcal{P}) - L(f, \mathcal{P}) < \epsilon.$$

To begin, fix $\epsilon > 0$. Since f is bounded, there exists an $M \in \mathbb{R}$ such that $|f(x)| \leq M$ for all $x \in [a, b]$. Choose $n \in \mathbb{N}$ such that $\frac{1}{n}(b - a) < \frac{1}{2}\epsilon$. Since $D_n(f)$ has Lebesgue measure zero, there exists a collection $\{I_k\}_{k=1}^\infty$ of open intervals such that $D_n(f) \subseteq \bigcup_{k=1}^\infty I_k$ and

$$\lambda\left(\bigcup_{k=1}^\infty I_k\right) \leq \frac{\epsilon}{2(M+1)}.$$

As $D_n(f)$ is closed and thus a compact subset of $[a,b]$, there exists an $m \in \mathbb{N}$ such that $D_n(f) \subseteq \bigcup_{k=1}^m I_k$ and thus

$$\lambda\left(\bigcup_{k=1}^m I_k\right) \leq \frac{\epsilon}{2(M+1)}.$$

Consider $F = [a,b] \cap (\bigcup_{k=1}^m I_k)^c$. Then F is a finite union of closed intervals in $[a,b]$ such that $F \subseteq D_n(f)^c$. Hence if $x \in F \subseteq D_n(f)^c$ there exists an open neighbourhood U_x of x such that if $y, z \in U_x$ then $|f(y) - f(z)| < \frac{1}{n}$. Since F is a closed subset of a compact set and thus compact, we can cover F with a finite number of these open intervals. Hence one can form a partition $\mathcal{P}$ of F such that the difference between the upper and lower Riemann sums of f with respect to $\mathcal{P}$ on each interval is at most the length of the interval times $\frac{1}{n}$.

Notice $\mathcal{P}$ can then also be viewed as a partition on $[a,b]$ (by adding in a and/or b if necessary). Then the intervals described by the partition that intersect F contribute at most $\frac{1}{n}(b-a)$ to the difference of the upper and lower Riemann sums. Furthermore, the intervals described by the partition that do not intersect F contribute at most $2M\lambda\left(\bigcup_{k=1}^m I_k\right)$ to the difference of the upper and lower Riemann sums. Hence

$$U(f,\mathcal{P}) - L(f,\mathcal{P}) \leq \frac{1}{n}(b-a) + 2M\lambda\left(\bigcup_{k=1}^m I_k\right) < \epsilon$$

and the result follows. ∎

Corollary 4.2.12. *If $f : [a,b] \to [0,\infty)$ is Riemann integrable, then f is Lebesgue measurable.*

Proof. By Proposition 4.2.11, f is continuous almost everywhere. Hence there exists a Lebesgue measurable subset A of $[a,b]$ such that $\lambda(A^c) = 0$ and $f|_A : A \to [0,\infty)$ is continuous (where A is a metric space with the metric induced by the absolute value).

To show that f is Lebesgue measurable, we will apply Corollary 3.1.9. To begin, let $\alpha \in \mathbb{R}$ be arbitrary. Then

$$f^{-1}((\alpha,\infty)) = \left(f^{-1}((\alpha,\infty)) \cap A^c\right) \cup f|_A^{-1}((\alpha,\infty)).$$

Since

$$\left(f^{-1}((\alpha,\infty)) \cap A^c\right) \subseteq A^c$$

and $\lambda(A^c) = 0$, we obtain from the completeness of λ that $f^{-1}((\alpha,\infty)) \cap A^c)$ is measurable. Hence it suffices to show that $f|_A^{-1}((\alpha,\infty))$ is measurable.

Since $f|_A$ is continuous, $V = f|_A^{-1}((\alpha,\infty))$ is an open subset of $(A, |\cdot|)$. Since V is open, for each $x \in V$ there exists an $r_x > 0$ such that $(x - r_x, x + r_x) \cap A \subseteq V$. Let

$$U = \bigcup_{x \in V} (x - r_x, x + r_x).$$

Clearly U is an open subset of $\mathbb{R}$ such that $V = U \cap A$. Therefore, since U is measurable (as U is open) and A is measurable, we obtain that V is measurable. ■

Now that we know every Riemann integrable function is Lebesgue measurable, we can compare the integrals.

Proposition 4.2.13. *If $f : [a, b] \to [0, \infty)$ is Riemann integrable, then*

$$\int_a^b f(x)\,dx = \int_{[a,b]} f\,d\lambda.$$

Proof. Let $\mathcal{P} = \{t_k\}_{k=0}^n$ be a arbitrary partition of $[a, b]$. Clearly if for each $k \in \{1, \ldots, n\}$ we define

$$m_k = \inf\{f(x) \mid x \in [t_{k-1}, t_k]\} \quad \text{and} \quad M_k = \sup\{f(x) \mid x \in [t_{k-1}, t_k]\}$$

and we let

$$\varphi_{\mathcal{P}} = \sum_{k=1}^n m_k \chi_{(t_{k-1}, t_k)} \quad \text{and} \quad \psi_{\mathcal{P}} = \sum_{k=1}^n M_k \chi_{[t_{k-1}, t_k]}$$

then φ and ψ are simple functions such that $\varphi_{\mathcal{P}} \leq f \leq \psi_{\mathcal{P}}$. Furthermore, we clearly see by Theorem 4.2.9 that

$$L(f, \mathcal{P}) = \int_{[a,b]} \varphi_{\mathcal{P}}\,d\lambda \leq \int_{[a,b]} f\,d\mu \leq \int_{[a,b]} \psi_{\mathcal{P}}\,d\lambda = U(f, \mathcal{P})$$

as $\varphi_{\mathcal{P}} \leq f \leq \psi_{\mathcal{P}}$ almost everywhere and a set of measure zero does not contribute to the integral. Therefore, as the Riemann integral of f is supremum of $L(f, \mathcal{P})$ over all partitions and the infimum of $U(f, \mathcal{P})$ over all partitions, we obtain that

$$\int_a^b f(x)\,dx \leq \int_{[a,b]} f\,d\mu \leq \int_a^b f(x)\,dx.$$

■

Hence the Riemann and Lebesgue integrals agree whenever the Riemann integral exists! Hence we truly have a more general notion of an integral for non-negative measurable functions!

4.3 The Monotone Convergence Theorem

In order to resolve the additivity of integrals, we turn to Theorem 3.2.4 which says each measurable function is the pointwise limit of an increasing sequence of simple functions. If we knew that integrals preserved these limits, then we would easily be able to show that

$$\int_X f + g \, d\mu = \int_X f \, d\mu + \int_X g \, d\mu$$

by approximating by simple functions and taking limits. Our next results will enable us to do this and demonstrates the power of the integrals under consideration as the analogous result fails in the Riemann integral is used.

Theorem 4.3.1 (The Monotone Convergence Theorem). *Let $(X, \mathcal{A}, \mu)$ be a measure space. For each $n \in \mathbb{N}$ let $f_n : X \to [0, \infty]$ be a measurable function such that $f_n \leq f_{n+1}$ for all $n \in \mathbb{N}$. If $f : X \to [0, \infty]$ is the pointwise limit of $(f_n)_{n \geq 1}$, then f is measurable and for all $A \in \mathcal{A}$*

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

Proof. First note that f is measurable by Proposition 3.1.17 as f is the pointwise limit of measurable functions. Next note Remark 4.2.10 implies we may assume that $A = X$ as multiplying by a characteristic function will preserve measurability and pointwise limits.

Since $f_n \leq f$ for all $n \in \mathbb{N}$ by construction, Theorem 4.2.9 implies that

$$\int_X f_n \, d\mu \leq \int_X f \, d\mu$$

for all $n \in \mathbb{N}$. Hence

$$\limsup_{n \to \infty} \int_X f_n \, d\mu \leq \int_X f \, d\mu.$$

To see the reverse inequality, let $\varphi : X \to [0, \infty)$ be an arbitrary simple function such that $\varphi \leq f$. In order to facilitate some 'wiggle room', fix an arbitrary $\alpha \in (0, 1)$. We will show that

$$\alpha \int_X f \, d\mu \leq \liminf_{n \to \infty} \int_X f_n \, d\mu$$

from which the result will follow as α was arbitrary so we may take the limit $\alpha \to 1$.

Notice $\alpha\varphi$ is a simple function such that $\alpha\varphi \leq f$. For each $n \in \mathbb{N}$, let

$$A_n = \{x \in X \mid f_n(x) - \alpha\varphi(x) \geq 0\}.$$

Since each $f_n - \alpha\varphi$ is a measurable function, A_n is measurable for all $n \in \mathbb{N}$. Furthermore, as $f_n \leq f_{n+1}$ for all $n \in \mathbb{N}$, clearly $A_n \subseteq A_{n+1}$ for all $n \in \mathbb{N}$. Notice by Theorem 4.2.9 that for all $n \in \mathbb{N}$

$$\alpha \int_{A_n} \varphi \, d\mu = \int_{A_n} \alpha\varphi \, d\mu \leq \int_{A_n} f_n \, d\mu \leq \int_X f_n \, d\mu \leq \liminf_{k \to \infty} \int_X f_n \, d\mu$$

with the first inequality following as $\alpha\varphi\chi_{A_n} \leq f_n\chi_{A_n}$, the second inequality following as $A_n \subseteq X$, and the third inequality following as $f_k \leq f_{k+1}$ for all $k \in \mathbb{N}$. Thus if $\alpha \int_{A_n} \varphi \, d\mu$ could be replaced with $\alpha \int_X \varphi \, d\mu$ the proof will be complete.

We claim that

$$X = \bigcup_{n \geq 1} A_n.$$

To see this, let $x \in X$ be arbitrary. If $f(x) = 0$ then $f_n \leq f$ and $\varphi \leq f$ implies that $f_n(x) = 0 = \alpha\varphi(x)$ and thus $x \in A_n$ for all $n \in \mathbb{N}$. Otherwise, if $f(x) > 0$, then we notice $\varphi \leq f$ implies that $f(x) > \alpha\varphi(x)$ as $\alpha < 1$. Hence, as $\lim_{n\to\infty} f_n(x) = f(x)$, there exists an $N \in \mathbb{N}$ such that $f(x) \geq f_N(x) > \alpha\varphi(x)$ and thus $x \in A_N$. Hence $X = \bigcup_{n\geq 1} A_n$.

Recall from Remark 4.2.2 that if we define $\nu : X \to [0, \infty]$ by

$$\nu(A) = \int_A \varphi \, d\mu$$

then ν is a measure as φ is a simple function. Therefore, as $\{A_n\}_{n=1}^\infty$ is an increasing sequence of measurable sets with $X = \bigcup_{n\geq 1} A_n$, the Monotone Convergence Theorem for measures (Theorem 2.1.18) implies that

$$\int_X \varphi \, d\mu = \nu(X) = \lim_{n\to\infty} \nu(A_n) = \lim_{n\to\infty} \int_{A_n} \varphi \, d\mu.$$

Hence as

$$\alpha \int_{A_n} \varphi \, d\mu \leq \liminf_{k\to\infty} \int_X f_n \, d\mu$$

for all $n \in \mathbb{N}$, we obtain by taking a limit over n that

$$\alpha \int_X \varphi \, d\mu \leq \liminf_{k\to\infty} \int_X f_n \, d\mu.$$

Therefore, as $\alpha \in (0, 1)$ was arbitrary, we obtain that

$$\int_X \varphi \, d\mu \leq \liminf_{k\to\infty} \int_X f_n \, d\mu.$$

Hence, as φ was an arbitrary simple function less than f, we obtain by definition that

$$\int_X f \, d\mu \leq \liminf_{k\to\infty} \int_X f_n \, d\mu.$$

Hence the result is complete. ∎

Remark 4.3.2. Finally, note that the Riemann integral is not defined for unbounded functions and Example 4.1.11 demonstrated that the Monotone Convergence Theorem 4.3.1 fails for the Riemann integral. In particular, our new notion of an integral is far better, at least for non-negative measurable functions.

Using the Monotone Convergence Theorem (Theorem 4.3.1), we easily obtain the following final properties of integrals of positive functions we desire.

Theorem 4.3.3. *Let $(X, \mathcal{A}, \mu)$ be a measure space, let $A \in \mathcal{A}$, and let $f, g : X \to [0, \infty]$ be measurable functions. Then:*

1. $\int_A f + g \, d\mu = \int_A f \, d\mu + \int_A g \, d\mu$.
2. *If $f = g$ a.e., then $\int_X f \, d\mu = \int_X g \, d\mu$.*

Proof. To see (1), note by Theorem 3.2.4 there exists increasing sequences of simple functions $(\varphi_n)_{n\geq 1}$ and $(\psi_n)_{n\geq 1}$ on X that converge pointwise to f and g respectively such that $\varphi_n \leq f$ and $\psi_n \leq g$ for all $n \in \mathbb{N}$. Therefore $(\varphi_n + \psi_n)_{n\geq 1}$ is an increasing sequence of simple functions that converges to $f + g$ pointwise such that $\varphi_n + \psi_n \leq f + g$. Therefore, by applying the Monotone Convergence Theorem (Theorem 4.3.1) and the additivity of integrals of simple functions from Theorem 4.2.5, we obtain that

$$\begin{aligned}\int_A f + g \, d\mu &= \lim_{n\to\infty} \int_A \varphi_n + \psi_n \, d\mu \\ &= \lim_{n\to\infty} \int_A \varphi_n \, d\mu + \int_A \psi_n \, d\mu \\ &= \int_A f \, d\mu + \int_A g \, d\mu\end{aligned}$$

thereby completing (1).

To see (2), let $B \in \mathcal{A}$ be such that $f(x) = g(x)$ for all $x \in B$ and $\mu(B^c) = 0$. Thus $f\chi_B = g\chi_B$. Since $\mu(B^c) = 0$, Theorem 4.2.5 implies that

$$\int_{B^c} f \, d\mu = \int_{B^c} g \, d\mu = 0$$

Hence we see that

$$\begin{aligned}\int_X f \, d\mu &= \int_X f\chi_B \, d\mu + \int_X f\chi_{B^c} \, d\mu \\ &= \int_X f\chi_B \, d\mu + \int_{B^c} f \, d\mu \\ &= \int_X g\chi_B \, d\mu + \int_{B^c} g \, d\mu \\ &= \int_X g\chi_B \, d\mu + \int_X g\chi_{B^c} \, d\mu = \int_X g \, d\mu\end{aligned}$$

as desired. ∎

Remark 4.3.4. Using part (2) of Theorem 4.3.3 and the fact that the integral of any positive measurable function against a set of measure zero is zero, the Monotone Convergence Theorem (Theorem 4.3.1) also holds if the condition that "$f : X \to [0, \infty]$ is the pointwise limit of $(f_n)_{n\geq 1}$" is replaced with the condition that "$f(x) = \lim_{n\to\infty} f_n(x)$ almost everywhere" provided we know f is measurable (which is the case when μ is complete).

The Monotone Convergence Theorem (Theorem 4.3.1) can also be used to prove several interesting properties of integrals of non-negative measurable functions.

Corollary 4.3.5. *Let $(X, \mathcal{A}, \mu)$ be a measure space. For each $n \in \mathbb{N}$ let $f_n : X \to [0, \infty]$ be a measurable function. If $f : X \to [0, \infty]$ is a measurable function such that $f(x) = \sum_{n=1}^{\infty} f_n(x)$ for almost every $x \in X$ (note f is automatically measurable if μ is complete), then for all $A \subseteq \mathcal{A}$*

$$\int_A f \, d\mu = \sum_{n=1}^{\infty} \int_A f_n \, d\mu.$$

Proof. For each $m \in \mathbb{N}$, let $g_m : X \to [0, \infty]$ be defined by $g_m = \sum_{n=1}^{m} f_n$. Clearly $(g_m)_{m\geq 1}$ is an increasing sequence of non-negative measurable functions that converges to f pointwise almost everywhere. Hence the Monotone Convergence Theorem (Theorem 4.3.1) implies

$$\int_A f \, d\mu = \lim_{m\to\infty} \int_A g_m \, d\mu = \lim_{m\to\infty} \sum_{n=1}^{m} \int_A f_n \, d\mu = \sum_{n=1}^{\infty} \int_A f_n \, d\mu$$

as desired. ∎

As Remark 4.2.2, which demonstrated that integrating against a simple function gave rise to a measure, was instrumental in the proof of the Monotone Convergence Theorem (Theorem 4.3.1), we note the following extension to integrating against non-negative measurable functions.

Corollary 4.3.6. *Let $(X, \mathcal{A}, \mu)$ be a measure space and let $f : X \to [0, \infty]$ be measurable. Define $\nu : \mathcal{A} \to [0, \infty]$ by*

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

for all $A \in \mathcal{A}$. Then ν is a measure on $(X, \mathcal{A})$. Furthermore, if $\mu(A) = 0$ for some $A \in \mathcal{A}$, then $\nu(A) = 0$.

Proof. Recall Theorem 4.2.5 that if $A \in \mathcal{A}$ and $\mu(A) = 0$, then

$$\nu(A) = \int_A f \, d\mu = 0.$$

Hence clearly $\nu(\emptyset) = 0$. Finally, to see that ν is a measure and to complete the proof, notice if $\{A_n\}_{n=1}^{\infty}$ is a collection of pairwise disjoint measurable sets in $(X, \mathcal{A})$, then

$$\begin{aligned}
\nu\left(\bigcup_{n=1}^\infty A_n\right) &= \int_{\bigcup_{n=1}^\infty A_n} f\, d\mu \\
&= \int_X f\chi_{\bigcup_{n=1}^\infty A_n}\, d\mu \\
&= \int_X \sum_{n=1}^\infty \chi_{A_n} f\, d\mu \quad \text{as } \{A_n\}_{n=1}^\infty \text{ are pairwise disjoint} \\
&= \sum_{n=1}^\infty \int_X \chi_{A_n} f\, d\mu \\
&= \sum_{n=1}^\infty \int_{A_n} f\, d\mu \\
&= \sum_{n=1}^\infty \nu(A_n)
\end{aligned}$$

by Corollary 4.3.5. Hence ν is a measure as desired. ∎

The property described in Corollary 4.3.6 is quite useful when discussing relations between two measure. Consequently, we make the following definition.

Definition 4.3.7. Let $(X, \mathcal{A})$ be a measurable space and let μ, ν be measures on $(X, \mathcal{A})$. It is said that ν is *absolutely continuous* with respect to μ is whenever $A \in \mathcal{A}$ and $\mu(A) = 0$, it must be the case that $\nu(A) = 0$.

We will not discuss absolutely continuous measures at this time (perhaps in later chapters). Of course, one interesting question is, "Does every absolutely continuous measure with respect to μ arise via Corollary 4.3.6; that is, is integration against a non-negative measurable function?"

To complete this section, we look at three remarks relating back to the Riemann integral.

Remark 4.3.8. Of course, one may ask why in Definition 4.2.7 we did define

$$\int_A f\, d\mu = \inf\left\{ \int_A \varphi\, d\mu \;\middle|\; \varphi : X \to [0, \infty) \text{ simple}, f \leq \varphi \right\}?$$

That is, in the Riemann integral we used supremums so why not in this integral? Well, if f is bounded and $\mu(X) < \infty$, then these two notions are equal!

To see this, suppose $f : X \to [0, \infty)$ is such that there exists an $M > 0$ with $f(x) \leq M$ for all $x \in X$. We first desire to reduce the number of possible simple functions in the supremum.

Suppose $\varphi : X \to [0, \infty)$ is a simple function such that $f \leq \varphi$. If $B = \varphi^{-1}((M, \infty))$, then B is a measurable set as φ is measurable. Thus if we define

$$\varphi_0 = \varphi \chi_{B^c} + M\chi_B,$$

then $\varphi_0 : X \to [0, M]$ is a simple function such that $f \leq \varphi_0 \leq \varphi$ so

$$\int_A \varphi_0 \, d\mu \leq \int_A \varphi \, d\mu.$$

Hence

$$\begin{aligned}
&\inf \left\{ \int_A \varphi \, d\mu \,\middle|\, \varphi : X \to [0, \infty) \text{ simple}, f \leq \varphi \right\} \\
&= \inf \left\{ \int_A \varphi \, d\mu \,\middle|\, \varphi : X \to [0, M] \text{ simple}, f \leq \varphi \right\}.
\end{aligned}$$

Next it is clear that that if $\varphi : X \to [0, M]$ is a simple function such that $f \leq \varphi$ then $M - \varphi : X \to [0, M]$ is also a simple function such that $M - \varphi \leq M - f$. Furthermore, we see as $M\chi_X$ is a simple function that

$$M\mu(A) = \int_A M\chi_X \, d\mu = \int_A \varphi + (M - \varphi) \, d\mu = \int_A \varphi \, d\mu + \int_A M - \varphi \, d\mu$$

so, as $M\mu(A) < \infty$, we obtain that

$$\int_A \varphi \, d\mu = M\mu(A) - \int_A M - \varphi \, d\mu.$$

Conversely, if $\psi : X \to [0, \infty)$ is a simple function such that $\psi \leq M - f$ (and thus $\psi : X \to [0, M]$), then $M - \psi : X \to [0, M]$ is a simple function such that $f \leq M - \psi$. Hence, due to these relations on simple functions, we see that

$$\begin{aligned}
&\inf \left\{ \int_A \varphi \, d\mu \,\middle|\, \varphi : X \to [0, \infty) \text{ simple}, f \leq \varphi \right\} \\
&= \inf \left\{ M\mu(A) - \int_A M - \varphi \, d\mu \,\middle|\, \varphi : X \to [0, M] \text{ simple}, f \leq \varphi \right\} \\
&= M\mu(A) - \sup \left\{ \int_A M - \varphi \, d\mu \,\middle|\, \varphi : X \to [0, M] \text{ simple}, f \leq \varphi \right\} \\
&= M\mu(A) - \sup \left\{ \int_A \psi \, d\mu \,\middle|\, \varphi : X \to [0, M] \text{ simple}, \psi \leq M - f \right\} \\
&= M\mu(A) - \int_A M - f \, d\mu.
\end{aligned}$$

However, as f and $M - f$ are non-negative measurable functions, we see from Theorem 4.3.3 that

$$M\mu(A) = \int_A M\chi_X \, d\mu = \int_A f + (M - f) \, d\mu = \int_A f \, d\mu + \int_A M - f \, d\mu$$

so, as $M\mu(A) < \infty$, we obtain that

$$M\mu(A) - \int_A M - f\,d\mu = \int_A f\,d\mu$$

thereby completing the claim.

In general, if μ is not finite or f is not bounded, the above formula need not hold. For example, consider $f : (0,1] \to (0,\infty)$ by $f(x) = \frac{1}{\sqrt{x}}$. Using Proposition 4.2.13 together with the Monotone Convergence Theorem (Theorem 4.3.1), we see that

$$\int_{(0,1]} f\,d\lambda = \lim_{n\to\infty} \int_{(0,1]} f\chi_{(\frac{1}{n},1]}\,d\lambda = \lim_{n\to\infty} \int_{\frac{1}{n}}^{1} \frac{1}{\sqrt{x}}\,dx = \lim_{n\to\infty} 2 - 2\sqrt{\frac{1}{n}} = 2.$$

However, if $\varphi : (0,1] \to [0,\infty)$ is a simple function, it is not possible for $f \leq \varphi$ (and it we allowed φ to take the value ∞, then $f \leq \varphi$ implies φ takes the value ∞ on a set of positive measure and thus $\int_{(0,1]} \varphi\,d\lambda = \infty$. Note this example also shows us why in elementary calculus the indefinite integrals are defined the way they are.

4.4 The Lebesgue Integral

As we have developed a suitable notion of integration for non-negative measurable functions, we now turn to extended this notion to complex-valued measurable functions. However, there is one caveat that needs to be dealt with. If f_1 and f_2 are non-negative measurable functions with infinity integrals, how should we define the integral of $f_1 - f_2$? This is problematic as we desire our integral to be linear and what should $\infty - \infty$ be defined to be?

To solve (by avoiding) this problem, we will remove the possibility of such functions by focusing on the following class of functions.

Definition 4.4.1. Let $(X, \mathcal{A}, \mu)$ be a measure space. A measurable function $f : X \to \mathbb{C}$ is said to be *integrable* if

$$\int_X |f|\,d\mu < \infty.$$

In the case that $(X, \mathcal{A}, \mu) = (\mathbb{R}, \mathcal{M}(\mathbb{R}), \lambda)$, it is said that f is *Lebesgue integrable*.

Before defining the integral of a integrable function, we note some important properties of integrable functions.

Remark 4.4.2. Notice if $(X, \mathcal{A}, \mu)$ and $f : X \to \mathbb{C}$ is integrable, then for all $A \in \mathcal{A}$

$$\int_A |f|\,d\mu = \int_X |f\chi_A|\,d\mu \leq \int_X |f|\,d\mu < \infty.$$

Hence the integral of $|f|$ with respect to μ against any set is finite.

Remark 4.4.3. Let $(X, \mathcal{A}, \mu)$ be a measure space and let $f : X \to [-\infty, \infty]$ be measurable. We can again define f to be integrable if

$$\int_X |f| \, d\mu < \infty.$$

Since f is measurable, the set $B = \{x \in X \mid |f(x)| = \infty\}$ is measurable. Furthermore, if $\mu(B) > 0$, it is elementary to see by the definition of the integral that $\int_X |f| \, d\mu = \infty$, which would be a contradiction. Hence f can only be integrable if there exists a measurable function $g : X \to \mathbb{R}$ such that $f = g$ almost everywhere. Thus it suffices to consider real-valued integrable functions in lieu of extended real-valued measurable functions.

Remark 4.4.4. Let $(X, \mathcal{A}, \mu)$ be a measure space and let $f : X \to \mathbb{C}$ be a measurable function. Notice

$$f = \mathrm{Re}(f) + i\mathrm{Im}(f) = (\mathrm{Re}(f)_+ - \mathrm{Re}(f)_-) + i(\mathrm{Im}(f)_+ - \mathrm{Im}(f)_-)$$

(see Remarks 3.1.13 and 3.1.14 for definitions) where $\mathrm{Re}(f)_+$, $\mathrm{Re}(f)_-$, $\mathrm{Im}(f)_+$, and $\mathrm{Im}(f)_-$ are all measurable by Remarks 3.1.13 and 3.1.14. As

$$\mathrm{Re}(f)_+, \mathrm{Re}(f)_-, \mathrm{Im}(f)_+, \mathrm{Im}(f)_- \leq |f|,$$

clearly if f is integrable then $\mathrm{Re}(f)_+$, $\mathrm{Re}(f)_-$, $\mathrm{Im}(f)_+$, and $\mathrm{Im}(f)_-$ are integrable. Conversely, as

$$\begin{aligned} |f| = \sqrt{\mathrm{Re}(f)^2 + \mathrm{Im}(f)^2} &\leq |\mathrm{Re}(f)| + |\mathrm{Im}(f)| \\ &= \mathrm{Re}(f)_+ + \mathrm{Re}(f)_- + \mathrm{Im}(f)_+ + \mathrm{Im}(f)_-, \end{aligned}$$

we see that f is integrable if and only if $\mathrm{Re}(f)_+$, $\mathrm{Re}(f)_-$, $\mathrm{Im}(f)_+$, and $\mathrm{Im}(f)_-$ are all integrable. In particular, if $f : X \to \mathbb{R}$, then f is integrable if and only if f_+ and f_- are integrable.

Based on Remark 4.4.4, we make the following definition of the integral of an integrable function.

Definition 4.4.5. Let $(X, \mathcal{A}, \mu)$ be a measure space, let $A \in \mathcal{A}$, and let $f : X \to \mathbb{C}$ be integrable. The *integral of f over A against μ* is defined to be

$$\int_A f \, d\mu = \int_A \mathrm{Re}(f)_+ \, d\mu - \int_A \mathrm{Re}(f)_- \, d\mu + i \int_A \mathrm{Im}(f)_+ \, d\mu - i \int_A \mathrm{Im}(f)_- \, d\mu.$$

In the case that $(X, \mathcal{A}, \mu) = (\mathbb{R}, \mathcal{M}(\mathbb{R}), \lambda)$, the above is called the *Lebesgue integral of f*.

The way we constructed Definition 4.4.5, given a complex-valued measurable function f, we first check that $|f|$ has finite integral so that the integrals of $\mathrm{Re}(f)_+$, $\mathrm{Re}(f)_-$, $\mathrm{Im}(f)_+$, and $\mathrm{Im}(f)_-$ are finite, and then we define the integral of f to be the appropriate linear combination of the integrals of $\mathrm{Re}(f)_+$, $\mathrm{Re}(f)_-$, $\mathrm{Im}(f)_+$, and $\mathrm{Im}(f)_-$. Thus we have constructed our integral by demonstrating the theory of integration can be completely reduced to non-negative measurable functions!

Of course, we still need to verify that this integral is linear, which happens to be a rather annoying task technically.

Theorem 4.4.6. *Let $(X, \mathcal{A}, \mu)$ be a measure space. The set of integrable functions from X to $\mathbb{K}$ is a vector space over $\mathbb{K}$. In particular, if $f, g : X \to \mathbb{K}$ are integrable and $\alpha, \beta \in \mathbb{K}$, then*

$$\int_A \alpha f + \beta g \, d\mu = \alpha \int_A f \, d\mu + \beta \int_A g \, d\mu$$

for all $A \in \mathcal{A}$.

Proof. We will focus on the case $\mathbb{K} = \mathbb{C}$ as the case $\mathbb{K} = \mathbb{R}$ follows as a subcase.

First if $f, g : X \to \mathbb{C}$ are integrable and $\alpha, \beta \in \mathbb{C}$, then

$$\int_X |\alpha f + \beta g| \, d\mu \le \int_X |\alpha||f| + |\beta||g| \, d\mu = |\alpha| \int_X |f| \, d\mu + |\beta| \int_X |g| \, d\mu < \infty$$

as $\int_X |f| \, d\mu, \int_X |g| \, d\mu < \infty$. Hence $\alpha f + \beta g$ is integrable. Thus the set of integrable functions from X to $\mathbb{C}$ is a vector space over $\mathbb{C}$.

In order to show the linearity of the integral, it suffices to consider $A = X$ by multiplying the functions by χ_A if necessary (see Remark 4.4.2). Next, as a technical lemma, we claim that if $h_1, h_2, h_3, h_4 : X \to [0, \infty)$ are integrable functions, then

$$\int_X h_1 - h_2 + ih_3 - ih_4 \, d\mu = \int_X h_1 \, d\mu - \int_A h_2 \, d\mu + i \int_A h_3 \, d\mu - i \int_A h_4 \, d\mu.$$

To see this, let $h = h_1 - h_2 + ih_3 - ih_4$. Hence

$$h_1 - h_2 + ih_3 - ih_4 = \mathrm{Re}(h)_+ - \mathrm{Re}(h)_- + i\mathrm{Im}(h)_+ - i\mathrm{Im}(h)_-.$$

Thus

$$(h_1 + \mathrm{Re}(h)_-) + i(h_3 + \mathrm{Im}(h)_-) = (\mathrm{Re}(h)_+ + h_2) + i(\mathrm{Im}(h)_+ + h_4).$$

Therefore, by equating real and imaginary parts, we see that

$$h_1 + \mathrm{Re}(h)_- = \mathrm{Re}(h)_+ + h_2 \qquad \text{and} \qquad h_3 + \mathrm{Im}(h)_- = \mathrm{Im}(h)_+ + h_4.$$

Since $h_1, h_2, \mathrm{Re}(h)_+$, and $\mathrm{Re}(h)_-$ are non-negative measurable functions, we see that

$$\begin{aligned} \int_X h_1 \, d\mu + \int_X \mathrm{Re}(h)_- \, d\mu &= \int_X h_1 + \mathrm{Re}(h)_- \, d\mu \\ &= \int_X \mathrm{Re}(h)_+ + h_2 \, d\mu \\ &= \int_X h_2 \, d\mu + \int_X \mathrm{Re}(h)_+ \, d\mu. \end{aligned}$$

Therefore, since $h_1, h_2, \mathrm{Re}(h)_+$, and $\mathrm{Re}(h)_-$ are integrable so each integral is finite, we see that

$$\int_X h_1\,d\mu - \int_X h_2\,d\mu = \int_X \mathrm{Re}(h)_+\,d\mu - \int_X \mathrm{Re}(h)_-\,d\mu.$$

Similarly

$$\int_X h_3\,d\mu - \int_X h_4\,d\mu = \int_X \mathrm{Im}(h)_+\,d\mu - \int_X \mathrm{Im}(h)_-\,d\mu.$$

Hence

$$\begin{aligned}\int_A h\,d\mu &= \int_A \mathrm{Re}(h)_+\,d\mu - \int_A \mathrm{Re}(h)_-\,d\mu + i\int_A \mathrm{Im}(h)_+\,d\mu - i\int_A \mathrm{Im}(h)_-\,d\mu \\ &= \int_X h_1\,d\mu - \int_A h_2\,d\mu + i\int_A h_3\,d\mu - i\int_A h_4\,d\mu\end{aligned}$$

as claimed.

To proceed with the proof, for notational simplicity let

$$\begin{aligned} f_1 &= \mathrm{Re}(f)_+, \quad f_2 = \mathrm{Re}(f)_- \quad f_3 = \mathrm{Im}(f)_+ \quad f_4 = \mathrm{Im}(f)_- \\ g_1 &= \mathrm{Re}(g)_+, \quad g_2 = \mathrm{Re}(g)_- \quad g_3 = \mathrm{Im}(g)_+ \quad g_4 = \mathrm{Im}(g)_-.\end{aligned}$$

Since all f_i and g_j are positive integrable functions by Remark 4.4.4, we obtain by our above technical result that

$$\begin{aligned}\int_X f+g\,d\mu &= \int_X (f_1+g_1) - (f_2+g_2) + i(f_3+g_3) - i(f_4+g_4)\,d\mu \\ &= \int_X f_1\,d\mu + \int_X g_1\,d\mu - \int_X f_2\,d\mu - \int_X g_2\,d\mu \\ &\quad + i\int_X f_3\,d\mu + i\int_X g_3\,d\mu - i\int_X f_4\,d\mu - i\int_X fg_4\,d\mu \\ &= \int_X f\,d\mu + \int_X g\,d\mu.\end{aligned}$$

Hence the integral is additive. Thus, due to the properties of linear maps, it suffices to prove that the integral preserves scalar multiplication.

To begin, let $a \in \mathbb{R}$ be arbitrary. If $a \geq 0$, then af_1, af_2, af_3 and af_4 are positive integral functions

$$\begin{aligned}\int_X af\,d\mu &= \int_X (af_1) - (af_2) + i(af_3) - i(af_4)\,d\mu \\ &= \int_X af_1\,d\mu - \int_X af_2\,d\mu + i\int_X af_3\,d\mu - i\int_X af_4\,d\mu \\ &= a\int_X f_1\,d\mu - a\int_X f_2\,d\mu + ai\int_X f_3\,d\mu - ai\int_X f_4\,d\mu \\ &= a\int_X f\,d\mu\end{aligned}$$

and if $a < 0$, then $-a > 0$ and $(-a)f_1, (-a)f_2, (-a)f_3$ and $(-a)f_4$ are positive integral functions so

$$\begin{aligned}\int_X af\,d\mu &= \int_X ((-a)f_2) - ((-a)f_1) + i((-a)f_4) - i((-a)f_3)\,d\mu \\ &= \int_X (-a)f_2\,d\mu - \int_X (-a)f_1\,d\mu + i\int_X (-a)f_4\,d\mu - i\int_X (-a)f_3\,d\mu \\ &= (-a)\int_X f_2\,d\mu - (-a)\int_X f_1\,d\mu + (-a)i\int_X f_4\,d\mu - (-a)i\int_X f_2\,d\mu \\ &= a\int_X f\,d\mu.\end{aligned}$$

Furthermore, as f_1, f_2, f_3, and f_4 are positive integrable functions

$$\begin{aligned}\int_X if\,d\mu &= \int_X f_4 - f_3 + if_1 - if_2\,d\mu \\ &= \int_X f_4\,d\mu - \int_X f_3\,d\mu + i\int_X f_1\,d\mu - i\int_X f_2\,d\mu \\ &= i\left(-i\int_X f_4\,d\mu\right) + i\left(i\int_X f_3\,d\mu\right) + i\int_X f_1\,d\mu - i\int_X f_2\,d\mu \\ &= i\int_X f\,d\mu.\end{aligned}$$

Combining all of the above, we see that if $\alpha = a + bi$ where $a, b \in \mathbb{R}$, then

$$\begin{aligned}\int_X (a+bi)f\,d\mu &= \int_X af + b(if)\,d\mu \\ &= \int_X af\,d\mu + \int_X b(if)\,d\mu \\ &= a\int_X f\,d\mu + b\int_X if\,d\mu \\ &= a\int_X f\,d\mu + bi\int_X f\,d\mu = \alpha\int_X f\,d\mu.\end{aligned}$$

Hence the result follows. ∎

Remark 4.4.7. It is now not difficult to see that if $f : [a, b] \to \mathbb{R}$ is bounded and Riemann integrable then $\int_a^b f(x)\,dx = \int_{[a,b]} f\,d\lambda$. Indeed, as f is Riemann integrable, $|f|$ is Riemann integrable and thus f_+ and f_- are Riemann integrable. By Proposition 4.2.13, this implies that f_+ and f_- are Lebesgue integrable. Hence

$$\begin{aligned}\int_a^b f(x)\,dx &= \int_a^b f_+(x)\,dx - \int_a^b f_-(x)\,dx \\ &= \int_{[a,b]} f_+\,d\lambda - \int_{[a,b]} f_-\,d\lambda \\ &= \int_{[a,b]} f\,d\lambda.\end{aligned}$$

by Proposition 4.2.13 and Theorem 4.4.6. Thus the Lebesgue integral is truly an extension of the Riemann integral (and works for complex functions too!).

Remark 4.4.8. If $(X, \mathcal{A}, \mu)$ is a measure space and $f : X \to \mathbb{C}$ is integrable, then clearly the function $\overline{f} : X \to \mathbb{C}$ defined by $\overline{f}(x) = \overline{f(x)}$ is integrable as

$$\mathrm{Re}(\overline{f}) = \mathrm{Re}(f), \quad \mathrm{Im}(\overline{f})_+ = \mathrm{Im}(f)_-, \quad \text{and} \quad \mathrm{Im}(\overline{f})_- = \mathrm{Im}(f)_+.$$

Furthermore, from this we easily obtain that

$$\int_X \overline{f}\, d\mu = \overline{\int_X f\, d\mu}.$$

Thus our notion of integration plays well with respect to complex conjugation.

As we now have our integral of complex-valued integrable functions, we begin our study of the uses of this integral. To begin, we note two simple yet essential results, the first of which is reminiscent of a property of the Riemann integral.

Theorem 4.4.9. *Let $(X, \mathcal{A}, \mu)$ be a measure space. If $f : X \to \mathbb{C}$ is integrable, then*

$$\left|\int_A f\, d\mu\right| \leq \int_A |f|\, d\mu$$

for all $A \in \mathcal{A}$.

Proof. By properties of complex numbers, there exists a $z \in \mathbb{C}$ such that $|z| = 1$ and

$$z \int_A f\, d\mu = \left|\int_A f\, d\mu\right| \geq 0$$

(i.e. rotate the complex number $\int_A f\, d\mu$ until it is positive). Hence zf is integrable and

$$0 \leq \left|\int_A f\, d\mu\right| = \int_A zf\, d\mu = \int_A \mathrm{Re}(zf)\, d\mu + i \int_A \mathrm{Im}(zf)\, d\mu.$$

However, since $\int_A \mathrm{Re}(zf)\, d\mu, \int_A \mathrm{Im}(zf)\, d\mu \in \mathbb{R}$, it must be the case that $\int_A \mathrm{Im}(zf)\, d\mu = 0$. Hence

$$\begin{aligned}
\left|\int_A f\, d\mu\right| &= \int_A \mathrm{Re}(zf)\, d\mu \\
&= \int_A \mathrm{Re}(zf)_+\, d\mu - \int_A \mathrm{Re}(zf)_-\, d\mu \\
&\leq \int_A \mathrm{Re}(zf)_+\, d\mu + \int_A \mathrm{Re}(zf)_-\, d\mu \\
&= \int_A \mathrm{Re}(zf)_+ + \mathrm{Re}(zf)_-\, d\mu \\
&= \int_A |\mathrm{Re}(zf)|\, d\mu \\
&\leq \int_A |zf|\, d\mu = \int_A |f|\, d\mu
\end{aligned}$$

as desired. ∎

Theorem 4.4.10. *Let $(X, \mathcal{A}, \mu)$ be a measure space, let $f : X \to \mathbb{C}$ be integrable, and let $g : X \to \mathbb{C}$ be measurable. If $f = g$ a.e., then g is integrable and $\int_X f\, d\mu = \int_X g\, d\mu$.*

Proof. Since $f = g$ a.e., it is easy to see that

$$\mathrm{Re}(f)_+ = \mathrm{Re}(g)_+, \qquad \mathrm{Re}(f)_- = \mathrm{Re}(g)_-,$$
$$\mathrm{Im}(f)_+ = \mathrm{Im}(f)_+, \qquad \mathrm{Im}(f)_- = \mathrm{Im}(g)_-$$

almost everywhere. Hence g is integrable. Thus we trivially obtain that $\int_X f\, d\mu = \int_X g\, d\mu$. ∎

For our final result pertaining to integrals, we note Proposition 2.3.8 implies the Lebesgue integral behaves well with respect to the properties of $\mathbb{R}$.

Proposition 4.4.11 (Translation Invariance). *Let $f : \mathbb{R} \to \mathbb{C}$ be Lebesgue integrable. For each $y \in \mathbb{R}$ let $f_y : \mathbb{R} \to \mathbb{C}$ be defined by $f_y(x) = f(x - y)$. Then f_y is integrable and*

$$\int_{\mathbb{R}} f_y\, d\lambda = \int_{\mathbb{R}} f\, d\lambda.$$

Proof. As every L_1-function can be written as a linear combination of four Lebesgue measurable non-negative functions, we may assume that $f : \mathbb{R} \to [0, \infty)$. Hence $f_y : \mathbb{R} \to [0, \infty)$. As $f_y^{-1}([\alpha, \infty))$ is a translation of $f^{-1}([\alpha, \infty))$ for all $\alpha \in \mathbb{R}$ and thus measurable by Proposition 2.3.8 and the fact that f is measurable, we obtain that f_y is measurable.

To see that

$$\int_{\mathbb{R}} f_y\, d\lambda = \int_{\mathbb{R}} f\, d\lambda$$

(and thus f_y is integrable), first consider $A, B \subseteq \mathbb{R}$ and $y \in \mathbb{R}$ such that $B = y + A$. Hence Proposition 2.3.8 implies B is measurable if and only if A is measurable and

$$\lambda(B) = \lambda(A).$$

By the above, we obtain for all measurable $A \subseteq \mathbb{R}$ that

$$\int_{\mathbb{R}} (\chi_A)_y\, d\lambda = \int_{\mathbb{R}} \chi_A\, d\lambda.$$

Therefore, as the above also shows us that φ is a simple function such that $\varphi \leq f$ if and only if φ_y is a simple function with $\varphi_y \leq f_y$ and as by the linearity of the integral we know that that

$$\int_{\mathbb{R}} \varphi_y\, d\lambda = \int_{\mathbb{R}} \varphi\, d\lambda,$$

the result follows using the definition of the integral of a non-negative measurable function. ∎

By replacing Proposition 2.3.8 with Proposition 2.3.9 and repeating the above proof, we easily obtain the following.

Proposition 4.4.12 (Inversion Invariance). *Let $f : \mathbb{R} \to \mathbb{C}$ be Lebesgue integrable. Let $f : \mathbb{R} \to \mathbb{C}$ be defined by $\check{f}(x) = f(-x)$. Then $\check{f}$ is integrable and*

$$\int_{\mathbb{R}} \check{f}\, d\lambda = \int_{\mathbb{R}} f\, d\lambda.$$

4.5 Fatou's Lemma

To complete this chapter, we desire to discuss two limit theorems for integrals. The first is another limit theorem for non-negative measurable functions.

Theorem 4.5.1 (Fatou's Lemma). *Let $(X, \mathcal{A}, \mu)$ be a measure space. For each $n \in \mathbb{N}$ let $f_n : X \to [0, \infty]$ be a measurable function. Then for each $A \in \mathcal{A}$*

$$\int_A \liminf_{n\to\infty} f_n\, d\mu \leq \liminf_{n\to\infty} \int_A f_n\, d\mu.$$

Proof. For each $k \in \mathbb{N}$, define $g_k : X \to [0, \infty]$ by

$$g_k(x) = \inf\{f_n(x) \mid n \geq k\}$$

for all $x \in X$. By Proposition 3.1.17 each g_k is a measurable function. Furthermore, for all $k \in \mathbb{N}$ and for all $n \geq k$ we see that $g_k \leq f_n$ so

$$\int_A g_k\, d\mu \leq \int_A f_n\, d\mu$$

for all $n \geq k$ and thus

$$\int_A g_k\, d\mu \leq \liminf_{n\to\infty} \int_A f_n\, d\mu$$

for all $k \in \mathbb{N}$. However, it is elementary to see that $(g_k)_{k\geq 1}$ is an increasing sequence of measurable functions that converges to $\liminf_{n\to\infty} f_n$ pointwise. Thus the Monotone Convergence Theorem (Theorem 4.3.1) implies that

$$\int_A \liminf_{n\to\infty} f_n\, d\mu = \lim_{n\to\infty} \int_A g_k\, d\mu \leq \liminf_{n\to\infty} \int_A f_n\, d\mu$$

as desired. ∎

Remark 4.5.2. It is not difficult to see that the inequality in Fatou's Lemma (Theorem 4.5.1) may be strict. Indeed if $f_n = \frac{1}{n}\chi_{[0,n]}$ for all $n \in \mathbb{N}$, it is easy to see that $\int_{\mathbb{R}} f_n\, d\lambda = 1$ for all $n \in \mathbb{N}$ yet $(f_n)_{n\geq 1}$ converges to zero pointwise almost everywhere so $\int_{\mathbb{R}} \liminf_{n\to\infty} f_n\, d\lambda = 0$.

4.6 The Dominated Convergence Theorem

Finally, we arrive at the most powerful notion of limit theorem for integrals of arbitrary integrable functions.

Theorem 4.6.1 (Dominated Convergence Theorem). *Let $(X, \mathcal{A}, \mu)$ be a measure space and let $g : X \to [0, \infty)$ be an integrable function. For each $n \in \mathbb{N}$ let $f_n : X \to \mathbb{C}$ be a measurable function such that $|f_n| \leq g$ almost everywhere. If $f : X \to \mathbb{C}$ is such that $(f_n)_{n\geq 1}$ converges to f pointwise almost everywhere and f is measurable (i.e. if μ is complete), then f is integrable with*

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

for all $A \in \mathcal{A}$.

Proof. Since for each $n \in \mathbb{N}$ we have $|f_n| \leq g$ almost everywhere and since $(f_n)_{n\geq 1}$ converges to f pointwise almost everywhere, we see that $|f| \leq g$. Hence, as g is integrable and as f and each f_n is measurable, f and each f_n is integrable by Theorem 4.2.9. Furthermore, as $|f - f_n|$ is measurable and as $|f - f_n| \leq |f| + |f_n| \leq 2g$, we also see that $|f - f_n|$ is integrable for all $n \in \mathbb{N}$.

Notice that for each $n \in \mathbb{N}$ that $2g - |f - f_n| \geq 0$ and that $(2g - |f - f_n|)_{n\geq 1}$ converges to $2g$ pointwise almost everywhere. Therefore Fatou's Lemma (Theorem 4.5.1) implies that

$$\begin{aligned}
\int_A 2g \, d\mu &= \int_A \liminf_{n\to\infty} 2g - |f - f_n| \, d\mu \\
&\leq \liminf_{n\to\infty} \int_A 2g - |f - f_n| \, d\mu \\
&= \liminf_{n\to\infty} \int_A 2g \, d\mu - \int_A |f - f_n| \, d\mu \\
&= \int_A 2g \, d\mu - \limsup_{n\to\infty} \int_A |f - f_n| \, d\mu.
\end{aligned}$$

Hence

$$\limsup_{n\to\infty} \int_A |f - f_n| \, d\mu = 0.$$

Therefore, by Theorem 4.4.9, we see that

$$\limsup_{n\to\infty} \left| \int_A f - f_n \, d\mu \right| \leq \limsup_{n\to\infty} \int_A |f - f_n| \, d\mu = 0$$

Hence

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

so the result follows. ∎

Remark 4.6.2. Notice that the proof of the Dominated Convergence Theorem (Theorem 4.6.1) actually produced that

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

This stronger claim will be made of use later.

Remark 4.6.3. The necessity of the existence of an integrable function $g : X \to [0, \infty)$ such that $|f_n| \leq g$ in the Dominated Convergence Theorem (Theorem 4.6.1) can be seen via the same example as in Remark 4.5.2.

To conclude, we note a result similar to part of the proof of Corollary 4.3.6 extends.

Corollary 4.6.4. *Let $(X, \mathcal{A}, \mu)$ be a measure space and let $f : X \to \mathbb{C}$ be an integrable function. If $\{A_n\}_{n=1}^{n} \subseteq \mathcal{A}$ are pairwise disjoint with $A = \bigcup_{n=1}^{\infty} A_n$, then*

$$\int_A f \, d\mu = \sum_{n=1}^{\infty} \int_{A_n} f \, d\mu.$$

Proof. Let $\{A_n\}_{n=1}^{\infty}$ be a collection of pairwise disjoint measurable sets in $(X, \mathcal{A})$ with $A = \bigcup_{n=1}^{\infty} A_n$. For each $m \in \mathbb{N}$ let $g_m = f\chi_{\bigcup_{n=1}^{m} A_n}$. Clearly g is measurable since f is measurable and $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{A}$. Furthermore, as $|g_m| \leq |f|$, f is integrable, and $(g_m)_{m\geq 1}$ converge to $f\chi_A$ pointwise, we obtain by the Dominated Convergence Theorem (Theorem 4.6.1) that

$$\begin{aligned}
\int_A f \, d\mu &= \int_X f\chi_A \, d\mu \\
&= \lim_{m\to\infty} \int_X g_m \, d\mu \\
&= \lim_{m\to\infty} \int_X f\chi_{\bigcup_{n=1}^{m} A_n} \, d\mu \\
&= \lim_{m\to\infty} \int_X \sum_{n=1}^{m} f\chi_{A_n} \, d\mu \\
&= \lim_{m\to\infty} \sum_{n=1}^{m} \int_X f\chi_{A_n} \, d\mu \\
&= \lim_{m\to\infty} \sum_{n=1}^{m} \int_{A_n} f \, d\mu \\
&= \sum_{n=1}^{\infty} \int_{A_n} f \, d\mu
\end{aligned}$$

as desired. ∎

Chapter 5

Differentiation and Integration

With our construction of integrals with respect to measures complete, we can turn our attention to studying the relation between our integral to other objects. As the relationship between integration and differentiation is the centrepiece of any undergraduate calculus course, it makes sense we analyze whether we have similar results in our theory. Thus this section is devoted to understanding the relationship between the Lebesgue integral and differentiation of measurable functions.

After a technical lemma pertaining to covering subsets of $\mathbb{R}$ with intervals of small size, we will demonstrate that every monotone function is differentiable λ-almost everywhere and obtain a bound for the integral of the derivative. We then turn our attention to seeing if there is a version of the Fundamental Theorems of Calculus for the Lebesgue integral. In particular, if $f \in L_1([a, b], \lambda)$, what can we say about the function $F : [a, b] \to \mathbb{R}$ defined by

$$F(x) = \int_{[a,x]} f \, d\lambda?$$

5.1 Vitali Coverings

Clearly given a subset of $\mathbb{R}$ there are many ways to cover the set with intervals of arbitrary small size. These coverings have many important properties, especially if we are dealing with open intervals covering a compact subsets for which a finite subcover can be chosen. However, as we are often dealing with measurable sets instead of compact subsets, it is useful to to study collections of intervals and how they behave with respect to the Lebesgue measure. Consequently, we will focus on the following collections of intervals in this section.

Definition 5.1.1. A collection $\mathcal{I}$ of intervals of $\mathbb{R}$ containing no singleton points is said to be a *Vitali covering* of a set $X \subseteq \mathbb{R}$ if for all $\delta > 0$ and $x \in X$ there exists an $I \in \mathcal{I}$ such that $x \in I$ and $\lambda(I) < \delta$.

Example 5.1.2. Clearly the set of all open intervals of $\mathbb{R}$ is a Vitali covering of $\mathbb{R}$ whereas the set of all intervals with length at least 1 is not a Vitali covering of $\mathbb{R}$.

The important property of Vitali coverings is the following reminiscent of the ability to choose a finite subcover of any open cover of a compact set.

Theorem 5.1.3 (Vitali Covering Lemma). *Let $X \subseteq \mathbb{R}$ be such that $\lambda^*(X) < \infty$. If $\mathcal{I}$ is a Vitali covering of X then for all $\epsilon > 0$ there exists a finite, pairwise disjoint collection $\{I_k\}_{k=1}^n \subseteq \mathcal{I}$ such that*

$$\lambda^*\left(X \setminus \bigcup_{k=1}^n I_k\right) < \epsilon.$$

Proof. First we will demonstrate we can assume $\mathcal{I}$ has some specific properties. To begin, first note that as $\lambda^*(X) < \infty$ there exists an open subset $U \subseteq \mathbb{R}$ such that $X \subseteq U$ and $\lambda(U) < \infty$ by the definition of the Lebesgue outer measure.

We claim that

$$\mathcal{J} = \left\{\overline{I} \mid I \in \mathcal{I}, \overline{I} \subseteq U\right\}$$

is a Vitali covering of X. To see this, first notice that $\mathcal{J}$ consists of intervals of $\mathbb{R}$ that are not singletons. Next let $\delta > 0$ and $x \in X$ be arbitrary. As $x \in X \subseteq U$ there exists an $\epsilon_x > 0$ such that $(x - \epsilon_x, x + \epsilon_x) \subseteq U$. However, as $x \in X$ and $\mathcal{I}$ is a Vitali covering of X, there exists an $I \in \mathcal{I}$ such that $x \in I$ and

$$\lambda(I) < \min\left\{\frac{1}{2}\delta, \frac{1}{2}\epsilon_x\right\}.$$

Since $x \in I$ and $\lambda(I) < \frac{1}{2}\epsilon_x$, one easily sees that

$$I \subseteq \left(x - \frac{1}{2}\epsilon_x, x + \frac{1}{2}\epsilon_x\right) \subseteq U.$$

Therefore $\overline{I} \subseteq (x - \epsilon_x, x + \epsilon_x) \subseteq U$ so $\overline{I} \in \mathcal{J}$. Hence $\overline{I} \in \mathcal{J}$, $x \in \overline{I}$, and $\lambda(\overline{I}) < \delta$. Hence, as $\delta > 0$ and $x \in X$ were arbitrary, $\mathcal{J}$ is a Vitali covering of X.

We claim it suffices to prove the result for $\mathcal{J}$ in place of $\mathcal{I}$. Indeed suppose given an $\epsilon > 0$ there exists a finite, pairwise disjoint collection $\{J_k\}_{k=1}^n \subseteq \mathcal{J}$ such that

$$\lambda^*\left(X \setminus \bigcup_{k=1}^n J_k\right) < \epsilon.$$

By the definition of $\mathcal{J}$ there exists a collection $\{I_k\}_{k=1}^n \subseteq \mathcal{I}$ such that $\overline{I_k} = J_k$ for all $k \in \{1, \ldots, n\}$. Therefore, as $\{J_k\}_{k=1}^n$ is pairwise disjoint and $\overline{I_k} = J_k$

for all $k \in \{1, \ldots, n\}$, clearly $\{I_k\}_{k=1}^n$ are pairwise disjoint and there exists a finite subset $Y \subseteq X$ such that

$$X \setminus \bigcup_{k=1}^n I_k = Y \cup \left(X \setminus \bigcup_{k=1}^n J_k \right).$$

Hence

$$\lambda^* \left(X \setminus \bigcup_{k=1}^n I_k \right) \leq \lambda^* \left(X \setminus \bigcup_{k=1}^n J_k \right) + \lambda(Y) < \epsilon + 0 = \epsilon$$

as desired. Therefore, it suffices to prove the result for $\mathcal{J}$ in place of $\mathcal{I}$. Note using $\mathcal{J}$ is more desirable due to the additional property that each interval in $\mathcal{J}$ is a closed interval contained in U.

Let $\epsilon > 0$ be arbitrary. Consider the following recursive process to create a pairwise disjoint collection $\{J_k\}_{k=1}^\infty \subseteq \mathcal{J}$ with certain properties. Let $J_1 \in \mathcal{J}$ be any interval (which must exist unless X is empty; a case which is trivial).

To proceed with the recursive step, suppose for some $n \in \mathbb{N}$ that $\{J_k\}_{k=1}^n \subseteq \mathcal{J}$ have been defined with certain properties. Notice if we ended up in the situation that $X \setminus \bigcup_{k=1}^n J_k = \emptyset$, then the result would be complete. Hence we assume that $X \setminus \bigcup_{k=1}^n J_k \neq \emptyset$. To construct J_{n+1} let

$$M_n = \sup\{\lambda(J) \mid J \in \mathcal{J}, J \cap J_k = \emptyset \text{ for all } k \in \{1, \ldots, n\}\}.$$

Notice since $J \subseteq U$ for all $J \in \mathcal{J}$ that $\lambda(J) \leq \lambda(U)$ for all $J \in \mathcal{J}$ so $M_n \leq \lambda(U) < \infty$.

To see that $M_n > 0$, recall that there exists an $x \in X \setminus \bigcup_{k=1}^n J_k$. As each element of $\mathcal{J}$ is closed, $\bigcup_{k=1}^n J_k$ is a closed set. Therefore, as $x \in X \setminus \bigcup_{k=1}^n J_k$,

$$\text{dist}\left(\{x\}, \bigcup_{k=1}^n J_k \right) > 0.$$

Since $\mathcal{J}$ is a Vitali covering of X, there exists a $J \in \mathcal{J}$ such that $x \in J$ and $\lambda(J) < \text{dist}\left(\{x\}, \bigcup_{k=1}^n J_k\right)$. Therefore $J \cap J_k = \emptyset$ for all $k \in \{1, \ldots, n\}$ so $M_n \geq \lambda(J) > 0$ as every element of $\mathcal{J}$ has positive length. Hence there exists a $J_{n+1} \in \mathcal{J}$ such that $J_{n+1} \cap J_k = \emptyset$ for all $k \in \{1, \ldots, n\}$ and

$$\lambda(J_{n+1}) > \frac{1}{2} M_n.$$

Using the above, we either end the proof after a finite number of steps, or we obtain a pairwise disjoint collection $\{J_k\}_{k=1}^\infty \subseteq \mathcal{J}$ such that each J_k is a closed interval contained in U such that $\lambda(J_{n+1}) > \frac{1}{2} M_n$ for all $n \in \mathbb{N}$. Notice

$$\sum_{k=1}^\infty \lambda(J_k) = \lambda\left(\bigcup_{k=1}^\infty J_k \right) \leq \lambda(U) < \infty.$$

Hence $\lim_{k\to\infty} \lambda(J_k) = 0$ so there exists an $N \in \mathbb{N}$ such that

$$\sum_{k=N+1}^{\infty} \lambda(J_k) < \frac{\epsilon}{5}.$$

For each $k \in \mathbb{N}$, let I_k denote the unique interval with the same midpoint as J_k and $\lambda(I_k) = 5\lambda(J_k)$. We claim that

$$X \setminus \bigcup_{k=1}^{N} J_k \subseteq \bigcup_{k=N+1}^{\infty} I_k.$$

To see this, let $x \in X \setminus \bigcup_{k=1}^{N} J_k$ be arbitrary. As $\mathcal{J}$ is a Vitali covering of X and as $\bigcup_{k=1}^{N} J_k$ is a closed set disjoint from $\{x\}$, the above demonstrates there exists a $J_x \in \mathcal{J}$ such that $x \in J_x$ and $J_x \cap J_k = \emptyset$ for all $k \in \{1, \ldots, N\}$. If $J_x \cap J_k = \emptyset$ for all $k \in \{1, \ldots, n\}$ for some $n \geq N$, then the definition of M_n implies that

$$0 < \lambda(J_x) \leq M_n < 2\lambda(J_{n+1}).$$

However, as $\lim_{n\to\infty} \lambda(J_n) = 0$, it must be the case that there exists an $n > N$ such that $J_x \cap J_n \neq \emptyset$. Let n_x be the least natural number such that $J_x \cap J_{n_x} \neq \emptyset$. Hence $n_x > N$. Since $J_x \cap J_k = \emptyset$ for all $k \in \{1, \ldots, n_x - 1\}$, the above computation shows that

$$0 < \lambda(J_x) \leq M_{n_x - 1} < 2\lambda(J_{n_x}).$$

Furthermore, as $x \in J_x$ and $J_x \cap J_{n_x} \neq \emptyset$, we see that the distance between x and the midpoint of J_{n_x} is at most

$$\lambda(J_x) + \frac{1}{2}\lambda(J_{n_x}) \leq 2\lambda(J_{n_x}) + \frac{1}{2}\lambda(J_{n_x}) = \frac{5}{2}\lambda(J_{n_x}).$$

Hence $x \in I_{n_x} \subseteq \bigcup_{k=N+1}^{\infty} I_k$ by the definition of I_{n_x}. As $x \in X \setminus \bigcup_{k=1}^{N} J_k$ was arbitrary, the claim follows.

Combining the above, we see that

$$\begin{aligned} \lambda^*\left(X \setminus \bigcup_{k=1}^{n} J_k\right) &\leq \lambda\left(\bigcup_{k=N+1}^{\infty} I_k\right) \\ &\leq \sum_{k=N+1}^{\infty} \lambda(I_k) \\ &\leq 5 \sum_{k=N+1}^{\infty} \lambda(J_k) < \epsilon \end{aligned}$$

as desired. ∎

5.2 The Lebesgue Differentiation Theorem

With the proof of the Vitali Covering Lemma (Theorem 5.1.3) out of the way, we can turn our attention to the first use of the lemma in demonstrating that monotone measurable functions are differentiable and obtaining a bound on their integrals. To begin, we set some notation and redefine the derivative of a function.

Definition 5.2.1. Let $f : \mathbb{R} \to \mathbb{R}$. For each $x \in \mathbb{R}$ define

$$\begin{aligned} D^+f(x) &= \limsup_{h\to 0^+} \frac{f(x+h)-f(x)}{h} \\ D_+f(x) &= \liminf_{h\to 0^+} \frac{f(x+h)-f(x)}{h} \\ D^-f(x) &= \limsup_{h\to 0^-} \frac{f(x+h)-f(x)}{h} \\ D_-f(x) &= \liminf_{h\to 0^-} \frac{f(x+h)-f(x)}{h} \end{aligned}$$

and note that $D_+f(x) \leq D^+f(x)$ and $D_-f(x) \leq D^-f(x)$. It is said that f is *differentiable* at x if

$$D^+f(x) = D_+f(x) = D^-f(x) = D_-f(x) \in \mathbb{R}.$$

If f is differentiable at x, then the *derivative of f at x*, denoted $f'(x)$, is $f'(x) = D^+f(x) = D_+f(x) = D^-f(x) = D_-f(x)$.

The main goal of this section is to prove the following.

Theorem 5.2.2 (Lebesgue Differentiation Theorem). *If $f : [a,b] \to \mathbb{R}$ is a non-decreasing function, then f is differentiable λ-almost everywhere, f' is Lebesgue measurable, $f' \geq 0$ λ-almost everywhere, and*

$$\int_{[a,b]} f' \, d\lambda \leq f(b) - f(a).$$

Proof. First, for notational simplicity, if $x < a$ define $f(x) = f(a)$ and if $x > b$ defined $f(x) = f(b)$. Moreover, under these assumptions, it is easy to see for all $c \in \mathbb{R}$ that $f^{-1}([c,\infty))$ is of the form (y,∞) or $[y,\infty)$ for some $y \in \mathbb{R} \cup \{\pm\infty\}$ as f is non-decreasing. Hence f is measurable.

To see that f is differentiable almost everywhere, we desire to show that for all $s,t \in \{+,-\}$ that

$$\begin{aligned} &\{x \in [a,b] \mid D^sf(x) \neq D^tf(x)\} \\ &\{x \in [a,b] \mid D^sf(x) \neq D_tf(x)\} \\ &\{x \in [a,b] \mid D_sf(x) \neq D_tf(x)\} \end{aligned}$$

are Lebesgue measurable with Lebesgue measure zero. Here we will only demonstrate that

$$X = \{x \in [a, b] \mid D^+ f(x) > D_+ f(x)\}$$

is Lebesgue measurable with Lebesgue measure zero as the proofs of the remainders will be similar (and have next to no changes in that which follows).

For each $p, q \in \mathbb{R}$ let

$$E_{p,q} = \{x \in [a, b] \mid D^+ f(x) > p > q > D_+ f(x)\}.$$

Clearly

$$X = \bigcup_{p,q\in\mathbb{Q}} E_{p,q}.$$

Therefore, if it can be demonstrated that $\lambda^*(E_{p,q}) = 0$ for all $p, q \in \mathbb{Q}$, then $\lambda^*(X) = 0$ as $\mathbb{Q}$ is countable and thus X is measurable as the Lebesgue measure is complete by Remark 2.2.8.

Fix $p, q \in \mathbb{Q}$ with $p > q$. Let $r = \lambda^*(E_{p,q}) \leq \lambda^*([a, b]) < \infty$ and let $\epsilon > 0$ be arbitrary. By the definition of the Lebesgue measure, there exists an open subset $U \subseteq \mathbb{R}$ such that $E_{p,q} \subseteq U$ and

$$\lambda(U) \leq \lambda^*(E_{p,q}) + \epsilon = r + \epsilon.$$

Notice if $x \in E_{p,q}$ then $D_+ f(x) < q$ so

$$\sup_{\delta>0} \inf_{0<h<\delta} \frac{f(x + h) - f(x)}{h} = \liminf_{h\to 0^+} \frac{f(x + h) - f(x)}{h} < q.$$

Hence, for each $x \in E_{p,q}$ and $\delta > 0$ there exists an interval of the form $[x, x + h)$ such that $[x, x + h) \subseteq U$, $h < \delta$, and $f(x + h) - f(x) < qh$. As the collection of such intervals forms a Vitali covering of $E_{p,q}$, the Vitali Covering Lemma (Theorem 5.1.3) implies there exists an $n \in \mathbb{N}$, $x_1, \ldots, x_n \in E_{p,q}$, and $h_1, \ldots, h_n > 0$ such that if $I_k = (x_k, x_k + h_k)$ for all $k \in \{1, \ldots, n\}$, then $\{I_k\}_{k=1}^n$ are pairwise disjoint subsets of U such that $f(x_k + h_k) - f(x_k) < qh_k$ for all $k \in \{1, \ldots, n\}$ and

$$\lambda^*\left(E_{p,q} \setminus \bigcup_{k=1}^n I_k\right) < \epsilon.$$

Notice this implies

$$\begin{aligned}\sum_{k=1}^{n} f(x_k + h_k) - f(x_k) &< q \sum_{k=1}^{n} h_k \\ &= q \sum_{k=1}^{n} \lambda(I_k) \\ &= q\lambda\left(\bigcup_{k=1}^{n} I_k\right) \\ &\leq q\lambda(U) \leq q(r+\epsilon).\end{aligned}$$

Let

$$A = E_{p,q} \cap \left(\bigcup_{k=1}^{n} I_k\right) \subseteq E_{p,q}.$$

Thus

$$E_{p,q} = A \cup \left(E_{p,q} \setminus \bigcup_{k=1}^{n} I_k\right)$$

so $r = \lambda^*(E_{p,q}) \leq \lambda^*(A) + \epsilon$. Hence $\lambda^*(A) \geq r - \epsilon$.

Notice if $x \in A$ then $D^+ f(x) > p$ so

$$\inf_{\delta > 0} \sup_{0 < h < \delta} \frac{f(x+h) - f(x)}{h} = \limsup_{h \to 0^+} \frac{f(x+h) - f(x)}{h} > p.$$

Hence, as $A \subseteq \bigcup_{k=1}^{n} I_k$ and $\{I_k\}_{k=1}^{n}$ are pairwise disjoint open intervals, for each $x \in A$ and $\delta > 0$ there exists an interval of the form $[x, x+h)$ such that $[x, x+h) \subseteq I_k$ for some k, $h < \delta$, and $f(x+h) - f(x) > ph$. As the collection of such intervals forms a Vitali covering of A, the Vitali Covering Lemma (Theorem 5.1.3) implies there exists an $m \in \mathbb{N}$, $y_1, \ldots, y_m \in A$, and $s_1, \ldots, s_m > 0$ such that if $J_k = (y_k, y_k + s_k)$ for all $k \in \{1, \ldots, m\}$, then $\{J_k\}_{k=1}^{m}$ are pairwise disjoint subsets such that each J_k is contained in a single I_j, $f(y_k + s_k) - f(y_k) > ps_k$ for all $k \in \{1, \ldots, m\}$ and

$$\lambda^*\left(A \setminus \bigcup_{k=1}^{m} J_k\right) < \epsilon.$$

Let

$$B = A \cap \left(\bigcup_{k=1}^{m} J_k\right) \subseteq \bigcup_{k=1}^{m} J_k.$$

Thus

$$A = B \cup \left(A \setminus \bigcup_{k=1}^{m} J_k\right)$$

so $\lambda^*(B) > \lambda^*(A) - \epsilon > r - 2\epsilon$. Furthermore

$$\begin{aligned}\sum_{k=1}^m f(y_k + s_k) - f(y_k) &> p \sum_{k=1}^m s_k \\ &= p \sum_{k=1}^m \lambda(J_k) \\ &= p\lambda\left(\bigcup_{k=1}^m J_k\right) \\ &\geq p\lambda^*(B) \\ &\geq p(r - 2\epsilon).\end{aligned}$$

However, as each J_k is contained in a single I_j and as f is non-decreasing, we obtain for each $j \in \{1, \ldots, n\}$ that

$$\sum_{k \text{ such that} J_k \subseteq I_j} f(y_k + s_k) - f(y_k) \leq f(x_j + h_j) - f(x_j).$$

Therefore

$$p(r - 2\epsilon) \leq \sum_{k=1}^m (y_k + s_k) - f(y_k) \leq \sum_{j=1}^n f(x_j + h_j) - f(x_j) \leq q(r + \epsilon).$$

However, as $\epsilon > 0$ was arbitrary, the above implies $pr \leq qr$. Hence, as $p > q$ and $r \geq 0$, we obtain that $r = 0$ as desired.

By the above

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

exists almost everywhere provided we allow $\pm\infty$ as limits. Note as f is non-decreasing, the limit is always non-negative and thus never $-\infty$.

For each $n \in \mathbb{N}$, let $g_n : [a, b] \to [0, \infty)$ be defined by

$$g_n(x) = n\left(f\left(x + \frac{1}{n}\right) - f(x)\right)$$

for all $x \in [a, b]$ (where $f(y) = f(b)$ for all $y > b$). Note each g_n maps into $[0, \infty)$ as f is non-decreasing. By the above and Proposition 4.4.11 $(g_n)_{n\geq 1}$ is a sequence of measurable functions that converge pointwise almost everywhere to a Lebesgue measurable function $g : [a, b] \to [0, \infty]$ (which will be f' provided $g(x) < \infty$ for almost every x). Furthermore, as $g_n : [a, b] \to [0, \infty)$ and as f bounded (being non-decreasing) and thus $f \in L_1([a, b], \lambda))$, we obtain by Fatou's Lemma (Theorem 4.5.1), Proposition 4.4.11, and Corollary

4.6.4 that

$$\begin{aligned}
\int_{[a,b]} g\, d\lambda &= \int_{[a,b]} \liminf_{n\to\infty} g_n\, d\lambda \\
&\leq \liminf_{n\to\infty} \int_{[a,b]} g_n\, d\lambda \\
&= \liminf_{n\to\infty} n \int_{[a,b]} f\left(x + \frac{1}{n}\right) - f(x)\, d\lambda(x) \\
&= \liminf_{n\to\infty} n \int_{[a+\frac{1}{n}, b+\frac{1}{n}]} f\left(x\right)\, d\lambda(x) - n \int_{[a,b]} f(x)\, d\lambda(x) \\
&= \liminf_{n\to\infty} n \int_{\left[b, b+\frac{1}{n}\right]} f(x)\, d\lambda(x) - n \int_{\left[a, a+\frac{1}{n}\right]} f(x)\, d\lambda(x) \\
&= \liminf_{n\to\infty} f(b) - n \int_{\left[a, a+\frac{1}{n}\right]} f(x)\, d\lambda(x) \\
&= f(b) - \limsup_{n\to\infty} n \int_{\left[a, a+\frac{1}{n}\right]} f(x)\, d\lambda(x) \\
&\leq f(b) - f(a)
\end{aligned}$$

as for all $n \in \mathbb{N}$

$$n \int_{\left[a, a+\frac{1}{n}\right]} f(x)\, d\lambda(x) \geq n \int_{\left[a, a+\frac{1}{n}\right]} f(a)\, d\lambda(x) = f(a).$$

Therefore, as $f(b) - f(a) < \infty$, it must be the case that $g(x) < \infty$ for almost every x. Hence f' exists almost everywhere and $f' = g$ almost everywhere. Therefore, since λ is complete and g is Lebesgue measurable, f' is Lebesgue measurable thereby completing the proof. ∎

Remark 5.2.3. Notice if $f : [a, b] \to \mathbb{R}$ is non-increasing, then $-f$ is non-decreasing and thus differentiable λ-almost everywhere with $(-f)' \geq 0$ almost everywhere. Hence f is differentiable λ-almost everywhere with $f' \leq 0$ almost everywhere.

To conclude this section, we present a vital example in Lebesgue integration theory to demonstrate that the inequality in the Lebesgue Differentiation Theorem (Theorem 5.2.2) can be strict.

Definition 5.2.4 (The Cantor Ternary Function). Given a sequence $\vec{a} = (a_n)_{n\geq 1}$ of elements of $\{0, 1, 2\}$, define

$$K_{\vec{a}} = \begin{cases} N & \text{if } a_N = 1 \text{ and } a_k \neq 1 \text{ for all } k < N \\ \infty & \text{otherwise} \end{cases}$$

and define a sequence $\vec{b}_{\vec{a}} = (b_n)_{n\geq 1}$ of elements of $\{0, 1\}$ by

$$b_n = \begin{cases} \frac{a_n}{2} & \text{if } n \leq K_{\vec{a}} \\ 1 & \text{if } n = K_{\vec{a}} \\ 0 & \text{otherwise} \end{cases}.$$

The *Cantor ternary function* is the function $f : [0,1] \to [0,1]$ defined as follow: if $x \in [0,1]$, if $x = \sum_{n=1}^{\infty} \frac{a_n}{3^n}$ for a sequence $\vec{a} = (a_n)_{n\geq 1}$ of elements of $\{0,1,2\}$, and if $\vec{b}_{\vec{a}} = (b_n)_{n\geq 1}$ is the sequence of elements of $\{0,1\}$ as defined above, then

$$f(x) = \sum_{n=1}^{\infty} \frac{b_n}{2^n}.$$

(i.e. Write a ternary expansion of x. If N is the first index where a 1 occurs, replace $\frac{1}{3^N}$ with $\frac{1}{2^N}$, change all terms of index greater than N to zero, and replace each $\frac{0}{3^n}$ with $n < N$ to $\frac{0}{2^n}$ and each $\frac{2}{3^n}$ with $n < N$ to $\frac{1}{2^n}$).

Lemma 5.2.5. *The Cantor ternary function is well-defined.*

Proof. Let f denote the Cantor ternary function. Fix $x \in [0,1]$. To show that $f(x)$ is well-defined, we must demonstrate the value of $f(x)$ does not depend on the ternary representation of x. Thus to see that $f(x)$ is well-defined we need only analyze following two cases:

(1) There exists an $m \in \mathbb{N}$ and $a_1, \ldots, a_{m-1} \in \{0,1,2\}$ such that

$$x = \sum_{k=1}^{m-1} \frac{a_k}{3^k} + \frac{0}{3^m} + \sum_{k=m+1}^{\infty} \frac{2}{3^k} = \sum_{k=1}^{m-1} \frac{a_k}{3^k} + \frac{1}{3^m} + \sum_{k=m+1}^{\infty} \frac{0}{3^k}.$$

(we do not need to include $m = 0$ here as $\sum_{k=1}^{\infty} \frac{2}{3^k}$ is the only ternary expansion of 1 we need to consider in this problem).

(2) There exists an $m \in \mathbb{N}$ and $a_1, \ldots, a_{m-1} \in \{0,1,2\}$ such that

$$x = \sum_{k=1}^{m-1} \frac{a_k}{3^k} + \frac{1}{3^m} + \sum_{k=m+1}^{\infty} \frac{2}{3^k} = \sum_{k=1}^{m-1} \frac{a_k}{3^k} + \frac{2}{3^m} + \sum_{k=m+1}^{\infty} \frac{0}{3^k}.$$

We begin with case (1). Let $\vec{a}_1$ be the sequence corresponding to the first ternary expansion of x and let $\vec{a}_2$ be the sequence corresponding to the second ternary expansion of x; that is,

$$\vec{a}_1 = (a_1, a_2, \ldots, a_{m-1}, 0, 2, 2, 2, \ldots)$$
$$\vec{a}_2 = (a_1, a_2, \ldots, a_{m-1}, 1, 0, 0, 0, \ldots).$$

If $\vec{b}_{\vec{a}_1} = (b_k)_{k\geq 1}$ and $\vec{b}_{\vec{a}_2} = (c_k)_{k\geq 1}$ are as defined as above, then it suffices to show that

$$\sum_{k=1}^{\infty} \frac{b_k}{2^k} = \sum_{k=1}^{\infty} \frac{c_k}{2^k}.$$

Notice if there exists a $n \in \{1, \ldots, m-1\}$ such that $a_n = 1$, then $b_k = c_k$ for all $k \in \mathbb{N}$ by definition (as the sequence becomes 0 after n and thus does not depend on the differences in $\vec{a}_1$ and $\vec{a}_2$) thereby completing the case. Otherwise suppose that $a_n \neq 1$ for all $n \in \{1, \ldots, m-1\}$. Hence

$$\vec{b}_{\vec{a}_1} = \left(\frac{a_1}{2}, \frac{a_2}{2}, \ldots, \frac{a_{m-1}}{2}, 0, 1, 1, 1, \ldots\right)$$
$$\vec{b}_{\vec{a}_2} = \left(\frac{a_1}{2}, \frac{a_2}{2}, \ldots, \frac{a_{m-1}}{2}, 1, 0, 0, 0, \ldots\right)$$

by definition. Hence we easily see that

$$\sum_{k=1}^{\infty} \frac{b_k}{2^k} = \sum_{k=1}^{\infty} \frac{c_k}{2^k}$$

thereby completing case (1).

For case (2), let $\vec{a}_1$ be the sequence corresponding to the first ternary expansion of x and let $\vec{a}_2$ be the sequence corresponding to the second ternary expansion of x; that is,

$$\vec{a}_1 = (a_1, a_2, \ldots, a_{m-1}, 1, 2, 2, 2, \ldots)$$
$$\vec{a}_2 = (a_1, a_2, \ldots, a_{m-1}, 2, 0, 0, 0, \ldots).$$

If $\vec{b}_{\vec{a}_1} = (b_k)_{k\geq 1}$ and $\vec{b}_{\vec{a}_2} = (c_k)_{k\geq 1}$ are as defined as above, then it suffices to show that

$$\sum_{k=1}^{\infty} \frac{b_k}{2^k} = \sum_{k=1}^{\infty} \frac{c_k}{2^k}.$$

Notice if there exists a $n \in \{1, \ldots, m-1\}$ such that $a_n = 1$, then $b_k = c_k$ for all $k \in \mathbb{N}$ by definition (as the sequence becomes 0 after n and thus does not depend on the differences in $\vec{a}_1$ and $\vec{a}_2$). Otherwise suppose that $a_n \neq 1$ for all $n \in \{1, \ldots, m-1\}$. Hence

$$\vec{b}_{\vec{a}_1} = \left(\frac{a_1}{2}, \frac{a_2}{2}, \ldots, \frac{a_{m-1}}{2}, 1, 0, 0, 0, \ldots\right)$$
$$\vec{b}_{\vec{a}_2} = \left(\frac{a_1}{2}, \frac{a_2}{2}, \ldots, \frac{a_{m-1}}{2}, 1, 0, 0, 0, \ldots\right)$$

by definition. Hence we easily see that

$$\sum_{k=1}^{\infty} \frac{b_k}{2^k} = \sum_{k=1}^{\infty} \frac{c_k}{2^k}$$

thereby completing case (2) and the proof. ∎

Lemma 5.2.6. *Let $\mathcal{C}$ denote the Cantor set and let f denote the Cantor ternary function. Then f is a non-decreasing continuous function which is constant on each interval of $\mathcal{C}^c$. Furthermore $f(C) = [0, 1]$.*

Proof. Exercise. ∎

Remark 5.2.7. As demonstrated by Lemma 5.2.6, if $f : [0,1] \to \mathbb{R}$ denotes the Cantor ternary function then f is non-decreasing on $[0,1]$ and constant on $\mathcal{C}^c$. As $\mathcal{C}^c$ is a finite union of open sets, we easily see by Definition 5.2.1 that f is differentiable at each element of $\mathcal{C}^c$ with $f'(x) = 0$ for all $x \in \mathcal{C}^c$. Therefore f is differentiable almost everywhere with $f' = 0$ almost everywhere as $\lambda(\mathcal{C}) = 0$. Thus, as

$$\int_{[0,1]} f'\, d\lambda = 0 < 1 = f(1) - f(0),$$

we see that the inequality in the Lebesgue Differentiation Theorem (Theorem 5.2.2) may be strict.

5.3 Bounded Variation

As we have seen above, the Cantor ternary function is a function that cannot be recovered from its derivative via integration as its derivative is zero almost everywhere. Therefore, if we desire to better understand the relationship between the Lebesgue integral and differentiation, we need to restrict ourselves to certain classes of functions. In particular, we are interested trying to understand functions that arise in a 'Fundamental Theorem of Calculus' manner. As functions that 'wiggle' too much are notorious for having derivatives that are not well-behaved (and probably not Lebesgue integrable), we begin by analyzing the following type of functions.

Definition 5.3.1. A function $f : [a,b] \to \mathbb{C}$ is said to be of *bounded variation* if there exists an $M \in \mathbb{R}$ such that whenever $\{x_k\}_{k=0}^n$ is a partition of $[a,b]$ (that is, $a = x_0 < x_1 < \cdots < x_n = b$) then

$$\sum_{k=1}^n |f(x_k) - f(x_{k-1})| \leq M.$$

Of course, the Cantor ternary function is of bounded variation due to the following.

Remark 5.3.2. It is elementary to see that if f is monotone then f is of bounded variation as

$$\sum_{k=1}^n |f(x_k) - f(x_{k-1})| = |f(b) - f(a)|$$

for any partition $\{x_k\}_{k=0}^n$ of $[a,b]$. Similarly if f and g are both of bounded variation, it is elementary that any linear combination of f and g is of bounded variation by the triangle inequality. Furthermore, clearly the restriction of a function f of bounded variation to a closed interval contained in the domain of f is also of bounded variation.

Remark 5.3.3. If $f : [a, b] \to \mathbb{C}$ it is clear that f is of bounded variation if and only if $\mathrm{Re}(f)$ and $\mathrm{Im}(f)$ are of bounded variation. Thus we will focus on real-valued functions of bounded variation.

Although the Cantor ternary function is perhaps a function we do not want to include by previous discussions, the collection of functions of bounded variation does contain some nice functions we wish to study.

Example 5.3.4. Suppose $f : [a, b] \to \mathbb{R}$ is differentiable on $[a, b]$ and there exist an $M \in \mathbb{N}$ such that $|f'(x)| \leq M$ for all $x \in (a, b)$. Then f is of bounded variation. Indeed suppose $\{x_k\}_{k=0}^{n}$ is a partition of $[a, b]$. Then $|f(x_k) - f(x_{k-1})| \leq M|x_k - x_{k-1}|$ by the Mean Value Theorem. Hence

$$\sum_{k=1}^{n} |f(x_k) - f(x_{k-1})| \leq \sum_{k=1}^{n} M|x_k - x_{k-1}| = M|b-a| < \infty$$

as desired.

Going back to our motivation for functions of bounded variation, if a function ‘wiggles’ too much, then the function is not of bounded variation.

Example 5.3.5. The continuous function $f : [0, 1] \to [-1, 1]$ defined by

$$f(x) = x\cos\left(\frac{\pi}{2x}\right)$$

(so $f(0) = 0$) is not of bounded variation. Indeed for each $n \in \mathbb{N}$ consider the partition $\{x_k\}_{k=0}^{2n+1}$ of $[0, 1]$ where $x_0 = 0$ and

$$x_k = \frac{1}{2n+2-k}.$$

Notice that

$$|f(x_k)| = \begin{cases} 0 & \text{if } k \text{ is odd} \\ \frac{1}{2n+2-k} & \text{if } k \text{ is even} \end{cases}$$

and thus

$$\sum_{k=0}^{2n+1} |f(x_k) - f(x_{k-1})| = 2\sum_{j=1}^{n} \frac{1}{2n+2-2j} = \sum_{j=1}^{n} \frac{1}{j}.$$

Therefore, as $\lim_{n\to\infty} \sum_{j=1}^{n} \frac{1}{j} = \infty$, it follows that f is not of bounded variation.

Of course, of interest when dealing with a function of bounded variation is the smallest constant that works in Definition 5.3.1.

Definition 5.3.6. Let $f : [a, b] \to \mathbb{R}$ be of bounded variation. The *total variation* of f, denoted $V_f(a, b)$, is

$$V_f(a,b) = \sup \left\{ \sum_{k=1}^{n} |f(x_k) - f(x_{k-1})| \,\middle|\, \substack{n\in\mathbb{N},\\ \{x_k\}_{k=1}^n \text{ a partition of } [a,b]} \right\}.$$

If $f : [a, b] \to \mathbb{R}$ is of bounded variation, then for all $x, y \in (a, b)$ such that $x < y$ the restriction of f to $[x, y]$ is of bounded variation so $V_f(x, y)$ makes sense. Using this, we are able to prove the following.

Theorem 5.3.7 (Jordan Decomposition Theorem). *Let $f : [a, b] \to \mathbb{R}$ be of bounded variation. Define $V, D : [a, b] \to \mathbb{R}$ by $V(x) = V_f(a, x)$ (with $V(a) = 0$) and $D(x) = V(x) - f(x)$ for all $x \in [a, b]$. Then V and D are non-decreasing functions such that $f = V - D$.*

In particular, by Remark 5.3.2, a function is of bounded variation if and only if it is the difference of two non-decreasing functions.

Proof. To see that V is non-decreasing, let $x, y \in [a, b]$ with $x < y$ be arbitrary. To see that $V(x) \leq V(y)$, we claim that

$$V_f(a, y) = V_f(a, x) + V_f(x, y).$$

To see this, first notice that if $\{x_k\}_{k=0}^n$ is a partition of $[a, x]$ and $\{y_k\}_{k=0}^m$ is a partition of $[x, y]$, then $\{x_k\}_{k=0}^n \cup \{y_k\}_{k=0}^m$ is a partition of $[a, y]$ (with $x_n = y_0$). As this implies

$$\sum_{k=0}^{n} |f(x_k) - f(x_{k-1})| + \sum_{k=0}^{m} |f(y_k) - f(y_{k-1})| \leq V_f(a, y)$$

and as $\{x_k\}_{k=0}^n$ and $\{y_k\}_{k=0}^m$ were arbitrary partitions of $[a, x]$ and $[x, y]$ respectively, we obtain that

$$V_f(a, x) + V_f(x, y) \leq V_f(a, y)$$

by the definition of the total variation. For the other inequality, let $\{z_k\}_{k=0}^n$ be an arbitrary partition of $[a, y]$. Then $\mathcal{P} = \{z_k\}_{k=0}^n \cup \{x\}$ is a potentially larger partition such that $\mathcal{P} \cap [a, x]$ is a partition of $[a, x]$ and $\mathcal{P} \cap [x, y]$ is a partition of $[x, y]$. Therefore, if $\mathcal{P} = \{w_k\}_{k=0}^m$ is the standard way to write $\mathcal{P}$, then, by at most one application of the triangle inequality,

$$\begin{aligned}
\sum_{k=1}^{n} |f(z_k) - f(z_{k-1})| &\leq \sum_{k=1}^{m} |f(w_k) - f(w_{k-1})| \\
&= \sum_{k \text{ such that } w_k \in [a,x]} |f(w_k) - f(w_{k-1})| \\
&\qquad + \sum_{k \text{ such that } w_{k-1} \in [x,y]} |f(w_k) - f(w_{k-1})| \\
&\leq V_f(a, x) + V_f(x, y).
\end{aligned}$$

Therefore, as $\{z_k\}_{k=0}^n$ was an arbitrary partition of $[a, y]$, the claim follows. Hence

$$V(y) - V(x) = V_f(a, y) - V_f(a, x) = V_f(x, y) \geq 0.$$

Thus V is non-decreasing as desired.

Clearly $f = V - D$ by construction. To see that D is non-decreasing, notice for all $x, y \in [a, b]$ with $x < y$ that

$$D(y) - D(x) = V(y) - V(x) - (f(y) - f(x)) = V_f(x, y) - (f(y) - f(x)) \geq 0$$

since clearly $|f(y) - f(x)| \leq V_f(x, y)$ by using the trivial partition $\{x, y\}$ in the definition of the total variation. Hence the first part of the proof is complete. ∎

Of course, combining the Lebesgue Differentiation Theorem (Theorem 5.2.2) with the Jordan Decomposition Theorem (Theorem 5.3.7) immediately implies that functions of bounded variation have derivatives with a specific integration property.

Corollary 5.3.8. *If $f : [a, b] \to \mathbb{R}$ is of bounded variation, then f is differentiable λ-almost everywhere, f' is Lebesgue measurable, and $f' \in L_1([a, b], \lambda)$.*

Proof. Since f is of bounded variation, there exists non-decreasing functions $V, D : [a, b] \to \mathbb{R}$ such that $f = V - D$ by the Jordan Decomposition Theorem (Theorem 5.3.7). As every non-decreasing function is differentiable with Lebesgue measurable derivatives by the Lebesgue Differentiation Theorem (Theorem 5.2.2), we clearly see that f is differentiable with $f' = V' - D'$ being Lebesgue measurable. As V and D are non-decreasing, we see that $V', D' \geq 0$ almost everywhere and thus $|f'| \leq V' + D'$. Hence

$$\int_{[a,b]} |f'| \, d\lambda \leq \int_{[a,b]} V' + D' \, d\lambda \leq V(b) + D(b) - V(a) - D(a) < \infty$$

by the Lebesgue Differentiation Theorem (Theorem 5.2.2). Hence $f \in L_1([a, b], \lambda)$. ∎

5.4 Absolutely Continuous Functions

Although functions of bounded variation are nice in terms of having derivatives that are Lebesgue integrable, these are not the functions we are looking for as they include the Cantor ternary function. In fact, the collection of functions we are looking are contained in the set of continuous functions, but contain functions that are not differentiable everywhere.

Definition 5.4.1. A function $f : [a, b] \to \mathbb{C}$ is said to be *absolutely continuous* if for all $\epsilon > 0$ there exists a $\delta > 0$ such that whenever $\{a_k\}_{k=1}^n, \{b_k\}_{k=1}^n \subseteq [a, b]$ are such that

$$a \leq a_1 < b_1 \leq a_2 < b_1 \leq \cdots \leq a_n < b_n \leq b \qquad \text{and} \qquad \sum_{k=1}^n |b_k - a_k| < \delta$$

then

$$\sum_{k=1}^n |f(b_k) - f(a_k)| < \epsilon.$$

Remark 5.4.2. Again, it is not difficult to see using the triangle inequality that a function $f : [a, b] \to \mathbb{C}$ is absolutely continuous if and only if $\mathrm{Re}(f)$ and $\mathrm{Im}(f)$ are absolutely continuous. Thus we will mainly focus on real-valued functions.

Of course, continuously differentiable functions are absolutely continuous.

Example 5.4.3. Suppose $f : [a, b] \to \mathbb{R}$ is differentiable on $[a, b]$ and there exist an $M \in \mathbb{N}$ such that $|f'(x)| \leq M$ for all $x \in (a, b)$. Then f is absolutely continuous. Indeed let $\epsilon > 0$ be arbitrary and let $\delta = \frac{\epsilon}{M+1}$. If $\{a_k\}_{k=1}^n, \{b_k\}_{k=1}^n \subseteq [a, b]$ are such that

$$a \leq a_1 < b_1 \leq a_2 < b_1 \leq \cdots \leq a_n < b_n \leq b \qquad \text{and} \qquad \sum_{k=1}^n |b_k - a_k| < \delta$$

then $|f(b_k) - f(a_k)| \leq M|b_k - a_k|$ for all k by the Mean Value Theorem. Hence

$$\sum_{k=1}^n |f(b_k) - f(a_k)| \leq \sum_{k=1}^n M|b_k - a_k| \leq M\delta < \epsilon.$$

Thus f is absolutely continuous.

Example 5.4.4. If $f : [0, 1] \to [0, 1]$ is the Cantor ternary function (see Definition 5.2.4) then f is uniformly continuous on $[0, 1]$ and of bounded variation, but not absolutely continuous. Indeed f is non-decreasing and continuous by Lemma 5.2.6 and thus uniformly continuous $[0, 1]$ and of bounded variation. The fact that f is not absolutely continuous follows from Proposition 5.4.7 along with the fact that f is non-constant yet $f' = 0$ almost everywhere.

Our next result says that absolutely continuous functions are of bounded variation and thus the results of the previous section apply.

Proposition 5.4.5. *Every real-valued absolutely continuous function is continuous and of bounded variation.*

Proof. Let $f : [a, b] \to \mathbb{R}$ be absolutely continuous. It easily follows from definition that f is continuous (i.e. take $n = 1$ in Definition 5.4.1).

To see that f is of bounded variation, recall as f is absolutely continuous that if $\epsilon = 1 > 0$ then there exists a $\delta > 0$ such that if $\{a_k\}_{k=1}^n, \{b_k\}_{k=1}^n \subseteq [a, b]$ are such that

$$a \leq a_1 < b_1 \leq a_2 < b_1 \leq \cdots \leq a_n < b_n \leq b \qquad \text{and} \qquad \sum_{k=1}^n |b_k - a_k| < \delta$$

then

$$\sum_{k=1}^n |f(b_k) - f(a_k)| < \epsilon.$$

Let $\ell = \left\lfloor \frac{2(b-a)}{\delta} \right\rfloor$. We claim f is of bounded variation with total variation at most $(\ell + 1)\epsilon$. To see this, let $\{x_k\}_{k=0}^n$ be an arbitrary partition of $[a, b]$ and consider the partition

$$\mathcal{P} = \{x_k\}_{k=0}^n \cup \left\{a + \frac{1}{2}k\delta\right\}_{k=1}^{\ell}.$$

Clearly $\mathcal{P}$ is a partition of $[a, b]$. Write $\{z_k\}_{k=0}^m$ as the standard form of $\mathcal{P}$ and for each $j \in \{0, 1, \ldots, \ell + 1\}$ let $p_j \in \{0, \ldots, m\}$ be such that

$$z_{p_j} = \min\left\{a + \frac{1}{2}j\delta, b\right\}.$$

Notice if we let

$$z_{p_j} = a_1 < z_{p_j+1} = b_1 = a_2 < z_{p_j+2} = b_2 = a_3 < \cdots \leq z_{p_{j+1}},$$

then, as $|z_{p_{j+1}} - z_{p_j}| < \delta$, we obtain by our choice of δ via absolutely continuity that

$$\sum_{k=p_j+1}^{p_{j+1}} |f(z_k) - f(z_{k-1})| < \epsilon.$$

Hence

$$\begin{aligned}
\sum_{k=1}^n |f(x_k) - f(x_{k-1})| &\leq \sum_{k=1}^m |f(z_k) - f(z_{k-1})| \\
&= \sum_{j=0}^{\ell} \sum_{k=p_j+1}^{p_{j+1}} |f(z_k) - f(z_{k-1})| \\
&\leq (\ell + 1)\epsilon < \infty.
\end{aligned}$$

Therefore, as $\{x_k\}_{k=0}^n$ was an arbitrary partition of $[a, b]$, f is of bounded variation. ∎

Corollary 5.4.6. *If $f : [a, b] \to \mathbb{R}$ is absolutely continuous, then f is differentiable λ-almost everywhere, f' is Lebesgue measurable, and $f' \in L_1([a, b], \lambda)$.*

Proof. As every absolutely continuous function is of bounded variation by Proposition 5.4.5, the result follows from Corollary 5.3.8. ∎

In Example 5.4.4 we claimed the Cantor ternary function was not absolutely continuous without proof. The following interesting proposition demonstrates (via Remark 5.2.7) why this is the case.

Proposition 5.4.7. *If $f : [a, b] \to \mathbb{R}$ is absolutely continuous and $f' = 0$ λ-almost everywhere, then f is constant.*

Proof. To see that f is constant on $[a, b]$, let $c \in (a, b]$ be arbitrary. We claim that $f(c) = f(a)$. To begin, let $\epsilon > 0$ and recall that since $f' = 0$ almost everywhere, there exists a Lebesgue measurable set $X \subseteq [a, c]$ such that $f'(x) = 0$ for all $x \in X$ and $\lambda([a, c] \setminus X) = 0$. As f is absolutely continuous, there exists a $\delta > 0$ such that if $\{a_k\}_{k=1}^n, \{b_k\}_{k=1}^n \subseteq [a, c]$ are such that

$$a \leq a_1 < b_1 \leq a_2 < b_1 \leq \cdots \leq a_n < b_n \leq c \qquad \text{and} \qquad \sum_{k=1}^n |b_k - a_k| < \delta$$

then

$$\sum_{k=1}^n |f(b_k) - f(a_k)| < \epsilon.$$

Note we can even allow $a_k = b_k$ in the above as the interval $[a_k, b_k]$ then contributes zero to both sums.

Let $x \in X \setminus \{c\}$ be arbitrary. Then

$$0 = f'(x) = \lim_{h\to 0} \frac{f(x + h) - f(x)}{h}.$$

Therefore, for any $\delta_0 > 0$ there exists an $h > 0$ such that $\lambda([x, x + h)) < \delta_0$, $[x, x+h) \subseteq [a, c)$, and $|f(x+h)-f(x)| < \epsilon h$. As the collection of such intervals forms a Vitali covering of $X \setminus \{c\}$, the Vitali Covering Lemma (Theorem 5.1.3) implies there exists an $n \in \mathbb{N}$, $x_1, \ldots, x_n \in X \setminus \{c\}$, and $h_1, \ldots, h_n > 0$ such that if $I_k = (x_k, x_k + h_k)$ for all $k \in \{1, \ldots, n\}$, then $\{I_k\}_{k=1}^n$ are pairwise disjoint subsets of $[a, c)$ such that $|f(x_k + h_k) - f(x_k)| < \epsilon h_k$ for all $k \in \{1, \ldots, n\}$ and

$$\lambda^*\left([a, c] \setminus \bigcup_{k=1}^n I_k\right) = \lambda^*\left((X \setminus \{c\}) \setminus \bigcup_{k=1}^n I_k\right) < \delta.$$

Let $y_0 = a$, $x_{n+1} = c$, and $y_k = x_k + h_k$ for all $k \in \{1, \ldots, n\}$. Then

$$a \leq y_0 \leq x_1 < y_1 \leq x_2 < y_2 \leq \cdots \leq x_n < y_n \leq x_{n+1} = c.$$

Therefore, as

$$\sum_{k=0}^{n} |x_{k+1} - y_k| = \lambda\left(\bigcup_{k=0}^{n} [y_k, x_{k+1})\right) = \lambda^*\left([a, c] \setminus \bigcup_{k=1}^{n} I_k\right) < \delta,$$

we obtain by our choice of δ via absolute continuity that

$$\sum_{k=0}^{n} |f(x_{k+1}) - f(y_k)| < \epsilon.$$

However, note in addition that

$$\sum_{k=1}^{n} |f(y_k) - f(x_k)| < \sum_{k=1}^{n} \epsilon h_k \leq (c - a)\epsilon.$$

Therefore, by the triangle inequality,

$$|f(c) - f(a)| \leq \sum_{k=0}^{n} |f(x_{k+1}) - f(y_k)| + \sum_{k=1}^{n} |f(y_k) - f(x_k)| < (c - a + 1)\epsilon.$$

Hence, as $\epsilon > 0$ was arbitrary, we obtain that $f(c) = f(a)$. Hence, as $c \in (a, b]$ was arbitrary, the result follows. ∎

To conclude this section, we desire to show that functions defined by integrating against an L_1-function are absolutely continuous and thus the collection of absolutely continuous functions include those defined in a 'Fundamental Theorem of Calculus'-like manner. This is achieved via the following lemma.

Lemma 5.4.8. *Let $(X, \mathcal{A}, \mu)$ be a measure space and let $f \in L_1(X, \mu)$. Then for all $\epsilon > 0$ there exists a $\delta > 0$ such that if $A \in \mathcal{A}$ and $\mu(A) < \delta$ then*

$$\int_A |f| \, d\mu < \epsilon.$$

Proof. Exercise. ∎

Proposition 5.4.9. *Let $f \in L_1([a, b], \lambda)$. If $F : [a, b] \to \mathbb{C}$ is defined by*

$$F(x) = \int_{[a,x]} f \, d\lambda$$

for all $x \in [a, b]$, then F is absolutely continuous.

Proof. First notice that F is well-defined as $f \in L_1([a, b], \lambda)$.

To see that F is absolutely continuous, let $\epsilon > 0$. Since $f \in L_1([a, b], \lambda)$, by Lemma 5.4.8 there exists a $\delta > 0$ such that if $A \in \mathcal{A}$ and $\mu(A) < \delta$ then

$$\int_A |f| \, d\mu < \epsilon.$$

To see that this δ satisfies the requirements of Definition 5.4.1, let

$$\{a_k\}_{k=1}^n, \{b_k\}_{k=1}^n \subseteq [a,b]$$

be such that

$$a \leq a_1 < b_1 \leq a_2 < b_1 \leq \cdots \leq a_n < b_n \leq b \qquad \text{and} \qquad \sum_{k=1}^n |b_k - a_k| < \delta.$$

Therefore, since

$$\lambda\left(\bigcup_{k=1}^n [a_k, b_k]\right) = \sum_{k=1}^n |b_k - a_k| < \delta,$$

we obtain that

$$\begin{aligned}
\sum_{k=1}^n |F(b_k) - F(a_k)| &= \sum_{k=1}^n \left| \int_{[a,b_k]} f \, d\lambda - \int_{[a,a_k]} f \, d\lambda \right| \\
&= \sum_{k=1}^n \left| \int_{\mathbb{R}} f\chi_{[a,b_k]} - f\chi_{[a,a_k]} \, d\lambda \right| \\
&= \sum_{k=1}^n \left| \int_{\mathbb{R}} f\chi_{[a_k,b_k]} \, d\lambda \right| \\
&= \sum_{k=1}^n \left| \int_{[a_k,b_k]} f \, d\lambda \right| \\
&\leq \sum_{k=1}^n \int_{[a_k,b_k]} |f| \, d\lambda \\
&= \int_{\bigcup_{k=1}^n [a_k,b_k]} |f| \, d\lambda < \epsilon.
\end{aligned}$$

Hence F is absolutely continuous as desired. ∎

5.5 The Fundamental Theorems of Calculus

Due to the examples of absolutely continuous functions in Proposition 5.4.9 resembling functions analyzed in undergraduate calculus in relation to the Fundamental Theorems of Calculus, it is natural to ask what the derivatives of the functions defined in Proposition 5.4.9 are and whether all absolutely continuous functions are of the above form. Both of these questions will be answered in this section thereby generalizing the Fundamental Theorems of Calculus!

To begin, we note the following technical lemma.

Lemma 5.5.1. *Let $f \in L_1([a,b],\lambda)$ be real-valued and define $F : [a,b] \to \mathbb{R}$ by*

$$F(x) = \int_{[a,x]} f \, d\lambda$$

for all $x \in [a,b]$. If F is non-decreasing, then $f(x) \geq 0$ for almost every x.

Proof. Let

$$X = \{x \in [a,b] \mid f(x) < 0\},$$

which is a measurable set as f is measurable. Clearly it suffices to prove that $\lambda(X) = 0$. To see this, suppose to the contrary that $\lambda(X) > 0$. Due to the regularity of the Lebesgue measure from Proposition 2.3.11, there exists a compact subset $K \subseteq X$ such that $\lambda(K) > 0$. Therefore, as $f(x) < 0$ for all $x \in K \subseteq X$ and as $\lambda(K) > 0$, we obtain that

$$\int_K f \, d\lambda < 0.$$

Notice if $V = (a,b) \setminus K$, then

$$F(b) - F(a) = F(b) = \int_{[a,b]} f \, d\lambda = \int_K f \, d\lambda + \int_V f \, d\lambda < \int_V f \, d\lambda.$$

However, as V is an open subset of (a,b) and as every open subset of $\mathbb{R}$ is a countable union of disjoint open intervals, we may write

$$V = \bigcup_{k=1}^\infty (a_k, b_k)$$

where $(a_k, b_k) \subseteq (a,b)$ for all $k \in \mathbb{N}$ and $\{(a_k, b_k)\}_{k=1}^\infty$ are pairwise disjoint. Therefore, if $f_k = f\chi_{(a_k,b_k)}$ for each $k \in \mathbb{N}$, then

$$\int_V f \, d\lambda = \int_{\mathbb{R}} \sum_{k=1}^\infty f_k \, d\lambda.$$

Notice if $S_n = \sum_{k=1}^n f_k$ for each $n \in \mathbb{N}$, then $|S_n| \leq |f|$. Hence, as $f \in L_1([a,b],\lambda)$, we obtain by the Dominated Convergence Theorem (Theorem 4.6.1) that

$$\begin{aligned} F(b) - F(a) &< \int_V f \, d\lambda \\ &= \lim_{n\to\infty} \int_{\mathbb{R}} \sum_{k=1}^n f_k \, d\lambda \\ &= \lim_{n\to\infty} \sum_{k=1}^n F(b_k) - F(a_k) \\ &\leq F(b) - F(a) \end{aligned}$$

as F is non-decreasing. As this clearly is a contradiction, we obtain that $\lambda(X) = 0$ as desired. ∎

Corollary 5.5.2. *Let $f \in L_1([a,b],\lambda)$ be real-valued and define $F : [a,b] \to \mathbb{R}$ by*

$$F(x) = \int_{[a,x]} f \, d\lambda$$

for all $x \in [a,b]$. If $F(x) = 0$ for all $x \in [a,b]$, then $f = 0$ almost everywhere.

Proof. Since F is constant, F is non-decreasing. Hence Lemma 5.5.1 implies that $f \geq 0$ almost everywhere. Hence

$$0 = F(b) = \int_{[a,b]} f \, d\lambda = \|f\|_1 .$$

Therefore $f = 0$ in $L_1([a,b],\lambda)$ and thus $f = 0$ almost everywhere. ∎

Using all of the above, we arrive at our Fundamental Theorems of Calculus which complete characterize absolutely continuous functions.

Theorem 5.5.3 (Fundamental Theorem of Calculus, I). *Let $f \in L_1([a,b],\lambda)$ be real-valued. If $F : [a,b] \to \mathbb{R}$ is defined by*

$$F(x) = \int_{[a,x]} f \, d\lambda$$

for all $x \in [a,b]$, then F' exists almost everywhere and $F'(x) = f(x)$ for almost every x.

Proof. To begin, note F is absolutely continuous (and thus measurable) by Proposition 5.4.9. Hence F' exists almost everywhere and is an element of $L_1([a,b],\lambda)$ by Corollary 5.4.6. To demonstrate that $F'(x) = f(x)$ for almost every x, we divide the proof into three cases.

Case 1: f is bounded. Thus there exists an $M \geq 0$ such that $|f(x)| \leq M$ for all $x \in [a,b]$. For notational simplicity, for all $t \geq b$ define $F(t) = F(b)$. Furthermore, for each $n \in \mathbb{N}$, let $F_n : [a,b] \to \mathbb{R}$ be defined by

$$F_n(x) = n\left(F\left(x + \frac{1}{n}\right) - F(x)\right) = n \int_{\left[x, x+\frac{1}{n}\right]} f \, d\lambda$$

for all $x \in [a,b]$. Clearly each F_n is a measurable function since F is measurable and by Proposition 2.3.8. Furthermore, notice for each $n \in \mathbb{N}$ and $x \in [a,b]$ that

$$|F_n(x)| \leq n \int_{\left[x, x+\frac{1}{n}\right]} |f| \, d\lambda \leq n\left(\frac{1}{n}M\right) = M.$$

Since $M\chi_{[a,b]} \in L_1([a,b],\lambda)$, since $\lim_{n\to\infty} F_n(x) = F'(x)$ for almost every $x \in [a,b]$, and since $|F_n| \leq M\chi_{[a,b]}$, we obtain by the Dominated Convergence Theorem (Theorem 4.6.1) that

$$\int_{[a,c]} F' \, d\lambda = \lim_{n\to\infty} \int_{[a,c]} F_n \, d\lambda$$

for all $c \in [a, b]$. Hence

$$\begin{aligned}\int_{[a,c]} F' \, d\lambda &= \lim_{n\to\infty} n \int_{[a,c]} F\left(x + \frac{1}{n}\right) - F(x) \, d\lambda(x) \\ &= \lim_{n\to\infty} n \left(\int_{\left[c, c+\frac{1}{n}\right]} F \, d\lambda - \int_{\left[a, a+\frac{1}{n}\right]} F \, d\lambda \right)\end{aligned}$$

for all $c \in [a, b]$.

We claim that

$$\lim_{n\to\infty} n \int_{\left[c, c+\frac{1}{n}\right]} F \, d\lambda = F(c)$$

for all $c \in [a, b]$. Indeed since F is absolutely continuous, F is continuous. Therefore, as $c \in [a, b]$, for every $\epsilon > 0$ there exists an $N_c \in \mathbb{N}$ such that $|F(x) - F(c)| < \epsilon$ for all $x \in \left[c, c + \frac{1}{N_c}\right]$. Hence for all $n \geq N_c$ we obtain that

$$\begin{aligned}\left| F(c) - n \int_{\left[c, c+\frac{1}{n}\right]} F(x) \, d\lambda(x) \right| &= \left| n \int_{\left[c, c+\frac{1}{n}\right]} F(c) - F(x) \, d\lambda(x) \right| \\ &\leq n \int_{\left[c, c+\frac{1}{n}\right]} |F(c) - F(x)| \, d\lambda(x) \\ &\leq n \int_{\left[c, c+\frac{1}{n}\right]} \epsilon \, d\lambda(x) = \epsilon.\end{aligned}$$

Hence the claim follows.

Therefore, by applying the above limit twice (once with $c = a$), we obtain for all $c \in [a, b]$ that

$$\int_{[a,c]} F' \, d\lambda = F(c) - F(a) = F(c) = \int_{[a,c]} f \, d\lambda.$$

Therefore, as $F', f \in L_1([a, b], \lambda)$, we obtain that

$$\int_{[a,x]} F' - f \, d\lambda = 0$$

for all $x \in [a, b]$. However, as $F' - f \in L_1([a, b], \lambda)$, Corollary 5.5.2 implies that $F' - f = 0$ almost everywhere. Hence $F' = f$ almost everywhere as desired.

Case 2: $f \geq 0$. For each $n \in \mathbb{N}$, define $f_n : [a, b] \to [0, n]$ by $f_n(x) = \min\{f(x), n\}$ for all $x \in [a, b]$. Clearly each f_n is a measurable function such that $|f_n| \leq n$ (so $f_n \in L_1([a, b], \lambda)$) and such that $\lim_{n\to\infty} f_n(x) = f(x)$ for all $x \in [a, b]$.

We claim for all $n \in \mathbb{N}$ that $F'(x) \geq f_n(x)$ for almost every x. To see this, for each $n \in \mathbb{N}$ define $F_n, G_n : [a, b] \to \mathbb{R}$ by

$$F_n(x) = \int_{[a,x]} f_n \, d\lambda \qquad \text{and} \qquad G_n(x) = \int_{[a,x]} f - f_n \, d\lambda$$

for all $x \in [a,b]$. As $f_n, f - f_n \in L_1([a,b],\lambda)$, we see that F_n and G_n are well-defined, $F = F_n + G_n$, and F_n and G_n are differentiable almost everywhere. Furthermore, as f_n is bounded, the first case of this proof implies that $F_n' = f_n$ almost everywhere. Moreover, since $f - f_n \geq 0$ by construction, G_n is non-decreasing so $G_n'(x) \geq 0$ for almost every x. Hence for almost every $x \in [a,b]$,

$$F'(x) = F_n'(x) + G_n'(x) \geq F_n'(x) = f_n(x)$$

as claimed.

Since $F'(x) \geq f_n(x)$ for almost every x and $\lim_{n\to\infty} f_n(x) = f(x)$ for all $x \in [a,b]$, we obtain that $F'(x) \geq f(x)$ for almost every $x \in [a,b]$. Furthermore, as $f(x) \geq 0$ for almost every $x \in [a,b]$, $F' \geq 0$ and F is non-decreasing on $[a,b]$. Hence by the Lebesgue Differentiation Theorem (Theorem 5.2.2), we obtain that

$$F(b) - F(a) \geq \int_{[a,b]} F' \, d\lambda \geq \int_{[a,b]} f \, d\lambda = F(b) - F(a).$$

Hence $F' \in L_1([a,b],\lambda)$ and

$$\int_{[a,b]} F' - f \, d\lambda = 0.$$

Therefore, as $F' - f \geq 0$, the above integral implies that $F' = f$ almost everywhere.

Case 3: f arbitrary. Recall that we may write

$$f = f_1 - f_2$$

where $f_1, f_2 \in L_1([a,b],\lambda)$ are such that $f_k \geq 0$. Therefore, if $F_k : [a,b] \to \mathbb{R}$ is defined by

$$F_k(x) = \int_{[a,x]} f_k \, d\lambda,$$

then Case 2 implies that F_k is a well-defined function such that $F_k' = f_k$ almost everywhere. As clearly $F = F_1 - F_2$ by linearity, we obtain that

$$F' = F_1' - F_2' = f_1 - f_2 = f$$

almost everywhere as desired. ■

Theorem 5.5.4 (Fundamental Theorem of Calculus, II). *If $F : [a,b] \to \mathbb{R}$ is absolutely continuous, then $F' \in L_1([a,b],\lambda)$ and*

$$F(x) = F(a) + \int_{[a,x]} F' \, d\lambda$$

for all $x \in [a,b]$.

Proof. To begin, recall that if $F : [a, b] \to \mathbb{R}$ is absolutely continuous, then F is differentiable almost everywhere with $F' \in L_1([a, b], \lambda)$ by Corollary 5.4.6. Define $G : [a, b] \to \mathbb{R}$ by

$$G(x) = \int_{[a,x]} F' \, d\lambda$$

for all $x \in [a, b]$. Then G is absolutely continuous by Proposition 5.4.9 and $G' = F'$ almost everywhere by the First Fundamental Theorem of Calculus (Theorem 5.5.3). Thus $F - G$ is absolutely continuous and $(F - G)' = F' - G' = 0$ almost everywhere. Hence Proposition 5.4.7 implies that $F - G$ is constant. Therefore, as $(F - G)(a) = F(a)$, we obtain that $F(x) - G(x) = F(a)$ for all $x \in [a, b]$ so

$$F(x) = F(a) + \int_{[a,x]} F' \, d\lambda$$

for all $x \in [a, b]$ as desired. ∎

Chapter 6

Another Undecided Chapter

Probably Fubini's Theorem for Lebesgue integrals.

Index