MATH 4012
Real Analysis IIIB
Lebesgue Measure Theory

Paul Skoufranis

April 28, 2025

Preface:

These are the first edition of these lecture notes for MATH 4012 (Real Analysis IIIB: Lebesgue Measure Theory). Consequently, 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

Motivation for this Course

It was seen in MATH 2001 and MATH 3001 that the Riemann integral is an important tool in analysis with many properties and applications. Since the Riemann integral works incredible well for continuous functions and interacts naturally with respect to differentiation, the Riemann integral is ideal for calculus and science. However, the Riemann integral does have its flaws.

One such flaw comes when trying to determine which functions are Riemann integrable. In MATH 2001 it was shown that continuous functions on closed intervals are uniformly continuous and thus Riemann integrable. However it can be very difficult to verify whether or not a given function is Riemann integrable. For example, although the function $\chi_{\mathbb{Q}} : \mathbb{R} \to \mathbb{R}$ defined by

$$\chi_{\mathbb{Q}}(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \notin \mathbb{Q} \end{cases}$$

is easily seen to be not Riemann integrable on $[0, 1]$ (see Example A.2.1), the function $d : \mathbb{R} \to \mathbb{R}$ defined by

$$d(x) = \begin{cases} 0 & \text{if } x \notin \mathbb{Q} \\ 1 & \text{if } x = 0 \\ \frac{1}{q} & \text{if } x \neq 0 \text{ and } x = \frac{p}{q} \text{ where } p, q \in \mathbb{Z},\ q > 0, \text{ and } \gcd(p, q) = 1 \end{cases}$$

is Riemann integral on $[0, 1]$ with $\int_0^1 d(x)\, dx = 0$ (see Example A.3.4).

Another flaw of the Riemann integral occurs with respect to limits; the concept at the heart analysis. In MATH 3001 it was shown that if $(f_n)_{n\geq 1}$ was a sequence of Riemann integrable functions that converged to f uniformly on $[a, b]$, then f is Riemann integrable and

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

However, the Riemann integral does not behave well with respect to pointwise limits. For one example define $f_n : [0, 1] \to \mathbb{R}$ by

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

It is elementary to verify that $(f_n)_{n\geq 1}$ converges to 0 pointwise yet $\int_0^1 f_n(x)\,dx = \frac{1}{2}$ for all n thereby showing that

$$\int_0^1 f(x)\,dx = 0 \neq \frac{1}{2} = \lim_{n\to\infty}\int_0^1 f_n(x)\,dx.$$

Another example occurs by considering $\chi_{\mathbb{Q}}$. Indeed, since $\mathbb{Q}$ is countable, we can enumerate $\mathbb{Q}\cap[0,1]$ as $\mathbb{Q}\cap[0,1] = \{r_n \mid n \in \mathbb{N}\}$. Consequently, if we define $f_n : [0,1] \to \mathbb{R}$ by

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

then f_n has a finite number of discontinuities and thus is Riemann integrable, yet converges pointwise to $\chi_{\mathbb{Q}}$ which is not even Riemann integrable!

The main goal of this course is to understand the analysis concepts and techniques that allow us to improve on the Riemann integral. After all, since $\mathbb{R}\setminus\mathbb{Q}$ is "much larger" than $\mathbb{Q}$, we would expect that $\int_0^1 \chi_{\mathbb{Q}}(x)\,dx = 0$ as $\chi_{\mathbb{Q}}$ is zero "almost everywhere". Of course, we need to make mathematically precise what we mean by "much larger". For example, we could consider cardinality where $\mathbb{R}\setminus\mathbb{Q}$ is uncountable whereas $\mathbb{Q}$ is countable. However, if we want an appropriate notion of size of a subset of $\mathbb{R}$ to develop a better integral, we need to recall the motivating aspect of integration: the area under the curve. Since $\chi_{\mathbb{Q}}$ is 1 at each rational and 0 at each irrational, the integral of $\chi_{\mathbb{Q}}$ on $[0,1]$ should just be the "length" of $\mathbb{Q}\cap[0,1]$. Similarly, if $A \subseteq \mathbb{R}$ and $\chi_A : \mathbb{R} \to \mathbb{R}$ is defined by

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

then we would expect that the integral of χ_A should be the "length" of A. Thus, the question remains, how do we "measure" the "length" of a subset of $\mathbb{R}$?

Chapter 1

The Lebesgue Measure

Our first goal is to develop a good notion of "length" or "measure" for subsets of $\mathbb{R}$. Such a notion should be a function $\ell : \mathcal{P}(\mathbb{R}) \to [0, \infty]$ that assigns a length to each subset of $\mathbb{R}$. To make this a well-defined notion of length, we want ℓ to have specific properties we associate to length including:

(L1) $\ell(\emptyset) = 0$ (i.e. the empty set has no length),

(L2) $\ell([a, b]) = \ell((a, b]) = \ell([a, b)) = \ell((a, b)) = b - a$ for all $a, b \in \mathbb{R}$ (i.e. the length of intervals is correct),

(L3) if $A \subseteq B \subseteq \mathbb{R}$, then $\ell(A) \leq \ell(B)$ (i.e. larger sets have large length),

(L4) if $x \in \mathbb{R}$, $A \subseteq \mathbb{R}$, and $x + A = \{x + a \mid a \in A\}$, then $\ell(x + A) = \ell(A)$ (i.e. the translation of a set along the number line preserves the length),

(L5) if $\alpha \in \mathbb{R}$, $A \subseteq \mathbb{R}$, and $\alpha A = \{\alpha a \mid a \in A\}$, then $\ell(\alpha A) = |\alpha|\ell(A)$ (i.e. scaling a set scales the length of the set), and

(L6) if $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{P}(\mathbb{R})$ are pairwise disjoint (i.e. $A_n \cap A_m = \emptyset$ whenever $n \neq m$), then $\ell\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \ell(A_n)$ (i.e. the length of a union of disjoint sets is the sum of their lengths).

One may question whether (L6) should only contain finite unions of disjoint sets. However, to perform analysis, one requires the ability to take limits. Therefore, since

$$\sum_{n=1}^{\infty} \ell(A_n) = \lim_{N\to\infty} \sum_{n=1}^{N} \ell(A_n),$$

if (L6) held for finite limits, we would expect

$$\sum_{n=1}^{\infty} \ell(A_n) = \lim_{N\to\infty} \ell\left(\bigcup_{n=1}^{N} A_n\right)$$

and thus we would expect this limit to be $\ell\left(\bigcup_{n=1}^{\infty} A_n\right)$.

Now that we have some properties we would expect of a length function, the question is "How do we construct such a function?"

1.1 What Goes Wrong...

Perhaps unexpectedly, constructing such a function is impossible as the following result demonstrates.

Theorem 1.1.1. *There does not exists a function $\ell : \mathcal{P}(\mathbb{R}) \to [0, \infty]$ such that (L1), (L2), (L4), and (L6) hold.*

Proof. Suppose for the sake of a contradiction that there exists a function $\ell : \mathcal{P}(\mathbb{R}) \to [0, \infty]$ such that (L1), (L2), (L4), and (L6) hold. Note by taking $A_n = \emptyset$ for all $n > N$ in (L6), we obtain by (L1) that

$$\ell\left(\bigcup_{n=1}^{N} A_n\right) = \sum_{n=1}^{N} \ell(A_n)$$

whenever $\{A_n\}_{n=1}^{N} \subseteq \mathcal{P}(\mathbb{R})$ are pairwise disjoint. We will now proceed in constructing a very problematic subset of $\mathbb{R}$.

Define a relation $\sim$ on $\mathbb{R}$ by $x \sim y$ if and only if $x - y \in \mathbb{Q}$. It is not difficult to verify that $\sim$ is an equivalence relation on $\mathbb{R}$.

We claim that every element of $\mathbb{R}$ is $\sim$-equivalent to some element in $[0, 1)$. Indeed if $x \in \mathbb{R}$, then x is the sum of its integer part $\lfloor x \rfloor$ and its fractional part $\{x\}$. Since $x - \{x\} = \lfloor x \rfloor \in \mathbb{Q}$, we obtain that $x \sim \{x\}$. Therefore, since $\{x\} \in [0, 1)$, x is $\sim$-equivalent to some element in $[0, 1)$.

Consequently every equivalence class under $\sim$ has an element in $[0, 1)$. Let $A \subseteq [0, 1)$ be a set that contains precisely one element from each equivalence class of $\sim$. Note the existence of A follows from the Axiom of Choice B.2.8 (see Remark 1.1.2). Our goal is to use A to show that ℓ cannot possible satisfy (L2), (L4), and (L6).

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).

We claim that $\{A_n\}_{n=1}^{\infty}$ are disjoint with union $[0, 1)$. To see this, note if $x \in [0, 1)$ then there exists a unique $y \in A \subseteq [0, 1)$ such that $x \sim y$. Thus $x - y \in \mathbb{Q} \cap (-1, 1)$. If $x - y \in \mathbb{Q} \cap [0, 1)$ then $x - y = r_n$ for some n and thus $x = r_n + y \in A_n$. Otherwise if $x - y \in \mathbb{Q} \cap (-1, 0)$ then $(x + 1) - y \in (0, 1)$. Thus $(x + 1) - y = r_n$ for some n and thus $x = r_n + y - 1 \in A_n$. Hence

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

To see that $\{A_n\}_{n=1}^\infty$ are pairwise disjoint, assume $x \in A_n \cap A_m$ for some $n, m \in \mathbb{N}$. By definition, there exists $y, z \in A$ and $k, l \in \{0, 1\}$ such that $x+k = r_n+y$ and $x+l = r_m+z$. Therefore $y-z = r_m-r_n+k-l \in \mathbb{Q}$ so $y \sim z$. Hence $y = z$ as A contains exactly on element from each equivalence class of $\sim$. Thus $0 = r_m - r_n + k - l$. Since $k - l \in \{-1, 0, 1\}$ and $r_n - r_m \in (-1, 1)$, $0 = r_m - r_n + k - l$ can only occur when $r_n = r_m$ in which case $n = m$. 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 $A_n = B_{n,1} \cup B_{n,2}$ since $r_n + A \subseteq [0, 2)$ for all n.

We claim that $B_{n,1} \cap B_{n,2} = \emptyset$. To see this, suppose for the sake of a contradiction that $b \in B_{n,1} \cap B_{n,2}$. By definition there exists $x, y \in A$ such that $r_n + x \in [0, 1)$, $r_n + y \in [1, 2)$, and $b = r_n + x = -1 + r_n + y$. Clearly $x - y = -1 \in \mathbb{Q}$ so $x \neq y$ and $x \sim y$. Therefore, since 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, note that

$$\begin{aligned} 1 &= \ell([0, 1)) \\ &= \ell\left(\bigcup_{n=1}^\infty A_n\right) \\ &= \sum_{n=1}^\infty \ell(A_n) && \text{by (L6) as } \{A_n\}_{n=1}^\infty \text{are disjoint} \\ &= \sum_{n=1}^\infty \ell(B_{n,1} \cup B_{n,2}) \\ &= \sum_{n=1}^\infty \ell(B_{n,1}) + \ell(B_{n,2}) && \text{by (L1) and (L6) as } B_{n,1} \cap B_{n,2} = \emptyset \\ &= \sum_{n=1}^\infty \ell((r_n + A) \cap [0, 1)) + \ell\left(((r_n + A) \cap [1, 2)\right) && \text{by (L4)} \\ &= \sum_{n=1}^\infty \ell((r_n + A) \cap [0, 2)) && \text{by (L1) and (L6)} \\ &= \sum_{n=1}^\infty \ell(r_n + A) && \text{since } r_n + A \subseteq [0, 2) \\ &= \sum_{n=1}^\infty \ell(A) && \text{by (L4).} \end{aligned}$$

This yields our contradiction since $\ell(A) \in [0, \infty]$ yet no number in $[0, \infty]$ when summed an infinite number of times produces 1. Hence ℓ cannot possibly exist. ∎

Remark 1.1.2. In the proof of Theorem 1.1.1, the existence of the set A containing exactly one element from each equivalence class may seem natural; just pick one element from each equivalence class. However, the construction of the set A does not follow from axioms of Zermelo–Fraenkel set theory and requires the Axiom of Choice (Axiom B.2.8). To construct A, note if $\{E_k\}_{k\in I}$ are the equivalence classes of $\sim$, then the Axiom of Choice implies there is a function $f : I \to \bigcup_{k\in I} E_k$ such that $f(k) \in E_k$ for all $k \in I$. The set A is the range of f.

Having established that there is no length function on all subsets of $\mathbb{R}$ that has our desired properties, what can we do to define a notion of length on subsets of $\mathbb{R}$ in order to generalized and improve on the Riemann integral?

1.2 σ-Algebras

The reason there does not exist a length function on all subsets of $\mathbb{R}$ with our required properties is that there are just too many subsets of $\mathbb{R}$. Consequently, our only hope is to reduce the collection of subsets of $\mathbb{R}$ that will form the domain of our length function. This leads to the question of what types of sets will we allow in our collection and what set operations will be permitted on this collection?

Properties (L1)-(L6) give us some clue about what set operations we want to be able to perform on our collection. However, it turns out that we can simply focus on a few of these properties and the remaining will automatically follow. The essential properties we will focus on are derived by considering probability theory and modelling the collection of sets we can measure the length of as events we can compute the probability of. This leads us to the following notion.

Definition 1.2.1. Let X be a non-empty set. A *σ-algebra on X* is a subset $\mathcal{A} \subseteq \mathcal{P}(X)$ such that

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

Remark 1.2.2. The 'σ' in 'σ-algebra' refers to the object being closed under 'countable' unions. Note that σ-algebras are also closed under finite unions.

Indeed by taking $A_n = \emptyset$ for all $n > N$, we see that if $A_1, \ldots, A_N \in \mathcal{A}$, then $\bigcup_{n=1}^N A_n = \bigcup_{n=1}^\infty A_n \in \mathcal{A}$.

Remark 1.2.3. The property that σ-algebras are closed under complements might seem odd to include when we think about lengths; after all, if a subset of $\mathbb{R}$ has a finite length, its complement would be expected to have infinite length. However, complements allow us to guarantee the intersection of any countable collection of measurable sets to be measurable. Indeed if $\{A_n\}_{n=1}^\infty$ are elements of a σ-algebra $\mathcal{A}$, 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 measurable sets are measurable. Furthermore,by taking $A_n = X$ for all $n > N$, we see that if $A_1, \ldots, A_N \in \mathcal{A}$, then $\bigcap_{n=1}^N A_n = \bigcap_{n=1}^\infty A_n \in \mathcal{A}$.

Remark 1.2.4. The property that σ-algebras are closed under complements and thus closed under intersections also means that σ-algebras are closed under set differences. Indeed if A and B are elements in a σ-algebra $\mathcal{A}$, then $B^c \in \mathcal{A}$ so

$$A \setminus B = A \cap B^c \in \mathcal{A}.$$

Of course, there are some obvious σ-algebras that work for every set.

Example 1.2.5. Let X be a non-empty set. Then $\mathcal{P}(X)$ is a σ-algebra.

Example 1.2.6. Let X be a non-empty set. Then $\{\emptyset, X\}$ is a σ-algebra.

We have seen with respect to measuring the length of subsets of $\mathbb{R}$ that $\mathcal{P}(\mathbb{R})$ is too large of a σ-algebra and clearly $\{\emptyset, \mathbb{R}\}$ is too small as we will want every interval to be assigned a length. Thus we may need a way to construct a σ-algebra on $\mathbb{R}$ that contains the intervals. This is accomplished via the following.

Lemma 1.2.7. *Let X be a non-empty set and let $A \subseteq \mathcal{P}(X)$. There exists a smallest (with respect to inclusion) σ-algebra $\sigma(A)$ of X such that $A \subseteq \sigma(A)$. We call $\sigma(A)$ is the* σ-algebra generated by *A.*

Proof. Let

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

Note $\mathcal{P}(X) \in I$ so I is non-empty.

Let

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

It follows from the definition of a σ-algebra that $\sigma(A)$ is a σ-algebra. Moreover, by the definition of I, we obtain that $A \subseteq \sigma(A)$. Finally, since $\sigma(A)$ is the intersection of all σ-algebras of X that contain A, clearly $\sigma(A)$ is the smallest (with respect to inclusion) σ-algebra of X containing A. ∎

Definition 1.2.8. The σ-algebra generated by the open subsets of $\mathbb{R}$ is called the *Borel σ-algebra* and is denoted $\mathfrak{B}(\mathbb{R})$. Elements of $\mathfrak{B}(\mathbb{R})$ are called *Borel sets*.

Remark 1.2.9. Although we have defined $\mathfrak{B}(\mathbb{R})$ to be the σ-algebra generated by the open subsets of $\mathbb{R}$, there are other collections of sets that generate $\mathfrak{B}(\mathbb{R})$. For example, we claim that the σ-algebra generated by

$$\mathcal{I} = \{(a,b) \mid a,b \in \mathbb{R}, a < b\}$$

is also $\mathfrak{B}(\mathbb{R})$. Indeed, as each element of $\mathcal{I}$ is open, $\mathcal{I} \subseteq \mathfrak{B}(\mathbb{R})$. Therefore, since $\mathfrak{B}(\mathbb{R})$ is a σ-algebra and since $\sigma(\mathcal{I})$ is the smallest σ-algebra that contains $\mathcal{I}$, we obtain that $\sigma(\mathcal{I}) \subseteq \mathfrak{B}(\mathbb{R})$.

To see the other inclusion, recall from MATH 2001 that every open subset of $\mathbb{R}$ is a countable union of open intervals. Clearly $\mathcal{I}$ contains all the open intervals except the open intervals of infinite length. Therefore, since $\mathfrak{B}(\mathbb{R})$ is the smallest σ-algebra containing the open sets, since $\sigma(\mathcal{I})$ is an σ-algebra containing $\mathcal{I}$, and since σ-algebras are closed under countable unions, to show that $\mathfrak{B}(\mathbb{R}) \subseteq \mathcal{I}$, it suffices to show that $(a,\infty) \in \mathcal{I}$ and $(-\infty, b) \in \sigma(\mathcal{I})$ for all $a, b \in \mathbb{R}$.

Note for all $a, b \in \mathbb{R}$ that

$$(a,\infty) = \bigcup_{n=1}^{\infty} (a,n) \qquad \text{and} \qquad (-\infty,b) = \bigcup_{n=1}^{\infty} (-n,b).$$

Therefore, (a,∞) and $(-\infty,b)$ are countable unions of elements of $\mathcal{I} \subseteq \sigma(\mathcal{I})$. Therefore, since $\sigma(\mathcal{I})$ is closed under countable unions, we obtain that $(a,\infty) \in \mathcal{I}$ and $(-\infty,b) \in \sigma(\mathcal{I})$ for all $a,b \in \mathbb{R}$. Hence $\mathfrak{B}(\mathbb{R})$ is also the σ-algebra generated by $\mathcal{I}$.

By using similar arguments as above together with computations like

$$(a,b] = \bigcap_{n=1}^{\infty} \left(a, b + \frac{1}{n}\right) \qquad \text{and} \qquad (a,b) = \bigcup_{n=1}^{\infty} \left(a, b - \frac{1}{n}\right)$$

and the fact that σ-algebras are closed under countable unions, countable intersections, and complements, it is possible to show that each of the following sets generate $\mathfrak{B}(\mathbb{R})$:

a) $\{F \subseteq \mathbb{R} \mid F \text{ is closed}\}$

b) $\{(a,b] \mid a,b \in \mathbb{R}, a < b\}$

c) $\{[a,b) \mid a,b \in \mathbb{R}, a < b\}$

d) $\{[a,b] \mid a,b \in \mathbb{R}, a < b\}$

e) $\{(-\infty,b) \mid b \in \mathbb{R}\}$

f) $\{(-\infty, b] \mid b \in \mathbb{R}\}$

g) $\{(a, \infty) \mid a \in \mathbb{R}\}$

h) $\{[a, \infty) \mid a \in \mathbb{R}\}$

In particular, all intervals are Borel sets and $\mathfrak{B}(\mathbb{R})$ is the smallest σ-algebra containing the intervals. Hence, as we hope to measure the lengths of all sets in a σ-algebra containing the intervals, we are really hoping to measure the length of all Borel sets.

Remark 1.2.10. 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 $\mathbb{R}$. Since Cantor's Theorem (Theorem B.7.6) implies that $|\mathbb{R}| < |\mathcal{P}(\mathbb{R})|$, there are by far many more subsets of $\mathbb{R}$ than there are Borel subsets of $\mathbb{R}$. Thus, it is far more likely we can measure the length of all Borel sets. However, this might be disappointing as there are far more subsets of $\mathbb{R}$ than Borel sets.

1.3 Measure Spaces

With the construction of the Borel sets complete, we turn our attention to whether or not we can measure the length of every Borel set via a function that satisfies (L1)-(L6) as listed at the start of this chapter. In mathematics, it is always useful to see which properties can be derived from other properties. As such, we will start with functions on σ-algebras with a minimal number of properties and see what can be derived from those properties. Since most σ-algebras need not be related to the real numbers, it is best to start with just (L1) and (L6) as these make sense for any σ-algebra. Of course we could also add in (L3). However, we will see how (L3) actually immediately follows from (L1) and (L6).

In the following definition and 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 1.3.1. Let $(X, \mathcal{A})$ be a measurable space. A *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, 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 of measures, we note some trivial properties of our measures.

Remark 1.3.2. 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 countable additivity on disjoint subsets with $A_k = \emptyset$ for all $k > n$. Thus measures are finite additive on disjoint subsets.

Remark 1.3.3. Let $(X, \mathcal{A}, \mu)$ be a measure space and let $E, F \in \mathcal{A}$. Assume $E \subseteq F$. Since $F \setminus E \in \mathcal{A}$ and since $F \setminus E$ is disjoint from E, we obtain by finite additivity on disjoint subsets that

$$\mu(F) = \mu(E \cup (F \setminus E)) = \mu(E) + \mu(F \setminus E) \geq \mu(E) + 0 = \mu(E).$$

In particular, if $\mathcal{A}$ is ordered by inclusion, then μ is monotone with respect to this inclusion (i.e. (L3) holds). Consequently, if $\mu(F) < \infty$ then $\mu(E) < \infty$. Moreover, notice if $\mu(E) < \infty$ the above computation implies that we may subtract $\mu(E)$ from both sides in order to obtain that $\mu(F \setminus E) = \mu(F) - \mu(E)$.

Remark 1.3.4. Let $(X, \mathcal{A}, \mu)$ be a measure space and let $A, B \in \mathcal{A}$. Assume $\mu(A \cap B) < \infty$. Since $A \in \mathcal{A}$ and $B \setminus (B \cap A) \in \mathcal{A}$ are disjoint, we obtain finite additivity on disjoint subsets and Remark 1.3.3 that

$$\begin{aligned} \mu(A \cup B) = \mu(A \cup (B \setminus (B \cap A)) &= \mu(A) + \mu(B \setminus (B \cap A)) \\ &= \mu(A) + \mu(B) - \mu(A \cap B) \end{aligned}$$

The above formula is probably very familiar in the context of probability. In fact, the basic objects in probability theory can be modelled as follows.

Definition 1.3.5. 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 1.3.3 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.

Using some intuition from probability, we obtain some basic examples of measures.

Example 1.3.6. 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

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

for all $A \in \mathcal{P}(X)$. It is elementary to verify that δ_x is a measure.

Example 1.3.7. Let $X = \{1, 2, 3, 4, 5, 6\}$ and define $\mu : \mathcal{P}(X) \to [0, 1]$ by

$$\mu(A) = \frac{|A|}{6}$$

for all $A \in \mathcal{P}(X)$. It is elementary to verify that μ is a measure. One can think of μ as the probability measure associated to rolling a unweighted 6-sided die.

Example 1.3.8. 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.

Unfortunately the above do not help us construct a measure that will measure the length of every Borel subset of $\mathbb{R}$. To gain some intuition on how we might construct such a measure, it is best to analyze the analytic properties of arbitrary measures. We begin with the following that shows all measures must behave well with respect to monotone sequences of sets. And yes, this is as important to this course as the Monotone Convergence Theorem for sequences is important to MATH 2001.

Theorem 1.3.9 (Monotone Convergence Theorem, Measures). *Let $(X, \mathcal{A}, \mu)$ be a measure space and let $\{A_n\}_{n=1}^\infty \subseteq \mathcal{A}$.*

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

b) 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 a) is true, 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 $n \in \mathbb{N}$. Hence

$$\begin{aligned}
\mu\left(\bigcup_{n=1}^\infty A_n\right) &= \mu\left(\bigcup_{n=1}^\infty B_n\right) \\
&= \sum_{k=1}^\infty \mu(B_k) && \{B_k\}_{k=1}^\infty \text{ pairwise disjoint} \\
&= \lim_{n\to\infty} \sum_{k=1}^n \mu(B_k) && \text{definition of series} \\
&= \lim_{n\to\infty} \mu\left(\bigcup_{k=1}^n B_k\right) && \{B_k\}_{k=1}^\infty \text{ pairwise disjoint} \\
&= \lim_{n\to\infty} \mu(A_n)
\end{aligned}$$

as desired.

To see b) is true, notice if $B_n = A_1 \setminus A_n$ for all $n \in \mathbb{N}$, 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

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

we obtain by part a) that

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

Since $\mu(A_1) < \infty$, Remark 1.3.3 implies that $\mu(A_1 \setminus E) = \mu(A_1) - \mu(E)$ for all $E \in \mathcal{A}$ with $E \subseteq A_1$. Hence

$$\begin{aligned}
\mu(A_1) - \mu\left(\bigcap_{n=1}^\infty A_n\right) &= \mu\left(A_1 \setminus \left(\bigcap_{n=1}^\infty A_n\right)\right) \\
&= \lim_{n\to\infty} \mu(A_1 \setminus A_n) \\
&= \lim_{n\to\infty} \mu(A_1) - \mu(A_n) \\
&= \mu(A_1) - \lim_{n\to\infty} \mu(A_n).
\end{aligned}$$

Hence, by subtracting $\mu(A_1) < \infty$ from both sides, the result follows. ∎

The strategy for the proof of the first part of the Monotone Convergence Theorem was to take our increasing sequence of sets and make them pairwise disjoint while preserving the union via set operations that preserve measurable sets. This is a strategy we will often apply. In particular, by a more advanced version of "make sets pairwise disjoint", we can prove the following which shows how measures behave on countable unions of sets that may not be pairwise disjoint.

Proposition 1.3.10 (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. Let $E_1 = A_1$. For each $n \in \mathbb{N}$ with $n \geq 2$ let

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

Since $\{A_n\}_{n=1}^\infty \subseteq \mathcal{A}$, by the properties of 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 $n \in \mathbb{N}$, and

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

Hence by the definition and monotonicity of measures (Remark 1.3.3), we obtain that

$$\begin{aligned}
\mu\left(\bigcup_{n=1}^\infty A_n\right) &= \mu\left(\bigcup_{n=1}^\infty E_n\right) \\
&= \sum_{n=1}^\infty \mu(E_n) && \{E_n\}_{n=1}^\infty \text{ pairwise disjoint} \\
&\leq \sum_{n=1}^\infty \mu(A_n) && \text{monotonicity of measures}
\end{aligned}$$

as desired. ∎

1.4 The Lebesgue Outer Measure

With our knowledge of the properties of measures, we turn our attention to attempting to construct a measure that will measure the length of every Borel subset of $\mathbb{R}$ and satisfy (L1)-(L6) from the start of this chapter. The question is, how can we do this?

Our approach will be motivated by the notion of a compact sets. Recall a subset of $\mathbb{R}$ is said to be compact if every open cover has a finite subcover. Our goal is to use the collection of all open covers of a set to measure the length of the set. Since every open set is a disjoint union of open intervals and we know what we want the length of an open interval to be, we know what we want the length of an open set to be. By adding up the lengths of the open sets in an open covering of a set A, we obtain an upper bound for what the length of A should be. To obtain a best approximation for the length of A, we should take the least upper bound. We formalize this as follows.

Definition 1.4.1. 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

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

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

Of course λ^* cannot be the measure we are looking for since it is defined on all of $\mathcal{P}(\mathbb{R})$ and thus cannot satisfy all of (L1)-(L6) by Theorem 1.1.1. However, in order to see how close we are, let's see which of (L1)-(L6) and which properties of measures λ^* satisfies. In particular, λ^* has all of the desired properties except that we only have countable subadditivity instead of additivity on countable pairwise disjoint sets. Since measures must be countable subadditivity, perhaps we are not too far off.

Theorem 1.4.2. *The Lebesgue outer measure satisfies the following:*

a) $\lambda^*(\emptyset) = 0$,

b) $\lambda^*(I) = \ell(I)$ *for all intervals* I,

c) if $A \subseteq B \subseteq \mathbb{R}$, *then* $\lambda^*(A) \leq \lambda^*(B)$,

d) if $x \in \mathbb{R}$ *and* $A \subseteq \mathbb{R}$, *then* $\lambda^*(x + A) = \lambda^*(A)$,

e) if $\alpha \in \mathbb{R}$ *and* $A \subseteq \mathbb{R}$, *then* $\lambda^*(\alpha A) = |\alpha|\lambda^*(A)$, *and*

f) if $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{P}(\mathbb{R})$, *then* $\lambda^*\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \lambda^*(A_n)$.

Proof. It is clear from the definition of λ^* that $\lambda^*(\emptyset) = 0$. Moreover, if $A \subseteq B$, then every collection of open intervals that covers B must also cover A. Therefore, since $\lambda^*(A)$ and $\lambda^*(B)$ are computed via infimums, we obtain that $\lambda^*(A) \leq \lambda^*(B)$ if $A \subseteq B$. Hence a) and c) hold.

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

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

Therefore, since $\epsilon > 0$ 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 such that $I \subseteq \bigcup_{n=1}^{\infty} I_n$. Hence $\{I_n \mid n \in \mathbb{N}\}$ is an open cover of I. Therefore, since I is compact, there must exists a finite subcover of

$\{I_n \mid n \in \mathbb{N}\}$ for I. 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 algorithm here. Otherwise $b_1 \leq b$ so $b_1 \in I$. Since $I \subseteq \bigcup_{k=1}^{m} I_k$, 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 $a_2 < b_1 < b_2$. If $b < b_2$, terminate this algorithm 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 $a_1 < a < b_1$, $a_{k+1} < b_k < b_{k+1}$ for all $1 \leq k \leq m' - 1$, and $a_{m'} < b < b_{m'}$. 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, since $\{I_n \mid n \in \mathbb{N}\}$ was arbitrary, we obtain that $\lambda^*(I) \geq b - a$. Hence $\lambda^*(I) = b - a$ as desired.

To complete the proof of b), first assume $I \subseteq \mathbb{R}$ is an interval of finite length. Thus $I \in \{(a, b), [a, b), (a, b], [a, b]\}$ for some $a, b \in \mathbb{R}$ with $a \leq b$. Hence $\ell(I) = b - a$. Let $\overline{I} = [a, b]$ so that $I \subseteq \overline{I}$ and $\lambda^*(\overline{I}) = \ell(\overline{I}) = b - a$ by the previous case. Next, for any $\epsilon > 0$ with $\epsilon < \frac{b-a}{2}$, let $J_\epsilon = \left[a + \frac{\epsilon}{2}, b - \frac{\epsilon}{2}\right]$. Thus $J_\epsilon \subseteq I$ and $\lambda^*(J_\epsilon) = \ell(J_\epsilon) = b - a - \epsilon$ for all $\epsilon > 0$. Therefore by c) we obtain for all $\epsilon > 0$ that

$$b - a - \epsilon = \lambda^*(J_\epsilon) \leq \lambda^*(I) \leq \lambda^*(\overline{I}) = b - a.$$

Hence $\lambda^*(I) = b - a$ as desired.

Otherwise, assume I is an infinite interval. Since I is an infinite interval, for all $M > 0$ there exists a closed interval J_M such that $J_M \subseteq I$ and $\lambda^*(J_M) = \ell(J_M) = M$. Hence c) implies

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

Therefore, since $M > 0$ was arbitrary, we obtain that $\lambda^*(I) = \infty = \ell(I)$ as desired. Hence b) holds.

To see that d) holds, note $\{I_n \mid n \in \mathbb{N}\}$ is a collection of open intervals such that $A \subseteq \bigcup_{n=1}^{\infty} I_n$ if and only if $\{x + I_n \mid n \in \mathbb{N}\}$ is a collection of open intervals such that $x + A \subseteq \bigcup_{n=1}^\infty x + I_n$. Therefore, since $\ell(x + I) = \ell(I)$ for all open intervals I, it follows that $\lambda^*(x + A) = \lambda^*(A)$ for all $A \in \mathcal{P}(\mathbb{R})$. Hence d) holds.

To see that e) holds, first consider the case that $\alpha = 0$. If $A = \emptyset$ then $\alpha A = \emptyset$ and thus $\lambda^*(\alpha A) = 0 = 0\lambda^*(A) = |\alpha|\lambda^*(A)$. Otherwise, if $A \neq \emptyset$ we obtain that $\alpha A = \{0\}$ so

$$\lambda^*(\alpha A) = \lambda^*(\{0\}) = \lambda^*([0, 0]) = 0 = |\alpha|\lambda^*(A)$$

by b). Otherwise, if $\alpha \neq 0$, we see that $\{I_n \mid n \in \mathbb{N}\}$ is a collection of open intervals such that $A \subseteq \bigcup_{n=1}^\infty I_n$ if and only if $\{\alpha I_n \mid n \in \mathbb{N}\}$ is a collection of open intervals such that $\alpha A \subseteq \bigcup_{n=1}^\infty \alpha I_n$. Therefore, since $\ell(\alpha I) = |\alpha|\ell(I)$ for all open intervals I, it follows that $\lambda^*(\alpha A) = |\alpha|\lambda^*(A)$ for all $A \in \mathcal{P}(\mathbb{R})$. Hence e) holds.

Finally, to see that f) holds, let $\{A_n\}_{n=1}^\infty \subseteq \mathcal{P}(\mathbb{R})$ 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

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

Since countable unions of countable sets are countable (see Appendix B.5.3), $\{I_{n,k} \mid n, k \in \mathbb{N}\}$ is a countable collection of open intervals such that

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

Hence the definition of λ^* implies that

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

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

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

as desired. ∎

1.5 The Carathéodory Method

Theorem 1.4.2 shows us that the Lebesgue outer measure is really close to the object we want; we only have countable subadditivity oppose to countable additivity on disjoint sets. Perhaps the Lebesgue outer measure is good enough for us to do analysis? Well, no because we need countable additivity on disjoint sets to prove the Monotone Convergence Theorem (Theorem 1.3.9), which is definitely something we want to hold.

Thus our only hope to use the Lebesgue outer measure is to restrict the domain from $\mathcal{P}(\mathbb{R})$ to a σ-algebra containing the Borel sets. The question is, how do we do this?

To answer is to invoke a technique known as the Carathéodory Method. This technique only requires three specific properties of the Lebesgue outer measure, which we encapsulate in the following definition to prove the most general result possible.

Definition 1.5.1. Let X be a non-empty set. A function $\mu^* : \mathcal{P}(X) \to [0, \infty]$ is said to be an *outer measure* if

- $\mu^*(\emptyset) = 0$,
- if $A \subseteq B \subseteq X$, then $\mu^*(A) \leq \mu^*(B)$, and
- if $\{A_n\}_{n=1}^\infty \subseteq \mathcal{P}(X)$, then $\mu^*\left(\bigcup_{n=1}^\infty A_n\right) \leq \sum_{n=1}^\infty \mu^*(A_n)$.

Example 1.5.2. By Theorem 1.4.2, the Lebesgue outer measure is an example of an outer measure.

Our hope is to take an outer measure μ^* and form a σ-algebra $\mathcal{A}$ such that $\mu^*|_{\mathcal{A}}$ is a measure. This requires us to describe which sets should be 'measurable'.

Definition 1.5.3. 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 all $B \in \mathcal{P}(X)$ we have

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

Remark 1.5.4. The reason we are interested in outer measurable sets is that if $A \subseteq X$ has the property that

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

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.

Remark 1.5.5. 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 Carathéodory Method of constructing a measure is as follows: construct an outer measure μ^*, and apply the following to get a σ-algebra $\mathcal{A}$ such that $\mu^*|_{\mathcal{A}}$ is a measure.

Theorem 1.5.6. *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. To see that $\mathcal{A}$ is a σ-algebra, 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 $\mathcal{A}$ is closed under compliments and $X \in \mathcal{A}$.

In order to demonstrate that $\mathcal{A}$ is closed under countable unions, let's first verify that $\mathcal{A}$ is closed under finite unions. To verify that $\mathcal{A}$ is closed under finite unions, it suffices to verify that if $A_1, A_2 \in \mathcal{A}$ then $A_1 \cup A_2 \in A$ as we can then apply recursion to take arbitrary finite unions of element of $\mathcal{A}$. Thus let $A_1, A_2 \in \mathcal{A}$ be arbitrary. 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 know 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, since $B \subseteq X$ was arbitrary, we obtain that $A_1 \cup A_2 \in \mathcal{A}$. Hence $\mathcal{A}$ is closed under finite unions.

Since $\mathcal{A}$ is also closed under complements, we also obtain that $\mathcal{A}$ is closed under finite intersections using a similar argument to that used in Remark 1.2.3.

To see that $\mathcal{A}$ is closed under countable unions, let $\{A_n\}_{n=1}^\infty \subseteq \mathcal{A}$ be arbitrary. Let $E_1 = A_1$ and for $n \geq 1$ let

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

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 $\mathcal{A}$ since $\mathcal{A}$ is closed under finite unions. Therefore, since F_n is outer measurable, since $F_n \subseteq E$ so $E^c \subseteq F_n^c$, and since μ^* is monotone, we obtain 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 $n \in \mathbb{N}$.

Since $(F_n)_{n\geq 1}$ are a increasing sequence of sets with union E, we would like to take the limit of the right-hand side of the above inequality to obtain that $\mu^*(B) \geq \mu^*(B \cap E) + \mu^*(B \cap E^c)$ thereby obtaining that E is outer measurable. However, since we do not know the Monotone Convergence Theorem (Theorem 1.3.9) works for outer measures (i.e. the proof required countable additivity on disjoint sets, which we don't have), we will need another approach to taking the limit.

Notice that $F_n = F_{n-1} \cup E_n$ and $F_{n-1} \cap E_n = \emptyset$ by construction. Therefore, since $E_n \in \mathcal{A}$, we obtain that

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

for all $n \in \mathbb{N}$. Therefore recursion implies 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 subadditivity implies that

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

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 $\mu^*|_{\mathcal{A}}$ is a measure, first notice that $\mu^*(\emptyset) = 0$ by design. To check the other property of Definition 1.3.1, 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

$$\mu^*(E) \geq \mu^*(E \cap E^c) + \sum_{k=1}^{\infty} \mu^*(E \cap E_k) = 0 + \sum_{k=1}^{\infty} \mu^*(E_k) = \sum_{k=1}^{\infty} \mu^*(E_k).$$

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 $\mu^*|_{\mathcal{A}}$ is a measure as desired. ∎

Before moving on to studying what Theorem 1.5.6 yields when applied to the Lebesgue outer measure, it is useful to note all measures constructed via the Carathéodory Method have one property in common.

Definition 1.5.7. 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 1.5.8. *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.*

Proof. Assume $A \in \mathcal{P}(X)$ is such that $\mu^*(A) = 0$. 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, $A \in \mathcal{A}$.

To see that $(X, \mathcal{A}, \mu^*|_{\mathcal{A}})$ is complete, let $A \in \mathcal{A}$ and $B \in \mathcal{P}(X)$ be such that $B \subseteq A$ and $\mu^*(A) = 0$. Hence monotonicity implies that $\mu^*(B) = 0$. Thus the first part of this proof implies that $B \in \mathcal{A}$ as desired. ∎

1.6 The Lebesgue Measure

With Carathéodory's Method and the Lebesgue outer measure, we can finally construct the object we desire.

Definition 1.6.1. Let λ^* be the Lebesgue outer measure from Definition 1.4.1. By Theorem 1.5.6 the collection $\mathcal{M}(\mathbb{R})$ of λ^*-measurable sets is a σ-algebra and $\lambda^*|_{\mathcal{M}(\mathbb{R})}$ is a measure. We call $\lambda = \lambda^*|_{\mathcal{M}(\mathbb{R})}$ the *Lebesgue measure* on $\mathbb{R}$ and elements of $\mathcal{M}(\mathbb{R})$ *Lebesgue measurable sets.*

For the Lebesgue measure to be the measure we seek, we still need to verify that λ satisfies (L1)-(L6) and we want to ensure that $\mathcal{M}(\mathbb{R})$ contains the Borel sets. Of course, (L1), (L3), and (L6) automatically hold for λ since λ is a measure. To verify (L2) (i.e. that every interval is Lebesgue measurable and has measure equal to the length), we need only verify that every interval is Lebesgue measurable as Theorem 1.4.2 already implies $\lambda^*(I) = \ell(I)$ for every interval I. Thus we begin with the following.

Theorem 1.6.2. *For each $a \in \mathbb{R}$, (a, ∞) is Lebesgue measurable.*

Proof. To see that (a, ∞) is Lebesgue measurable, 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 countable collections of intervals such that $B_1 \subseteq \bigcup_{n=1}^\infty I'_n$ and $B_2 \subseteq \bigcup_{n=1}^\infty I''_n$. Hence

$$\begin{aligned}
&\lambda^*(B \cap (a,\infty)) + \lambda^*(B \cap (a,\infty)^c) \\
&= \lambda^*(B_1) + \lambda^*(B_2) \\
&\leq \sum_{n=1}^\infty \lambda^*(I'_n) + \sum_{n=1}^\infty \lambda^*(I''_n) && \text{subadditivity} \\
&= \sum_{n=1}^\infty \ell(I'_n) + \sum_{n=1}^\infty \ell(I''_n) && \text{Theorem 1.4.2} \\
&= \sum_{n=1}^\infty \ell(I_n) \\
&\leq \lambda^*(B) + \epsilon.
\end{aligned}$$

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

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

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

Corollary 1.6.3. *Every Borel subset of $\mathbb{R}$ is Lebesgue measurable.*

Proof. Since $\mathcal{M}(\mathbb{R})$ is a σ-algebra, since Theorem 1.6.2 implies $(a,\infty) \in \mathcal{M}(\mathbb{R})$ for all $a \in \mathbb{R}$, and since $\{(a,\infty) \mid a \in \mathbb{R}\}$ generated $\mathfrak{B}(\mathbb{R})$ as a σ-algebra by Remark 1.2.9, it follows that $\mathfrak{B}(\mathbb{R}) \subseteq \mathcal{M}(\mathbb{R})$. ∎

Corollary 1.6.4. *Every interval $I \subseteq \mathbb{R}$ is Lebesgue measurable with $\lambda(I) = \ell(I)$.*

Proof. Since every interval is a Borel subset, every interval is Lebesgue measurable by Corollary 1.6.3. Moreover, if I is an interval, Theorem 1.4.2 implies that $\lambda(I) = \lambda^*(I) = \ell(I)$. ∎

With the above, it remains only to verify that the Lebesgue measure behaves well with respect to translation and scaling.

Proposition 1.6.5. *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. Let $A \in \mathcal{M}(\mathbb{R})$ and $x \in \mathbb{R}$. To see that $x + A$ is Lebesgue measurable, let $B \subseteq \mathbb{R}$ be arbitrary. Since the Lebesgue outer measure is translation invariant, we obtain that

$$\begin{aligned}
\lambda^*(B) &= \lambda^*(-x+B) && \text{by Theorem 1.4.2} \\
&= \lambda^*((-x+B) \cap A) + \lambda^*((-x+B) \cap A^c) && \text{since } A \in \mathcal{M}(\mathbb{R}) \\
&= \lambda^*(B \cap (x+A)) + \lambda^*(B \cap (x+A^c)) && \text{by Theorem 1.4.2} \\
&= \lambda^*(B \cap (x+A)) + \lambda^*(B \cap (x+A)^c).
\end{aligned}$$

Therefore, since $B \subseteq \mathbb{R}$ was arbitrary, $x+A \in \mathcal{M}(\mathbb{R})$. Hence $\lambda(x+A) = \lambda(A)$ by the translation invariance of the Lebesgue outer measure. ∎

Proposition 1.6.6. *If $A \in \mathcal{M}(\mathbb{R})$ and $\alpha \in \mathbb{R}$, then $\alpha A \in \mathcal{M}(\mathbb{R})$ and $\lambda(\alpha A) = |\alpha|\lambda(A)$.*

Proof. Let $A \in \mathcal{M}(\mathbb{R})$ and $\alpha \in \mathbb{R}$. To see that $\alpha A \in \mathcal{M}(\mathbb{R})$, note that if $\alpha = 0$, then $\alpha A = \emptyset$ if $A = \emptyset$ and $\alpha A = \{0\}$ otherwise. In either case $\alpha A \in \mathcal{M}(\mathbb{R})$ when $\alpha = 0$.

If $\alpha \neq 0$, let $B \subseteq \mathbb{R}$ be arbitrary. Then Theorem 1.4.2 implies

$$\begin{aligned}
\lambda^*(B) &= |\alpha|\lambda^*(\alpha^{-1}B) && \text{by Theorem 1.4.2} \\
&= |\alpha|\lambda^*((\alpha^{-1}B) \cap A) + |\alpha|\lambda^*((\alpha^{-1}B) \cap A^c) && A \in \mathcal{M}(\mathbb{R}) \\
&= \lambda^*(\alpha((\alpha^{-1}B) \cap A)) + \lambda^*(\alpha((\alpha^{-1}B) \cap A^c)) && \text{by Theorem 1.4.2} \\
&= \lambda^*(B \cap (\alpha A)) + \lambda^*(B \cap (\alpha A^c)) \\
&= \lambda^*(B \cap (\alpha A)) + \lambda^*(B \cap (\alpha A)^c).
\end{aligned}$$

Therefore, since $B \subseteq \mathbb{R}$ was arbitrary, $\alpha A \in \mathcal{M}(\mathbb{R})$.

Finally, note $\lambda(\alpha A) = |\alpha|\lambda(A)$ by Theorem 1.4.2. ∎

By combining the properties of measures together with Corollary 1.6.3, Corollary 1.6.4, Proposition 1.6.5, and Proposition 1.6.6, we see that the Lebesgue measure satisfies all the properties and conditions we desired! Moreover Proposition 1.5.8 implies that λ is a complete measure.

With the above out of the way, we desire to better understand the Lebesgue measure. In particular, we will examine some additional sets and properties to gain intuition about this measure.

Proposition 1.6.7. *Let $A \subseteq \mathbb{R}$ be countable. Then $A \in \mathcal{M}(\mathbb{R})$ and $\lambda(A) = 0$. Hence $\mathbb{Q} \in \mathcal{M}(\mathbb{R})$ and $\lambda(\mathbb{Q}) = 0$.*

Proof. Let $A \subseteq \mathbb{R}$ be countable. By Proposition 1.5.8 (i.e. the Lebesgue outer measure is complete), it suffices to prove that $\lambda^*(A) = 0$.

To see that $\lambda^*(A) = 0$, let $\epsilon > 0$ be arbitrary. Since A is countable, we can 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, since $\epsilon > 0$ was arbitrary, we obtain that $\lambda^*(A) = 0$ as desired. ∎

For another interesting example, we turn our attention to a very interesting set.

Definition 1.6.8. 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*.

In fact, the Cantor set can be described via the ternary expansion of elements of $[0, 1]$.

Lemma 1.6.9. *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. Let $x \in \mathcal{C}$. Hence $x \in P_n$ for all $n \in \mathbb{N}$. By the recursive construction of the P_n, there exists a sequence $(a_n)_{n\geq 1} \subseteq \{0, 2\}$ such that

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

for all $n \in \mathbb{N}$. To see that $x = \lim_{n\to\infty} \sum_{k=1}^{n} \frac{a_k}{3^k}$, we notice that

$$\left| x - \sum_{k=1}^{n} \frac{a_k}{3^k} \right| \leq \left| \left(\frac{1}{3^n} + \sum_{k=1}^{n} \frac{a_k}{3^k} \right) - \sum_{k=1}^{n} \frac{a_k}{3^k} \right| = \frac{1}{3^n}.$$

Therefore, since $\lim_{n\to\infty} \frac{1}{3^n} = 0$, we obtain that $x = \lim_{n\to\infty} \sum_{k=1}^{n} \frac{a_k}{3^k}$ as desired.

Conversely, assume $x \in \mathbb{R}$ is such that there exists a sequence $(a_n)_{n\geq 1}$ with $a_n \in \{0, 2\}$ for all $n \in \mathbb{N}$ such that $x = \lim_{n\to\infty} \sum_{k=1}^{n} \frac{a_k}{3^k}$. For each $n \in \mathbb{N}$, let $s_n = \sum_{k=1}^{n} \frac{a_k}{3^k}$. Hence, by the description of P_n, we obtain that

$s_n \in P_n$ for all n. In fact, upon closer examination, we see that $s_m \in P_n$ whenever $m \geq n$. Indeed if $m \geq n$ then

$$\begin{aligned}\sum_{k=1}^{n} \frac{a_k}{3^k} \leq \sum_{k=1}^{m} \frac{a_k}{3^k} = s_m &\leq \sum_{k=1}^{n} \frac{a_k}{3^k} + \sum_{k=n+1}^{m} \frac{2}{3^k} \\ &= \sum_{k=1}^{n} \frac{a_k}{3^k} + \frac{2}{3^{n+1}} \frac{1-\left(\frac{1}{3}\right)^{m-n}}{1-\frac{1}{3}} \\ &= \sum_{k=1}^{n} \frac{a_k}{3^k} + \frac{1-\left(\frac{1}{3}\right)^{m-n}}{3^n} \\ &\leq \sum_{k=1}^{n} \frac{a_k}{3^k} + \frac{1}{3^n}.\end{aligned}$$

Since each P_n is a closed set, since $x = \lim_{m\to\infty} s_m$, and since $s_m \in P_n$ whenever $m \geq n$, we obtain that $x \in P_n$ for each $n \in \mathbb{N}$ by the sequential description of closed sets. Hence $x \in \bigcap_{n\geq 1} P_n = \mathcal{C}$. ∎

Remark 1.6.10. The Cantor set has many interesting properties. In particular, the Cantor set is an uncountable set (see Theorem B.5.10) that is compact with empty interior. Since the Cantor set is a closed set and thus a Borel set, the Cantor set is Lebesgue measurable.

We claim that the Cantor set has Lebesgue measure zero. To see this, recall 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, since $\lim_{n\to\infty} \frac{2^n}{3^{n+1}} = 0$, we obtain that $\lambda(C) = 0$ as desired.

Remark 1.6.11. Note Corollary 1.6.3 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 (Theorem B.7.6) implies that $|\mathbb{R}| < |\mathcal{P}(\mathbb{R})|$. 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 Remark 1.6.10 that the Cantor set $\mathcal{C}$ is Lebesgue measurable with $\lambda(\mathcal{C}) = 0$. Hence every subset of the Cantor set must be Lebesgue measurable as the Lebesgue measure is complete. Moreover, since $|\mathcal{C}| = |\mathbb{R}|$ by Theorem B.5.10, we obtain that $|\mathcal{P}(\mathcal{C})| = |\mathcal{P}(\mathbb{R})|$. Therefore, since $\mathcal{P}(\mathcal{C}) \subseteq \mathcal{M}(\mathbb{R}) \subseteq \mathcal{P}(\mathbb{R})$ and since, $|\mathcal{P}(\mathcal{C})| = |\mathcal{P}(\mathbb{R})|$, we obtain that $|\mathcal{M}(\mathbb{R})| = |\mathcal{P}(\mathbb{R})|$.

Thus, in terms of cardinality, the set of Lebesgue measurable subsets of $\mathbb{R}$ is as large as possible.

Of course $\mathcal{M}(\mathbb{R}) \neq \mathcal{P}(\mathbb{R})$ since Theorem 1.1.1 implies there exists (and explicitly constructs) a set $A \subseteq [0, 1)$ that is not Lebesgue measurable. Using this set, we can show there exists $|\mathcal{P}(\mathbb{R})|$ subsets of $\mathbb{R}$ that are not Lebesgue measurable. Indeed $A' = 2 + A \subseteq [2, 3)$ is not Lebesgue measurable being the translation of a set that is not Lebesgue measurable. If $A' \cup \mathcal{C}$ was Lebesgue measurable, then since $A' \cap \mathcal{C} = \emptyset$ we would have $(A' \cup \mathcal{C}) \cap \mathcal{C}^c = A'$ being the intersection of Lebesgue measurable sets and thus being Lebesgue measurable. Since this is a contradiction, we have that $A' \cup \mathcal{C}$ is not Lebesgue measurable. Similarly, if $S \subseteq \mathcal{C}$ then $A' \cup S$ is not Lebesgue measurable. Therefore, since $A' \cap \mathcal{C} = \emptyset$ and as there are $|\mathcal{P}(\mathcal{C})| = |\mathcal{P}(\mathbb{R})|$ subsets of $\mathcal{C}$, we obtain that there are $|\mathcal{P}(\mathbb{R})|$ subsets of $\mathbb{R}$ that are not measurable.

To conclude our initial discussion of the Lebesgue measure, we list several approximation properties.

Proposition 1.6.12. *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* $A \subseteq U\}$. *This property of λ is known as* outer regularity*.*

b) $\lambda(A) = \sup\{\lambda(K) \mid K \subseteq \mathbb{R}$ *is a compact set such that* $K \subseteq A\}$. *This property of λ is known as* inner regularity*.*

Proof. To see that a) is true, let $A \in \mathcal{M}(\mathbb{R})$. Clearly if $U \subseteq \mathbb{R}$ is an open subset such that $A \subseteq U$ then $\lambda(A) \leq \lambda(U)$ by the monotonicity of measures and thus

$$\lambda(A) \leq \inf\{\lambda(U) \mid U \subseteq \mathbb{R} \text{ is an open set such that } A \subseteq U\}.$$

To see the other inequality let $\epsilon > 0$. Since $A \in \mathcal{M}(\mathbb{R})$, we know that $\lambda(A) = \lambda^*(A)$. Hence there exists a countable collection $\{I_n\}_{n=1}^\infty$ of open intervals such that $A \subseteq \bigcup_{n=1}^\infty I_n$ and

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

Therefore, if $U = \bigcup_{n=1}^\infty I_n$, then U is an open subset of $\mathbb{R}$ such that $A \subseteq U$ and

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

Hence

$$\inf\{\lambda(U) \mid U \subseteq \mathbb{R} \text{ is an open set such that } A \subseteq U\} \leq \lambda(A) + \epsilon.$$

Therefore, since $\epsilon > 0$ was arbitrary, we obtain the desire inequality.

To see that b) is true, first note that the difficulty in using a) to directly prove this result is that we have no control of measure of the complement of a set with infinite measure. Thus fix $A \in \mathcal{M}(\mathbb{R})$. Clearly if $K \subseteq \mathbb{R}$ is a compact such that $K \subseteq A$ then $\lambda(K) \leq \lambda(A)$ by the monotonicity of measures and thus

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

For the other direction, for each $n \in \mathbb{N}$ let

$$A_n = A \cap [-n, n].$$

Clearly $A_n \in \mathcal{M}(\mathbb{R})$ and

$$\lambda(A_n) \leq \lambda([-n, n]) \leq 2n < \infty$$

by the monotonicity of measures. Furthermore, since $A = \bigcup_{n=1}^{\infty} A_n$ and $A_n \subseteq A_{n+1}$ for all $n \in \mathbb{N}$, we obtain by the Monotone Convergence Theorem (Theorem 1.3.9) that

$$\lambda(A) = \lim_{n\to\infty} \lambda(A_n).$$

For each $n \in \mathbb{N}$, let $B_n = A_n^c \cap [-n, n]$. Clearly $\lambda(B_n) \leq \lambda([-n, n]) \leq 2n < \infty$ by the monotonicity of measures. By part a) there exists an open subset $U_n \subseteq \mathbb{R}$ such that $B_n \subseteq U_n$ and

$$\lambda(U_n) \leq \lambda(B_n) + \frac{1}{2^n}.$$

Hence, since $\lambda(B_n) < \infty$ so $\lambda(U_n) < \infty$, we obtain that $U_n \cap [-n, n] \in \mathcal{M}(\mathbb{R})$ and

$$0 \leq \lambda(U_n \cap [-n, n]) - \lambda(B_n) \leq \lambda(U_n) - \lambda(B_n) \leq \frac{1}{2^n}.$$

For each $n \in \mathbb{N}$, let $K_n = U_n^c \cap [-n, n]$. Clearly K_n is closed being the intersection of two closed sets and is bounded by n. Hence K_n is compact and $K_n \in \mathcal{M}(\mathbb{R})$. Moreover, since $B_n \subseteq U_n$, we have $K_n = U_n^c \cap [-n, n] \subseteq B_n^c \cap [-n, n] = A_n$. Since

$$[-n, n] = K_n \cup (U_n \cap [-n, n]) \qquad \text{and} \qquad [-n, n] = A_n \cup B_n$$

are disjoint unions of measurable sets, we obtain that

$$\lambda(K_n) + \lambda(U_n \cap [-n, n]) = 2n = \lambda(A_n) + \lambda(B_n)$$

so

$$\lambda(A_n) \leq \lambda(K_n) + \lambda(U_n \cap [-n, n]) - \lambda(B_n) \leq \lambda(K_n) + \frac{1}{2^n}.$$

Therefore, since

$$\lambda(A) = \lim_{n\to\infty} \lambda(A_n) \leq \liminf_{n\to\infty} \lambda(K_n) + \frac{1}{2^n} = \liminf_{n\to\infty} \lambda(K_n),$$

we have that

$$\lambda(A) \leq \sup\{\lambda(K) \mid K \subseteq \mathbb{R} \text{ is a compact set such that } K \subseteq A\}$$

as desired. ■

Proposition 1.6.13. *Let $A \subseteq \mathbb{R}$. The following are equivalent:*

a) $A \in \mathcal{M}(\mathbb{R})$.

b) For all $\epsilon > 0$ there exists an open subset $U \subseteq \mathbb{R}$ such that $A \subseteq U$ and $\lambda^*(U \setminus A) < \epsilon$.

c) For all $\epsilon > 0$ there exists a closed subset $F \subseteq \mathbb{R}$ such that $F \subseteq A$ and $\lambda^*(A \setminus F) < \epsilon$.

d) There exists a G_δ set $G \subseteq \mathbb{R}$ (i.e. G is a countable intersection of open sets) such that $A \subseteq G$ and $\lambda^*(G \setminus A) = 0$.

e) There exists an F_σ set $F \subseteq \mathbb{R}$ (i.e. F is a countable union of closed sets) such that $F \subseteq A$ and $\lambda^*(A \setminus F) = 0$.

Proof. We will show that a), b), and d) are equivalent whereas the equivalence of a), c), and e) will follow by taking complements (i.e. it is easy to see that b) holds if and only if c) holds, and d) holds if and only if e) holds).

Fix $A \subseteq \mathbb{R}$ and assume that d) holds. Notice if $G \subseteq \mathbb{R}$ is a G_δ-set such that $A \subseteq G$ and $\lambda^*(G \setminus A) = 0$, we obtain that $G \setminus A \in \mathcal{M}(\mathbb{R})$ since the Lebesgue measure is complete. Furthermore, since G is G_δ, we obtain that G is Borel and thus $G \in \mathcal{M}(\mathbb{R})$. Therefore, since

$$A = (G \setminus A)^c \cap G$$

and since $\mathcal{M}(\mathbb{R})$ is closed under complements and intersections, we obtain that $A \in \mathcal{M}(\mathbb{R})$. Thus d) implies a).

Next, assume that a) holds so that $A \in \mathcal{M}(\mathbb{R})$. For each $n \in \mathbb{Z}$, let

$$A_n = A \cap [n, n+1].$$

By Proposition 1.6.12 for each $n \in \mathbb{Z}$ and $k \in \mathbb{N}$ there exists an open set $U_{n,k}$ such that $A_n \subseteq U_{n,k}$ and

$$0 \leq \lambda(U_{n,k}) \leq \lambda(A_n) + \frac{1}{k2^{-|n|}}.$$

Hence, since $0 \leq \lambda(A_n) \leq \lambda([n, n+1]) < \infty$ by the monotonicity of measures, we obtain that

$$\lambda(U_{n,k} \setminus A_n) \leq \frac{1}{k2^{-|n|}}.$$

For each $k \in \mathbb{N}$ let
$$U_k = \bigcup_{n\in\mathbb{Z}} U_{n,k}.$$
Clearly U_k is an open set being the countable union of open sets. Furthermore, since $U_k, A \in \mathcal{M}(\mathbb{R})$, we obtain by subadditivity and monotonicity of measures that
$$\begin{aligned}\lambda(U_k \setminus A) &= \lambda\left(\bigcup_{n\in\mathbb{Z}} (U_{n,k} \setminus A)\right)\\ &\leq \sum_{n\in\mathbb{Z}} \lambda(U_{n,k} \setminus A)\\ &\leq \sum_{n\in\mathbb{Z}} \lambda(U_{n,k} \setminus A_n)\\ &\leq \sum_{n\in\mathbb{Z}} \frac{1}{k2^{-|n|}}\\ &= \frac{3}{k}.\end{aligned}$$
Hence b) follows.

To see that b) implies d), note that b) implies for each $k \in \mathbb{N}$ there exists an open set U_k such that $A \subseteq U_k$ and $\lambda(U_k \setminus A) \leq \frac{3}{k}$. Let
$$G = \bigcap_{k=1}^{\infty} U_k.$$
Then G is a G_δ set being the countable intersection of open sets. Thus G is Borel so $G \in \mathcal{M}(\mathbb{R})$. Furthermore, notice for all $k \in \mathbb{N}$ that
$$0 \leq \lambda(G \setminus A) \leq \lambda(U_k \setminus A) \leq \frac{3}{k}$$
by the monotonicity of measures. Hence, since $\lim_{k\to\infty} \frac{3}{k} = 0$, we obtain
$$\lambda^*(G \setminus A) = \lambda(G \setminus A) = 0$$
as desired. ∎

There is far more examples and topics to discuss related to arbitrary measures. For example, we could generalize the Lebesgue measure to obtain the Hausdorff measures on $\mathbb{R}$. The Hausdorff measures can be used to define a dimension function on Borel subsets of the real numbers that give fractional dimensions. For example, it can be show that the Cantor set has dimension $\frac{\ln(2)}{\ln(3)}$. The construction of such objects can be found in Appendix C. For this course, we will be focusing on the theory of the Lebesgue measure and improving the Riemann integral.

Chapter 2

Measurable Functions

With the construction of the Lebesgue measure complete, we return our attention to improving the Riemann integral. However, we immediately run into an issue with the functions we will be able to integrate; just as not every function is Riemann integrable, we cannot expect all functions to be integrable with respect to our new integral. Recall if $A \subseteq \mathbb{R}$, we can define the function $\chi_A : \mathbb{R} \to \mathbb{R}$ by

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

Thinking of an integral as the area under the graph of the function, we would expect $\int_{\mathbb{R}} \chi_A(x)\, dx$ to be the length of A. However, as we have seen, not every subset of $\mathbb{R}$ is Lebesgue measurable. Thus, if A was a set that was not Lebesgue measurable, we would have no way to define $\int_{\mathbb{R}} \chi_A(x)\, dx$. Consequently, we need to examine which functions are 'suitably measurable' and the properties of said functions before we can improve on the Riemann integral.

2.1 Measurable Functions

To define what it means for a function $f : \mathbb{R} \to \mathbb{R}$ to be 'suitably measurable', let's for a moment keep things abstract and take motivation from the topological definition of a continuous function: a function is continuous if the inverse image of an open set is open. By replacing 'open' with 'measurable', we have a potential definition to make a function 'suitably measurable'.

Definition 2.1.1. 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 2.1.2. 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 $A \in \mathcal{A}_Y$.

Example 2.1.3. Let $(X, \mathcal{A}_X)$ and $(Y, \mathcal{A}_Y)$ be measurable spaces and let $f : X \to Y$. If $\mathcal{A}_X = \mathcal{P}(X)$, then f is automatically measurable as $f^{-1}(A) \in \mathcal{P}(X)$ for all $A \in \mathcal{A}_Y$.

Example 2.1.4. Let $(X, \mathcal{A}_X)$ and $(Y, \mathcal{A}_Y)$ be measurable spaces and let $f : X \to Y$. If $\mathcal{A}_Y = \{\emptyset, Y\}$, then f is automatically measurable as $f^{-1}(\emptyset) = \emptyset \in \mathcal{A}_X$ and $f^{-1}(Y) = X \in \mathcal{A}_Y$.

Based on the above examples, there are many options for the notion of measurable functions on $\mathbb{R}$. The question is, "What notion of measurable functions is the correct one to generalize the Riemann integral?"

First, by Example 2.1.4, we see the σ-algebra we place on the co-domain shouldn't be too small for otherwise all functions are forced to be measurable thereby hindering our efforts to construct an integral for measurable functions. Furthermore, provided the σ-algebra on the domain is not too small, we can see by considering the functions χ_A from the beginning of our chapter that the σ-algebra on the domain cannot be too big. As the function χ_A will be of use to us throughout this course, it is about time we give them a name.

Definition 2.1.5. 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 $x \in X$.

Remark 2.1.6. Let $(X, \mathcal{A})$ be a measurable space and let $A \subseteq X$. Notice for a subset $B \subseteq \mathbb{R}$ that

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

Therefore, if $\mathcal{B}$ is a σ-algebra on $\mathbb{R}$ such that there is a set $B \in \mathcal{B}$ with $0 \notin B$ and $1 \in B$ (so $B^c \in \mathcal{B}$, $0 \in B^c$, and $1 \notin B^c$), we see that χ_A is a measurable function from $(X, \mathcal{A})$ to $(\mathbb{R}, \mathcal{B})$ exactly when $A \in \mathcal{A}$.

Based on our goal of integrating measurable functions, the above implies that $\mathcal{M}(\mathbb{R})$ is the largest σ-algebra we should consider for the domain in our definition of measurable functions. The question remains, "What σ-algebra should we take for the co-domain?"

Perhaps we can take $\mathcal{M}(\mathbb{R})$ for the σ-algebra on the co-domain so that our definition is symmetric? Unfortunately, this is not the case. To see this, we require the following peculiar function.

Definition 2.1.7 (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 $x = \sum_{n=1}^{\infty} \frac{a_n}{3^n} \in [0,1]$ for a sequence $\vec{a} = (a_n)_{n\geq 1}$ of elements of $\{0,1,2\}$ and $\vec{b}_{\vec{a}} = (b_n)_{n\geq 1}$ is the sequence of elements of $\{0,1\}$ as defined above, then

$$f(x) = \sum_{n=1}^{\infty} \frac{b_n}{2^n};$$

(That is, write a ternary expansion of x. If N is the first index where a 1 occurs, replace each $\frac{0}{3^n}$ with $n < N$ with $\frac{0}{2^n}$, replace each $\frac{2}{3^n}$ with $n < N$ with $\frac{1}{2^n}$, replace $\frac{1}{3^N}$ with $\frac{1}{2^N}$, and change all terms of index greater than N to zero).

Lemma 2.1.8. *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

$$x = \sum_{k=1}^{m-1} \frac{a_k}{3^k} + \frac{0}{3^m} + \sum_{k=m+1}^{\infty} \frac{2}{3^k} = \sum_{k=1}^{m-1} \frac{a_k}{3^k} + \frac{1}{3^m} + \sum_{k=m+1}^{\infty} \frac{0}{3^k}$$

Note we do not need to include $m = 0$ since as $\sum_{k=1}^{\infty} \frac{2}{3^k}$ is the only ternary expansion of 1 we need to consider in the definition of f.

(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 assume 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 assume 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 2.1.9. *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 $f(\mathcal{C}) = [0,1]$.*

Proof. By Lemma 2.1.8 we know that f is well-defined. Hence for each point in $[0,1]$ with two ternary expansions we can select one to use throughout the proof.

To see that f is constant on $\mathcal{C}^c$, notice by the definition of $\mathcal{C}$ (Definition 1.6.8) that

$$\mathcal{C}^c = \bigcup_{n\geq 0} \bigcup_{a_1,\ldots,a_n\in\{0,2\}} I_{n;a_1,\ldots,a_n}$$

where

$$I_{n;a_1,\ldots,a_n} = \left\{ x = \sum_{k=1}^{\infty} \frac{a'_k}{3^{-k}} \;\middle|\; \substack{a'_k\in\{0,1,2\}, a'_{n+1}=1, \text{ and} \\ a'_k=a_k \text{ for all } k\in\{1,\ldots,n\}} \right\}.$$

Therefore, by the definition of f we see that

$$f(x) = \sum_{k=1}^{n} \frac{\frac{1}{2}a_n}{2^n} + \frac{1}{2^{n+1}}$$

for all $x \in I_{n;a_1,\ldots,a_n}$. Hence f is constant on each interval in $\mathcal{C}^c$.

To see that f is non-decreasing, let $x, y \in [0,1]$ be such that $x < y$ and write the ternary expansions of x and y as

$$x = \sum_{k=1}^{\infty} \frac{a_k(x)}{3^k} \qquad \text{and} \qquad y = \sum_{k=1}^{\infty} \frac{a_k(y)}{3^k}.$$

Since $x \neq y$, due to our assumed uniqueness of the ternary expansions there exists a $q \in \mathbb{N}$ such that $a_q(x) \neq a_q(y)$ and $a_k(x) = a_k(y)$ for all $k < q$. We claim that $a_q(x) < a_q(y)$. Indeed if $a_q(x) > a_q(y)$ then, since

$a_k(x), a_k(y) \in \{0, 1, 2\}$ for all $k \in \mathbb{N}$, we see that

$$\begin{aligned} y - x &= \sum_{k=1}^{\infty} \frac{a_k(y)}{3^k} - \sum_{k=1}^{\infty} \frac{a_k(x)}{3^k} \\ &= \frac{a_q(y) - a_q(x)}{3^q} + \sum_{k=q+1}^{\infty} \frac{a_k(y) - a_k(x)}{3^k} \\ &\leq \frac{-1}{3^q} + \sum_{k=q+1}^{\infty} \frac{a_k(y) - a_k(x)}{3^k} \\ &\leq \frac{-1}{3^q} + \sum_{k=q+1}^{\infty} \frac{2}{3^k} \\ &= 0, \end{aligned}$$

which is a contradiction. Hence $a_q(x) < a_q(y)$.

Using the index q we can show that $f(x) \leq f(y)$. To do this we divide the proof into three cases:

(1) There exists an $k < q$ such that $a_k(x) = a_k(y) = 1$.

(2) Case (1) does not occur and $a_q(x) = 0$ (and thus $a_q(y) \in \{1, 2\}$).

(3) Case (1) does not occur and $a_q(x) = 1$ (and thus $a_q(y) = 2$).

To begin, in all cases write

$$f(x) = \sum_{k=1}^{\infty} \frac{b_k(x)}{2^k} \qquad \text{and} \qquad f(y) = \sum_{k=1}^{\infty} \frac{b_k(y)}{2^k}$$

where the sequences $(b_k(x))_{k\geq 1}$ and $(b_k(y))_{k\geq 1}$ are determined from the sequences $(a_k(x))_{k\geq 1}$ and $(a_k(y))_{k\geq 1}$ via the construction of the Cantor ternary function.

In case (1), note that $(b_k(x))_{k\geq 1} = (b_k(y))_{k\geq 1}$ by definition. Hence $f(x) = f(y)$ as desired.

In case (2), note that $b_k(x) = b_k(y)$ for all $k < q$, that $b_q(x) = 0$, and that $b_q(y) = 1$. Therefore, since $b_k(x), b_k(y) \in \{0, 1\}$ for all $k \in \mathbb{N}$, we see that

$$\begin{aligned} f(y) - f(x) &= \sum_{k=1}^{\infty} \frac{b_k(y)}{2^k} - \sum_{k=1}^{\infty} \frac{b_k(x)}{2^k} \\ &= \frac{1}{2^q} + \sum_{k=q+1}^{\infty} \frac{b_k(y) - b_k(x)}{2^k} \\ &\geq \frac{1}{2^q} + \sum_{k=q+1}^{\infty} \frac{-1}{2^k} \\ &= 0. \end{aligned}$$

Hence $f(x) \leq f(y)$ in case (2).

Finally, in case (3), note that $b_k(x) = b_k(y)$ for all $k < q$, that $b_q(x) = 1$, that $b_k(x) = 0$ for all $k > q$, and that $b_q(y) = 1$. Therefore, since $b_k(x), b_k(y) \in \{0, 1\}$ for all $k \in \mathbb{N}$, we see that

$$\begin{aligned} f(y) - f(x) &= \sum_{k=1}^{\infty} \frac{b_k(y)}{2^k} - \sum_{k=1}^{\infty} \frac{b_k(x)}{2^k} \\ &= \sum_{k=q+1}^{\infty} \frac{b_k(y) - b_k(x)}{2^k} \\ &= \sum_{k=q+1}^{\infty} \frac{b_k(y)}{2^k} \\ &\geq 0. \end{aligned}$$

Hence $f(x) \leq f(y)$ in case (3). Therefore, by combining all of the cases, we obtain that f is non-decreasing and thus monotone.

To see that f is continuous, first notice that f is continuous at each point in $\mathcal{C}^c$ since f is constant on each open interval of $\mathcal{C}^c$. Thus it remains to demonstrate that f is continuous at each point in $\mathcal{C}$. To see this, fix $x \in \mathcal{C}$ and let $\epsilon > 0$. Choose $n \in \mathbb{N}$ such that $\frac{1}{2^n} < \epsilon$. By Definition 1.6.8 there exists $a_1, \ldots, a_n \in \{0, 2\}$ such that

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

Consider the open interval $I = (y, z)$ where

$$y = -\frac{1}{3^n} + \sum_{k=1}^{n} \frac{a_k}{3^k} \qquad \text{and} \qquad z = \frac{2}{3^n} + \sum_{k=1}^{n} \frac{a_k}{3^k}$$

Clearly $x \in I$. We divide the discussion into two cases based on the value of a_n.

Assume $a_n = 0$. Let m be the greatest natural number such that $a_k = 0$ for all $k \geq m$ yet $a_{m-1} \neq 0$ (so $a_{m-1} = 2$). Then

$$f(y) = f\left(\sum_{k=1}^{m-2} \frac{a_k}{3^k} + \frac{1}{3^{m-1}} + \sum_{k=m}^{n-1} \frac{2}{3^m} + \frac{1}{3^n} + \sum_{k=n+1}^{\infty} \frac{2}{3^n}\right) = \sum_{k=1}^{m-2} \frac{\frac{a_k}{2}}{2^k} + \frac{1}{2^{m-1}}$$

whereas

$$f(z) = \sum_{k=1}^{n-1} \frac{\frac{a_k}{2}}{2^k} + \frac{1}{2^n} = \sum_{k=1}^{m-1} \frac{\frac{a_k}{2}}{2^k} + \frac{1}{2^n} = f(y) + \frac{1}{2^n}$$

(since $a_k = 0$ for all $k \geq m$). Therefore, since f is non-decreasing, we see for all $q \in I$ that

$$f(y) \leq f(q) \leq f(z) = f(y) + \frac{1}{2^n}.$$

Hence $|f(x) - f(q)| < \frac{1}{2^n} < \epsilon$ for all $q \in I$ so f is continuous at x.

Otherwise $a_n = 2$. Let m be the greatest natural number such that $a_k = 2$ for all $k \geq m$ yet $a_{m-1} \neq 2$ (so $a_{m-1} = 0$). Then

$$f(z) = f\left(\sum_{k=1}^{m-2} \frac{a_k}{3^k} + \frac{1}{3^{m-1}} + \sum_{k=m}^{n-1} \frac{0}{3^m} + \frac{1}{3^n}\right) = \sum_{k=1}^{m-2} \frac{\frac{a_k}{2}}{2^k} + \frac{1}{2^{m-1}}$$

whereas

$$f(y) = \sum_{k=1}^{n} \frac{\frac{a_k}{2}}{2^k} = \sum_{k=1}^{m-2} \frac{\frac{a_k}{2}}{2^k} + \sum_{k=m}^{n} \frac{1}{2^k} = f(z) - \frac{1}{2^{m-1}} + \sum_{k=m}^{n} \frac{1}{2^k} = f(z) - \frac{1}{2^n}.$$

Therefore, since f is non-decreasing, we see for all $q \in I$ that

$$f(y) \leq f(q) \leq f(z) = f(y) + \frac{1}{2^n}.$$

Hence $|f(x) - f(q)| < \frac{1}{2^n} < \epsilon$ for all $q \in I$ so f is continuous at x. Hence f is continuous on $[0, 1]$.

Finally, clearly $f(0) = 0$ and $f(1) = 1$. Therefore, since f is non-decreasing, the Intermediate Value Theorem immediately implies that $f(\mathcal{C}) = [0, 1]$. ■

With the above properties of the Cantor ternary function, we can now demonstrate why we do not want to use the set of Lebesgue measurable functions for the σ-algebra of the co-domain of measurable functions.

Example 2.1.10. Let f be the Cantor ternary function and define $\psi : [0, 1] \to [0, 2]$ by $\psi(x) = x + f(x)$. Thus ψ is a strictly increasing continuous function.

We claim that $\psi(\mathcal{C})$ is Lebesgue measurable. Indeed since ψ is a continuous function and since $\mathcal{C}$ is compact that $\psi(\mathcal{C})$ is a compact set. Therefore, since compact sets are Lebesgue measurable, $\psi(\mathcal{C})$ is Lebesgue measurable.

Moreover, we claim that $\lambda(\psi(\mathcal{C})) > 0$. To see this, first notice since ψ is a strictly increasing continuous function that if $[a, b] \subseteq [0, 1]$ then $\psi([a, b]) = [\psi(a), \psi(b)]$. Therefore, if $(a, b) \subseteq \mathcal{C}^c$, then since $f(a) = f(b)$ as f is continuous and constant on each interval of $\mathcal{C}^c$ by construction, we obtain that

$$\lambda^*(\psi((a, b))) \leq \lambda^*(\psi([a, b])) = \lambda([\psi(a), \psi(b)]) = \psi(b) - \psi(a) = b - a.$$

Since ψ is strictly increasing (and thus injective), we know that $[0, 2] = \psi(\mathcal{C}) \cup \psi(\mathcal{C}^c)$ and $\psi(\mathcal{C}) \cap \psi(\mathcal{C}^c) = \emptyset$. Therefore, $\mathcal{C}^c$ is a disjoint union of intervals whose sum of lengths is one, the above computation shows that

$$\lambda^*(\psi(\mathcal{C}^c)) \leq 1$$

so $\lambda(\psi(\mathcal{C})) \geq 1 > 0$.

By similar arguments to Theorem 1.1.1 there exists a subset $A \subseteq \mathcal{C}$ such that $B = \psi(A)$ is not Lebesgue measurable.

Since ψ is a strictly increasing continuous function, MATH 2001 implies that $\varphi = \psi^{-1} : [0, 2] \to [0, 1]$ is continuous. However, note A is Lebesgue measurable since $A \subseteq \mathcal{C}$, $\lambda(\mathcal{C}) = 0$, and λ is complete, yet $\varphi^{-1}(A) = \psi(A)$ is not Lebesgue measurable. Hence there is a continuous function on $\mathbb{R}$ such that the inverse image of a Lebesgue measurable set is not Lebesgue measurable.

As continuous functions are the nicest functions we have in analysis and are Riemann integrable, we definitely want the continuous functions to be measurable. Therefore, by Example 2.1.10, we see that using the Lebesgue measurable sets for the co-domain is not the correct notion of a measurable function since it would exclude certain continuous functions from being measurable and thus integrable.

The problem is that the Lebesgue measurable sets is just too large of a σ-algebra to consider for the domain. Thus, to make continuous functions measurable we need to consider a smaller σ-algebra for the co-domain. Since we want to do analysis, it turns out the best thing to do is to take the smallest σ-algebra that contains the open sets; namely the Borel sets.

Definition 2.1.11. Let $A \in \mathcal{M}(\mathbb{R})$. A function $f : A \to \mathbb{R}$ is said to be *Lebesgue measurable* if $f^{-1}(B)$ is Lebesgue measurable for every Borel set B.

Example 2.1.12. Let $A \subseteq \mathbb{R}$. Then $\chi_A : \mathbb{R} \to \mathbb{R}$ is Lebesgue measurable if and only if $A \in \mathcal{M}(\mathbb{R})$ by Remark 2.1.6.

Example 2.1.13. Let $c \in \mathbb{R}$. For an interval $A \in \mathcal{M}(\mathbb{R})$, let $f : A \to \mathbb{R}$ be defined by $f(x) = c$ for all $x \in A$. Then f is Lebesgue measurable since

$$f^{-1}(B) = \begin{cases} \emptyset & \text{if } c \notin B \\ A & \text{if } c \in B \end{cases}$$

and since $\emptyset$ and A are Lebesgue measurable sets.

Example 2.1.13 shows why we want the domain of definition for a Lebesgue measurable function to be a Lebesgue measurable set for otherwise constant functions will not be Lebesgue measurable.

Of course, we still need to verify that continuous functions are Lebesgue measurable. Since open sets are Borel and thus Lebesgue measurable, we know that the inverse image of an open set under a continuous function is open and thus Lebesgue measurable. However, the definition of a continuous function does not provide us with any information about the inverse image of Borel sets. Luckily, the Borel sets are generated as a σ-algebra by the open sets so we can use the following.

Proposition 2.1.14. *Let $E \in \mathcal{M}(\mathbb{R})$, let $f : E \to \mathbb{R}$, and let $A \subseteq \mathfrak{B}(\mathbb{R})$ be such that $\mathfrak{B}(\mathbb{R}) = \sigma(A)$. Then f is Lebesgue measurable if and only if*

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

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

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

$$\mathcal{A} = \{B \subseteq \mathbb{R} \mid f^{-1}(B) \in \mathcal{M}(\mathbb{R})\} \subseteq \mathcal{P}(\mathbb{R}).$$

Thus $A \subseteq \mathcal{A}$ by assumption.

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

$$f^{-1}\left(\bigcup_{n=1}^\infty B_n\right) = \bigcup_{n=1}^\infty f^{-1}(B_n) \in \mathcal{M}(\mathbb{R})$$

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 $\mathfrak{B}(\mathbb{R}) = \sigma(A)$, we obtain that $\mathfrak{B}(\mathbb{R}) \subseteq \mathcal{A}$. Hence f is Lebesgue measurable by definition. ∎

Corollary 2.1.15. *Let $E \in \mathcal{M}(\mathbb{R})$ and let $f : E \to \mathbb{R}$. The following are equivalent:*

a) f is Lebesgue measurable.

b) $f^{-1}(U) \in \mathcal{M}(\mathbb{R})$ for all open subsets $U \subseteq \mathbb{R}$.

c) $f^{-1}((a, \infty)) = \{x \in X \mid f(x) > a\} \in \mathcal{M}(\mathbb{R})$ for all $a \in \mathbb{R}$.

d) $f^{-1}([a, \infty)) = \{x \in X \mid f(x) \geq a\} \in \mathcal{M}(\mathbb{R})$ for all $a \in \mathbb{R}$.

e) $f^{-1}((-\infty, a)) = \{x \in X \mid f(x) < a\} \in \mathcal{M}(\mathbb{R})$ for all $a \in \mathbb{R}$.

f) $f^{-1}((-\infty, a)) = \{x \in X \mid f(x) \leq a\} \in \mathcal{M}(\mathbb{R})$ for all $a \in \mathbb{R}$.

g) $f^{-1}((a, b)) = \{x \in X \mid a < f(x) < b\} \in \mathcal{M}(\mathbb{R})$ for all $a, b \in \mathbb{R}$.

Proof. The result follows from Proposition 2.1.14 since Remark 1.2.9 implies each of the sets used in the inverse images generate $\mathfrak{B}(\mathbb{R})$. ∎

Corollary 2.1.16. *Let $E \in \mathcal{M}(\mathbb{R})$. Every continuous function on E is Lebesgue measurable.*

Proof. Let $f : E \to \mathbb{R}$ be continuous. Since $f^{-1}((a, \infty))$ is the intersection of an open set with E and thus Lebesgue measurable for all $a \in \mathbb{R}$, Corollary 2.1.15 implies f is Lebesgue measurable. ∎

To conclude the basics of Lebesgue measurable functions, we can actually assume the domain of definition of Lebesgue measurable functions is $\mathbb{R}$.

Proposition 2.1.17. *Let $E \in \mathcal{M}(\mathbb{R})$ and let $f : E \to \mathbb{R}$. Define $g : \mathbb{R} \to \mathbb{R}$ by*

$$g(x) = \begin{cases} f(x) & \text{if } x \in E \\ 0 & \text{if } x \notin E \end{cases}.$$

Then f is Lebesgue measurable if and only if g is Lebesgue measurable.

Proof. Note for any Borel set B that

$$g^{-1}(B) = \begin{cases} f^{-1}(B) & \text{if } 0 \notin B \\ f^{-1}(B) \cup E^c & \text{if } 0 \in B \end{cases}.$$

Since $E^c \in \mathcal{M}(\mathbb{R})$ and since $f^{-1}(B) \subseteq E$ for any Borel set B, we see that $f^{-1}(B) \in \mathcal{M}(\mathbb{R})$ implies $f^{-1}(B) \cup E^c \in \mathcal{M}(\mathbb{R})$ and $g^{-1}(B) \in \mathcal{M}(\mathbb{R})$ implies $f^{-1}(B) = (f^{-1}(B) \cup E^c) \cap E \in \mathcal{M}(\mathbb{R})$. Hence $f^{-1}(B) \in \mathcal{M}(\mathbb{R})$ if and only if $g^{-1}(B) \in \mathcal{M}(\mathbb{R})$ for all Borel sets B. Hence the result follows. ∎

By Proposition 2.1.17, we will assume all of our Lebesgue measurable functions are defined on $\mathbb{R}$ unless otherwise specified.

2.2 Operations on Measurable Functions

Before attempt to construct our generalization of the Riemann integrable, it is useful and important to obtain as much information about the Lebesgue measurable functions as possible. After all, Lebesgue measurable functions will be as important to this course as continuous functions were important in MATH 2001 and MATH 3001. Thus we begin by seeing which algebraic and analytic operations preserve the set of Lebesgue measurable functions. This will allow us to extend our known collection of Lebesgue measurable functions.

We begin by showing that the Lebesgue measurable functions form a vector subspace of the real-valued functions that is also closed under multiplication.

Theorem 2.2.1. *Let $f, g : \mathbb{R} \to \mathbb{R}$ be Lebesgue measurable functions. Then*

a) cf is Lebesgue measurable for all $c \in \mathbb{R}$,

b) $f + g$ is Lebesgue measurable, and *c) fg is Lebesgue measurable.*

Proof. To see that a) holds, first note that if $c = 0$ then $cf = 0$ which is Lebesgue measurable. Otherwise, if $c \neq 0$, then for all $a \in \mathbb{R}$ we see that

$$\{x \in \mathbb{R} \mid cf(x) > a\} = \begin{cases} \{x \in \mathbb{R} \mid f(x) > \frac{a}{c}\} & \text{if } c > 0 \\ \{x \in \mathbb{R} \mid f(x) < \frac{a}{c}\} & \text{if } c < 0 \end{cases}.$$

Therefore, since $\{x \in \mathbb{R} \mid f(x) > \frac{a}{c}\}$ and $\{x \in \mathbb{R} \mid f(x) < \frac{a}{c}\}$ are Lebesgue measurable for all $a, c \in \mathbb{R}$ since f is Lebesgue measurable, it follows that cf is Lebesgue measurable.

To see that b) holds, let $a \in \mathbb{R}$ be arbitrary. Notice for $x \in \mathbb{R}$ that $(f+g)(x) > a$ if and only if $f(x) > a - g(x)$. Since the rational numbers are dense in $\mathbb{R}$, we obtain that $(f+g)(x) > a$ if and only if there exists an $r \in \mathbb{Q}$ such that $f(x) > r > a - g(x)$. Since $r > a - g(x)$ if and only if $g(x) > a - r$, we see that

$$\begin{aligned}&\{x \in \mathbb{R} \mid (f+g)(x) > a\} \\ &\qquad = \bigcup_{r \in \mathbb{Q}} \left(\{x \in \mathbb{R} \mid f(x) > r\} \cap \{x \in \mathbb{R} \mid g(x) > a - r\}\right).\end{aligned}$$

Since f and g are Lebesgue measurable, we know that $\{x \in \mathbb{R} \mid f(x) > r\}$ and $\{x \in \mathbb{R} \mid g(x) > a - r\}$ are Lebesgue measurable sets for all $r \in \mathbb{Q}$. Therefore since $\mathcal{M}(\mathbb{R})$ is closed under countable unions and intersections, and since $\mathbb{Q}$ is countable, we obtain that $\{x \in \mathbb{R} \mid (f+g)(x) > a\} \in \mathcal{M}(\mathbb{R})$. Hence, since $a \in \mathbb{R}$ was arbitrary, $f + g$ is Lebesgue measurable.

To see that c) holds, first we claim that if $h : \mathbb{R} \to \mathbb{R}$ is Lebesgue measurable, then h^2 is Lebesgue measurable. Indeed for $a \in \mathbb{R}$ we see that

$$\begin{aligned}&\{x \in \mathbb{R} \mid (h(x))^2 > a\} \\ &\qquad = \begin{cases} \mathbb{R} & \text{if } a < 0 \\ \{x \in \mathbb{R} \mid h(x) > \sqrt{a}\} \cup \{x \in \mathbb{R} \mid h(x) < -\sqrt{a}\} & \text{if } a \geq 0 \end{cases}.\end{aligned}$$

Since h is Lebesgue measurable, we see that $\{x \in \mathbb{R} \mid (h(x))^2 > a\} \in \mathcal{M}(\mathbb{R})$ for all $a \in \mathbb{R}$. Hence h^2 is Lebesgue measurable.

To see that fg is Lebesgue measurable, note since f and g are Lebesgue measurable that $f + g$ is Lebesgue measurable by part b). Hence f^2, g^2, and $(f+g)^2 = f^2 + 2fg + g^2$ are Lebesgue measurable. Therefore, by parts a) and b), we obtain that

$$fg = \frac{1}{2}\left((f+g)^2 - f^2 - g^2\right)$$

is Lebesgue measurable as desired. ∎

Like with continuous functions, we can obtain some information about the composition of Lebesgue measurable functions being Lebesgue measurable.

Proposition 2.2.2. *If $f : \mathbb{R} \to \mathbb{R}$ is Lebesgue measurable, if V is an open set containing the range of f, and if $g : V \to \mathbb{R}$ is continuous, then $g \circ f : \mathbb{R} \to \mathbb{R}$ is Lebesgue measurable.*

Proof. To see that $g \circ f$ is Lebesgue measurable, it suffices by Corollary 2.1.15 to show that

$$(g \circ f)^{-1}(U) = f^{-1}(g^{-1}(U)) \in \mathcal{M}(\mathbb{R})$$

for all open sets U. Thus let $U \subseteq \mathbb{R}$ be an arbitrary open set. Since g is continuous, we obtain that $g^{-1}(U)$ is the intersection of an open set with V and thus is open. Therefore, since f is Lebesgue measurable, $f^{-1}(g^{-1}(U)) \in \mathcal{M}(\mathbb{R})$. Therefore, since U was an arbitrary open set, $g \circ f$ is Lebesgue measurable. ∎

Remark 2.2.3. Clearly the proof of Proposition 2.2.2 breaks down when g is only Lebesgue measurable since the inverse image of a Lebesgue measurable set under a Lebesgue measurable function need not be Lebesgue measurable by Example 2.1.10. To exhibit an example where Proposition 2.2.2 fails when g is only Lebesgue measurable, let φ and A be as in Example 2.1.10 and let $f = \varphi$ and $g = \chi_A$. Since A is Lebesgue measurable, g is Lebesgue measurable by Example 2.1.12. Moreover, since φ is continuous, f is Lebesgue measurable by Corollary 2.1.16. However, note that $\{1\}$ is a Borel set, yet

$$(g \circ f)^{-1}(\{1\}) = f^{-1}(g^{-1}(\{1\})) = f^{-1}(A) = \varphi^{-1}(A)$$

is not Lebesgue measurable. Hence $g \circ f$ is not Lebesgue measurable.

Using Proposition 2.2.2, we easily obtain the following operation preserve the set of Lebesgue measurable functions.

Corollary 2.2.4. *Let $f : \mathbb{R} \to \mathbb{R}$ be a Lebesgue measurable function. Then*

a) $|f|$ is Lebesgue measurable, and

b) $\frac{1}{f}$ is Lebesgue measurable provided $f(x) \neq 0$ for all $x \in \mathbb{R}$.

Proof. Since the functions $a : \mathbb{R} \to \mathbb{R}$ and $q : \mathbb{R} \setminus \{0\} \to \mathbb{K} \setminus \{0\}$ defined by $a(z) = |z|$ and $q(z) = \frac{1}{z}$ are continuous on an open set containing the range of f, Proposition 2.2.2 implies that $|f| = a \circ f$ and $\frac{1}{f} = q \circ f$ are Lebesgue measurable. ∎

Remark 2.2.5. Using Theorem 2.2.1 and Corollary 2.2.4, the theory of Lebesgue measurable functions can often be reduced to analyzing non-negative Lebesgue measurable functions. Indeed if $f : \mathbb{R} \to \mathbb{R}$, define $f_+, f_- : \mathbb{R} \to [0, \infty)$ 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

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

for all $x \in \mathbb{R}$. Hence $|f|(x) = f_+(x) + f_-(x)$ and $f(x) = f_+(x) - f_-(x)$ for all $x \in \mathbb{R}$. Moreover $f_+(x)f_-(x) = 0$ for all $x \in \mathbb{R}$ (i.e. for any $x \in \mathbb{R}$, only one of $f_+(x)$ and $f_-(x)$ can be non-zero). Finally f is Lebesgue measurable if and only if f_+ and f_- are Lebesgue measurable by Theorem 2.2.1 and Corollary 2.2.4. Thus every Lebesgue measurable function is a linear combination of non-negative Lebesgue measurable functions.

Definition 2.2.6. Given a function $f : \mathbb{R} \to \mathbb{R}$, the functions f_+ and f_- in Remark 2.2.5 are call the *positive and negative parts of* f respectively.

Our next goal is to examine how the set of Lebesgue measurable functions behave with respect to limits. Of course, when dealing with limits of functions, often the sequence of functions diverge to $\pm\infty$ at specific points. Consequently, it is useful to extend the notion of measurable functions to allow for infinite values.

Definition 2.2.7. An extended real-valued function $f : \mathbb{R} \to \mathbb{R} \cup \{\pm\infty\}$ is said to be *Lebesgue measurable* if

$$f^{-1}(\{-\infty\}), f^{-1}(\{\infty\}) \in \mathcal{M}(\mathbb{R})$$

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

Remark 2.2.8. It is not difficult to see that characterizations c)-g) of Lebesgue measurable real-valued functions from Corollary 2.1.15 extends to extended real-valued functions. Indeed characterization c) of Corollary 2.1.15 extends since

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

Another reason to use extended real-valued functions is it enables us to take supremums and infimums of functions without worrying about pointwise boundedness. Using limit infimums and supremums, we obtain information on how Lebesgue measurable functions are preserved under limits.

Proposition 2.2.9. *For each $n \in \mathbb{N}$, let $f_n : \mathbb{R} \to [-\infty,\infty]$ be a Lebesgue measurable function. Then the functions*

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

are Lebesgue 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 $f : \mathbb{R} \to [-\infty,\infty]$ is such that $f(x) = \lim_{n\to\infty} f_n(x)$ (that is, f_n converge to f pointwise), then f is Lebesgue measurable.*

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

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

Hence $\sup_{n\geq 1} f_n$ is Lebesgue measurable by Corollary 2.1.15. Similarly, to see that $\inf_{n\geq 1} f_n$ is Lebesgue measurable, notice for all $a \in \mathbb{R}$ that

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

Hence $\inf_{n\geq 1} f_n$ is Lebesgue measurable by Corollary 2.1.15.

To obtain the result for $\limsup_{n\geq 1} f_n$ and $\liminf_{n\geq 1} f_n$, for each $k \in \mathbb{N}$ let

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

Note g_k and h_k are Lebesgue measurable for all k by the above. 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 Lebesgue measurable.

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

Remark 2.2.10. Recall from MATH 2001/3001 that the set of continuous functions was preserved under uniform limits, but not under pointwise limits. Thus, Proposition 2.2.9 shows that, in some sense, the Lebesgue measurable functions all more well-behaved with respect to limits than continuous functions. This will play a role in our generalization of the Riemann integral.

It still may be that pointwise convergence at every point is a lot to ask. However, we are dealing with the Lebesgue measure which determine the length of a set. As sets with zero Lebesgue measure have 'no length', we can imagine these sets would not deter us. Thus it is natural to ask, "Does a sequence of functions that pointwise convergence except on a set of zero Lebesgue measure still yields a Lebesgue measurable function?" This leads us to the following notion.

Definition 2.2.11. Let P be a property that at each point in $\mathbb{R}$ is either true or false. It is said that P holds *almost everywhere* (abbreviated a.e.) if there exists a set $A \in \mathcal{M}(\mathbb{R})$ such that $P(x)$ is true for all $x \in A$ and $\lambda(A^c) = 0$.

Remark 2.2.12. For example, two functions $f, g : \mathbb{R} \to \mathbb{R}$ are said to be equal almost everywhere if there exists a set $A \in \mathcal{M}(\mathbb{R})$ such that $f(x) = g(x)$ for all $x \in A$ and $\lambda(A^c) = 0$. Note this is not necessarily the same as saying

$$\lambda(\{x \in \mathbb{R} \mid f(x) \neq g(x)\}) = 0$$

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

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

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

Example 2.2.13. 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 Lebesgue measure, then $\chi_A = 0$ almost everywhere.

As we hoped for, Lebesgue measurable functions behave well if properties only hold almost everywhere.

Proposition 2.2.14. *Let $f, g : \mathbb{R} \to [-\infty, \infty]$ be such that $f = g$ almost everywhere. If f is Lebesgue measurable, then g is Lebesgue measurable.*

Proof. Let $f, g : \mathbb{R} \to [-\infty, \infty]$ be such that f is Lebesgue measurable and $f = g$ almost everywhere. Hence there exists a set $A \in \mathcal{M}(\mathbb{R})$ such that $f(x) = g(x)$ for all $x \in A$ and $\lambda(A^c) = 0$. Let $B \in \mathfrak{B}(\mathbb{R}) \cup \{\{\infty\}, \{-\infty\}\}$ be arbitrary. Notice

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

since $f(x) = g(x)$ for all $x \in A$. Since $A^c \cap g^{-1}(B) \subseteq A^c$, since $A^c \in \mathcal{M}(\mathbb{R})$, and since $\lambda(A^c) = 0$, we obtain that $A^c \cap g^{-1}(B) \in \mathcal{M}(\mathbb{R})$ since λ is complete. Furthermore, since f is Lebesgue measurable, $f^{-1}(B) \in \mathcal{M}(\mathbb{R})$. Hence, we obtain that $A \cap f^{-1}(B) \in \mathcal{M}(\mathbb{R})$. Hence $g^{-1}(B) \in \mathcal{M}(\mathbb{R})$. Therefore, since $B \in \mathcal{M}(\mathbb{R})$ was arbitrary, g is measurable. ∎

Corollary 2.2.15. *For each $n \in \mathbb{N}$ let $f_n : \mathbb{R} \to [-\infty, \infty]$ be a Lebesgue measurable function. If $f : \mathbb{R} \to [-\infty, \infty]$ 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 Lebesgue measurable.*

Proof. Since $f(x) = \lim_{n\to\infty} f_n(x)$ for a.e. $x \in \mathbb{R}$, there exists a set $A \in \mathcal{M}(\mathbb{R})$ such that $f(x) = \lim_{n\to\infty} f_n(x)$ for all $x \in A$ and $\lambda(A^c) = 0$. Consider the sequence of functions $(f_n\chi_A)_{n\geq 1}$. Clearly $f_n\chi_A$ is Lebesgue measurable for all $n \in \mathbb{N}$ by Theorem 2.2.1 since f_n is Lebesgue measurable and χ_A is Lebesgue measurable as $A \in \mathcal{M}(\mathbb{R})$. Therefore, since

$f(x)\chi_A(x) = \lim_{n\to\infty} f_n(x)\chi_A(x)$ for all $x \in \mathbb{R}$, $f\chi_A$ is Lebesgue measurable by Proposition 2.2.9. Therefore, since $A \in \mathcal{M}(\mathbb{R})$, $f(x)\chi_A(x) = f(x)$ for all $x \in A$, and $\lambda(A^c) = 0$, we see that $f = f\chi_A$ almost everywhere. Hence Proposition 2.2.14 implies that f is Lebesgue measurable. ∎

2.3 Simple Functions

We desire to study Lebesgue measurable functions beyond the properties developed above. We will focus on the non-negative Lebesgue measurable functions since Remark 2.2.5 shows that every Lebesgue measurable function is a linear combination of non-negative Lebesgue measurable functions. However, since Lebesgue measurable functions may appear on the surface to be difficult to describe, it is useful to have a 'simple' collection of Lebesgue measurable functions that are easy to construct and understand. We find such a collection in the following definition.

Definition 2.3.1. A function $\varphi : \mathbb{R} \to [0, \infty)$ is said to be *simple* if there exists an $n \in \mathbb{N}$, non-empty pairwise disjoint sets $\{A_k\}_{k=1}^n \subseteq \mathcal{M}(\mathbb{R})$ such that $\mathbb{R} = \bigcup_{k=1}^n A_k$, and distinct $\{a_k\}_{k=1}^n \subseteq [0, \infty)$ (i.e. $a_i \neq a_j$ whenever $i \neq j$) such that

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

Remark 2.3.2. Note all simple functions are Lebesgue measurable by Example 2.1.12 and Theorem 2.2.1.

Simple functions are the correct Lebesgue measure theoretic analogues of certain functions students may have used in other courses to approximate continuous functions:

Example 2.3.3. 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 $[a, b]$. Clearly every step function is a simple function.

Similar to how every continuous function can be approximated with step functions, our goal is to show that non-negative Lebesgue measurable functions can be approximated by simple functions. Before we get to that, we can discuss another definition for simple functions

Remark 2.3.4. Let $\varphi : \mathbb{R} \to [0, \infty)$ be a Lebesgue measurable with finite range. We claim that φ is a simple function. Indeed write $\varphi(\mathbb{R}) = \{b_1, \ldots, b_m\}$. Since φ is Lebesgue measurable, $A_k = \varphi^{-1}(\{b_k\}) \in \mathcal{M}(\mathbb{R})$ 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 $\{A_k\}_{k=1}^n \subseteq \mathcal{M}(\mathbb{R})$ pairwise disjoint non-empty with $\mathbb{R} = \bigcup_{k=1}^n A_k$.

Since every simple function has finite range, we see that the set of simple functions is precisely the set of Lebesgue measurable functions with finite non-negative range. In particular, the simple functions are closed under addition and non-negative scalar multiplication.

Consequently, if $g : \mathbb{R} \to [0, \infty)$ is such that $g = \sum_{k=1}^{n} a_k \chi_{A_k}$ where $\{A_k\}_{k=1}^{n} \subseteq \mathcal{M}(\mathbb{R})$ and $\{a_k\}_{k=1}^{n} \subseteq [0, \infty)$, then g has finite range and thus is a simple function. Note the description of g differs from that in Definition 2.3.1 since conditions are lacking on $\{A_k\}_{k=1}^{n}$ and on $\{a_k\}_{k=1}^{n}$. The representation of a simple function given in Definition 2.3.1 is called the *canonical representation of a simple function*.

The following demonstrates how simple functions can be used to approximate non-negative Lebesgue measurable functions. Moreover, given two functions $f, g : \mathbb{R} \to \mathbb{R}$, we will use $f \leq g$ to denote that $f(x) \leq g(x)$ for all $x \in \mathbb{R}$.

Theorem 2.3.5. *Let $f : \mathbb{R} \to [0, \infty]$. Then f is Lebesgue measurable if and only if 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.*

Proof. Assume 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 Lebesgue measurable, we obtain that f is Lebesgue measurable by Proposition 2.2.9.

Conversely assume f is Lebesgue measurable. We will proceed by recursively approximating f by dividing up the range of f into interval regions of length $\frac{1}{2^n}$ and approximating f from below. This is accomplished as follows.

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 Lebesgue measurable since f is a Lebesgue measurable function. Moreover, clearly $\{A_{n,k}\}_{k=1}^{n2^n}$ are pairwise disjoint. Furthermore, notice that $x \in B_n$ 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 $n \in \mathbb{N}$ let $\varphi_n : \mathbb{R} \to [0, \infty)$ be defined by

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

Clearly φ_n is a simple function. Moreover $\varphi_n \leq \varphi_{n+1}$ 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 $x \in \mathbb{R}$. If $f(x) < \infty$ then for all $n \in \mathbb{N}$ such that $f(x) < n$ we see that $|f(x) - \varphi_n(x)| \leq \frac{1}{2^n}$

since $f(x) < n$ 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. ∎

Theorem 2.3.5 will be essential to us since having every non-negative Lebesgue measurable function as a pointwise increasing limit of simple functions is quite powerful. However, as pointwise convergence can be weak, it is often useful to have a strong convergence.

2.4 Egoroff's Theorem

In this and the subsequent two sections, we will look at the three Littlewood principles which give us more control over the behaviour of Lebesgue measurable sets and functions. The following Littlewood principle (which is actually the third of Littlewood's principles) enables us to deduce that outside of a set of small Lebesgue measure, pointwise convergence implies uniform convergence.

Theorem 2.4.1 (Egoroff's Theorem). *Let $a, b \in \mathbb{R}$ be such that $a < b$. For each $n \in \mathbb{N}$ let $f_n : [a, b] \to \mathbb{R}$ be a Lebesgue measurable function. If $f : [a, b] \to \mathbb{R}$ is 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 $(f_n)_{n\geq 1}$ converges uniformly to f on $[a, b] \setminus B$.*

Proof. Note f is Lebesgue measurable by Proposition 2.2.9.

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

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

(i.e. $B_{m,k}$ are the 'bad' sets that might prevent uniform convergence). Since f and $(f_n)_{n\geq 1}$ are Lebesgue measurable functions, we see that $B_{m,k}$ is Lebesgue measurable for all $m, k \in \mathbb{N}$. Notice that $B_{m+1,k} \subseteq B_{m,k}$ for all $m, k \in \mathbb{N}$. Moreover, since $f(x) = \lim_{n\to\infty} f_n(x)$ for all $x \in [a, b]$, we see that

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

for all $k \in \mathbb{N}$. Therefore, since $\lambda(\emptyset) = 0$ and $\lambda([a, b]) < \infty$, the Monotone Convergence Theorem (Theorem 1.3.9) implies that

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

for all $k \in \mathbb{N}$. Hence for each $k \in \mathbb{N}$, 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 Lebesgue measurable being the countable union of Lebesgue measurable sets. Furthermore

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

We claim that $(f_n)_{n\geq 1}$ converges uniformly to f on $[a,b] \setminus B$ thereby completing the proof. To see that $(f_n)_{n\geq 1}$ converges uniformly to f on $[a,b] \setminus B$, let $\epsilon > 0$ be arbitrary. Choose $k \in \mathbb{N}$ 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, since $\epsilon > 0$ was arbitrary, we obtain that $(f_n)_{n\geq 1}$ converges uniformly to f on $[a,b] \setminus B$ as desired. ∎

Remark 2.4.2. If in the statement of Egoroff's Theorem (Theorem 2.4.1) one only knew that $f(x) = \lim_{n\to\infty} f_n(x)$ almost everywhere, then the conclusions still hold. Indeed, assume $\delta > 0$ 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 2.4.1) as stated, there exists a Lebesgue measurable set B such that $\lambda(B) < \delta$ and $(\chi_A f_n)_{n\geq 1}$ converges uniformly to $\chi_A f$ on $[a,b] \setminus B$. Hence, if $C = B \cup A^c$, then C is Lebesgue measurable, $\lambda(C) < \delta$, and f_n convergences uniformly to f on $[a,b] \setminus C$ as desired.

Example 2.4.3. The conclusions of Egoroff's Theorem (Theorem 2.4.1) 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

$$|f_n(x)| = |f_n(x) - 0| < \epsilon = 1$$

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 $\lambda(B) = \infty$.

2.5 Littlewood's First Principle

Our next goal in this course is to proof Lusin's Theorem (Theorem 2.6.1), 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 the need for 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.

Theorem 2.5.1 (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 disjoint 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 1.6.12 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, by Remark 1.3.3,

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

Since every open subset of $\mathbb{R}$ is a countable disjoint union of open intervals, we can write $V = \bigcup_{k=1}^\infty I_k$ where each I_k is an open interval and $I_k \cap I_j = \emptyset$ if $k \neq j$. By the Monotone Convergence Theorem for measures (Theorem 1.3.9), 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 implies $\lambda(V \setminus U) < \frac{\epsilon}{2}$. Hence

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

and

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

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

2.6 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 Lebesgue measurable functions. This principle roughly states that 'every Lebesgue measurable function is continuous except on a set of small measure' which is remarkable considering the behaviours and examples of Lebesgue measurable functions we have studied! Formally, we have the following.

Theorem 2.6.1 (Lusin's Theorem). *Let $a, b \in \mathbb{R}$ with $a < b$, and let $f : [a, b] \to \mathbb{R}$ be Lebesgue 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 $g : [a, b] \to \mathbb{R}$ such that

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

and

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

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 of the first part of Lusin's Theorem.

Theorem 2.6.2 (Tietze's Extension Theorem on $\mathbb{R}$). *Let $F \subseteq \mathbb{R}$ be closed and let $h : F \to \mathbb{R}$ be continuous. There exists a continuous function $g : \mathbb{R} \to \mathbb{R}$ 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, F^c is a countable union of disjoint non-empty open intervals. Thus $F^c = \bigcup_{n=1}^{\infty}(a_n, b_n)$ for some $a_n, b_n \in \mathbb{R}$ with $a_n < b_n$. Define $g : \mathbb{R} \to \mathbb{R}$ by

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

for all $x \in \mathbb{R}$. Thus g agrees with h on F and is linear on each (a_n, b_n). Thus 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 prove Lusin's Theorem (Theorem 2.6.1), we will build up the collection of Lebesgue measurable functions for which the conclusions hold starting with the simplest case.

Lemma 2.6.3. *Lusin's Theorem (Theorem 2.6.1) 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 1.6.12, 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 subsets of $[a,b]$ since $\{A_k\}_{k=1}^N$ are pairwise disjoint subsets of $[a,b]$.

Let $F = \bigcup_{k=1}^N F_k \subseteq [a,b]$. Then F is compact (and thus closed) being the finite union of compact (and thus closed) sets. Moreover, since $\lambda([a,b]) < \infty$, $\{A_k\}_{k=1}^N$ are pairwise disjoint, and $\{F_k\}_{k=1}^N$ are pairwise disjoint, we obtain 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, assume $(x_n)_{n\geq 1}$ 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, since $x_n \in F_{k_0}$ 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.

The Tietz Extension Theorem (Theorem 2.6.2) then implies the second conclusion of Lusin's Theorem holds for simple functions. ∎

Using our knowledge of simple functions, we are in a position to prove Lusin's Theorem (Theorem 2.6.1).

Proof of Lusin's Theorem (Theorem 2.6.1). Let $f : [a,b] \to \mathbb{R}$ be an arbitrary Lebesgue measurable function and fix $\epsilon > 0$. Write $f = f_+ - f_-$ where f_+ and f_- are the positive and negative parts of f. Recall that f_+ and f_- are Lebesgue measurable.

By Theorem 2.3.5 there exist sequences of simple functions $(\varphi_{+,n})_{n\geq 1}$ and $(\varphi_{-,n})_{n\geq 1}$ that converge pointwise to f_+ and f_- respectively. For each $n \in \mathbb{N}$ let $f_n = \varphi_{+,n} - \varphi_{-,n}$. Thus $(f_n)_{n\geq 1}$ is a sequence of Lebesgue measurable functions that converge to f pointwise.

Since $(f_n)_{n\geq 1}$ converges pointwise to f, Egoroff's Theorem (Theorem 2.4.1) implies there exists a Lebesgue measurable set $B \subseteq [a,b]$ such that $\lambda(B) < \frac{\epsilon}{4}$ and $(f_n)_{n\geq 1}$ converges uniformly to f on $[a,b]\setminus B$. By Proposition 1.6.12 there exists an open set U 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

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

For each $n \in \mathbb{N}$, Lemma 2.6.3 implies there exists closed sets $F_{+,n}, F_{-,n} \subseteq [a,b]$ such that

$$\lambda([a,b]\setminus F_{+,n}) < \frac{\epsilon}{2^{n+2}} \quad \text{and} \quad \lambda([a,b]\setminus F_{-,n}) < \frac{\epsilon}{2^{n+2}}$$

and continuous functions $g_{1,n}, g_{2,n} : [a,b] \to \mathbb{R}$ such that $g_{1,n}(x) = \varphi_{+,n}(x)$ for all $x \in F_{+,n}$ and $g_{2,n}(x) = \varphi_{-,n}(x)$ for all $x \in F_{-,n}$. Let $F_n = F_{+,n} \cap F_{-,n}$ and let $g_n : [a,b] \to \mathbb{R}$ be defined by $g_n = g_{1,n} - g_{2,n}$. Then F_n is a closed set such that

$$\begin{aligned}
\lambda([a,b]\setminus F_n) &= (b-a) - \lambda(F_{+,n} \cap F_{-,n}) \\
&= (b-a) - (\lambda(F_{+,n}) + \lambda(F_{-,n}) - \lambda(F_{+,n} \cup F_{-,n})) \\
&= (b-a) + \lambda(F_{+,n} \cup F_{-,n}) - \lambda(F_{+,n}) - \lambda(F_{-,n}) \\
&\leq (b-a) + (b-a) - \lambda(F_{+,n}) - \lambda(F_{-,n}) \\
&= \lambda([a,b]\setminus F_{+,n}) + \lambda([a,b]\setminus F_{-,n}) \\
&< \frac{\epsilon}{2^{n+2}} + \frac{\epsilon}{2^{n+2}} = \frac{\epsilon}{2^{n+1}}
\end{aligned}$$

and g_n is a continuous function such that $g_n(x) = f_n(x)$ for all $x \in F_n$.

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

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

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 $(g_n)_{n\geq 1}$ converge uniformly to f on F. Hence $f|_F$ is continuous as desired. ∎

Although Lusin's Theorem (Theorem 2.6.1) appears to rely on the finiteness of the measure used, especially with the use of Egoroff's Theorem (Theorem 2.4.1) in the proof, this is not required as the following result demonstrates.

Theorem 2.6.4 (Lusin's Theorem, Lebesgue measure on $\mathbb{R}$). *Let $f : \mathbb{R} \to \mathbb{R}$ 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

$$\sup(\{|g(x)| \mid x \in \mathbb{R}\}) \leq \sup(\{|f(x)| \mid x \in \mathbb{R}\})$$

and

$$\lambda(\{x \in \mathbb{R} \mid f(x) \neq g(x)\}) < \epsilon.$$

Proof. For each $n \in \mathbb{Z}$, let $A_n = [n, n+1]$. Then $\bigcup_{n\in\mathbb{Z}} A_n = \mathbb{R}$. We will apply Lusin's Theorem (Theorem 2.6.1) to each A_n and stitch together the results.

Let $\epsilon > 0$. Since Lusin's Theorem (Theorem 2.6.1) holds finite closed intervals, for each $n \in \mathbb{Z}$ there exists a closed subset $F_n \subseteq [n, n+1]$ such that $f|_{F_n}$ is continuous and

$$\lambda(A_n \setminus F_n) < \frac{\epsilon}{2^{3+|n|}}.$$

It would be nice to say that f is continuous on $\bigcup_{n\in\mathbb{Z}} F_n$. However, for each $n \in \mathbb{Z}$, $f|_{F_n}$ and $f_{F_{n-1}}$ might have different limits at x. To solve this, we introduce some distance between F_n and F_{n-1}.

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

$$I_n = \left[n + \frac{\epsilon}{2^{4+|n|}}, n + 1 - \frac{\epsilon}{2^{4+|n|}}\right] \quad \text{and} \quad F'_n = F_n \cap I_n.$$

Then F'_n is a closed subset of F_n such that $f|_{F'_n}$ is continuous and

$$\lambda(A_n \setminus F'_n) = \lambda((A_n \setminus F_n) \cup (A_n \setminus I_n)) < \frac{\epsilon}{2^{3+|n|}} + \frac{\epsilon}{2^{3+|n|}} = \frac{\epsilon}{2^{2+|n|}}.$$

Let $F = \bigcup_{n\in\mathbb{Z}} F'_n$. Although a countable union of closed sets need not be closed, F is a closed set. To see this, let $(x_n)_{n\geq 1}$ be a sequence in F that converges to some $x \in \mathbb{R}$. Choose $M \in \mathbb{N}$ such that $x \in (M-1, M+1)$. Thus, since $(x_n)_{n\geq 1}$ converges to x, there exists an $N \in \mathbb{N}$ such that $x_n \in F \cap (M-1, M+1) \subseteq F'_{M-1} \cup F'_M$ for all $n \geq N$. Therefore, since $F'_{M-1} \cup F'_M$ is closed, we must have that $x \in F'_{M-1} \cup F'_M \subseteq F$. Moreover, since the pairwise disjoint closed intervals subsets $\{I_n\}_{n\in\mathbb{Z}}$ have positive separation from one another, since $F'_n \subseteq I_n$, and since $f|_{F'_n}$ is continuous for all n, it follows that $f|_F$ is continuous (i.e. any sequence that is in F must eventually completely lie in I_{n_0} for some n_0 and thus has distance at least $\frac{\epsilon}{2^{4+|n_0|}}$ from any other I_n). Finally, since

$$\lambda(F^c) = \lambda\left(\bigcup_{n\in\mathbb{Z}} A_n \setminus F'_n\right) \leq \sum_{n\in\mathbb{Z}} \lambda(A_n \setminus F'_n) = \sum_{n\in\mathbb{Z}} \frac{\epsilon}{2^{2+|n|}} < \epsilon,$$

the result follows. ∎

Chapter 3

The Lebesgue Integral

With the above study of Lebesgue measurable functions complete, we turn our attention to generalizing the Riemann integral. Naturally, we will call the integral, the *Lebesgue integral.* However, the question remains, "How do we construct the Lebesgue integrals?"

Of course, our approach to the Lebesgue integral must be different than that of the Riemann integral in order to obtain a new integral. Going back to our motivation for studying the length of subsets of the real numbers, we know what we want the Lebesgue integral of χ_A to be for a Lebesgue measurable set A; we want $\lambda(A)$. Consequently, since we want the Lebesgue integral to be linear, we know what we want the Lebesgue integral of a simple function to be. Hopefully this let's us define the Lebesgue integral of a non-negative Lebesgue measurable function via simple functions and Theorem 2.3.5. Thus, since every Lebesgue measurable function is the different of two non-negative Lebesgue measurable functions, and since we want the Lebesgue integral to be linear, we arrive at a definition of the Lebesgue integral. This is the approach we should take.

However, the above approach may be riddled with issues. In particular, if we define the Lebesgue integral this way, we are going to need to be very careful in checking that the Lebesgue integral has the desired properties. It is the complexity of the definition and verification of properties of the Lebesgue integral that causes the Riemann integral to be taught in first-year calculus over the Lebesgue integral.

3.1 The Integral of Simple Functions

Since we want the Lebesgue integral of χ_A to be $\lambda(A)$ and the Lebesgue integral to be linear, we know what we want the Lebesgue integral of a simple function to be.

Definition 3.1.1. Let $\varphi : \mathbb{R} \to [0, \infty)$ be a simple function with canonical representation $\varphi = \sum_{k=1}^{n} a_k \chi_{A_k}$. For every $A \in \mathcal{M}(\mathbb{R})$, we define the

Lebesgue integral of φ over A to be

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

where

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

Example 3.1.2. 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

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

Note that Definition 3.1.1 is a bit cumbersome to use in Example 3.1.2 since we need to know the canonical representation of a simple function. This causes some immediate issues when we attempt to verify that the Lebesgue integral of simple functions has properties we would expect of an integral. For example, if φ and ψ are simple functions, we know that $\varphi + \psi$ will be a simple function by Remark 2.3.4 but the canonical form of $\varphi + \psi$ need not be the sum of the canonical forms. Thus our goal is to show that the formula in Definition 3.1.1 does not depend on the representation of the simple function and Lebesgue integral of simple functions has the desired properties. We begin as follows.

Remark 3.1.3. Let $g : \mathbb{R} \to [0, \infty)$ be such that $g = \sum_{k=1}^n a_k \chi_{A_k}$ where $\{A_k\}_{k=1}^n \subseteq \mathcal{M}(\mathbb{R})$ are pairwise disjoint possibly empty sets with union $\mathbb{R}$, and $\{a_k\}_{k=1}^n \subseteq [0, \infty)$. By Remark 2.3.4 we know that g is a simple function. In particular, Remark 2.3.4 shows that if $g(\mathbb{R}) = \{b_1, \ldots, b_m\}$ and $B_j = g^{-1}(\{b_j\})$ then $g = \sum_{j=1}^m b_j \chi_{B_j}$ is the canonical representation of g. Thus, 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 $A \in \mathcal{M}(\mathbb{R})$, then

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

Hence in Definition 3.1.1 it is not necessary for the $\{a_k\}_{k=1}^n \subseteq [0, \infty)$ to be distinct nor for the A_k to be non-empty.

With Remark 3.1.3, we can verify the Lebesgue integral of simple functions has the desired properties of an integral.

Theorem 3.1.4. *Let $A \in \mathcal{M}(\mathbb{R})$ and let $\varphi, \psi : \mathbb{R} \to [0, \infty)$ be simple functions. Then:*

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

b) $\varphi + \psi$ is a simple function with $\int_A \varphi + \psi \, d\lambda = \int_A \varphi \, d\lambda + \int_A \psi \, d\lambda$.

c) If $B \in \mathcal{M}(\mathbb{R})$ and $B \subseteq A$, then $\int_B \varphi \, d\lambda \leq \int_A \varphi \, d\lambda$.

d) $\varphi\chi_A$ is a simple function with $\int_{\mathbb{R}} \chi_A \varphi \, d\lambda = \int_A \varphi \, d\lambda$.

e) If $\varphi\chi_A \leq \psi\chi_A$, then $\int_A \varphi \, d\lambda \leq \int_A \psi \, d\lambda$.

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 Lebesgue measurable sets with union $\mathbb{R}$ and $\{B_k\}_{k=1}^m$ are pairwise disjoint Lebesgue measurable sets with union $\mathbb{R}$.

To see that a) holds, notice the result is trivial if $c = 0$. Otherwise, if $c > 0$ then

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

is the canonical representation of $c\varphi$. Hence, by definitions,

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

To see that b) holds, 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 Lebesgue measurable sets with union $\mathbb{R}$ 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 3.1.3,

$$\begin{aligned}
\int_A \varphi + \psi \, d\lambda &= \sum_{i=1}^{n}\sum_{j=1}^{m}(a_i + b_j)\lambda(C_{i,j} \cap A) \\
&= \sum_{i=1}^{n} a_i \sum_{j=1}^{m} \lambda(C_{i,j} \cap A) + \sum_{j=1}^{m} b_j \sum_{i=1}^{n} \lambda(C_{i,j} \cap A) \\
&= \sum_{i=1}^{n} a_i\lambda\left(\left(\bigcup_{j=1}^{m} C_{i,j}\right) \cap A\right) + \sum_{j=1}^{m} b_j\lambda\left(\left(\bigcup_{i=1}^{n} C_{i,j}\right) \cap A\right) \\
&= \sum_{i=1}^{n} a_i\lambda(A_i \cap A) + \sum_{j=1}^{m} b_j\lambda(B_j \cap A) \\
&= \int_A \varphi \, d\lambda + \int_A \psi \, d\lambda.
\end{aligned}$$

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

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

as desired.

To see that d) holds, we notice that

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

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

To see that e) holds, note that $\psi\chi_A - \varphi\chi_A$ has finite range and is non-negative and thus a simple function by Remark 2.3.4. Hence by b)

$$\int_{\mathbb{R}} \psi\chi_A \, d\lambda = \int_{\mathbb{R}} \varphi\chi_A + (\psi\chi_A - \varphi\chi_A) \, d\lambda = \int_{\mathbb{R}} \varphi\chi_A \, d\lambda + \int_{\mathbb{R}} \psi\chi_A - \varphi\chi_A \, d\lambda.$$

Therefore, since $\int_X \psi\chi_A - \varphi\chi_A \, d\lambda \geq 0$, the result follows by d). ∎

Using Theorem 3.1.4, we can conclude the representation of a simple function does not effect the Lebesgue integral.

Corollary 3.1.5. *Let $\varphi : \mathbb{R} \to [0, \infty]$ be such that $\varphi = \sum_{k=1}^n a_k \chi_{A_k}$ where $\{A_k\}_{k=1}^n \subseteq \mathcal{M}(\mathbb{R})$ and $\{a_k\}_{k=1}^n \subseteq [0, \infty)$. Then for all $A \in \mathcal{M}(\mathbb{R})$,*

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

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

Proof. If $E \in \mathcal{M}(\mathbb{R})$ and $E \notin \{\mathbb{R}, \emptyset\}$, then $\chi_E = 1\chi_E + 0\chi_{E^c}$ is the canonical representation of χ_E. Therefore

$$\int_A \chi_E \, d\lambda = 1\lambda(E \cap A) + 0\lambda(E^c \cap A) = \lambda(E \cap A).$$

Since similar equations hold with $E \in \{\mathbb{R}, \emptyset\}$, the result then follows from the fact that the Lebesgue integral for simple functions is additive and respects non-negative scalar multiplication (i.e. parts a) and b) of Theorem 3.1.4). ∎

3.2 The Integral of Non-Negative Functions

Our next goal is to extend the Lebesgue integral of simple functions to non-negative Lebesgue measurable functions. To do so, we must use some form of approximation. Although Riemann integral was obtained by approximating the area under the curve from above and below, we will just use Theorem 2.3.5 and approximate from below.

Definition 3.2.1. Let $A \in \mathcal{M}(\mathbb{R})$ and let $f : \mathbb{R} \to [0, \infty]$ be Lebesgue measurable. The *Lebesgue integral of f over A* is defined to be

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

Remark 3.2.2. One incredibly subtlety that we need to be careful of is that every simple function is a non-negative Lebesgue measurable function and thus we have two definitions for the Lebesgue integral of a simple function:

Definition 3.1.1 and Definition 3.2.1. We better make sure these definitions agree.

Let $A \in \mathcal{M}(\mathbb{R})$ and let $\psi : \mathbb{R} \to [0,\infty)$ be a simple function. Let $\alpha = \int_A \psi \, d\lambda$ when we evaluate the integral viewing ψ as a simple function and let $\beta = \int_A \psi \, d\lambda$ when we evaluate the integral viewing ψ as a non-negative Lebesgue measurable function. By Definition 3.2.1, we see using $\varphi = \psi$ that $\alpha \leq \beta$. However, if $\varphi : \mathbb{R} \to [0,\infty)$ is a simple function such that $\varphi \leq \psi$, we obtain by part e) of Theorem 3.1.4 that

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

Hence taking the supremum in Definition 3.2.1 yields $\beta \leq \alpha$ so $\alpha = \beta$. Thus the two definitions for the integral of a simple function are equal.

Using Theorem 3.1.4, several properties of integrating simple functions transfer to integrating non-negative measurable functions.

Theorem 3.2.3. *Let $A \in \mathcal{M}(\mathbb{R})$ and let $f, g : \mathbb{R} \to [0,\infty]$ be Lebesgue measurable functions. Then:*

a) if $c \geq 0$ then $\int_A cf \, d\lambda = c \int_A f \, d\lambda$.

b) If $B \in \mathcal{M}(\mathbb{R})$ and $B \subseteq A$, then $\int_B f \, d\lambda \leq \int_A f \, d\lambda$.

c) $\int_{\mathbb{R}} \chi_A f \, d\lambda = \int_A f \, d\lambda$.

d) If $f\chi_A \leq g\chi_A$, then $\int_A f \, d\lambda \leq \int_A g \, d\lambda$.

e) $\int_A f \, d\lambda = 0$ if and only if $\lambda(\{x \in \mathbb{R} \mid f(x) > 0\} \cap A) = 0$.

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

Proof. Note a) clearly holds when $c = 0$. Otherwise if $c > 0$, it is clear that if $\varphi : \mathbb{R} \to [0,\infty)$ is a simple function and $\varphi \leq f$ then $c\varphi$ is a simple function and $c\varphi \leq cf$. Hence, since Theorem 3.1.4 implies

$$c \int_A \varphi \, d\lambda = \int_A c\varphi \, d\lambda \leq \int_A cf \, d\lambda,$$

we obtain that $c \int_A f \, d\lambda \leq \int_A cf \, d\lambda$. Similarly, if $\varphi : \mathbb{R} \to [0,\infty)$ is a simple function and $\varphi \leq cf$ then $\frac{1}{c}\varphi$ is a simple function and $\frac{1}{c}\varphi \leq f$. Hence, since Theorem 3.1.4 implies

$$\frac{1}{c} \int_A \varphi \, d\lambda = \int_A \frac{1}{c}\varphi \, d\lambda \leq \int_A f \, d\lambda \qquad \text{so} \qquad \int_A \varphi \, d\lambda \leq c \int_A f \, d\lambda$$

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

Note b) clearly follows from Theorem 3.1.4 and d) follows by Definition 3.2.1 once c) is complete. Moreover f) follows from e).

To see c) holds, notice by Theorem 3.1.4

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

as desired. [Note the third equality holds since if φ is a simple function and $\varphi \leq f$, then $\psi = \chi_A\varphi$ is a simple function and $\psi \leq \chi_A f$, and if ψ is a simple function and $\psi \leq \chi_A f$, then $\psi(x) = 0$ for all $x \notin A$ so $\psi = \psi\chi_A$ is a simple function and $\psi \leq f$.]

To see that e) is true, let $B = \{x \in \mathbb{R} \mid f(x) > 0\} \cap A$. Notice B is Lebesgue measurable as A is measurable and f is Lebesgue measurable.

Assume $\int_A f\,d\lambda = 0$. For each $n \in \mathbb{N}$ let

$$A_n = \left\{x \in \mathbb{R} \,\middle|\, f(x) > \frac{1}{n}\right\}.$$

Since f is Lebesgue measurable, A_n is Lebesgue 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 Lebesgue integral implies that

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

Hence $\lambda(A_n \cap A) = 0$ for all $n \in \mathbb{N}$. Since f is non-negative, we know that

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

Hence, by the subadditivity of measures (Proposition 1.3.10), we obtain that $\lambda(B) = 0$.

Conversely, assume $\lambda(B) = 0$. Suppose $\varphi : \mathbb{R} \to [0,\infty)$ 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 $k \in \{1,\ldots,n\}$. Since $\varphi \leq f$, we see that

$$A_k \subseteq \{x \in \mathbb{R} \mid f(x) > 0\}.$$

Thus the monotonicity of measures implies that

$$\lambda(A_k \cap A) \leq \lambda(B) = 0.$$

Hence

$$\int_A \varphi\,d\lambda = \sum_{k=1}^n a_k\lambda(A_k \cap A) = 0.$$

Therefore, by the definition of the Lebesgue integral, $\int_A f\,d\lambda = 0$. ∎

Remark 3.2.4. Since Theorem 3.2.3 implies that $\int_{\mathbb{R}} \chi_A f \, d\lambda = \int_A f \, d\lambda$, it suffices to consider only integrals over all of $\mathbb{R}$ when developing the theory of the Lebesgue integral as the results for integrating over an arbitrary Lebesgue measurable set A will then follow from multiplying the functions under consideration by χ_A. Note multiplying by χ_A is linear and preserves pointwise limits.

One omission in Theorem 3.2.3 is the additivity of integrals:

$$\int_{\mathbb{R}} f + g \, d\lambda = \int_{\mathbb{R}} f \, d\lambda + \int_{\mathbb{R}} g \, d\lambda.$$

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 3.1.4 clearly implies

$$\int_X f \, d\lambda + \int_X g \, d\lambda \leq \int_X f + g \, d\lambda.$$

However, difficulty occurs with the reverse inequality since 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 $\varphi_1 + \varphi_2 = \varphi$?

3.3 The Monotone Convergence Theorem

In order to try and demonstrate the additivity of the Lebesgue integral of non-negative functions, we turn our attention to Theorem 2.3.5. We know every non-negative Lebesgue measurable function is the pointwise limit of an increasing sequence of simple functions. If we knew that the Lebesgue integral preserved these limits, then we would obtain

$$\int_{\mathbb{R}} f \, d\lambda + \int_{\mathbb{R}} g \, d\lambda = \int_{\mathbb{R}} f + g \, d\lambda$$

for all Lebesgue measurable functions $f, g : \mathbb{R} \to [0, \infty]$ since the Lebesgue integral is additive for simple functions, and since the limit of a sum is the sum of the limit. Thus our first goal is to show that the Lebesgue integral for non-negative Lebesgue measurable functions preserves monotone limits; that is, we want a Monotone Convergence Theorem for the Lebesgue integral of non-negative Lebesgue measurable functions.

To prove our Monotone Convergence Theorem, we will make use of the Monotone Convergence Theorem for measures (Theorem 1.3.9) since, as it turns out, integration against a simple function is a measure!

Lemma 3.3.1. *Let $\varphi : \mathbb{R} \to [0, \infty)$ be a simple function. If $\mu : \mathcal{M}(\mathbb{R}) \to [0, \infty]$ is defined by*

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

for all $A \in \mathcal{M}(\mathbb{R})$, then μ is a measure on $(\mathbb{R}, \mathcal{M}(\mathbb{R}))$.

Proof. Note part f) of Theorem 3.2.3 implies that $\mu(\emptyset) = 0$.

To see that μ is countably additive on pairwise disjoint subsets, let $\{B_m\}_{m=1}^\infty \subseteq \mathcal{M}(\mathbb{R})$ be pairwise disjoint. Since φ is a simple function, there exists $\{A_k\}_{k=1}^n \subseteq \mathcal{M}(\mathbb{R})$ and let $\{a_k\}_{k=1}^n \in [0, \infty)$ such that

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

Therefore

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

Hence μ is a measure as desired. ∎

It is time to use the Monotone Convergence Theorem to prove the Monotone Convergence Theorem! Moreover, this is likely the first non-contrived result students will have seen where a limit is shown to exist and we compute its value by making use of the limit infimum and limit supremum.

Theorem 3.3.2 (The Monotone Convergence Theorem). *For each $n \in \mathbb{N}$ let $f_n : \mathbb{R} \to [0, \infty]$ be a Lebesgue measurable function such that $f_n \leq f_{n+1}$ for all $n \in \mathbb{N}$. If $f : \mathbb{R} \to [0, \infty]$ is the pointwise limit of $(f_n)_{n\geq 1}$, then f is Lebesgue measurable and for all $A \in \mathcal{M}(\mathbb{R})$*

$$\int_A f \, d\lambda = \lim_{n\to\infty} \int_A f_n \, d\lambda.$$

Proof. First note since f is the pointwise limit of Lebesgue measurable functions that f is Lebesgue measurable by Proposition 2.2.9. Next note Remark 3.2.4 implies we may assume that $A = \mathbb{R}$ since multiplying by a characteristic function will preserve measurability, pointwise limits, and the value of the Lebesgue integral by Theorem 3.2.3.

Since $f_n \leq f$ for all $n \in \mathbb{N}$, Theorem 3.2.3 implies that

$$\int_{\mathbb{R}} f_n \, d\lambda \leq \int_{\mathbb{R}} f \, d\lambda$$

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

$$\limsup_{n\to\infty} \int_{\mathbb{R}} f_n \, d\lambda \leq \int_{\mathbb{R}} f \, d\lambda.$$

Thus, to complete the proof, it suffices to show that

$$\int_{\mathbb{R}} f \, d\lambda \leq \liminf_{n\to\infty} \int_{\mathbb{R}} f_n \, d\lambda.$$

In order to facilitate some 'wiggle room', we will show that

$$\alpha \int_{\mathbb{R}} f \, d\lambda \leq \liminf_{n\to\infty} \int_{\mathbb{R}} f_n \, d\lambda$$

for all $\alpha \in (0, 1)$ from which the desired inequality will follow by take the limit $\alpha \to 1$.

To obtain the desired inequality, fix $\alpha \in (0, 1)$. Let $\varphi : \mathbb{R} \to [0, \infty)$ be an arbitrary simple function such that $\varphi \leq f$. Thus, if we can prove that

$$\alpha \int_{\mathbb{R}} \varphi \, d\lambda \leq \liminf_{n\to\infty} \int_{\mathbb{R}} f_n \, d\lambda,$$

the proof will be complete by the definition of the Lebesgue integral of f (Definition 3.2.1).

Notice $\alpha\varphi$ is a simple function such that $\alpha\varphi \leq f$. For each $n \in \mathbb{N}$, let

$$A_n = \{x \in \mathbb{R} \mid f_n(x) - \alpha\varphi(x) \geq 0\}.$$

Since each $f_n - \alpha\varphi$ is a Lebesgue measurable function, A_n is Lebesgue measurable for all $n \in \mathbb{N}$. Moreover, by Theorem 3.2.3, we have for all $n \in \mathbb{N}$ that

$$\begin{aligned}
\alpha \int_{A_n} \varphi \, d\lambda &= \int_{A_n} \alpha\varphi \, d\lambda && \text{by Theorem 3.2.3, part a)} \\
&\leq \int_{A_n} f_n \, d\lambda && \text{since } \alpha\varphi\chi_{A_n} \leq f_n\chi_{A_n} \\
&\leq \int_{\mathbb{R}} f_n \, d\lambda && \text{since } A_n \subseteq \mathbb{R} \\
&\leq \liminf_{k\to\infty} \int_{\mathbb{R}} f_k \, d\lambda && \begin{array}{c}\text{since } f_k \leq f_{k+1} \text{ so } \left(\int_{\mathbb{R}} f_k \, d\lambda\right)_{k\geq 1} \\ \text{is an increasing sequence.}\end{array}
\end{aligned}$$

Thus, to complete the proof, it suffices to replace A_n with $\mathbb{R}$ in the above inequality.

Since $f_n \leq f_{n+1}$ for all $n \in \mathbb{N}$, clearly $A_n \subseteq A_{n+1}$ for all $n \in \mathbb{N}$. We claim that

$$\mathbb{R} = \bigcup_{n\geq 1} A_n.$$

To see this, let $x \in \mathbb{R}$ 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 $f(x) > \alpha\varphi(x)$ since $\alpha < 1$ (this is why we needed the wiggle room). Hence, since $\lim_{n\to\infty} f_n(x) = f(x)$, 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 $\mathbb{R} = \bigcup_{n\geq 1} A_n$.

Let $\mu : \mathbb{R} \to [0, \infty]$ be defined by

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

for all $A \in \mathcal{M}(\mathbb{R})$. Since φ is a simple function, Lemma 3.3.1 implies that μ is a measure on $(\mathbb{R}, \mathcal{M}(\mathbb{R}))$. Therefore, since $\{A_n\}_{n=1}^\infty$ is an increasing sequence of Lebesgue measurable sets with $\mathbb{R} = \bigcup_{n\geq 1} A_n$, the Monotone Convergence Theorem for measures (Theorem 1.3.9) implies that

$$\begin{aligned}
\alpha \int_{\mathbb{R}} \varphi \, d\lambda &= \alpha\mu(\mathbb{R}) \\
&= \alpha \lim_{n\to\infty} \mu(A_n) \\
&= \alpha \lim_{n\to\infty} \int_{A_n} \varphi \, d\lambda \\
&\leq \liminf_{k\to\infty} \int_{\mathbb{R}} f_k \, d\lambda.
\end{aligned}$$

Hence the proof is complete. ∎

Using the Monotone Convergence Theorem (Theorem 3.3.2), we easily obtain the following final properties of the Lebesgue integrals of non-negative functions that we desired.

Theorem 3.3.3. *Let $f, g : \mathbb{R} \to [0, \infty]$ be Lebesgue measurable functions. Then:*

a) $\int_{\mathbb{R}} f + g \, d\lambda = \int_{\mathbb{R}} f \, d\lambda + \int_{\mathbb{R}} g \, d\lambda$.

b) If $f = g$ a.e., then $\int_{\mathbb{R}} f \, d\lambda = \int_{\mathbb{R}} g \, d\lambda$.

Proof. To see that a) holds, note by Theorem 2.3.5 there exists increasing sequences of simple functions $(\varphi_n)_{n\geq 1}$ and $(\psi_n)_{n\geq 1}$ 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 3.3.2) twice along with the additivity of integrals of simple functions from Theorem 3.1.4, we obtain that

$$\begin{aligned}\int_{\mathbb{R}} f + g \, d\lambda &= \lim_{n\to\infty} \int_{\mathbb{R}} \varphi_n + \psi_n \, d\lambda \\ &= \lim_{n\to\infty} \int_{\mathbb{R}} \varphi_n \, d\lambda + \int_{\mathbb{R}} \psi_n \, d\lambda \\ &= \int_{\mathbb{R}} f \, d\lambda + \int_{\mathbb{R}} g \, d\lambda.\end{aligned}$$

To see that b) holds, let $B \in \mathcal{M}(\mathbb{R})$ be such that $f(x) = g(x)$ for all $x \in B$ and $\lambda(B^c) = 0$. Thus $f\chi_B = g\chi_B$. Since $\lambda(B^c) = 0$, Theorem 3.1.4 implies that

$$\int_{B^c} f \, d\lambda = \int_{B^c} g \, d\lambda = 0$$

Hence we see that

$$\begin{aligned}\int_{\mathbb{R}} f \, d\lambda &= \int_{\mathbb{R}} f(\chi_B + \chi_{B^c}) \, d\lambda \\ &= \int_{\mathbb{R}} f\chi_B \, d\lambda + \int_{\mathbb{R}} f\chi_{B^c} \, d\lambda \\ &= \int_{\mathbb{R}} f\chi_B \, d\lambda + \int_{B^c} f \, d\lambda \\ &= \int_{\mathbb{R}} g\chi_B \, d\lambda + \int_{B^c} g \, d\lambda \\ &= \int_{\mathbb{R}} g\chi_B \, d\lambda + \int_{\mathbb{R}} g\chi_{B^c} \, d\lambda = \int_{\mathbb{R}} g \, d\lambda\end{aligned}$$

as desired. ∎

Remark 3.3.4. By using part b) of Theorem 3.3.3 and the fact that the integral of any non-negative Lebesgue measurable function against a set of Lebesgue measure zero is zero by part f) of Theorem 3.2.3, it follows that the Monotone Convergence Theorem (Theorem 3.3.2) also holds if the condition "$f : \mathbb{R} \to [0, \infty]$ is the pointwise limit of $(f_n)_{n\geq 1}$" is replaced with the condition "$f(x) = \lim_{n\to\infty} f_n(x)$ almost everywhere".

The Monotone Convergence Theorem (Theorem 3.3.2) can also be used demonstrate the Lebesgue integral behaves well with respect to series of non-negative Lebesgue measurable functions.

Corollary 3.3.5. *For each $n \in \mathbb{N}$ let $f_n : \mathbb{R} \to [0, \infty]$ be a Lebesgue measurable function. If $f : \mathbb{R} \to [0, \infty]$ is such that*

$$f(x) = \sum_{n=1}^{\infty} f_n(x)$$

for almost every $x \in \mathbb{R}$, then f is Lebesgue measurable and

$$\int_{\mathbb{R}} f \, d\lambda = \sum_{n=1}^{\infty} \int_{\mathbb{R}} f_n \, d\lambda.$$

Proof. Note f is Lebesgue measurable by Corollary 2.2.15 being the almost everywhere pointwise limit of Lebesgue measurable functions.

For each $m \in \mathbb{N}$, let $g_m : \mathbb{R} \to [0, \infty]$ be defined by $g_m = \sum_{n=1}^m f_n$. Clearly $(g_m)_{m\geq 1}$ is an increasing sequence of non-negative Lebesgue measurable functions that converges to f pointwise almost everywhere. Hence the Monotone Convergence Theorem (Theorem 3.3.2) implies

$$\int_{\mathbb{R}} f \, d\lambda = \lim_{m\to\infty} \int_{\mathbb{R}} g_m \, d\lambda = \lim_{m\to\infty} \sum_{n=1}^{m} \int_{\mathbb{R}} f_n \, d\lambda = \sum_{n=1}^{\infty} \int_{\mathbb{R}} f_n \, d\lambda$$

as desired. ∎

Corollary 3.3.6. *If $f : \mathbb{R} \to [0, \infty]$ is Lebesgue measurable and $\{A_n\}_{n=1}^\infty \subseteq \mathcal{M}(\mathbb{R})$ are pairwise disjoint, then*

$$\sum_{n=1}^{\infty} \int_{A_n} f \, d\lambda = \int_{\bigcup_{n=1}^\infty A_n} f \, d\lambda.$$

In particular, since $\int_\emptyset f \, d\lambda = 0$, the function $\mu : \mathcal{M}(\mathbb{R}) \to [0, \infty]$ defined by

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

for all $A \in \mathcal{M}(\mathbb{R})$ is a measure on $(\mathbb{R}, \mathcal{M}(\mathbb{R}))$.

Proof. Since $\{A_n\}_{n=1}^\infty \subseteq \mathcal{M}(\mathbb{R})$ are pairwise disjoint, we see that

$$\chi_{\bigcup_{n=1}^\infty A_n} f = \sum_{n=1}^{\infty} \chi_{A_n} f.$$

Hence Corollary 3.3.5 implies that

$$\sum_{n=1}^{\infty} \int_{A_n} f \, d\lambda = \sum_{n=1}^{\infty} \int_{\mathbb{R}} \chi_{A_n} f \, d\lambda = \int_{\mathbb{R}} \chi_{\bigcup_{n=1}^\infty A_n} f \, d\lambda = \int_{\bigcup_{n=1}^\infty A_n} f \, d\lambda.$$ ∎

3.4 The Lebesgue Integral

As the above notion of the Lebesgue integral for non-negative Lebesgue measurable functions has all of our desired properties, we now turn to extended this notion to all Lebesgue measurable functions. We have see that if f is a Lebesgue measurable function, then we can write $f = f_+ - f_-$ where f_+ and f_- are non-negative Lebesgue measurable function. If we want the Lebesgue integral to be linear, we need to define the Lebesgue integral of f to be the difference of the Lebesgue integrals of f_+ and f_-. However, we run into an immediate issue: "what should $\infty - \infty$ be defined to be?" After all, we have allowed non-negative Lebesgue measurable functions to have infinite integrals.

To solve this problem, we will avoid this problem. Of course, it is never a good idea to ignore ones problems, but sometimes this is the best we can do in mathematics. We can solve/avoid this problem by restricting to a specific collection of the Lebesgue measurable functions so that we never end up in the "$\infty - \infty$" setting.

Definition 3.4.1. A Lebesgue measurable function $f : \mathbb{R} \to [-\infty, \infty]$ is said to be *Lebesgue integrable* if

$$\int_{\mathbb{R}} |f| \, d\lambda < \infty.$$

Before defining the Lebesgue integral of a Lebesgue integrable function, we note some important properties and simplifications of Lebesgue integrable functions.

Remark 3.4.2. Let $f : \mathbb{R} \to [-\infty, \infty]$ be Lebesgue measurable. Thus $|f|$ is Lebesgue measurable by Corollary 2.2.4 so Definition 3.4.1 is well-defined. Moreover, since $|f| = f_+ + f_-$, and since

$$\int_{\mathbb{R}} |f| \, d\lambda = \int_{\mathbb{R}} f_+ \, d\lambda + \int_{\mathbb{R}} f_- \, d\lambda,$$

we see that f is Lebesgue integrable if and only if

$$\int_{\mathbb{R}} f_+ \, d\lambda < \infty \qquad \text{and} \qquad \int_{\mathbb{R}} f_- \, d\lambda < \infty.$$

Since $|f_+| = f_+$ and $|f_-| = f_-$, we obtain f is Lebesgue integrable if and only if f_+ and f_- are Lebesgue integrable.

Remark 3.4.3. Notice if $f : \mathbb{R} \to [-\infty, \infty]$ is integrable, then for all $A \in \mathcal{M}(\mathbb{R})$ we have

$$\int_A |f| \, d\lambda = \int_{\mathbb{R}} |f\chi_A| \, d\lambda \leq \int_{\mathbb{R}} |f| \, d\lambda < \infty.$$

Hence the integral of $|f|$ with respect to λ against any Lebesgue measurable set is finite. Thus, by repeating the argument in Remark 3.4.2, we obtain that

$$\int_A f_+ \, d\lambda < \infty \qquad \text{and} \qquad \int_A f_- \, d\lambda < \infty$$

Using the positive and negative parts of f, we can now define the Lebesgue integral. Note this definition is well-defined by Remark 3.4.2.

Definition 3.4.4. Let $f : \mathbb{R} \to [-\infty, \infty]$ be Lebesgue integrable. For $A \in \mathcal{M}(\mathbb{R})$, the *Lebesgue integral of f over A against* λ is defined to be

$$\int_A f \, d\lambda = \int_A f_+ \, d\lambda - \int_A f_- \, d\lambda.$$

Remark 3.4.5. Note that if $f : \mathbb{R} \to [0, \infty]$ is Lebesgue integrable, then $f_+ = f$ and $f_- = 0$ so Definition 3.4.4 agrees with the definition of the Lebesgue integral of non-negative measurable functions from Definition 3.2.1.

Remark 3.4.6. If $f : \mathbb{R} \to [-\infty, \infty]$ is Lebesgue integrable and $A \in \mathcal{M}(\mathbb{R})$, then it is elementary to verify that

$$(f\chi_A)_+ = f_+\chi_A \qquad \text{and} \qquad (f\chi_A)_- = f_-\chi_A.$$

Moreover, by Remark 3.4.3, we know that $f\chi_A$ is Lebesgue integrable. Hence

$$\begin{aligned}\int_A f\,d\lambda &= \int_A f_+\,d\lambda - \int_A f_-\,d\lambda \\ &= \int_{\mathbb{R}} f_+\chi_A\,d\lambda - \int_{\mathbb{R}} f_-\chi_A\,d\lambda \\ &= \int_{\mathbb{R}} (f\chi_A)_+\,d\lambda - \int_{\mathbb{R}} (f\chi_A)_-\,d\lambda \\ &= \int_{\mathbb{R}} f\chi_A\,d\lambda.\end{aligned}$$

Therefore, when working with the Lebesgue integral and Lebesgue integrable functions, it suffices to just consider the Lebesgue integral over $\mathbb{R}$.

Moreover, it is possible to reduce the discussion of Lebesgue integrable functions down to those with values in $\mathbb{R}$. To see this, we first need the following.

Proposition 3.4.7. *Let $f : \mathbb{R} \to [-\infty, \infty]$ be Lebesgue integrable and let $g : \mathbb{R} \to [-\infty, \infty]$ be Lebesgue measurable. If $f = g$ a.e., then g is integrable and $\int_{\mathbb{R}} g\,d\lambda = \int_{\mathbb{R}} f\,d\lambda$.*

Proof. Since $f = g$ a.e., it is easy to see that

$$f_+ = g_+ \qquad \text{and} \qquad f_- = g_-$$

almost everywhere. Therefore, by Theorem 3.3.3, we obtain that

$$\int_{\mathbb{R}} g_+\,d\lambda = \int_{\mathbb{R}} f_+\,d\lambda < \infty \qquad \text{and} \qquad \int_{\mathbb{R}} g_-\,d\lambda = \int_{\mathbb{R}} f_-\,d\lambda < \infty.$$

Thus we trivially obtain that g is Lebesgue integrable since f is, and

$$\int_{\mathbb{R}} f\,d\lambda = \int_{\mathbb{R}} f_+\,d\lambda - \int_{\mathbb{R}} f_-\,d\lambda = \int_{\mathbb{R}} g_+\,d\lambda - \int_{\mathbb{R}} g_-\,d\lambda = \int_{\mathbb{R}} g\,d\lambda. \qquad \blacksquare$$

Remark 3.4.8. Let $f : \mathbb{R} \to [-\infty, \infty]$ be Lebesgue integrable. Since f is Lebesgue measurable, the set $B = \{x \in \mathbb{R} \mid |f(x)| = \infty\}$ is Lebesgue measurable. However, if $\lambda(B) > 0$, it is elementary to see by the definition of the integral that $\int_{\mathbb{R}} |f|\,d\lambda = \infty$, which would contradict the fact that

$$\int_{\mathbb{R}} |f|\,d\lambda < \infty.$$

Hence $\lambda(B) = 0$ so $f = \chi_{B^c} f$ almost everywhere. Since $\chi_{B^c} f : \mathbb{R} \to \mathbb{R}$, and since
$$\int_{\mathbb{R}} |\chi_{B^c} f| \, d\lambda = \int_{\mathbb{R}} |f| \, d\lambda < \infty$$
it suffices to only consider real-valued Lebesgue measurable functions when discussing Lebesgue integrable functions.

Of course, we still need to verify that this integral is linear, which happens to be a rather technical task.

Theorem 3.4.9. *The set of Lebesgue integrable functions from $\mathbb{R}$ to $\mathbb{R}$ is a vector space over $\mathbb{R}$. Moreover, the Lebesgue integral is linear; that is, if $f, g : \mathbb{R} \to \mathbb{R}$ are Lebesgue integrable and $\alpha, \beta \in \mathbb{R}$, then*
$$\int_{\mathbb{R}} \alpha f + \beta g \, d\lambda = \alpha \int_{\mathbb{R}} f \, d\lambda + \beta \int_{\mathbb{R}} g \, d\lambda.$$

Proof. Let $f, g : \mathbb{R} \to \mathbb{R}$ be Lebesgue integrable and let $\alpha, \beta \in \mathbb{R}$. Then
$$\int_{\mathbb{R}} |\alpha f + \beta g| \, d\lambda \leq \int_{\mathbb{R}} |\alpha||f| + |\beta||g| \, d\lambda = |\alpha| \int_{\mathbb{R}} |f| \, d\lambda + |\beta| \int_{\mathbb{R}} |g| \, d\lambda < \infty$$
since $\int_{\mathbb{R}} |f| \, d\lambda, \int_{\mathbb{R}} |g| \, d\lambda < \infty$. Hence $\alpha f + \beta g$ is Lebesgue integrable. Therefore the set of integrable functions from $\mathbb{R}$ to $\mathbb{R}$ is a vector space over $\mathbb{R}$.

In order to show the linearity of the Lebesgue integral, we claim that if $h_1, h_2 : \mathbb{R} \to [0, \infty)$ are Lebesgue integrable functions, then
$$\int_{\mathbb{R}} h_1 - h_2 \, d\lambda = \int_{\mathbb{R}} h_1 \, d\lambda - \int_{\mathbb{R}} h_2 \, d\lambda.$$
To see this, let $h = h_1 - h_2$. Hence
$$h_1 - h_2 = h = h_+ - h_-.$$
Thus
$$h_1 + h_- = h_+ + h_2$$
Since h_1, h_2, h_+, and h_- are non-negative Lebesgue measurable functions, we see that
$$\begin{aligned}\int_{\mathbb{R}} h_1 \, d\lambda + \int_{\mathbb{R}} h_- \, d\lambda &= \int_{\mathbb{R}} h_1 + h_- \, d\lambda \\ &= \int_{\mathbb{R}} h_+ + h_2 \, d\lambda \\ &= \int_{\mathbb{R}} h_+ \, d\lambda + \int_{\mathbb{R}} h_2 \, d\lambda.\end{aligned}$$
Therefore, since h_1, h_2, h_+, and h_- are Lebesgue integrable functions so each integral is finite, we obtain that
$$\int_{\mathbb{R}} h_1 \, d\lambda - \int_{\mathbb{R}} h_2 \, d\lambda = \int_{\mathbb{R}} h_+ \, d\lambda - \int_{\mathbb{R}} h_- \, d\lambda = \int_{\mathbb{R}} h \, d\lambda$$

as claimed.

To see the additivity of the integral, note since $f_+ + g_+$ and $f_- + g_-$ will be Lebesgue integrable that

$$\begin{aligned}\int_{\mathbb{R}} f + g \, d\lambda &= \int_{\mathbb{R}} (f_+ - f_-) + (g_+ - g_-) \, d\lambda \\ &= \int_{\mathbb{R}} (f_+ + g_+) - (f_- + g_-) \, d\lambda \\ &= \int_{\mathbb{R}} (f_+ + g_+) \, d\lambda - \int_{\mathbb{R}} (f_- + g_-) \, d\lambda && \text{by the claim} \\ &= \int_{\mathbb{R}} f_+ \, d\lambda + \int_{\mathbb{R}} g_+ \, d\lambda - \int_{\mathbb{R}} f_- \, d\lambda - \int_{\mathbb{R}} g_- \, d\lambda && \text{by Theorem 3.3.3} \\ &= \int_{\mathbb{R}} f_+ \, d\lambda - \int_{\mathbb{R}} f_- + \int_{\mathbb{R}} g_+ \, d\lambda \, d\lambda - \int_{\mathbb{R}} g_- \, d\lambda \\ &= \int_{\mathbb{R}} f \, d\lambda + \int_{\mathbb{R}} g \, d\lambda.\end{aligned}$$

Hence the integral is additive. Thus it remains only to prove that the Lebesgue integral preserves scalar multiplication.

To see the Lebesgue integral preserves scalar multiplication, let $a \in \mathbb{R}$ be arbitrary. If $a \geq 0$, then $(af)_+ = af_+$ and $(af)_- = af_-$. Thus we obtain that

$$\begin{aligned}\int_{\mathbb{R}} af \, d\lambda &= \int_{\mathbb{R}} (af)_+ \, d\lambda - \int_{\mathbb{R}} (af)_- \, d\lambda \\ &= \int_{\mathbb{R}} af_+ \, d\lambda - \int_{\mathbb{R}} af_- \, d\lambda \\ &= a \int_{\mathbb{R}} f_+ \, d\lambda - a \int_{\mathbb{R}} f_- \, d\lambda && \text{by Theorem 3.2.3} \\ &= a \int_{\mathbb{R}} f \, d\lambda.\end{aligned}$$

Otherwise, if $a < 0$ then $(af)_+ = (-a)f_-$ and $(af)_- = (-a)f_+$. Thus, since $-a > 0$, we obtain that

$$\begin{aligned}\int_{\mathbb{R}} af \, d\lambda &= \int_{\mathbb{R}} (af)_+ \, d\lambda - \int_{\mathbb{R}} (af)_- \, d\lambda \\ &= \int_{\mathbb{R}} (-a)f_- \, d\lambda - \int_{\mathbb{R}} (-a)f_+ \, d\lambda \\ &= (-a) \int_{\mathbb{R}} f_- \, d\lambda - (-a) \int_{\mathbb{R}} f_+ \, d\lambda && \text{by Theorem 3.2.3} \\ &= a \int_{\mathbb{R}} f_+ \, d\lambda - a \int_{\mathbb{R}} f_- \, d\lambda \\ &= a \int_{\mathbb{R}} f \, d\lambda.\end{aligned}$$

Hence the result follows. ∎

Not only is the Lebesgue integral linear, the Lebesgue integral has many properties similar to the Riemann integral. In particular, the Lebesgue integral behaves well with respect to absolute values, translation, and inversion of functions.

Theorem 3.4.10. *If $f : \mathbb{R} \to \mathbb{R}$ is Lebesgue integrable, then $|f|$ is Lebesgue integrable and*

$$\left|\int_{\mathbb{R}} f \, d\lambda\right| \leq \int_{\mathbb{R}} |f| \, d\lambda.$$

Proof. Since f is Lebesgue integrable, clearly $|f|$ is Lebesgue integrable. Moreover

$$\begin{aligned}\left|\int_{\mathbb{R}} f \, d\lambda\right| &= \left|\int_{\mathbb{R}} f_+ \, d\lambda - \int_{\mathbb{R}} f_- \, d\lambda\right| \\ &\leq \left|\int_{\mathbb{R}} f_+ \, d\lambda\right| + \left|\int_{\mathbb{R}} f_- \, d\lambda\right| \\ &= \int_{\mathbb{R}} f_+ \, d\lambda + \int_{\mathbb{R}} f_- \, d\lambda \\ &= \int_{\mathbb{R}} f_+ + f_- \, d\lambda \\ &= \int_{\mathbb{R}} |f| \, d\lambda\end{aligned}$$

as desired. ∎

Proposition 3.4.11 (Translation Invariance). *Let $f : \mathbb{R} \to \mathbb{R}$ be Lebesgue integrable. For each $y \in \mathbb{R}$ let $f_y : \mathbb{R} \to \mathbb{R}$ be defined by $f_y(x) = f(x - y)$. Then f_y is Lebesgue integrable and*

$$\int_{\mathbb{R}} f_y \, d\lambda = \int_{\mathbb{R}} f \, d\lambda.$$

Proof. To see that f_y is Lebesgue measurable, note $f_y^{-1}([\alpha, \infty)) = y + f^{-1}([\alpha, \infty))$ for all $\alpha \in \mathbb{R}$. Therefore, since f is Lebesgue measurable and since the translation of a Lebesgue measurable set is Lebesgue measurable by Proposition 1.6.5, we obtain that f_y is measurable.

To see that f_y is Lebesgue integrable and

$$\int_{\mathbb{R}} f_y \, d\lambda = \int_{\mathbb{R}} f \, d\lambda,$$

we will prove our result in the same way we built up our integral: first for characteristic functions, then for simple functions, then for non-negative functions, and finally for general functions. This is a common technique for proving facts about the Lebesgue integral.

First consider $A, B \subseteq \mathbb{R}$ and $y \in \mathbb{R}$ such that $B = y + A$. Hence Proposition 1.6.5 implies B is measurable if and only if A is measurable and

$$\lambda(B) = \lambda(A).$$

Hence, for all $A \in \mathcal{M}(\mathbb{R})$, we obtain that that

$$\int_{\mathbb{R}} (\chi_A)_y \, d\lambda = \int_{\mathbb{R}} \chi_{y+A} \, d\lambda = \lambda(y + A) = \lambda(A) = \int_{\mathbb{R}} \chi_A \, d\lambda.$$

Therefore, since

$$\left(\sum_{k=1}^{n} a_k \chi_{A_k} \right)_y = \sum_{k=1}^{n} a_k (\chi_{A_k})_y,$$

we obtain that

$$\int_{\mathbb{R}} \varphi_y \, d\lambda = \int_{\mathbb{R}} \varphi \, d\lambda$$

for all simple functions φ by the additivity of the Lebesgue integral.

Given a Lebesgue integrable function $f : \mathbb{R} \to [0, \infty)$, we note that φ is a simple function such that $\varphi \leq f$ if and only if φ_y is a simple function such that $\varphi_y \leq f_y$. Therefore, by the definition of the Lebesgue integral of a non-negative measurable function and the above result for simple functions, we obtain that

$$\int_{\mathbb{R}} f_y \, d\lambda = \int_{\mathbb{R}} f \, d\lambda$$

for all Lebesgue integrable functions $f : \mathbb{R} \to [0, \infty)$.

Finally, assume $f : \mathbb{R} \to \mathbb{R}$ is Lebesgue integrable. Since we know that $(f_y)_+ = (f_+)_y$ and $(f_y)_- = (f_-)_y$, we see that f_y is Lebesgue integrable since f is and

$$\begin{aligned} \int_{\mathbb{R}} f_y \, d\lambda &= \int_{\mathbb{R}} (f_y)_+ \, d\lambda - \int_{\mathbb{R}} (f_y)_- \, d\lambda \\ &= \int_{\mathbb{R}} (f_+)_y \, d\lambda - \int_{\mathbb{R}} (f_-)_y \, d\lambda \\ &= \int_{\mathbb{R}} f_+ \, d\lambda - \int_{\mathbb{R}} f_- \, d\lambda \\ &= \int_{\mathbb{R}} f \, d\lambda \end{aligned}$$

as desired. ∎

By replacing Proposition 1.6.5 with Proposition 1.6.6 and repeating the above proof, we easily obtain the following.

Proposition 3.4.12 (Inversion Invariance). *Let $f : \mathbb{R} \to \mathbb{R}$ be Lebesgue integrable. Let $\check{f} : \mathbb{R} \to \mathbb{R}$ be defined by $\check{f}(x) = f(-x)$. Then $\check{f}$ is Lebesgue integrable and*

$$\int_{\mathbb{R}} \check{f} \, d\lambda = \int_{\mathbb{R}} f \, d\lambda.$$

Proposition 3.4.13 (Scaling Invariance). *Let $f : \mathbb{R} \to \mathbb{R}$ be Lebesgue integrable and let $\alpha > 0$. Let $g : \mathbb{R} \to \mathbb{R}$ be defined by $g(x) = f(\alpha x)$. Then g is Lebesgue integrable and*

$$\int_{\mathbb{R}} g \, d\lambda = \frac{1}{\alpha} \int_{\mathbb{R}} f \, d\lambda.$$

Note Propositions 3.4.11, 3.4.12,and 3.4.13 are essential if one wanted to develop a notion of Fourier series for Lebesgue integrable functions.

3.5 Revisiting the Riemann Integral

Recall our goal was to generalize the Riemann integral in the hope of correcting many of the deficiencies of the Riemann integral. We still have yet to answer the questions: does the Lebesgue integral truly generalize the Riemann integral?

To answer this question, we must first understand the set of Riemann integrable functions. After all, if the Lebesgue integral is truly going to be a generalization of the Riemann integral, we need every Riemann integrable function to be Lebesgue integrable and thus Lebesgue measurable. To begin, we must understand the set of discontinuities of a function.

Lemma 3.5.1. *Let $a, b \in \mathbb{R}$ be such that $a < b$, let $f : [a, b] \to \mathbb{R}$, and let*

$$D(f) = \{x \in [a, b] \mid f \text{ is discontinuous at } x\}.$$

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

$$D_n(f) = \left\{ x \in [a, b] \,\middle|\, \begin{array}{c} \text{for every } \delta > 0 \text{ there exists } y, z \in [a, b] \text{ such that} \\ |x - y| < \delta, |x - z| < \delta, \text{ and } |f(y) - f(z)| \geq \frac{1}{n} \end{array} \right\}.$$

Then $D_n(f)$ is closed for each $n \in \mathbb{N}$ and $D(f) = \bigcup_{n=1}^{\infty} D_n(f)$. Hence the discontinuities of f is a countable union of closed sets.

Proof. Fix $m \in \mathbb{N}$. To see that $D_m(f)$ is closed, let $(x_n)_{n\geq 1}$ be an arbitrary sequence of elements of $D_m(f)$ that converges to some $x \in [a, b]$. To see that $x \in D_m(f)$, let $\delta > 0$ be arbitrary. Since $x = \lim_{n\to\infty} x_n$, there exists an $N \in \mathbb{N}$ such that $|x - x_N| < \frac{1}{2}\delta$. Furthermore, since $x_N \in D_m(f)$, there exists $y, z \in [a, b]$ such that $|x_N - y| < \frac{1}{2}\delta$, $|x_N - z| < \frac{1}{2}\delta$, and $|f(y) - f(z)| \geq \frac{1}{m}$. Since $|x - y| < \delta$ and $|x - z| < \delta$ by the triangle inequality, and since $|f(y) - f(z)| \geq \frac{1}{m}$, we obtain that $x \in D_m(f)$ as $\delta > 0$ was arbitrary. Hence, since $(x_n)_{n\geq 1}$ was arbitrary, $D_m(f)$ is closed.

To see that $D(f) = \bigcup_{n=1}^{\infty} D_n(f)$, first assume $x \in \bigcup_{n=1}^{\infty} D_n(f)$. Hence $x \in D_m(f)$ for some $m \in \mathbb{N}$. To see that f is discontinuous at x, suppose for the sake of a contradiction that f is continuous at x. Notice by the definition of $D_m(f)$ that for each $n \in \mathbb{N}$ there exists points $y_n, z_n \in [a, b]$ such that $|x - y_n| < \frac{1}{n}$, $|x - z_n| < \frac{1}{n}$, and $|f(y_n) - f(z_n)| \geq \frac{1}{m}$. Since

$$x = \lim_{n\to\infty} y_n = \lim_{n\to\infty} z_n,$$

the continuity of f implies

$$f(x) = \lim_{n\to\infty} f(y_n) = \lim_{n\to\infty} f(z_n),$$

which contradicts the fact that $|f(y_n) - f(z_n)| \geq \frac{1}{m}$ for all n. Hence we have obtained a contradiction so $x \in D(f)$. Hence $\bigcup_{n=1}^{\infty} D_n(f) \subseteq D(f)$.

For the other inclusion, notice if $x \in D(f)$ then f is discontinuous at x. Therefore there exists an $\epsilon > 0$ such that for all $\delta > 0$ there exists a $y \in [a,b]$ such that $|x-y| < \delta$ yet $|f(x) - f(y)| \geq \epsilon$. Choose $m \in \mathbb{N}$ such that $\frac{1}{m} < \epsilon$. By taking $z = x$ in the definition of $D_m(f)$, we see that $x \in D_m(f)$. Hence, since x was arbitrary, $D(f) \subseteq \bigcup_{n=1}^{\infty} D_n(f)$ thereby completing the proof. ∎

Using the characterization of the discontinuities of a function, we can provide an alternate description of the Riemann integrable functions beyond the descriptions given in MATH 2001.

Proposition 3.5.2. *A function $f : [a,b] \to \mathbb{R}$ is Riemann integrable if and only if f is bounded and continuous almost everywhere.*

Proof. To begin, assume f is Riemann integrable. Clearly this implies f is bounded by definition. To see that f is continuous almost everywhere (i.e. the set of discontinuities of f has Lebesgue measure zero), for each $n \in \mathbb{N}$ let

$$D_n(f) = \left\{ x \in [a,b] \;\middle|\; \begin{array}{c} \text{for every } \delta > 0 \text{ there exists } y, z \in [a,b] \text{ such that} \\ |x-y| < \delta, |x-z| < \delta, \text{ and } |f(y) - f(z)| \geq \frac{1}{n} \end{array} \right\}.$$

By Lemma 3.5.1 the discontinuities of f are $\bigcup_{n=1}^{\infty} D_n(f)$. Therefore, to show that f is continuous almost everywhere, it suffices to show that each $D_n(f)$ has Lebesgue measure zero by the subadditivity of the Lebesgue measure.

Suppose for the sake of a contradiction 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 $k \in \{1, \ldots, n\}$ we define

$$m_k = \inf\{f(x) \mid x \in [t_{k-1}, t_k]\} \quad \text{and} \quad M_k = \sup\{f(x) \mid x \in [t_{k-1}, t_k]\}$$

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 $k \in \{1, \ldots, n\}$ 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

$$D_q(f) \subseteq \bigcup_{\substack{k \in \{1, \ldots, n\} \\ I_k \cap D_q(f) \neq \emptyset}} I_k$$

we obtain that

$$\begin{aligned}\frac{1}{q}\lambda(D_q(f)) > \sum_{k=1}^{n}(M_k - m_k)(t_k - t_{k-1}) &\geq \sum_{\substack{k\in\{1,\ldots,n\}\\ I_k\cap D_q(f)\neq\emptyset}} (M_k - m_k)(t_k - t_{k-1}) \\ &\geq \sum_{\substack{k\in\{1,\ldots,n\}\\ I_k\cap D_q(f)\neq\emptyset}} \frac{1}{q}(t_k - t_{k-1}) \\ &\geq \frac{1}{q}\lambda\left(\bigcup_{\substack{k\in\{1,\ldots,n\}\\ I_k\cap D_q(f)\neq\emptyset}} I_k\right) \\ &\geq \frac{1}{q}\lambda(D_q(f)),\end{aligned}$$

which is a contradiction. Thus it must be the case that f is continuous almost everywhere.

Conversely, assume f is bounded and continuous almost everywhere. Thus $\lambda(D(f)) = 0$ so $\lambda(D_n(f)) = 0$ for all $n \in \mathbb{N}$. To see that f is Riemann integrable, we will 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 $\epsilon > 0$ and choose $n \in \mathbb{N}$ such that $\frac{1}{n}(b - a) < \frac{1}{2}\epsilon$. Since f is bounded, there exists an $M \in \mathbb{R}$ such that $|f(x)| \leq M$ for all $x \in [a, b]$. 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

$$\lambda\left(\bigcup_{k=1}^{\infty} I_k\right) \leq \frac{\epsilon}{2(M+1)}.$$

However, since $D_n(f)$ 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

$$\lambda\left(\bigcup_{k=1}^{m} I_k\right) \leq \lambda\left(\bigcup_{k=1}^{\infty} I_k\right) \leq \frac{\epsilon}{2(M+1)}.$$

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 3.5.3. *If $f : [a,b] \to \mathbb{R}$ is Riemann integrable, then f is Lebesgue measurable.*

Proof. By Proposition 3.5.2, f is continuous almost everywhere. Hence there exists a Lebesgue measurable subset A of $[a,b]$ such that $\lambda(A^c) = 0$ and f is continuous at each point in A.

To show that f is Lebesgue measurable, we will apply Corollary 2.1.15. To begin, let $\alpha \in \mathbb{R}$ be arbitrary. Then

$$f^{-1}((\alpha,\infty)) = \left(f^{-1}((\alpha,\infty)) \cap A^c\right) \cup \left(f^{-1}((\alpha,\infty)) \cap A\right).$$

Since

$$\left(f^{-1}((\alpha,\infty)) \cap A^c\right) \subseteq A^c$$

and since $\lambda(A^c) = 0$, we obtain from the completeness of λ that $f^{-1}((\alpha,\infty)) \cap A^c$ is Lebesgue measurable. Hence it suffices to show that $f^{-1}((\alpha,\infty)) \cap A$ is Lebesgue measurable.

Since f is continuous at each point in A and since (α,∞) is an open set, for each $x \in f^{-1}((\alpha,\infty)) \cap A$ there exists an $r_x > 0$ such that $(x-r_x, x+r_x) \subseteq f^{-1}((\alpha,\infty))$. Let

$$U = \bigcup_{x \in f^{-1}((\alpha,\infty)) \cap A} (x-r_x, x+r_x).$$

Clearly U is an open subset of $\mathbb{R}$ such that

$$U \cap A = f^{-1}((\alpha,\infty)) \cap A.$$

Therefore, since U is open and thus Lebesgue measurable, and since A is Lebesgue measurable, we obtain that $f^{-1}((\alpha,\infty)) \cap A$ is Lebesgue measurable. Hence f is Lebesgue measurable. ∎

We can now proceed to show that the Lebesgue integral generalizes the Riemann integral starting with the non-negative functions.

Proposition 3.5.4. *If $f : [a, b] \to [0, \infty)$ is Riemann integrable, then*

$$\int_a^b f(x)\, dx = \int_{[a,b]} f \, d\lambda.$$

Proof. By Corollary 3.5.3, we know that f is Lebesgue measurable. To see that the integrals agree, 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 3.2.3 that

$$L(f, \mathcal{P}) = \int_{[a,b]} \varphi_{\mathcal{P}} \, d\lambda \leq \int_{[a,b]} f \, d\lambda \leq \int_{[a,b]} \psi_{\mathcal{P}} \, d\lambda = U(f, \mathcal{P})$$

since $\varphi_{\mathcal{P}} \leq f \leq \psi_{\mathcal{P}}$ almost everywhere and a set of Lebesgue measure zero does not contribute to the Lebesgue integral. Therefore, since 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

$$\int_a^b f(x)\, dx \leq \int_{[a,b]} f \, d\lambda \leq \int_a^b f(x)\, dx. \qquad \blacksquare$$

Theorem 3.5.5. *If $f : [a, b] \to \mathbb{R}$ is Riemann integrable, then f is Lebesgue integrable and*

$$\int_a^b f(x)\, dx = \int_{[a,b]} f \, d\lambda.$$

Proof. By Corollary 3.5.3, we know that f is Lebesgue measurable. Since f is Riemann integrable, $|f|$ is Riemann integrable by Proposition A.4.6. Thus

$$f_+ = \frac{1}{2}(f + |f|) \qquad \text{and} \qquad f_- = \frac{1}{2}(|f| - f)$$

are Riemann integrable.

By Proposition 3.5.4, we have that

$$\int_{[a,b]} |f| \, d\lambda = \int_a^b |f(x)| \, dx < \infty$$

so $|f|$ is Lebesgue integrable. Therefore

$$\begin{aligned}
\int_a^b f(x)\,dx &= \int_a^b f_+(x) - f_-(x)\,dx \\
&= \int_a^b f_+(x)\,dx - \int_a^b f_-(x)\,dx && \text{Riemann integral is linear} \\
&= \int_{[a,b]} f_+\,d\lambda - \int_{[a,b]} f_-\,d\lambda && \text{by Proposition A.4.6} \\
&= \int_{[a,b]} f\,d\lambda
\end{aligned}$$

as desired. ∎

Hence the Riemann and Lebesgue integrals agree whenever the Riemann integral exists! Hence the Lebesgue integral is truly a generalization of the Riemann integral!

Remark 3.5.6. Of course, one may ask why in Definition 3.2.1 we didn't define the Lebesgue integral via

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

That is, in the Riemann integral we can use infimums so can we use infimums to define the Lebesgue integral? Well, if f is bounded and $\lambda(A) < \infty$, then these two notions are equal!

To see this, assume $f : A \to [0,\infty)$ is such that there exists an $M > 0$ with $f(x) \leq M$ for all $x \in A$. We first desire to reduce the number of simple functions we need to consider in the infimum.

Assume $\varphi : A \to [0,\infty)$ is a simple function such that $f \leq \varphi$. If $B = \varphi^{-1}((M,\infty))$, then B is a Lebesgue measurable set since φ is Lebesgue measurable. Thus if we define

$$\varphi_0 = \varphi\chi_{B^c} + M\chi_B,$$

then $\varphi_0 : A \to [0, M]$ is a simple function such that $f \leq \varphi_0 \leq \varphi$ so

$$\int_A \varphi_0\,d\lambda \leq \int_A \varphi\,d\lambda.$$

Hence

$$\begin{aligned}
&\inf\left\{ \int_A \varphi\,d\lambda \,\middle|\, \varphi : A \to [0,\infty) \text{ simple}, f \leq \varphi \right\} \\
&= \inf\left\{ \int_A \varphi\,d\lambda \,\middle|\, \varphi : A \to [0,M] \text{ simple}, f \leq \varphi \right\}.
\end{aligned}$$

To compare the above with the definition of the Lebesgue integral of a non-negative measurable function via the supremum, note $\varphi : A \to [0, M]$

is a simple function such that $f \leq \varphi$ if and only if $M - \varphi : A \to [0, M]$ is a simple function such that $M - \varphi \leq M - f$. Furthermore

$$M\lambda(A) = \int_A M \, d\lambda = \int_A \varphi + (M - \varphi) \, d\lambda = \int_A \varphi \, d\lambda + \int_A M - \varphi \, d\lambda.$$

Therefore, since $M\lambda(A) < \infty$, we obtain that

$$\int_A \varphi \, d\lambda = M\lambda(A) - \int_A M - \varphi \, d\lambda.$$

Hence

$$\begin{aligned}
&\inf \left\{ \int_A \varphi \, d\lambda \,\middle|\, \varphi : A \to [0, \infty) \text{ simple}, f \leq \varphi \right\} \\
&= \inf \left\{ M\lambda(A) - \int_A M - \varphi \, d\lambda \,\middle|\, \varphi : A \to [0, M] \text{ simple}, f \leq \varphi \right\} \\
&= M\lambda(A) - \sup \left\{ \int_A M - \varphi \, d\lambda \,\middle|\, \varphi : A \to [0, M] \text{ simple}, f \leq \varphi \right\} \\
&= M\lambda(A) - \sup \left\{ \int_A \psi \, d\lambda \,\middle|\, \psi : A \to [0, M] \text{ simple}, \psi \leq M - f \right\} \\
&= M\lambda(A) - \int_A M - f \, d\lambda.
\end{aligned}$$

Moreover, since f and $M - f$ are non-negative Lebesgue measurable functions, we see from Theorem 3.3.3 that

$$M\lambda(A) = \int_A M \, d\lambda = \int_A f + (M - f) \, d\lambda = \int_A f \, d\lambda + \int_A M - f \, d\lambda.$$

Therefore, since $M\lambda(A) < \infty$, we obtain that

$$M\lambda(A) - \int_A M - f \, d\lambda = \int_A f \, d\lambda$$

thereby completing the claim.

Remark 3.5.7. In general, if $\lambda(A) = \infty$ or if f is not bounded, then it need not be true that

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

For an example where $\lambda(A) = \infty$, let $A = [1, \infty)$ and let $f(x) = \frac{1}{x^2}$ for all $x \in A$. Note if $f_n = f\chi_{[1,n]}$ for all $n \in \mathbb{N}$, then $(f_n)_{n \geq 1}$ is an increasing sequence of non-negative Lebesgue measurable functions that converges to f pointwise. Therefore, using Proposition 3.5.4 together with the Monotone

Convergence Theorem (Theorem 3.3.2), we obtain that

$$\begin{aligned}
\int_{[1,\infty)} f\, d\lambda &= \lim_{n\to\infty} \int_{[1,\infty)} f\chi_{[1,n]}\, d\lambda \\
&= \lim_{n\to\infty} \int_{[1,n]} f\, d\lambda \\
&= \lim_{n\to\infty} \int_1^n \frac{1}{x^2}\, dx \\
&= \lim_{n\to\infty} 1 - \frac{1}{n} \\
&= 1.
\end{aligned}$$

However, we claim that if $\varphi : [1,\infty) \to [0,\infty)$ is a simple function such that $f \leq \varphi$, then $\int_{[1,\infty)} \varphi\, d\lambda = \infty$ thereby leading to the above infimum being infinity. To see this, note if $a = \min \varphi^{-1}(0,\infty)$, then $a > 0$ by the definition of a simple function. Moreover, if $f \leq \varphi$, then

$$\varphi^{-1}([a,\infty)) = \varphi^{-1}((0,\infty)) \supseteq f^{-1}((0,\infty)) = [1,\infty)$$

and thus

$$\int_{[1,\infty)} \varphi\, d\lambda \geq a\lambda(\varphi^{-1}([a,\infty)) = \infty.$$

For example where f is not bounded, let $A = (0,1]$ and let $f(x) = \frac{1}{\sqrt{x}}$ for all $x \in A$. Note if $f_n = f\chi_{[\frac{1}{n},1]}$ for all $n \in \mathbb{N}$, then $(f_n)_{n\geq 1}$ is an increasing sequence of non-negative Lebesgue measurable functions that converges to f pointwise. Therefore, using Proposition 3.5.4 together with the Monotone Convergence Theorem (Theorem 3.3.2), we obtain that

$$\begin{aligned}
\int_{(0,1]} f\, d\lambda &= \lim_{n\to\infty} \int_{(0,1]} f\chi_{[\frac{1}{n},1]}\, d\lambda \\
&= \lim_{n\to\infty} \int_{[\frac{1}{n},1]} f\, d\lambda \\
&= \lim_{n\to\infty} \int_{\frac{1}{n}}^1 \frac{1}{\sqrt{x}}\, dx \\
&= \lim_{n\to\infty} 2 - 2\sqrt{\frac{1}{n}} \\
&= 2.
\end{aligned}$$

However, if $\varphi : (0,1] \to [0,\infty)$ is a simple function, then it is not possible for $f \leq \varphi$ as φ has finite range whereas the range of f is $[1,\infty)$.

Remark 3.5.8. Note the computations in Remark 3.5.7 show why improper integrals are defined as they are in elementary calculus. Moreover, we see that all computations with improper integrals of non-negative Riemann integrable functions are valid by the Monotone Convergence Theorem.

3.6 Fatou's Lemma

Due to the use of the Monotone Convergence Theorem (Theorem 3.3.2) in the theory of the Lebesgue integral, we desire two more limit theorems to demonstrate how well-behaved the Lebesgue integral is with respect to limits. The first is another limit theorem for non-negative Lebesgue measurable functions. Note it is possible to prove this theorem before the Monotone Convergence Theorem and use it to prove the Monotone Convergence Theorem. However, we believe the approach we provided is the correct one.

Theorem 3.6.1 (Fatou's Lemma). *For each $n \in \mathbb{N}$ let $f_n : \mathbb{R} \to [0, \infty]$ be a Lebesgue measurable function. Then*

$$\int_{\mathbb{R}} \liminf_{n\to\infty} f_n \, d\lambda \leq \liminf_{n\to\infty} \int_{\mathbb{R}} f_n \, d\lambda.$$

Proof. Recall $\liminf_{n\to\infty} f_n$ is Lebesgue measurable by Proposition 2.2.9.

For each $k \in \mathbb{N}$, let $g_k : \mathbb{R} \to [0, \infty]$ be defined by

$$g_k(x) = \inf\{f_n(x) \mid n \geq k\}$$

for all $x \in \mathbb{R}$. By Proposition 2.2.9 each g_k is a Lebesgue measurable function. Furthermore, for all $k \in \mathbb{N}$ and for all $n \geq k$ we see that $g_k \leq f_n$. Therefore

$$\int_{\mathbb{R}} g_k \, d\lambda \leq \int_{\mathbb{R}} f_n \, d\lambda$$

for all $n \geq k$ by Theorem 3.2.3. Hence

$$\int_{\mathbb{R}} g_k \, d\lambda \leq \liminf_{n\to\infty} \int_{\mathbb{R}} f_n \, d\lambda$$

for all $k \in \mathbb{N}$.

However, it is elementary to see that $(g_k)_{k\geq 1}$ is an increasing sequence of Lebesgue measurable functions that converges to $\liminf_{n\to\infty} f_n$ pointwise. Therefore the Monotone Convergence Theorem (Theorem 3.3.2) implies that

$$\int_{\mathbb{R}} \liminf_{n\to\infty} f_n \, d\lambda = \lim_{n\to\infty} \int_{\mathbb{R}} g_k \, d\lambda \leq \liminf_{n\to\infty} \int_{\mathbb{R}} f_n \, d\lambda$$

as desired. ∎

Remark 3.6.2. It is not difficult to see that the inequality in Fatou's Lemma (Theorem 3.6.1) 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 $n \in \mathbb{N}$ whereas $(f_n)_{n\geq 1}$ converges to zero pointwise almost everywhere so $\int_{\mathbb{R}} \liminf_{n\to\infty} f_n \, d\lambda = 0$.

3.7 The Dominated Convergence Theorem

Finally, we arrive at the most powerful limit theorem for the Lebesgue integral.

Theorem 3.7.1 (Dominated Convergence Theorem). *Let $g : \mathbb{R} \to [0,\infty)$ be a Lebesgue integrable function. For each $n \in \mathbb{N}$ let $f_n : \mathbb{R} \to \mathbb{R}$ be a Lebesgue measurable function such that $|f_n| \leq g$ almost everywhere. If $f : \mathbb{R} \to \mathbb{R}$ is such that $(f_n)_{n\geq 1}$ converges to f pointwise almost everywhere, then f is Lebesgue integrable with*

$$\int_{\mathbb{R}} f \, d\lambda = \lim_{n\to\infty} \int_{\mathbb{R}} f_n \, d\lambda.$$

Proof. First note that f is Lebesgue measurable by Corollary 2.2.15 being the almost everywhere pointwise limit of Lebesgue measurable functions. Moreover, since $|f_n| \leq g$ almost everywhere for all $n \in \mathbb{N}$ and since $(f_n)_{n\geq 1}$ converges to f pointwise almost everywhere, we obtain that $|f| \leq g$ almost everywhere. Therefore, since g is Lebesgue integrable, the inequalities $|f| \leq g$ and $|f_n| \leq g$ almost everywhere imply that f and f_n are Lebesgue integrable for all $n \in \mathbb{N}$. Furthermore, since $|f - f_n|$ is Lebesgue measurable and since

$$|f - f_n| \leq |f| + |f_n| \leq 2g,$$

we also obtain that $|f - f_n|$ is Lebesgue integrable for all $n \in \mathbb{N}$.

Note that $2g - |f - f_n| \geq 0$ for all $n \in \mathbb{N}$ and that $(2g - |f - f_n|)_{n\geq 1}$ is a sequence of Lebesgue integrable functions that converges to $2g$ pointwise almost everywhere. Therefore Fatou's Lemma (Theorem 3.6.1) implies that

$$\begin{aligned}
\int_{\mathbb{R}} 2g \, d\lambda &= \int_{\mathbb{R}} \liminf_{n\to\infty} 2g - |f - f_n| \, d\lambda \\
&\leq \liminf_{n\to\infty} \int_{\mathbb{R}} 2g - |f - f_n| \, d\lambda \\
&= \liminf_{n\to\infty} \int_{\mathbb{R}} 2g \, d\lambda - \int_{\mathbb{R}} |f - f_n| \, d\lambda \\
&= \int_{\mathbb{R}} 2g \, d\lambda - \limsup_{n\to\infty} \int_{\mathbb{R}} |f - f_n| \, d\lambda.
\end{aligned}$$

Hence, since $0 \leq \int_{\mathbb{R}} 2g \, d\lambda < \infty$, we have that

$$\limsup_{n\to\infty} \int_{\mathbb{R}} |f - f_n| \, d\lambda = 0.$$

Therefore, by Theorem 3.4.10, we see that

$$\begin{aligned}
\limsup_{n\to\infty} \left| \int_A f \, d\lambda - \int_A f_n \, d\lambda \right| &= \limsup_{n\to\infty} \left| \int_{\mathbb{R}} f - f_n \, d\lambda \right| \\
&\leq \limsup_{n\to\infty} \int_{\mathbb{R}} |f - f_n| \, d\lambda \\
&= 0
\end{aligned}$$

so the result follows. ∎

Remark 3.7.2. Notice that the proof of the Dominated Convergence Theorem (Theorem 3.7.1) actually showed us that

$$\lim_{n\to\infty} \int_{\mathbb{R}} |f - f_n| \, d\lambda = 0.$$

This is actually a much stronger claim.

Remark 3.7.3. Note the necessity of the Lebesgue integrable function $g : \mathbb{R} \to [0, \infty)$ such that $|f_n| \leq g$ in the Dominated Convergence Theorem (Theorem 3.7.1) can be seen via the same example as used in Remark 3.6.2.

Chapter 4

Differentiation and Integration

Now that we have resolved many of the issues of the Riemann integral by using the Lebesgue integral, it is natural to ask, "what else do we get with the Lebesgue integral?" Since the relationship between integration and differentiation is the centrepiece of any undergraduate calculus course, it makes sense we analyze whether we have similar results when we use the Lebesgue integral. Of course, in MATH 2001 we saw that if a function f is differentiable, then f is continuous and thus the Riemann integral will suffice. However, perhaps the Lebesgue integral can handle functions that are differentiable almost everywhere and we can develop a deeper theory.

4.1 Vitali Coverings

To begin our study of differentiation using Lebesgue measure theory, we first need one if not the most technical results in this course. Clearly given a subset X of $\mathbb{R}$ there are many ways to cover X with intervals. 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 dealing with Lebesgue measurable sets instead of compact subsets, it is useful to to study various collections of intervals and how they behave with respect to the Lebesgue measure. The technical lemma that we need revolves around the following types of coverings where each point is covered by a set of arbitrarily small length.

Definition 4.1.1. 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 4.1.2. 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 $\mathbb{R}$.

Similar to how every open cover of a compact set has a finite subcover, the following, which is our technical lemma, shows that if we use a Vitali covering, we can almost choose a finite subcover. In fact, the finite almost subcover we obtain has some additional nice properties.

Theorem 4.1.3 (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. We begin by demonstrating that we can assume $\mathcal{I}$ has some additional properties. First note since $\lambda^*(X) < \infty$ that 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. To see the other property of a Vitali covering, let $\delta > 0$ and $x \in X$ be arbitrary. Since $x \in X \subseteq U$, there exists an $\epsilon_x > 0$ such that $(x - \epsilon_x, x + \epsilon_x) \subseteq U$. However, since $x \in X$ 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 $\lambda(\overline{I}) < \delta$. Therefore, since $\delta > 0$ 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, assume 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 $x \in X \setminus \bigcup_{k=1}^n J_k$. Since each element of $\mathcal{J}$ is closed, $\bigcup_{k=1}^n J_k$ is a closed set. Therefore, since $x \in X \setminus \bigcup_{k=1}^n J_k$,

$$\operatorname{dist}\left(\{x\}, \bigcup_{k=1}^{n} J_k \right) = \inf \left\{ |x - y| \,\middle|\, y \in \bigcup_{k=1}^{n} J_k \right\} > 0$$

(i.e. there is no sequence in $\bigcup_{k=1}^n J_k$ that converges to x). Since $\mathcal{J}$ is a Vitali covering of X, there exists a $J \in \mathcal{J}$ such that $x \in J$ and $\lambda(J) < \operatorname{dist}\left(\{x\}, \bigcup_{k=1}^n J_k\right)$. Hence $J \cap J_k = \emptyset$ for all $k \in \{1, \ldots, n\}$ so $M_n \geq \lambda(J) > 0$ as every element of $\mathcal{J}$ has positive length. Therefore 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

$$\lambda(J_{n+1}) > \frac{1}{2} M_n.$$

If we use the above process, either the process ends after a finite number of steps thereby completing the proof, 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. Since $\mathcal{J}$ is a Vitali covering of X and since $\bigcup_{k=1}^N J_k$ 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, since $\lim_{n\to\infty} \lambda(J_n) = 0$, 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, since $x \in J_x$ 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}. Therefore, since $x \in X \setminus \bigcup_{k=1}^N J_k$ 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. ∎

4.2 The Lebesgue Differentiation Theorem

With the technical proof of the Vitali Covering Lemma (Theorem 4.1.3) out of the way, we can turn our attention differentiation of Lebesgue measurable functions. The goal of this section is to demonstrate the Lebesgue Differentiation Theorem which tells us everything we want to know about differentiation monotone Lebesgue measurable functions. First we set some notation that is useful when discussing derivatives (that luckily could be avoided in MATH 2001).

Definition 4.2.1. Let $f : \mathbb{R} \to \mathbb{R}$. For each $x \in \mathbb{R}$ define

$$\begin{aligned} D^+f(x) &= \limsup_{h\to 0^+} \frac{f(x+h)-f(x)}{h}, \\ D_+f(x) &= \liminf_{h\to 0^+} \frac{f(x+h)-f(x)}{h}, \\ D^-f(x) &= \limsup_{h\to 0^-} \frac{f(x+h)-f(x)}{h}, \text{ and} \\ D_-f(x) &= \liminf_{h\to 0^-} \frac{f(x+h)-f(x)}{h}, \end{aligned}$$

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 $f'(x) = D^+f(x) = D_+f(x) = D^-f(x) = D_-f(x)$.

Theorem 4.2.2 (Lebesgue Differentiation Theorem). *If $f : [a, b] \to \mathbb{R}$ is a non-decreasing function, then f is differentiable almost everywhere, f' is Lebesgue measurable, $f' \geq 0$ almost everywhere, and*

$$\int_{[a,b]} f' \, d\lambda \leq f(b) - f(a).$$

Proof. For notational simplicity, if $x < a$ we define $f(x) = f(a)$ and if $x > b$ we define $f(x) = f(b)$. Clearly this extended definition of f is still non-decreasing. Thus for all $c \in \mathbb{R}$ we see that $f^{-1}([c, \infty))$ is of the form (y, ∞) or $[y, \infty)$ for some $y \in \mathbb{R} \cup \{\pm\infty\}$. Hence f is Lebesgue 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. In this write-up of the proof, we will only show 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 remaining facts are nearly identical.

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, we can demonstrate that $\lambda^*(E_{p,q}) = 0$ for all $p, q \in \mathbb{Q}$, then $\lambda^*(X) = 0$ since $\mathbb{Q}$ is countable and thus X is measurable as the Lebesgue measure is complete.

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 $f(x+h) - f(x) < qh$. Since the collection of such intervals forms a Vitali covering of $E_{p,q}$, the Vitali Covering Lemma (Theorem 4.1.3) 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 $x \in A \subseteq E_{p,q}$ 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, since $A \subseteq \bigcup_{k=1}^{n} I_k$ 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 $h < \delta$, $[x, x+h) \subseteq I_k$ for some k, and $f(x+h) - f(x) > ph$. Since the collection of such intervals forms a Vitali covering of A, the Vitali Covering Lemma (Theorem 4.1.3) 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 $\lambda^*(B) \geq \lambda^*(A) - \epsilon > r - 2\epsilon$. 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, since each J_k is contained in a single I_j and since 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

$$p(r - 2\epsilon) \leq \sum_{k=1}^{m} f(y_k + s_k) - f(y_k) \leq \sum_{j=1}^{n} f(x_j + h_j) - f(x_j) \leq q(r + \epsilon).$$

However, since $\epsilon > 0$ was arbitrary, the above implies $pr \leq qr$. Therefore, since $p > q$ 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 3.4.11, $(g_n)_{n \geq 1}$ is a sequence of Lebesgue 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, since $g_n : [a, b] \to [0, \infty)$ and since f is bounded (being non-decreasing) and thus Lebesgue integrable, we obtain by Fatou's Lemma (Theorem 3.6.1) and Proposition 3.4.11 that

$$\begin{aligned}
\int_{[a,b]} g \, d\lambda &= \int_{[a,b]} \liminf_{n \to \infty} g_n \, d\lambda \\
&\leq \liminf_{n \to \infty} \int_{[a,b]} g_n \, d\lambda \\
&= \liminf_{n \to \infty} n \int_{[a,b]} f\left(x + \frac{1}{n}\right) - f(x) \, d\lambda(x) \\
&= \liminf_{n \to \infty} n \int_{[a + \frac{1}{n}, b + \frac{1}{n}]} f \, d\lambda - n \int_{[a,b]} f \, d\lambda \\
&= \liminf_{n \to \infty} n \int_{\left[b, b + \frac{1}{n}\right]} f \, d\lambda - n \int_{\left[a, a + \frac{1}{n}\right]} f \, d\lambda \\
&= \liminf_{n \to \infty} f(b) - n \int_{\left[a, a + \frac{1}{n}\right]} f \, d\lambda \\
&= f(b) - \limsup_{n \to \infty} n \int_{\left[a, a + \frac{1}{n}\right]} f \, d\lambda \\
&\leq f(b) - f(a)
\end{aligned}$$

since, for all $n \in \mathbb{N}$,

$$n \int_{[a,a+\frac{1}{n}]} f \, d\lambda \geq n \int_{[a,a+\frac{1}{n}]} f(a) \, d\lambda = f(a).$$

Therefore, since $f(b) - f(a) < \infty$, 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 4.2.3. Note if $f : [a, b] \to \mathbb{R}$ is non-increasing, then $-f$ is non-decreasing and thus differentiable almost everywhere with $(-f)' \geq 0$ almost everywhere. Hence f is differentiable almost everywhere with $f' \leq 0$ almost everywhere.

Corollary 4.2.4. *If $f : \mathbb{R} \to \mathbb{R}$ is Lebesgue measurable and differentiable almost everywhere, then $f' : [a, b] \to \mathbb{R}$ is Lebesgue measurable.*

Proof. For each $n \in \mathbb{N}$, let $g_n : \mathbb{R} \to \mathbb{R}$ be defined by

$$g_n(x) = n\left(f\left(x + \frac{1}{n}\right) - f(x)\right)$$

for all $x \in \mathbb{R}$. By Proposition 3.4.11 $(g_n)_{n\geq 1}$ is a sequence of measurable functions that converge pointwise almost everywhere to f'. Hence f' is Lebesgue measurable. ∎

To conclude this section, we answer the question "Is the inequality in the Lebesgue Differentiation Theorem (Theorem 4.2.2) always an equality?" It turns out, the answer is no.

Remark 4.2.5. Let $f : [0, 1] \to [0, 1]$ be the Cantor ternary function. Thus f is non-decreasing on $[0, 1]$ and constant on $\mathcal{C}^c$. Since $\mathcal{C}^c$ is a finite union of open sets, we easily see by Definition 4.2.1 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 since $\lambda(\mathcal{C}) = 0$. However

$$\int_{[0,1]} f' \, d\lambda = 0 < 1 = f(1) - f(0).$$

Therefore the inequality in the Lebesgue Differentiation Theorem (Theorem 4.2.2) may be strict.

4.3 Bounded Variation

One nice result from MATH 2001 was the Fundamental Theorem of Calculus which showed the connection between integration and differentiation and that a differentiable function can be recovered from its derivative; that is

$$f(x) = f(a) + \int_a^x f'(y) \, dy.$$

However, as we have seen above, the Cantor ternary function is a function that cannot be recovered from its derivative via integration since 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 the set of functions we consider. Since 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 4.3.1. A function $f : [a, b] \to \mathbb{R}$ 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 $[a, b]$, then

$$\sum_{k=1}^{n} |f(x_k) - f(x_{k-1})| \leq M.$$

Example 4.3.2. Let $f : [a, b] \to \mathbb{R}$ be differentiable on $[a, b]$ for which 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 assume $\{x_k\}_{k=0}^n$ 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 4.3.3. The continuous function $f : [0, 1] \to [-1, 1]$ defined by

$$f(x) = x \cos\left(\frac{\pi}{2x}\right)$$

(with $f(0) = 0$) 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.

Unfortunately, these are not the functions we are looking for since the Cantor ternary function is of bounded variation by the following.

Remark 4.3.4. It is elementary to see that if f is monotone then f is of bounded variation since

$$\sum_{k=1}^{n} |f(x_k) - f(x_{k-1})| = |f(b) - f(a)|$$

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.

Even through functions of bounded variation are not the functions we are looking for, they do contain some nice functions we wish to study and the ideas and properties we develop will lead us to the correct collection of functions. To begin our study, we consider the smallest constant that works in Definition 4.3.1.

Definition 4.3.5. 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 4.3.6 (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 4.3.4, 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 $x_n = y_0$). Since 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 since $\{x_k\}_{k=0}^n$ 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,

$$\begin{aligned}
\sum_{k=1}^n |f(z_k) - f(z_{k-1})| &\leq \sum_{k=1}^m |f(w_k) - f(w_{k-1})| \\
&= \sum_{k \text{ such that } w_k \in [a,x]} |f(w_k) - f(w_{k-1})| \\
&\quad + \sum_{k \text{ such that } w_{k-1} \in [x,y]} |f(w_k) - f(w_{k-1})| \\
&\leq V_f(a,x) + V_f(x,y).
\end{aligned}$$

Therefore, since $\{z_k\}_{k=0}^n$ 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 proof is complete. ∎

By combining the Lebesgue Differentiation Theorem (Theorem 4.2.2) with the Jordan Decomposition Theorem (Theorem 4.3.6), we immediately obtain information about derivatives and integrals of functions of bounded variation.

Corollary 4.3.7. *If $f : [a,b] \to \mathbb{R}$ is of bounded variation, then f is differentiable almost everywhere and f' is Lebesgue integrable.*

Proof. Since f is of bounded variation, by the Jordan Decomposition Theorem (Theorem 4.3.6) there exists non-decreasing functions $V, D : [a,b] \to \mathbb{R}$ such that $f = V - D$. Since every non-decreasing function is differentiable with Lebesgue measurable derivatives by the Lebesgue Differentiation Theorem (Theorem 4.2.2), we clearly see that f is differentiable with $f' = V' - D'$

being Lebesgue measurable. Moreover, since V and D are non-decreasing, we see that $V', D' \geq 0$ almost everywhere and thus $|f'| \leq V' + D'$. Therefore

$$\int_{[a,b]} |f'| \, d\lambda \leq \int_{[a,b]} V' + D' \, d\lambda \leq V(b) + D(b) - V(a) - D(a) < \infty$$

by the Lebesgue Differentiation Theorem (Theorem 4.2.2). Hence f' is Lebesgue integrable. ∎

4.4 Absolutely Continuous Functions

Although the functions of bounded variation are not the functions we are looking for, the functions we desire are easy to describe and contain all differentiable functions with bounded derivatives.

Definition 4.4.1. A function $f : [a, b] \to \mathbb{R}$ 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

$$a \leq a_1 < b_1 \leq a_2 < b_2 \leq \cdots \leq a_n < b_n \leq b \qquad \text{and} \qquad \sum_{k=1}^n |b_k - a_k| < \delta$$

then

$$\sum_{k=1}^n |f(b_k) - f(a_k)| < \epsilon.$$

Example 4.4.2. Let $f : [a, b] \to \mathbb{R}$ be a differentiable on $[a, b]$ for which there exist an $M \in \mathbb{N}$ such that $|f'(x)| \leq M$ for all $x \in (a, b)$. We claim that f is absolutely continuous. To see this, 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

$$\sum_{k=1}^n |f(b_k) - f(a_k)| \leq \sum_{k=1}^n M|b_k - a_k| \leq M\delta < \epsilon.$$

Hence f is absolutely continuous.

Example 4.4.3. The Cantor ternary function is not absolutely continuous. To see this, let $f : [0, 1] \to [0, 1]$ be the Cantor ternary function and let $\{P_n\}_{n=0}^\infty$ be the sets from Definition 1.6.8 so that $\mathcal{C} = \bigcap_{n=0}^\infty P_n$ and P_n is a disjoint union of 2^n closed intervals such that $\lambda(P_n) = \left(\frac{2}{3}\right)^n$.

To see that f is not absolutely continuous, let $\epsilon = \frac{1}{2}$ and let $\delta > 0$ be arbitrary. Choose $N \in \mathbb{N}$ such that

$$\lambda(P_N) = \left(\frac{2}{3}\right)^N < \delta.$$

Since P_N is a disjoint union of 2^N closed intervals, we can write $P_N = \bigcup_{k=1}^{2^N} [a_k, b_k]$ where $b_k < a_{k+1}$ for all k. Thus

$$0 = a_1 < b_1 \leq a_2 < b_1 \leq \cdots \leq a_{2^N} < b_{2^N} = 1$$

and

$$\sum_{k=1}^{2^N} |b_k - a_k| = \lambda(P_N) < \delta.$$

However, since f is constant on each open interval in $\mathcal{C}^c$ and since $(b_k, a_{k+1}) \subseteq \mathcal{C}^c$ for all k, we obtain that $f(b_k) = f(a_{k+1})$ for all k and thus

$$\sum_{k=1}^{2^N} |f(b_k) - f(a_k)| = \sum_{k=1}^{2^N} f(b_k) - f(a_k) = f(b^{2^N}) - f(a_1) = f(1) - f(0) = 1 > \epsilon.$$

Therefore, since $\delta > 0$ was arbitrary, we see the definition of absolute continuity fails for f when $\epsilon = \frac{1}{2}$. Hence f is not absolutely continuous.

Unsurprisingly, absolutely continuous functions have some nice properties.

Proposition 4.4.4. *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 4.4.1).

To see that f is of bounded variation, recall since 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, since

$$\sum_{k=1}^{p_{j+1}-p_j} |z_{p_j+k} - z_{p_j+k-1}| = |z_{p_{j+1}} - z_{p_j}| < \delta,$$

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, since $\{x_k\}_{k=0}^n$ was an arbitrary partition of $[a,b]$, f is of bounded variation. ∎

Corollary 4.4.5. *If $f : [a,b] \to \mathbb{R}$ is absolutely continuous, then f is differentiable almost everywhere and f' is Lebesgue integrable.*

Proof. Since every absolutely continuous function is of bounded variation by Proposition 4.4.4, the result follows from Corollary 4.3.7. ∎

Of course, it is natural to ask whether the converse of Proposition 4.4.4 holds. To construct an example to show this is not the case, we require the following.

Proposition 4.4.6. *If $f : [a,b] \to \mathbb{R}$ is absolutely continuous and $f' = 0$ 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 see this, 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 $\lambda([a,c] \setminus X) = 0$. Since 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 $x \in X \cap [a,c)$ 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 $|f(x+h) - f(x)| < \epsilon h$. Since the collection of such intervals forms a Vitali covering of $X \cap [a,c)$, the Vitali Covering Lemma (Theorem 4.1.3) implies there exists an $n \in \mathbb{N}$, $x_1, \ldots, x_n \in X \cap [a,c)$ with $x_1 < x_2 < \cdots < x_n$, 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

$$\lambda^*\left([a,c] \setminus \bigcup_{k=1}^n I_k\right) \leq \lambda([a,c] \setminus X) + \lambda^*\left((X \setminus \{c\}) \setminus \bigcup_{k=1}^n I_k\right) < 0 + \delta = \delta.$$

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, since

$$\sum_{k=0}^n |x_{k+1} - y_k| = \lambda\left(\bigcup_{k=0}^n [y_k, x_{k+1})\right) = \lambda^*\left([a,c] \setminus \bigcup_{k=1}^n I_k\right) < \delta,$$

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 by our construction 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, since $\epsilon > 0$ was arbitrary, we obtain that $f(c) = f(a)$. Therefore, since $c \in (a, b]$ was arbitrary, the result follows. ∎

Example 4.4.7. If $f : [0, 1] \to [0, 1]$ is the Cantor ternary function, 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 2.1.9 and thus uniformly continuous $[0, 1]$ and of bounded variation. The fact that f is not absolutely continuous follows from Proposition 4.4.6 along with the fact that f is non-constant yet $f' = 0$ almost everywhere.

To conclude this section, we desire to construct some additional examples of absolutely continuous functions. To do so requires the following lemma.

Lemma 4.4.8. *Let $f : \mathbb{R} \to \mathbb{R}$ be Lebesgue integrable. Then for all $\epsilon > 0$ there exists a $\delta > 0$ such that if $A \in \mathcal{M}(\mathbb{R})$ and $\lambda(A) < \delta$, then*

$$\int_A |f| \, d\lambda < \epsilon.$$

Proof. Let $\epsilon > 0$ be arbitrary. Due to the definition of the Lebesgue integral of $|f|$ and the fact that $\int_{\mathbb{R}} |f| \, d\lambda < \infty$, there exists a simple function $\varphi : \mathbb{R} \to [0, \infty)$ such that $\varphi \leq |f|$ and

$$\int_{\mathbb{R}} |f| \, d\lambda \leq \int_{\mathbb{R}} \varphi \, d\lambda + \frac{\epsilon}{2}.$$

Since $0 \leq \varphi \leq |f|$ and $|f|$ is Lebesgue integrable, we obtain that φ is Lebesgue integrable with $f - \varphi \geq 0$. Hence for all $A \in \mathcal{M}(\mathbb{R})$ we obtain that

$$\int_A |f| \, d\lambda - \int_A \varphi \, d\lambda = \int_A (|f| - \varphi) \, d\lambda \leq \int_{\mathbb{R}} (|f| - \varphi) \, d\lambda \leq \frac{\epsilon}{2}.$$

Hence

$$\int_A |f| \, d\lambda \leq \int_A \varphi \, d\lambda + \frac{\epsilon}{2}.$$

for all $A \in \mathcal{M}(\mathbb{R})$.

Since φ is a simple function, we can write $\varphi = \sum_{k=1}^{n} a_k \chi_{A_k}$ where $n \in \mathbb{N}$, $\{a_k\}_{k=1}^{n} \subseteq [0, \infty)$, and $\{A_k\}_{k=1}^{n}$ are pairwise disjoint Lebesgue measurable sets. Let

$$M = \max(\{a_k\}_{k=1}^{n}) < \infty$$

and let $\delta = \frac{\epsilon}{2M+1}$. Then $\delta > 0$ and if $A \in \mathcal{M}(\mathbb{R})$ is such that $\lambda(A) < \delta$, then

$$\begin{aligned}
\int_A |f| \, d\lambda &\leq \frac{\epsilon}{2} + \int_A \varphi \, d\lambda \\
&= \frac{\epsilon}{2} + \sum_{k=1}^{n} a_k \lambda(A \cap A_k) \\
&\leq \frac{\epsilon}{2} + M \sum_{k=1}^{n} \lambda(A \cap A_k) \\
&\leq \frac{\epsilon}{2} + M\lambda\left(\bigcup_{k=1}^{n} A \cap A_k\right) \qquad \{A \cap A_k\}_{k=1}^{n} \text{ are pairwise disjoint} \\
&\leq \frac{\epsilon}{2} + M\delta \\
&= \frac{\epsilon}{2} + M\frac{\epsilon}{2M+1} < \epsilon.
\end{aligned}$$

Hence, since $\epsilon > 0$ was arbitrary, the result follows. ∎

Proposition 4.4.9. *Let $f : [a, b] \to \mathbb{R}$ be Lebesgue integrable. 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 is absolutely continuous.

Proof. First notice that F is clearly well-defined since f is Lebesgue integrable. To see that F is absolutely continuous, let $\epsilon > 0$. Since f is Lebesgue integrable, by Lemma 4.4.8 there exists a $\delta > 0$ such that if $A \in \mathcal{M}(\mathbb{R})$ and $\lambda(A) < \delta$ then

$$\int_A |f| \, d\lambda < \epsilon.$$

To see that this δ satisfies the requirements of Definition 4.4.1, 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. ∎

4.5 The Fundamental Theorems of Calculus

Due to the examples of absolutely continuous functions in Proposition 4.4.9 resembling the functions analyzed in MATH 2001 in relation to the Fundamental Theorems of Calculus, it is natural to ask what the derivatives of the functions defined in Proposition 4.4.9 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 4.5.1. *Let $f : [a, b] \to \mathbb{R}$ be Lebesgue integrable 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 Lebesgue measurable set since f is Lebesgue measurable. It suffices to prove that $\lambda(X) = 0$. To see that $\lambda(X) = 0$, suppose for the sake of a contradiction that $\lambda(X) > 0$. Due to the regularity of the Lebesgue measure from Proposition 1.6.12, there exists a compact subset $K \subseteq X$ such that

$\lambda(K) > 0$. Therefore, since $f(x) < 0$ 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, since V is an open and a subset of (a, b), and since 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

$$\int_V f \, d\lambda = \int_{\mathbb{R}} f\chi_V \, d\lambda = \int_{\mathbb{R}} \sum_{k=1}^{\infty} f_k \, d\lambda.$$

Notice if $S_n = \sum_{k=1}^{n} f_k$ for each $n \in \mathbb{N}$, then $|S_n| \leq |f|$. Hence, since f is Lebesgue integrable, we obtain by the Dominated Convergence Theorem (Theorem 3.7.1) that

$$\begin{aligned} F(b) - F(a) &< \int_V f \, d\lambda \\ &= \lim_{n\to\infty} \int_{\mathbb{R}} \sum_{k=1}^{n} f_k \, d\lambda \\ &= \lim_{n\to\infty} \sum_{k=1}^{n} F(b_k) - F(a_k) \\ &\leq F(b) - F(a) \end{aligned}$$

since F is non-decreasing. As this clearly is a contradiction, we obtain that $\lambda(X) = 0$ as desired. ∎

Corollary 4.5.2. *Let $f : [a, b] \to \mathbb{R}$ be Lebesgue integrable 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 4.5.1 implies that $f \geq 0$ almost everywhere. Similarly, since $-f$ is Lebesgue integrable and since

$$0 = (-F)(x) = \int_{[a,x]} -f \, d\lambda$$

for all $x \in [a, b]$, $-F$ is non-decreasing so Lemma 4.5.1 implies that $-f \geq 0$ almost everywhere. Hence $f = 0$ almost everywhere. ∎

Using all of the above, we arrive at the first version of our new Fundamental Theorems of Calculus which completely characterize absolutely continuous functions.

Theorem 4.5.3 (Fundamental Theorem of Calculus, I). *Let $f : [a, b] \to \mathbb{R}$ be Lebesgue integrable. 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 $F' = f$ almost everywhere.

Proof. To begin, note F is absolutely continuous (and thus Lebesgue measurable) by Proposition 4.4.9. Hence F' exists almost everywhere and is Lebesgue integrable by Corollary 4.4.5. To demonstrate that $F' = f$ almost everywhere we divide the proof into three cases.

Case 1: f is bounded. In this case 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

$$F_n(x) = n\left(F\left(x + \frac{1}{n}\right) - F(x)\right) = n \int_{[x, x+\frac{1}{n}]} f \, d\lambda$$

for all $x \in [a, b]$. Clearly each F_n is a Lebesgue measurable function by Proposition 1.6.5 since F is Lebesgue measurable. Furthermore, notice for each $n \in \mathbb{N}$ and $x \in [a, b]$ that

$$|F_n(x)| \leq n \int_{[x, x+\frac{1}{n}]} |f| \, d\lambda \leq n\left(\frac{1}{n} M\right) = M.$$

Since $M\chi_{[a,b]}$ is Lebesgue integrable, 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 3.7.1) 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

$$\begin{aligned} \int_{[a,c]} F' \, d\lambda &= \lim_{n\to\infty} n \int_{[a,c]} F\left(x + \frac{1}{n}\right) - F(x) \, d\lambda(x) \\ &= \lim_{n\to\infty} n \left(\int_{[c, c+\frac{1}{n}]} F \, d\lambda - \int_{[a, a+\frac{1}{n}]} F \, d\lambda \right) \end{aligned}$$

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 $c \in [a,b]$. To see this, recall that F is continuous since F is absolutely continuous. Therefore, since $c \in [a,b]$, 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, since F' and f are Lebesgue integrable, we obtain that

$$\int_{[a,x]} F' - f\, d\lambda = 0$$

for all $x \in [a,b]$. However, since $F' - f$ is Lebesgue integrable, Corollary 4.5.2 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 $x \in [a,b]$. Note each f_n is a Lebesgue measurable function being the infimum of two Lebesgue measurable functions. Moreover $|f_n| \leq n$ so f_n is Lebesgue integrable, and $\lim_{n\to\infty} f_n(x) = f(x)$ for all $x \in [a,b]$.

We claim for all $n \in \mathbb{N}$ that $F' \geq f_n$ almost everywhere. 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 $x \in [a,b]$. Since f_n and $f - f_n$ are Lebesgue integrable, we see that F_n and G_n are well-defined and absolutely continuous, $F = F_n + G_n$, and F_n and G_n are differentiable almost everywhere. Furthermore, since 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, since $f(x) \geq 0$ for almost every $x \in [a, b]$, we obtain that $F' \geq 0$ and F is non-decreasing on $[a, b]$. Therefore the Lebesgue Differentiation Theorem (Theorem 4.2.2) implies

$$F(b) - F(a) \geq \int_{[a,b]} F' \, d\lambda \geq \int_{[a,b]} f \, d\lambda = F(b) - F(a).$$

Hence F' is Lebesgue integrable and

$$\int_{[a,b]} F' - f \, d\lambda = 0.$$

Therefore, since $F' - f \geq 0$, the above integral implies that $F' = f$ almost everywhere by Theorem 3.2.3.

Case 3: f arbitrary. Recall that we may write

$$f = f_+ - f_-$$

where f_+ and f_- are non-negative Lebesgue integrable functions. Therefore, if $F_\pm : [a, b] \to \mathbb{R}$ are defined by

$$F_\pm(x) = \int_{[a,x]} f_\pm \, d\lambda,$$

then Case 2 implies that $F_\pm$ are well-defined functions such that $F_\pm' = f_\pm$ almost everywhere. Since 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. ∎

Using a proof of the second Fundamental Theorem of Calculus as a model, we obtain a Lebesgue measure theoretic version of the second Fundamental Theorem of Calculus.

Theorem 4.5.4 (Fundamental Theorem of Calculus, II). *If $F : [a, b] \to \mathbb{R}$ is absolutely continuous, then F' is Lebesgue integrable 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' Lebesgue integrable by Corollary 4.4.5. 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 4.4.9 and $G' = F'$ almost everywhere by the Fundamental Theorem of Calculus (Theorem 4.5.3). Thus $F - G$ is absolutely continuous and

$$(F - G)' = F' - G' = 0$$

almost everywhere. Hence Proposition 4.4.6 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. ∎

4.6 Leibniz Integral Rule

To finish our discussion of the connection between differentiation and the Lebesgue integral, we can prove the following very useful result from calculus with ease:

Theorem 4.6.1 (Leibniz Integral Rule). *Let $E \in \mathcal{M}(\mathbb{R})$ and let $f : E \times [c, d] \to \mathbb{R}$ be such that*

(I) for each $t \in [c, d]$, the function $g_t : E \to \mathbb{R}$ defined by $g(x) = f(x, t)$ is Lebesgue integrable,

(II) for almost every $x \in E$, the function $h_x : (c, d) \to \mathbb{R}$ defined by $h_x(t) = f(x, t)$ is differentiable on (c, d), and

(III) there exists a Lebesgue integrable function $\theta : E \to \mathbb{R}$ such that $|h'_x(t)| \leq \theta(x)$ for all $t \in (c, d)$ and almost every $x \in E$.

Then

$$\frac{d}{dt} \int_E f(x, t) \, d\lambda(x) = \int_E \frac{\partial f}{\partial t}(x, t) \, d\lambda(x)$$

for all $t \in (c, d)$.

Proof. To begin, let $I : (c, d) \to \mathbb{R}$ be defined by

$$I(t) = \int_E f(x, t) \, d\lambda(x)$$

for all $t \in (c,d)$. Therefore I is differentiable on (c,d) with

$$I'(t_0) = \lim_{h\to 0} \frac{I(t_0+h) - I(t_0)}{h} = \lim_{h\to 0} \int_E \frac{f(x,t_0+h) - f(x,t_0)}{h}\, dt$$

for all $t_0 \in (c,d)$ provided the limit exists.

Fix $t_0 \in (c,d)$. Suppose for the sake of a contradiction that the above limit does not exist or does not equal

$$\int_E \frac{\partial f}{\partial t}(x,t_0)\, d\lambda(x).$$

Hence, there exists a sequence $(h_n)_{n\geq 1}$ of non-zero real numbers such that $\lim_{n\to\infty} h_n = 0$ and

$$\lim_{n\to\infty} \int_E \frac{f(x,t_0+h_n) - f(x,t_0)}{h_n}\, dt$$

either does not exist or does not equal $\int_E \frac{\partial f}{\partial t}(x,t_0)\, d\lambda(x)$. For each $n \in \mathbb{N}$, let $g_n : E \to \mathbb{R}$ be defined by

$$g_n(x) = \frac{f(x,t_0+h_n) - f(x,t_0)}{h_n}$$

for all $x \in E$. Note that g_n is a Lebesgue integrable function by (I).

By (II) and the Mean Value Theorem, for almost every $x \in E$ for every $n \in \mathbb{N}$ there exists a $t_{x,n} \in (c,d)$ such that

$$|g_n(x)| = \left| \frac{f(x,t_0+h_n) - f(x,t_0)}{h_n} \right| = \left| h_x'(t_{x,n}) \right|.$$

Therefore (III) implies that

$$|g_n(x)| \leq \theta(x)$$

for almost every $x \in E$ for all $n \in \mathbb{N}$. Therefore, since

$$\lim_{n\to\infty} g_n(x) = \frac{\partial f}{\partial t}(x,t_0)$$

for almost every $x \in E$ and since θ is Lebesgue integrable, we obtain by the Dominated Convergence Theorem (Theorem 3.7.1) that

$$\begin{aligned}
\lim_{n\to\infty} \int_E \frac{f(x,t_0+h_n) - f(x,t_0)}{h_n}\, dt &= \lim_{n\to\infty} \int_E g_n(x)\, dt \\
&= \int_E \lim_{n\to\infty} g_n(x)\, dt \\
&= \int_E \frac{\partial f}{\partial t}(x,t_0)\, dt.
\end{aligned}$$

Hence we have a contradiction so the result follows. ∎

Remark 4.6.2. Note one can prove Theorem 4.6.1 without using a proof by contradiction by upgrading the Dominated Convergence Theorem (Theorem 3.7.1) from sequences of functions to a continuum of functions. The proof of such an upgrade is identical to the argument used in the proof of Theorem 4.6.1

Corollary 4.6.3. *Let* $f : [a, b] \times [c, d] \to \mathbb{R}$ *be such that*

- f *is continuous on* $[a, b] \times [c, d]$*, and*
- $\frac{\partial f}{\partial y}$ *exists and is continuous on* $[a, b] \times [c, d]$*.*

Then, for all $c < y < d$*,*

$$\frac{d}{dy} \int_a^b f(x, y)\, dx = \int_a^b \frac{\partial f}{\partial y}(x, y)\, dx.$$

Proof. This result follows immediately from the Leibniz's Integral Rule (Theorem 4.6.1) since every continuous function on a compact set is bounded and thus Riemann and Lebesgue integrable. ∎

Chapter 5

Higher Dimensional Lebesgue Integrals

Note that although Leibniz's Integral Rule (Theorem 4.6.1) does involve functions of two-variables, we only need to integrate against a single variable. Thus it is natural to ponder whether there is a Lebesgue integral for multivariate functions. The goal of this chapter is to show this is indeed the case and that the theory reduces to the theory developed previously in this course. To simplify the discussion, we will only consider the 2-dimensional Lebesgue integral and note a careful reading of the proofs along with the "obvious" modifications improves these results to higher dimensions.

5.1 The Two-Dimensional Lebesgue Measure

In order to define the Lebesgue integral for two-variable functions, we will first need an analogue of the Lebesgue measure that works on $\mathbb{R}^2$. Luckily, the process for constructing such a measure will follow easily from the results of Chapter 1 once we replace intervals with "rectangles".

Definition 5.1.1. The *Lebesgue measurable rectangles*, denoted $\mathcal{R}$, is the set

$$\mathcal{R} = \{I \times J \mid I, J \subseteq \mathbb{R}, I \text{ and } J \text{ are intervals}\}.$$

The *area function on* $\mathcal{R}$ is the function $\ell_2 : \mathcal{R} \to [0, \infty]$ defined by

$$\ell_2(I \times J) = \ell(I)\ell(J) = \lambda(I)\lambda(J)$$

for all $I \times J \in \mathcal{R}$.

By replacing intervals and their lengths with rectangles and their areas, we obtain a version of the Lebesgue outer measure for $\mathbb{R}^2$.

Definition 5.1.2. The *2-dimensional Lebesgue outer measure* is the function $\lambda_2^* : \mathcal{P}(\mathbb{R}^2) \to [0, \infty]$ defined by

$$\lambda_2^*(A) = \inf \left\{ \sum_{n=1}^{\infty} \ell_2(I_n \times J_n) \,\middle|\, \substack{\{I_n, J_n \,|\, n \in \mathbb{N}\} \text{ are open intervals of } \mathbb{R} \\ \text{such that } A \subseteq \bigcup_{n=1}^{\infty} I_n \times J_n} \right\}.$$

Of course, to apply the Carathéodory Method to obtain a measure from λ_2^*, we need only verify that λ_2^* is indeed an outer measure.

Theorem 5.1.3. *The 2-dimensional Lebesgue outer measure is an outer measure.*

Proof. It is clear from the definition of λ_2^* that $\lambda_2^*(\emptyset) = 0$. Moreover, if $A \subseteq B \subseteq \mathbb{R}^2$, then every collection of open rectangles that covers B must also cover A. Therefore, since $\lambda_2^*(A)$ and $\lambda_2^*(B)$ are computed via infimums, we obtain that $\lambda_2^*(A) \leq \lambda_2^*(B)$ if $A \subseteq B \subseteq \mathbb{R}$.

Let $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{P}(\mathbb{R}^2)$ and let $A = \bigcup_{n=1}^{\infty} A_n$. Fix $\epsilon > 0$. By the definition of λ_2^*, for each $n \in \mathbb{N}$ there exists a collection $\{I_{n,k}, J_{n,k} \mid k \in \mathbb{N}\}$ of open intervals such that $A_n \subseteq \bigcup_{k=1}^{\infty} I_{n,k} \times J_{n,k}$ and

$$\sum_{k=1}^{\infty} \ell_2(I_{n,k} \times J_{n,k}) \leq \lambda_2^*(A_n) + \frac{\epsilon}{2^n}.$$

Since countable unions of countable sets are countable (see Appendix B.5.3), $\{I_{n,k}, J_{n,k} \mid n, k \in \mathbb{N}\}$ is a countable collection of open intervals such that

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

Hence the definition of λ_2^* implies that

$$\lambda_2^*(A) \leq \sum_{n,k=1}^{\infty} \ell_2(I_{n,k} \times J_{n,k}) \leq \sum_{n=1}^{\infty} \lambda_2^*(A_n) + \frac{\epsilon}{2^n} = \epsilon + \sum_{n=1}^{\infty} \lambda^*(A_n).$$

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

$$\lambda_2^*(A) \leq \sum_{n=1}^{\infty} \lambda_2^*(A_n)$$

as desired. ∎

Definition 5.1.4. Let λ_2^* be the 2-dimensional Lebesgue outer measure from Definition 5.1.2. By Theorem 1.5.6 the collection $\mathcal{M}(\mathbb{R}^2)$ of λ_2^*-measurable sets is a σ-algebra and $\lambda_2^*|_{\mathcal{M}(\mathbb{R}^2)}$ is a measure. We call $\lambda_2 = \lambda_2^*|_{\mathcal{M}(\mathbb{R}^2)}$ the *2-dimensional Lebesgue measure* on $\mathbb{R}^2$ and elements of $\mathcal{M}(\mathbb{R}^2)$ *2-dimensional Lebesgue measurable sets.*

In order to use make use of the 2-dimensional Lebesgue measure and to show that it is actually a 'measure of area', we desire to show that Lebesgue measurable rectangles are 2-dimensional Lebesgue measurable sets and that their 2-dimensional Lebesgue measure is their area. Thus we proceed as we did in Chapter 1.

Proposition 5.1.5. *If $I, J \subseteq \mathbb{R}$ are intervals, then $\lambda_2^*(I \times J) = \lambda(I)\lambda(J)$.*

Proof. First suppose $I = [a, b]$ and $J = [c, d]$. To see that

$$\lambda_2^*(I) \leq (b-a)(d-c),$$

let $\epsilon > 0$ be arbitrary. Then $I' = (a-\epsilon, b+\epsilon)$ and $J' = (c-\epsilon, d+\epsilon)$ are open intervals such that $I \times J \subseteq I' \times J'$. Hence, by the definition of λ_2^* (using the empty set for all other open rectangles in our countable collection which covers $I \times J$), we obtain that

$$\lambda_2^*(I \times J) \leq \ell_2(I' \times J') = (b-a)(d-c) + 2\epsilon.$$

Therefore, as $\epsilon > 0$ was arbitrary, we obtain that $\lambda_2^*(I) \leq (b-a)(d-c)$.

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

Note $\ell_2(\overline{I_k} \times \overline{J_k}) = \ell_2(I_k \times J_k)$ and $I \times J \subseteq \bigcup_{k=1}^m \overline{I_k} \times \overline{J_k}$ where $\overline{J}$ denotes the closure of an interval J (i.e. add the endpoints). Since $\{\overline{I_k}\}_{k=1}^m$ is a finite set, we can write the set of all endpoints of all $\overline{I_k}$ between a and b as

$$a = a_0 < a_1 < a_2 < \cdots < a_p = b.$$

Similarly, since $\{\overline{J_k}\}_{k=1}^m$ is a finite set, we can write the set of all endpoints of all $\overline{J_k}$ between c and d as

$$c = c_0 < c_1 < c_2 < \cdots < c_q = d.$$

Clearly

$$\bigcup_{x=1}^{p} \bigcup_{y=1}^{q} [a_{x-1}, a_x] \times [c_{y-1}, c_y] = I \times J \subseteq \bigcup_{k=1}^{m} \overline{I_k} \times \overline{J_k}.$$

Moreover, since each $\overline{I_k} \times \overline{J_k} \cap ([a, b] \times [c, d])$ is a finite union of products of intervals of the form $[a_{x-1}, a_x] \times [c_{y-1}, c_y]$ such that the area of $\overline{I_k} \times \overline{J_k}$ is the sum of the areas of the intervals in the product, we obtain that

$$\begin{aligned}\sum_{k=1}^{\infty} \ell_2(I_k \times J_k) &\geq \sum_{k=1}^{m} \ell_2(I_k \times J_k) \\ &= \sum_{k=1}^{m} \ell_2(\overline{I_k} \times \overline{J_k}) \\ &\geq \sum_{x=1}^{p}\sum_{y=1}^{q} \ell_2([a_{x-1}, a_x] \times [c_{y-1}, c_y]) \\ &= \ell_2(I \times J) = (b-a)(d-c).\end{aligned}$$

Therefore, since $\{I_n, J_n \mid n \in \mathbb{N}\}$ was arbitrary, we obtain that

$$\lambda_2^*(I \times J) \geq (b-a)(d-c).$$

Hence $\lambda_2^*(I \times J) = (b-a)(d-c)$ as desired.

To complete the proof, first assume $I, J \subseteq \mathbb{R}$ are intervals of finite length. Thus $I \in \{(a,b), [a,b), (a,b], [a,b]\}$ for some $a, b \in \mathbb{R}$ with $a \leq b$ and $J \in \{(c,d), [c,d), (c,d], [c,d]\}$ for some $c, d \in \mathbb{R}$ with $c \leq d$. Hence $\ell_2(I \times J) = (b-a)(d-c)$. Let $\overline{I} = [a,b]$ and $\overline{J} = [c,d]$ so that $I \subseteq \overline{I}$, $J \subseteq \overline{J}$, and

$$\lambda_2^*(\overline{I} \times \overline{J}) = \ell_2(\overline{I} \times \overline{J}) = (b-a)(d-c)$$

by the previous case. For any $\epsilon > 0$ with

$$\epsilon < \min\left\{\frac{b-a}{2}, \frac{d-c}{2}\right\},$$

let $I_\epsilon = \left[a + \frac{\epsilon}{2}, b - \frac{\epsilon}{2}\right]$ and $J_\epsilon = \left[c + \frac{\epsilon}{2}, d - \frac{\epsilon}{2}\right]$. Thus $I_\epsilon \times J_\epsilon \subseteq I \times J$ and

$$\lambda_2^*(I_\epsilon \times J_\epsilon) = \ell_2(I_\epsilon \times J_\epsilon) = (b-a-\epsilon)(d-c-\epsilon)$$

for all $\epsilon > 0$. Therefore, since λ_2^* is an outer measure, we obtain for all $\epsilon > 0$ that

$$(b-a-\epsilon)(d-c-\epsilon) = \lambda_2^*(I_\epsilon \times J_\epsilon) \leq \lambda_2^*(I \times J) \leq \lambda_2^*(\overline{I} \times \overline{J}) = (b-a)(d-c).$$

Hence $\lambda_2^*(I \times J) = (b-a)(d-c)$ as desired.

Otherwise, if I is an infinite interval or J is an infinite interval, then there exists arbitrary large products of finite intervals contained in $I \times J$. Therefore, since λ_2^* is an outer measure and thus monotone, the result for products of intervals of finite length implies that

$$\lambda_2^*(I \times J) = \begin{cases} \infty = \ell_2(I \times J) & \text{if } I \text{ and } J \text{ are not singletons} \\ 0 = \ell_2(I \times J) & \text{if } I \text{ or } J \text{ is a singleton} \end{cases}.$$

Thus the proof is complete. ∎

Lemma 5.1.6. *For all $a \in \mathbb{R}$, $(a,\infty)\times\mathbb{R}$, $[a,\infty)\times\mathbb{R}$, $\mathbb{R}\times(a,\infty)$ and $\mathbb{R}\times[a,\infty)$ are 2-dimensional Lebesgue measurable.*

Proof. To see that $(a,\infty)\times\mathbb{R}$ is 2-dimensional Lebesgue measurable, let $B \subseteq \mathbb{R}^2$ be arbitrary. Therefore $B_1 = B \cap ((a,\infty)\times\mathbb{R})$ and

$$B_2 = B \cap ((a,\infty)\times\mathbb{R})^c = B \cap ((-\infty,a]\times\mathbb{R})$$

are disjoint sets such that $B = B_1 \cup B_2$.

Let $\epsilon > 0$ be arbitrary. By the definition of the 2-dimensional Lebesgue outer measure, there exists a collection $\{I_n, J_n \mid n \in \mathbb{N}\}$ of open intervals such that $B \subseteq \bigcup_{n=1}^\infty I_n \times J_n$ and

$$\sum_{n=1}^\infty \ell_2(I_n \times J_n) \leq \lambda_2^*(B) + \epsilon.$$

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 countable collections of intervals such that $B_1 \subseteq \bigcup_{n=1}^\infty I_n' \times J_n$ and $B_2 \subseteq \bigcup_{n=1}^\infty I_n'' \times J_n$. Hence

$$\begin{aligned}
&\lambda_2^*(B \cap ((a,\infty)\times\mathbb{R})) + \lambda^*(B \cap ((a,\infty)\times\mathbb{R})^c) \\
&= \lambda_2^*(B_1) + \lambda_2^*(B_2) \\
&\leq \sum_{n=1}^\infty \lambda_2^*(I_n' \times J_n) + \sum_{n=1}^\infty \lambda_2^*(I_n'' \times J_n) && \text{subadditivity} \\
&= \sum_{n=1}^\infty \ell(I_n')\ell(J_n) + \sum_{n=1}^\infty \ell(I_n'')\ell(J_n) && \text{by Proposition 5.1.5} \\
&= \sum_{n=1}^\infty \ell(I_n)\ell(J_n) \\
&\leq \lambda_2^*(B) + \epsilon.
\end{aligned}$$

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

$$\lambda_2^*(B \cap ((a,\infty)\times\mathbb{R})) + \lambda^*(B \cap ((a,\infty)\times\mathbb{R})^c) \leq \lambda_2^*(B).$$

Therefore, since $B \subseteq \mathbb{R}^2$ was arbitrary, $(a,\infty)\times\mathbb{R}$ is 2-dimensional Lebesgue measurable.

The proof that the remaining sets are 2-dimensional Lebesgue measurable is similar. ■

Theorem 5.1.7. *Every Lebesgue measurable rectangle $I \times J \in \mathbb{R}^2$ is 2-dimensional Lebesgue measurable with*

$$\lambda_2(I \times J) = \lambda(I)\lambda(J).$$

Proof. Since Lemma 5.1.6 implies $(a,\infty)\times\mathbb{R}$, $[a,\infty)\times\mathbb{R}$, $\mathbb{R}\times(c,\infty)$ and $\mathbb{R}\times[c,\infty)$ are 2-dimensional Lebesgue measurable for all $a,c\in\mathbb{R}$, and since the set of 2-dimensional Lebesgue measurable is a σ-algebra and thus closed under complements, we obtain that $(-\infty,b]\times\mathbb{R}$, $(-\infty,b)\times\mathbb{R}$, $\mathbb{R}\times(-\infty,d]$ and $\mathbb{R}\times(-\infty,d)$ are 2-dimensional Lebesgue measurable for all $b,d\in\mathbb{R}$. Therefore, since every Lebesgue measurable rectangle is the intersection of at most four of the above sets, and since set of 2-dimensional Lebesgue measurable is a σ-algebra and thus closed under intersections, we obtain that Lebesgue measurable rectangle is 2-dimensional Lebesgue measurable. Moreover, Proposition 5.1.5 immediately implies that $\lambda_2(I\times J)=\lambda(I)\lambda(J)$ for every Lebesgue measurable rectangle $I\times J\in\mathbb{R}^2$. ∎

Remark 5.1.8. By following the proofs from Chapter 1, Chapter 2, and Chapter 3, we obtain that

- The 2-dimensional Lebesgue measure is translation and inversion invariant in each variable,
- a function $f:\mathbb{R}^2\to\mathbb{R}$ is 2-dimensional Lebesgue measurable if and only if $f^{-1}((a,\infty))\in\mathcal{M}(\mathbb{R}^2)$ for all $a\in\mathbb{R}$,
- 2-dimensional Lebesgue measurable functions behave identically to Lebesgue measurable functions,
- Egoroff's Theorem (Theorem 2.4.1), Littlewood's First Principle (Theorem 2.5.1) where open intervals are replaced with open rectangles, and Lusin's Theorem (Theorem 2.6.1) all hold for 2-dimensional Lebesgue measurable functions (with the only gap in the proof is proving Tietz Extension Theorem (Theorem 2.6.2) for $\mathbb{R}^2$), and
- the 2-dimensional Lebesgue integral is defined in an analogous way and satisfies the Monotone Convergence Theorem (Theorem 3.3.2), Fatou's Lemma (Theorem 3.6.1), and the Dominated Convergence Theorem (Theorem 3.7.1).

Thus we can proceed with 2-dimensional Lebesgue measure theory identically to how we proceeded with Lebesgue measure theory.

Remark 5.1.9. It is also possible to define the 2-dimensional Lebesgue outer measure using open circles instead of open rectangles. We will not demonstrate the equivalence. One can see why we proceed with rectangles instead of circles in the next section.

5.2 Tonelli's and Fubini's Theorem

In this section, demonstrate two theorems that show that the integration theory for higher dimensional Lebesgue integrals reduces to the theory of the one-dimensional Lebesgue integral. We begin with the statements of the two theorems.

Theorem 5.2.1 (Fubini's Theorem). *If $f : \mathbb{R}^2 \to \mathbb{R}$ is 2-dimensional Lebesgue integrable, then:*

1. *for almost every $x \in \mathbb{R}$ the function $f_x : \mathbb{R} \to \mathbb{R}$ defined by $f_x(y) = f(x,y)$ for all $y \in \mathbb{R}$ is a well-defined Lebesgue integrable function and for almost every $y \in \mathbb{R}$ the function $f_y : \mathbb{R} \to \mathbb{R}$ defined by $f_y(x) = f(x,y)$ for all $x \in \mathbb{R}$ is a well-defined Lebesgue integrable function,*

2. *the function $\Phi : \mathbb{R} \to \mathbb{R}$ defined by $\Phi(x) = \int_Y f_x \, d\lambda$ is a well-defined Lebesgue integrable function and the function $\Psi : \mathbb{R} \to \mathbb{R}$ defined by $\Psi(y) = \int_X f_y \, d\lambda$ is a well-defined Lebesgue integrable function, and*

3. *$\int_{\mathbb{R}^2} f \, d\lambda_2 = \int_{\mathbb{R}} \Phi \, d\lambda = \int_{\mathbb{R}} \Psi \, d\lambda$; that is*

$$\begin{aligned}\int_{\mathbb{R}^2} f \, d\lambda_2 &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x,y) \, d\lambda(y) \right) \, d\lambda(x) \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x,y) \, d\lambda(x) \right) \, d\lambda(y).\end{aligned}$$

Theorem 5.2.2 (Tonelli's Theorem). *If $f : \mathbb{R}^2 \to [0,\infty]$ is 2-dimensional Lebesgue measurable, then:*

1. *for almost every $x \in \mathbb{R}$ the function $f_x : \mathbb{R} \to [0,\infty]$ defined by $f_x(y) = f(x,y)$ for all $y \in \mathbb{R}$ is a well-defined Lebesgue measurable function and for almost every $y \in \mathbb{R}$ the function $f_y : \mathbb{R} \to [0,\infty]$ defined by $f_y(x) = f(x,y)$ for all $x \in \mathbb{R}$ is a well-defined Lebesgue measurable function,*

2. *the function $\Phi : \mathbb{R} \to [0,\infty]$ defined by $\Phi(x) = \int_{\mathbb{R}} f_x \, d\lambda$ is a well-defined Lebesgue measurable function and the function $\Psi : \mathbb{R} \to [0,\infty]$ defined by $\Psi(y) = \int_{\mathbb{R}} f_y \, d\lambda$ is a well-defined Lebesgue measurable function, and*

3. *$\int_{\mathbb{R}^2} f \, d\lambda_2 = \int_{\mathbb{R}} \Phi \, d\lambda = \int_{\mathbb{R}} \Psi \, d\lambda$; that is*

$$\begin{aligned}\int_{\mathbb{R}^2} f \, d\lambda_2 &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x,y) \, d\lambda(y) \right) \, d\lambda(x) \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x,y) \, d\lambda(x) \right) \, d\lambda(y).\end{aligned}$$

Remark 5.2.3. It is clear that Fubini's Theorem (Theorem 5.2.1) is more general that Tonelli's Theorem (Theorem 5.2.2) except for positive functions that integrate to infinity. So what is the point of Tonelli's Theorem? The main use of Tonelli's Theorem is to show that the assumptions of Fubini's Theorem holds; that is, to check that f 2-dimensional Lebesgue integrable, one generally verifies that

$$\int_{\mathbb{R}^2} |f| \, d\lambda_2 < \infty$$

using Tonelli's Theorem!

Thus it remains to prove Fubini's Theorem (Theorem 5.2.1) and Tonelli's Theorem (Theorem 5.2.2). As the proofs are long and complicated, we will divided the proofs into several lemmata. The idea of the proof is similar to how we prove any complicated result for the Lebesgue integrals: we first prove the results for the characteristic functions, which by linearity gives the result for all simple function, which then gives the result for non-negative functions, which then gives the result for all integrable functions. Bizarrely enough, we first will prove Fubini's Theorem (i.e. the result for integrable functions) and use Fubini's Theorem to prove Tonelli's Theorem.

To verify that Fubini's Theorem holds for characteristic functions, we will need to make use of specific collections of 2-dimensional Lebesgue measurable sets.

Notation 5.2.4. Given $Z \subseteq \mathbb{R}^2$ and $x, y \in \mathbb{R}$ representing the x- and y-terms of a pair, denote

$$Z_x = \{w \in \mathbb{R} \mid (x, w) \in Z\} \qquad \text{and} \qquad Z_y = \{z \in \mathbb{R} \mid (z, y) \in Z\}.$$

Similarly, given a function $f : \mathbb{R} \to \mathbb{R}$, let $f_x : \mathbb{R} \to \mathbb{R}$ and $f_y : \mathbb{R} \to \mathbb{R}$ denote the functions defined by

$$f_x(w) = f(x, w) \qquad \text{and} \qquad f_y(w) = f(w, y)$$

for all $w \in \mathbb{R}$. Finally, let

$$\begin{aligned} \mathcal{R}_\sigma &= \left\{ Z \subseteq \mathbb{R}^2 \mid Z \text{ is a countable union of elements of } \mathcal{R} \right\} \\ \mathcal{R}_{\sigma\delta} &= \left\{ Z \subseteq \mathbb{R}^2 \mid Z \text{ is a countable intersection of elements of } \mathcal{R}_\sigma \right\} \end{aligned}$$

Note since $\mathcal{R} \subseteq \mathcal{M}(\mathbb{R}^2)$ and since $\mathcal{M}(\mathbb{R}^2)$ is a σ-algebra that

$$\mathcal{R} \subseteq \mathcal{R}_\sigma \subseteq \mathcal{R}_{\sigma\delta} \subseteq \mathcal{M}(\mathbb{R}^2).$$

Before we begin the proof of Fubini's Theorem for characteristic functions, we first need the following that shows that $\mathcal{R}_\sigma$ is well-behaved.

Lemma 5.2.5. *If $Z \in \mathcal{R}_\sigma$ then there exists a pairwise disjoint collection $\{R_k\}_{k=1}^\infty \subseteq \mathcal{R}$ such that $Z = \bigcup_{k=1}^\infty R_k$.*

Proof. Let $Z \in \mathcal{R}_\sigma$ be arbitrary. Hence we may write $Z = \bigcup_{k=1}^{\infty} I_k \times J_k$ where $\{I_k\}_{k=1}^{\infty}$ and $\{J_k\}_{k=1}^{\infty}$ are collections of intervals of $\mathbb{R}$. We will proceed by recursion on m to show that $\bigcup_{k=1}^{m} I_k \times J_k$ can be written as a disjoint union of elements of $\mathcal{R}$. Clearly the case $m = 1$ is trivial.

Assume it has been demonstrated for some $m \geq 1$ that $\bigcup_{k=1}^{m} I_k \times J_k = \bigcup_{k=1}^{M} I_k' \times J_k'$ where $\{I_k' \times J_k'\}_{k=1}^{M}$ are pairwise disjoint elements of $\mathcal{R}$. To see that $\bigcup_{k=1}^{m+1} I_k \times J_k$ can be written as a disjoint union of elements of $\mathcal{R}$, consider

$$X_1 = (I_{m+1} \times J_{m+1}) \setminus (I_1' \times J_1').$$

Since the set difference of one interval by another is the union of at most two disjoint intervals and since X_1 can be written as the disjoint union of

$$\begin{aligned} R_1 &= (I_{m+1} \setminus I_1') \times (J_{m+1} \cap J_1'), \\ R_2 &= (I_{m+1} \cap I_1') \times (J_{m+1} \setminus J_1'), \text{and} \\ R_3 &= (I_{m+1} \setminus I_1') \times (J_{m+1} \setminus J_1'), \end{aligned}$$

we obtain that X_1 can be written as the disjoint union of at most 8 Lebesgue measurable rectangles. By repeating this process, we see that

$$X_2 = X_1 \setminus (I_2' \times J_2') = (I_{m+1} \times J_{m+1}) \setminus ((I_1' \times J_1') \cup (I_2' \times J_2'))$$

can be written as the disjoint union of at most 64 Lebesgue measurable rectangles. Therefore, by repeating this process ad nauseum, we obtain that $\bigcup_{k=1}^{m+1} I_k \times J_k$ can be written as a disjoint union of elements of $\mathcal{R}$. ■

With Lemma 5.2.5 in hand, we can proceed with our proof that Fubini's Theorem holds for all characteristic functions. Our first goal is to show that Fubini's Theorem holds for all characteristic functions of elements of $\mathcal{R}_{\sigma\delta}$. Thus we begin with the following.

Lemma 5.2.6. *If $Z \in \mathcal{R}_{\sigma\delta}$, then $Z_x \in \mathcal{M}(\mathbb{R})$ for every $x \in \mathbb{R}$.*

Proof. The proof is divided into three cases of increasing generality.

Case 1: $Z \in \mathcal{R}$. In this case, we may write $Z = I \times J$ for some intervals A and B. Note for each $x \in \mathbb{R}$ that

$$Z_x = \begin{cases} \emptyset & \text{if } x \notin I \\ J & \text{if } x \in I \end{cases}.$$

Hence $Z_x \in \{\emptyset, B\} \subseteq \mathcal{M}(\mathbb{R})$ in this case.

Case 2: $Z \in \mathcal{R}_\sigma$. In this case $Z = \bigcup_{n=1}^{\infty} R_n$ for some collection $\{R_n\}_{n=1}^{\infty} \subseteq \mathcal{R}$. Since

$$Z_x = \bigcup_{n=1}^{\infty} (R_n)_x$$

for each $x \in \mathbb{R}$, and since $(R_n)_x \in \mathcal{M}(\mathbb{R})$ for all $n \in \mathbb{N}$ and $x \in \mathbb{R}$ by Case 1, we see that $Z_x \in \mathcal{M}(\mathbb{R})$ for all $x \in \mathbb{R}$ in this case.

Case 3: $Z \in \mathcal{R}_{\sigma\delta}$. In this case $Z = \bigcap_{n=1}^{\infty} Z_n$ for some collection $\{Z_n\}_{n=1}^{\infty} \subseteq \mathcal{R}_\sigma$. Since

$$Z_x = \bigcap_{n=1}^{\infty} (Z_n)_x$$

for each $x \in \mathbb{R}$, and since $(Z_n)_x \in \mathcal{M}(\mathbb{R})$ for all $n \in \mathbb{N}$ and $x \in \mathbb{R}$ by Case 2, we see that $Z_x \in \mathcal{M}(\mathbb{R})$ for all $x \in \mathbb{R}$. Thus the proof is complete. ∎

Lemma 5.2.7. *Let $Z \in \mathcal{R}_{\sigma\delta}$ be such that $\lambda_2(Z) < \infty$ and define $g : \mathbb{R} \to [0, \infty]$ by $g(x) = \lambda(Z_x)$ for all $x \in \mathbb{R}$. Then g is Lebesgue measurable and*

$$\int_{\mathbb{R}} g \, d\lambda = \lambda_2(Z) = \int_{\mathbb{R}^2} \chi_Z \, d\lambda_2.$$

In particular $\lambda(Z_x) < \infty$ for almost every $x \in \mathbb{R}$.

Proof. First note that g is well-defined since $Z_x \in \mathcal{M}(\mathbb{R})$ for all $x \in \mathbb{R}$ by Lemma 5.2.6. Furthermore, the equality

$$\lambda_2(Z) = \int_{\mathbb{R}^2} \chi_Z \, d\lambda_2$$

follows trivially by the definition of the 2-dimensional Lebesgue integral.

The remainder of the proof is divided into three cases of increasing generality.

Case 1: $Z \in \mathcal{R}$. In this case $Z = I \times J$ for some intervals I and J. Since

$$Z_x = \begin{cases} \emptyset & \text{if } x \notin I \\ J & \text{if } x \in I \end{cases}$$

for all $x \in \mathbb{R}$, we see that

$$g(x) = \lambda(Z_x) = \lambda(J)\chi_I(x)$$

for all $x \in \mathbb{R}$. Hence g is clearly Lebesgue measurable since $I \in \mathcal{M}(\mathbb{R})$ and

$$\int_{\mathbb{R}} g \, d\lambda = \lambda(J) \int_{\mathbb{R}} \chi_I \, d\lambda = \lambda(I)\lambda(J) = \lambda_2(Z)$$

by Theorem 5.1.7 as desired.

Case 2: $Z \in \mathcal{R}_\sigma$. In this case $Z = \bigcup_{n=1}^{\infty} R_n$ for some collection $\{R_n\}_{n=1}^{\infty} \subseteq \mathcal{R}$. By Lemma 5.2.5 we can assume that the collection $\{R_n\}_{n=1}^{\infty} \subseteq \mathcal{R}$ is pairwise disjoint. Hence $\{(R_n)_x\}_{n=1}^{\infty} \subseteq \mathcal{M}(\mathbb{R})$ is pairwise disjoint for all $x \in \mathbb{R}$. Therefore

$$g(x) = \lambda(Z_x) = \lambda\left(\bigcup_{n=1}^{\infty} (R_n)_x\right) = \sum_{n=1}^{\infty} \lambda((R_n)_x)$$

for all $x \in \mathbb{R}$. Therefore, by Case 1, g is a countable sum of non-negative Lebesgue measurable functions and hence is Lebesgue measurable by Proposition 2.2.9. Moreover, by Corollary 3.3.5,

$$\begin{aligned}\int_{\mathbb{R}} g\, d\lambda &= \int_{\mathbb{R}} \sum_{n=1}^{\infty} \lambda((R_n)_x)\, d\lambda(x) \\ &= \sum_{n=1}^{\infty} \int_{\mathbb{R}} \lambda((R_n)_x)\, d\lambda(x) \\ &= \sum_{n=1}^{\infty} \lambda_2(R_n) && \text{by Case 1} \\ &= \lambda_2\left(\bigcup_{n=1}^{\infty} R_n\right) \\ &= \lambda_2(Z)\end{aligned}$$

as desired.

Case 3: $Z \in \mathcal{R}_{\sigma\delta}$. In this case $Z = \bigcap_{n=1}^{\infty} Z_n$ for some collection $\{Z_n\}_{n=1}^{\infty} \subseteq \mathcal{R}_\sigma$. Since $\lambda_2(Z) < \infty$, by the definition of λ_2 there exists a $\{C_n\}_{n=1}^{\infty} \in \mathcal{R}$ such that $Z \subseteq \bigcup_{n=1}^{\infty} C_n$ and

$$\lambda_2(Z) \leq \sum_{n=1}^{\infty} \ell_2(C_n) < \infty.$$

Let $Z_0' = \bigcup_{n=1}^{\infty} C_n$. Then $Z_0' \in \mathcal{R}_\sigma$ and $\lambda_2(Z_0) < \infty$. Moreover, since the intersection of any two elements of $\mathcal{R}$ is an element of $\mathcal{R}$ (i.e. $(I_1 \times J_1) \cap (I_2 \times J_2) = (I_1 \cap I_2) \times (J_1 \cap J_2)$), since each element of $\mathcal{R}_\sigma$ is a countable union of elements of $\mathcal{R}$, and since the countable union of countable sets is countable, we note that if

$$Z_n' = Z_n \cap Z_{n-1}'$$

for all $n \in \mathbb{N}$, then $\{Z_n'\}_{n=0}^{\infty} \subseteq \mathcal{R}_\sigma$, $Z = \bigcap_{n=0}^{\infty} Z_n'$, $\lambda_2(Z_0') < \infty$, and $Z_n' \subseteq Z_{n-1}'$ for all $n \in \mathbb{N}$.

For each $n \in \mathbb{N} \cup \{0\}$, let $g_n : \mathbb{R} \to [0, \infty]$ be defined by $g_n(x) = \lambda((Z_n')_x)$ for all $x \in \mathbb{R}$. Clearly each g_n is Lebesgue measurable by Case 2. Moreover, Case 2 implies that

$$0 \leq \int_{\mathbb{R}} \lambda((Z_0')_x)\, d\lambda(x) = \int_{\mathbb{R}} g_0\, d\lambda = \lambda_2(Z_0') < \infty$$

and thus $\lambda((Z_0')_x) < \infty$ for almost every x.

Notice that $Z_x = \bigcap_{n=1}^{\infty} (Z_n')_x$ for all $x \in \mathbb{R}$ and, since $Z_n' \subseteq Z_{n-1}'$ for all $n \in \mathbb{N}$, that $(Z_n')_x \subseteq (Z_{n-1}')_x$ for all $n \in \mathbb{N}$ and $x \in \mathbb{R}$. Therefore, we obtain by the Monotone Convergence Theorem for Measures (Theorem 1.3.9) that

$$\lim_{n\to\infty} g_n(x) = \lim_{n\to\infty} \lambda((Z_n')_x) = \lambda(Z_x) = g(x)$$

for almost every x. Therefore Corollary 2.2.15 implies that g is Lebesgue measurable.

Since $(Z'_n)_x \subseteq (Z'_{n-1})_x$ for all $n \in \mathbb{N}$ and $x \in \mathbb{R}$, we see that $g_n(x) \leq g_0(x)$ for all $x \in \mathbb{R}$. However, since g_0 is non-negative and Lebesgue measurable, we see by Case 2 that

$$\int_{\mathbb{R}} g_0 \, d\lambda = \lambda_2(Z'_0) < \infty$$

and thus g_0 is Lebesgue integrable. Therefore, by the Dominated Convergence Theorem (Theorem 3.7.1) and Case 2, we obtain that

$$\begin{aligned}\int_{\mathbb{R}} g \, d\lambda &= \lim_{n\to\infty} \int_{\mathbb{R}} g_n \, d\lambda \\ &= \lim_{n\to\infty} \int_{\mathbb{R}} \lambda((Z'_n)_x) \, d\lambda(x) \\ &= \lim_{n\to\infty} \lambda_2(Z'_n).\end{aligned}$$

However, since $\lambda_2(Z'_0) < \infty$, $Z'_n \subseteq Z'_{n-1}$ for all $n \in \mathbb{N}$, and $Z = \bigcap_{n=1}^\infty Z'_n$, we obtain by the Monotone Convergence Theorem for Measures (Theorem 1.3.9) that

$$\int_{\mathbb{R}} g \, d\lambda = \lim_{n\to\infty} \lambda_2(Z'_n) = \lambda_2(Z)$$

as desired. Thus the proof is complete. ∎

Note Lemma 5.2.7 proves the desired result (i.e. Fubini's Theorem (Theorem 5.2.1)) for all characteristic functions of elements of $\mathcal{R}_{\sigma\delta}$. To extend this to all characteristic functions, we will require the following two lemmata. Note the proof of the first lemma is very similar to the proof of Proposition 1.6.13.

Lemma 5.2.8. *If $Z \in \mathcal{M}(\mathbb{R}^2)$ is such that $\lambda_2(Z) < \infty$, then there exists an $G \in \mathcal{R}_{\sigma\delta}$ such that $Z \subseteq G$ and $\lambda_2(G \setminus Z) = 0$.*

Proof. First, fix an $\epsilon > 0$. By the definition of λ_2 there exists a countable collection $\{R_n\}_{n=1}^\infty \subseteq \mathcal{R}$ such that $Z \subseteq \bigcup_{n=1}^\infty R_n$ and

$$\sum_{n=1}^\infty \lambda_2(R_n) = \sum_{n=1}^\infty \ell_2(R_n) \leq \lambda_2(Z) + \epsilon.$$

Let $G_\epsilon = \bigcup_{n=1}^\infty R_n \in R_\sigma$. Then clearly $Z \subseteq G_\epsilon$ and

$$\lambda_2(Z) \leq \lambda_2(G_\epsilon) \leq \sum_{n=1}^\infty \lambda_2(R_n) \leq \lambda_2(Z) + \epsilon.$$

Let $G = \bigcap_{n=1}^\infty G_{\frac{1}{n}} \in R_{\sigma\delta}$. Clearly $Z \subseteq G$ since $Z \subseteq G_{\frac{1}{n}}$ for all $n \in \mathbb{N}$. Moreover, clearly

$$\lambda_2(Z) \leq \lambda_2(G) \leq \lambda_2\left(G_{\frac{1}{n}}\right) \leq \lambda_2(Z) + \frac{1}{n}$$

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

$$\lambda_2(Z) = \lambda_2(G).$$

Therefore, since $\lambda_2(Z) < \infty$ and $Z \subseteq G$, we obtain by Remark 1.3.3 that

$$\lambda_2(G \setminus Z) = 0$$

as desired. ∎

Lemma 5.2.9. *If $Z \in \mathcal{M}(\mathbb{R}^2)$ is such that $\lambda_2(Z) = 0$, then Z_x is Lebesgue measurable with $\lambda(Z_x) = 0$ for almost every $x \in \mathbb{R}$.*

Proof. By Lemma 5.2.8 there exists an $G \in R_{\sigma\delta}$ such that $Z \subseteq G$ and $\lambda_2(G \setminus Z) = 0$. Hence $\lambda_2(G) = 0$.

By Lemma 5.2.7 if $g : \mathbb{R} \to [0, \infty]$ is defined by by $g(x) = \lambda(G_x)$ for all $x \in \mathbb{R}$, then g is Lebesgue measurable and

$$\int_{\mathbb{R}} g \, d\lambda = \lambda_2(G) = 0.$$

Therefore $0 = g(x) = \lambda(G_x)$ for almost every $x \in \mathbb{R}$ by Theorem 3.2.3. Since $Z \subseteq G$ so $Z_x \subseteq G_x$ for all $x \in \mathbb{R}$ and since the Lebesgue measure is complete, we obtain that Z_x is Lebesgue measurable with $\lambda(Z_x) = 0$ for almost every $x \in \mathbb{R}$. ∎

With the above lemmata, we can finally prove Fubini's Theorem for characteristic functions.

Lemma 5.2.10. *If $Z \in \mathcal{M}(\mathbb{R}^2)$ is such that $\lambda_2(Z) < \infty$, then Fubini's Theorem (Theorem 5.2.1) holds for the function $f = \chi_Z$.*

Proof. Fix $Z \in \mathcal{M}(\mathbb{R}^2)$ such that $\lambda_2(Z) < \infty$. By Lemma 5.2.8 there exists an $G \in \mathcal{R}_{\sigma\delta}$ such that $Z \subseteq G$ and $\lambda_2(G \setminus Z) = 0$.

Notice for all $x \in X$ that

$$Z_x = (G_x) \setminus (G \setminus Z)_x.$$

Since $\lambda_2(G) = \lambda_2(Z) < \infty$, we know that G_x is Lebesgue measurable for all $x \in \mathbb{R}$ by Lemma 5.2.7. Moreover, since $\lambda_2(G \setminus Z) = 0$, we know that $(G \setminus Z)_x$ is Lebesgue measurable for almost every $x \in \mathbb{R}$ by Lemma 5.2.9. Hence Z_x is Lebesgue measurable for almost every $x \in \mathbb{R}$. Moreover, by Lemma 5.2.9,

$$\lambda(Z_x) = \lambda(G_x) - \lambda((G \setminus Z)_x) = \lambda(G_x).$$

for almost every $x \in \mathbb{R}$.

Let $f = \chi_Z$ and notice that $f_x : \mathbb{R} \to [0, 1]$ is defined by

$$f_x(y) = f(x, y) = \chi_Z(x, y) = \chi_{Z_x}(y).$$

Therefore, since Z_x is Lebesgue measurable for almost every $x \in \mathbb{R}$, f_x is Lebesgue measurable for almost every $x \in \mathbb{R}$. Moreover

$$\int_{\mathbb{R}} f_x \, d\lambda = \int_{\mathbb{R}} \chi_{Z_x} \, d\lambda = \lambda(Z_x) = \lambda(G_x)$$

for almost every $x \in \mathbb{R}$. However, by Lemma 5.2.7,

$$\int_{\mathbb{R}} \lambda(G_x) \, d\lambda(x) = \lambda_2(G) < \infty$$

so that $\lambda(G_x) < \infty$ for almost every $x \in \mathbb{R}$. Hence $\int_{\mathbb{R}} f_x \, d\lambda < \infty$ for almost every $x \in \mathbb{R}$ so f_x is Lebesgue integrable for almost every $x \in \mathbb{R}$ as desired.

Next recall that $\Phi : \mathbb{R} \to [0, \infty]$ is defined by

$$\Phi(x) = \int_{\mathbb{R}} f_x \, d\lambda = \lambda(Z_x) = \lambda(G_x)$$

for all $x \in \mathbb{R}$. Therefore, by Lemma 5.2.7, Φ is Lebesgue measurable and

$$\int_{\mathbb{R}} \Phi \, d\lambda = \int_{\mathbb{R}} \lambda(G_x) \, d\lambda(x) = \lambda_2(G) < \infty.$$

Hence Φ is Lebesgue integrable as desired.

Finally, by Lemma 5.2.7,

$$\int_{\mathbb{R}} \Phi \, d\lambda = \lambda_2(G) = \lambda_2(Z) = \int_{\mathbb{R}^2} \chi_Z \, d\lambda_2$$

as desired. The remainder of the proof of Fubini's Theorem (Theorem 5.2.1) in this case holds by symmetry (i.e. repeat Lemmata 5.2.6, 5.2.7, and 5.2.9 with y in place of x). ∎

Finally, we can complete the proof of Fubini's Theorem (Theorem 5.2.1).!

Proof of Fubini's Theorem (Theorem 5.2.1). To begin, note Lemma 5.2.10 implies Fubini's Theorem holds for characteristic functions of finite 2-dimensional Lebesgue measure. Therefore, since simple functions are linear combinations of characteristic functions, it is elementary to see that Fubini's Theorem holds for 2-dimensional Lebesgue integrable simple functions.

Let f satisfy the assumptions of Fubini's Theorem. Recall that every 2-dimensional Lebesgue integrable function is a linear combination of two non-negative 2-dimensional Lebesgue integrable function. Therefore, since it is elementary to see that if Fubini's Theorem holds for a finite set of functions then Fubini's Theorem holds for all linear combinations of those functions, we may assume without loss of generality that f is non-negative.

Since f is non-negative, Theorem 2.3.5 implies there exists a sequence $(\varphi_n)_{n\geq 1}$ of simple functions on $(\mathbb{R}^2, \lambda_2)$ such that $\varphi_n \leq \varphi_{n+1}$ for all $n \in \mathbb{N}$

and $(\varphi_n)_{n\geq 1}$ converges to f pointwise. Hence the Monotone Convergence Theorem (Theorem 3.3.2) implies that

$$\lim_{n\to\infty}\int_{\mathbb{R}^2}\varphi_n\,d\lambda_2 = \int_{\mathbb{R}^2} f\,d\lambda_2 < \infty.$$

Moreover, since $0 \leq \varphi_n \leq f$, we see that φ_n are 2-dimensional Lebesgue integrable for all $n \in \mathbb{N}$. Therefore Fubini's Theorem holds for each φ_n.

To see that f_x is Lebesgue measurable for almost every $x \in \mathbb{R}$, notice by construction that

$$\lim_{n\to\infty}(\varphi_n)_x(y) = \lim_{n\to\infty}\varphi_n(x,y) = f(x,y) = f_x(y)$$

for all $(x,y) \in X \times Y$. Therefore, since the Lebesgue measure is complete and since $y \mapsto (\varphi_n)_x(y)$ is Lebesgue measurable for almost every $x \in \mathbb{R}$, we obtain by Proposition 2.2.9 that f_x is Lebesgue measurable for almost every $x \in \mathbb{R}$. Furthermore since $\varphi_n \leq \varphi_{n+1}$ implies that $(\varphi_n)_x(y) \leq (\varphi_{n+1})_x(y)$, the Monotone Convergence Theorem (Theorem 3.3.2) implies that

$$\Phi(x) = \int_{\mathbb{R}} f_x\,d\lambda = \lim_{n\to\infty}\int_{\mathbb{R}}(\varphi_n)_x\,d\lambda$$

for almost every $x \in \mathbb{R}$. Hence, since $\theta_n : \mathbb{R} \to [0,\infty]$ defined by

$$\theta_n(x) = \int_{\mathbb{R}}(\varphi_n)_x\,d\lambda$$

is Lebesgue measurable for every $n \in \mathbb{N}$, Proposition 2.2.9 implies that Φ is Lebesgue measurable. Moreover, since $\varphi_n \leq \varphi_{n+1}$ implies that $\theta_n \leq \theta_{n+1}$ for all $n \in \mathbb{N}$ and since $\lim_{n\to\infty}\varphi_n(x) = \Phi(x)$ for almost every $x \in \mathbb{R}$, we again obtain that

$$\begin{aligned}
\int_{\mathbb{R}}\Phi\,d\lambda &= \lim_{n\to\infty}\int_{\mathbb{R}}\theta_n\,d\lambda \quad \text{by the Monotone Convergence Theorem}\\
&= \lim_{n\to\infty}\int_{\mathbb{R}}\left(\int_{\mathbb{R}}(\varphi_n)_x(y)\,d\lambda(y)\right)d\lambda(x)\\
&= \lim_{n\to\infty}\int_{\mathbb{R}^2}\varphi_n\,d\lambda_2 \quad \text{since Fubini's Theorem holds for } \varphi_n\\
&= \int_{\mathbb{R}^2} f\,d\lambda_2 \quad \text{by the Monotone Convergence Theorem.}
\end{aligned}$$

Therefore, since $\int_{\mathbb{R}^2} f\,d\lambda_2 < \infty$, we see that Φ is Lebesgue integrable. Since Φ being Lebesgue integrable implies that $\Phi(x) < \infty$ for almost every $x \in \mathbb{R}$, we obtain that $\int_{\mathbb{R}} f_x\,d\lambda < \infty$ for almost every $x \in \mathbb{R}$. Hence f_x is Lebesgue integrable for almost every $x \in \mathbb{R}$ as desired.

The proof is then completed by interchanging x and y to obtain the results for f_y and Ψ. ∎

Proof of Tonelli's Theorem (Theorem 5.2.2). To begin, note that Fubini's Theorem holds for all 2-dimensional Lebesgue integrable simple functions. Hence Tonelli's Theorem holds for all 2-dimensional Lebesgue integrable simple functions.

Let f satisfy the assumptions of Tonelli's Theorem. Since f is non-negative, Theorem 2.3.5 implies there exists a sequence $(\varphi_n)_{n\geq 1}$ of simple functions on $(\mathbb{R}^2, \lambda_2)$ such that $\varphi_n \leq \varphi_{n+1}$ for all $n \in \mathbb{N}$ and $(\varphi_n)_{n\geq 1}$ converges to f pointwise.

Notice that if each φ_n is 2-dimensional Lebesgue integrable, then the proof of Fubini's Theorem carries forward verbatim to complete the proof. Hence it suffices to show we can take each φ_n to be 2-dimensional Lebesgue integrable.

For each $n \in \mathbb{N}$ let $\psi_n = \varphi_n \chi_{[-n,n]\times[-n,n]}$. Then $(\psi_n)_{n\geq 1}$ is a sequence of simple functions on $\mathbb{R}$ each of which vanishes off a set of finite 2-dimensional Lebesgue measure and thus is 2-dimensional Lebesgue integrable. Therefore, since by construction we have that $\psi_n \leq \psi_{n+1}$ for all $n \in \mathbb{N}$ and $(\psi_n)_{n\geq 1}$ converges to f pointwise, the proof of Fubini's Theorem carries forward verbatim using the sequence of 2-dimensional Lebesgue integrable simple functions $(\psi_n)_{n\geq 1}$. ∎

Appendix A

Review of the Riemann Integral

In this appendix chapter, we will recall the construction and properties of the Riemann integral presented in MATH 2001. The formal definition of the Riemann integral is modelled on trying to approximate the area under the graph of a function. The idea of approximating this area is to divide up the interval one wants to integrate over into small bits and approximate the area under the graph via rectangles. Thus we must make such constructions formal. Once this is done, we must decide whether or not these approximations are good approximations to the area. If they are, the resulting limit will be the Riemann integral.

A.1 Partitions and Riemann Sums

In order to 'divide up the interval into small bits', we will use the following notion.

Definition A.1.1. A *partition* of a closed interval $[a, b]$ is a finite list of real numbers $\{t_k\}_{k=0}^n$ such that

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

Eventually, we will want to ensure that $|t_k - t_{k-1}|$ is small for all k in order to obtain better and better approximations to the area under a graph. To obtain a lower bound for the area under a graph, we can choose our approximating rectangles to have the largest possible height while remaining completely under the graph. This leads us to the following notion.

Definition A.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 $k \in \{1, \ldots, n\}$,

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

Example A.1.3. If $f : [0,1] \to \mathbb{R}$ is defined by $f(x) = x$ for all $x \in [0,1]$ and if $\mathcal{P} = \{t_k\}_{k=0}^{n}$ is a partition of $[0,1]$, it is easy to see that

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

as f obtains its minimum on $[t_{k-1}, t_k]$ at t_{k-1}.

If it so happens that $t_k = \frac{k}{n}$ for all $k \in \{0, 1, \ldots, n\}$, we see that

$$\begin{aligned} L(f, \mathcal{P}) &= \sum_{k=1}^{n} \frac{k-1}{n}\left(\frac{k}{n} - \frac{k-1}{n}\right) \\ &= \sum_{k=1}^{n} \frac{1}{n^2}(k-1) \\ &= \frac{1}{n^2}\left(\sum_{j=1}^{n-1} j\right) \\ &= \frac{1}{n^2}\frac{n(n-1)}{2} = \frac{1 - \frac{1}{n}}{2} \end{aligned}$$

where the fact that $\sum_{j=1}^{n-1} j = \frac{n(n-1)}{2}$ follows by an induction argument. Clearly, as n tends to infinity, $L(f, \mathcal{P})$ tends to $\frac{1}{2}$ for this particular partitions, which happens to be the area under the graph of f on $[0,1]$.

Although lower Riemann sums accurately estimate the area under the graph of the function in the previous example, perhaps we also need an upper bound for the area under the graph. By choose our approximating rectangles to have the smallest possible height while remaining completely above the graph, we obtain the following notion.

Definition A.1.4. 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 *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 $k \in \{1, \ldots, n\}$,

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

Example A.1.5. If $f : [0,1] \to \mathbb{R}$ is defined by $f(x) = x$ for all $x \in [0,1]$ and if $\mathcal{P} = \{t_k\}_{k=0}^n$ is a partition of $[0,1]$, it is easy to see that

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

as f obtains its maximum on $[t_{k-1}, t_k]$ at t_k.

If it so happens that $t_k = \frac{k}{n}$ for all $k \in \{0, 1, \ldots, n\}$, we see that

$$\begin{aligned} U(f,\mathcal{P}) &= \sum_{k=1}^n \frac{k}{n}\left(\frac{k}{n} - \frac{k-1}{n}\right) \\ &= \sum_{k=1}^n \frac{1}{n^2} k \\ &= \frac{1}{n^2}\left(\sum_{k=1}^n k\right) \\ &= \frac{1}{n^2}\frac{n(n+1)}{2} = \frac{1 + \frac{1}{n}}{2} \end{aligned}$$

where the fact that $\sum_{k=1}^n k = \frac{n(n+1)}{2}$ follows by an induction argument. Clearly, as n tends to infinity, $U(f,\mathcal{P})$ tends to $\frac{1}{2}$ for this particular partitions, which happens to be the area under the graph of f on $[0,1]$.

Although we have been able to approximate the area under the graph of $f(x) = x$ using upper and lower Riemann sums, how do we know whether we can accurate do so for other functions? To analyze this question, we must first decide whether we can compare the upper and lower Riemann sums of a function. Clearly we have that $L(f,\mathcal{P}) \leq U(f,\mathcal{P})$ for any bounded function $f : [a,b] \to \mathbb{R}$ and any partition $\mathcal{P}$ of $[a,b]$. However, if $\mathcal{Q}$ is another partition of $[a,b]$, is it the case that $L(f,\mathcal{Q}) \leq U(f,\mathcal{P})$? Of course our intuition using 'areas under a graph' says this should be so, but how do we prove it?

To answer the above question and provide some 'sequence-like' structure to partitions, we define an ordering on the set of partitions.

Definition A.1.6. 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.

It is not difficult to check that refinement defines a partial ordering (Definition B.1.4) on the set of all partitions of $[a,b]$ (see Example B.1.5). Furthermore, the following says that if $\mathcal{Q}$ is a refinement of $\mathcal{P}$, then we should have better upper and lower bounds for the area under the graph of a function if we use $\mathcal{Q}$ instead of $\mathcal{P}$.

Lemma A.1.7. *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*

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

Proof. Note the inequality $L(f,\mathcal{Q}) \leq U(f,\mathcal{Q})$ is clear. Thus it remains only to show that $L(f,\mathcal{P}) \leq L(f,\mathcal{Q})$ and $U(f,\mathcal{Q}) \leq U(f,\mathcal{P})$. Write $\mathcal{P} = \{t_k\}_{k=0}^n$ where

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

To show the desired inequalities, we will first show that adding a single point to $\mathcal{P}$ does not decrease the lower Riemann sum and does not increase the upper Riemann sum. As there are only a finite number of points one needs to add to $\mathcal{P}$ to obtain $\mathcal{Q}$, the proof will follow.

To implement the above strategy, assume $\mathcal{Q} = \mathcal{P} \cup \{t'\}$ where $t' \in [a,b]$ is such that $t_{q-1} < t' < t_q$ for some $q \in \{1,\ldots,n\}$. For all $k \in \{1,\ldots,n\}$, let

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

Therefore

$$L(f,\mathcal{P}) = \sum_{k=1}^n m_k(t_k - t_{k-1}) \qquad \text{and} \qquad U(f,\mathcal{P}) = \sum_{k=1}^n M_k(t_k - t_{k-1}).$$

Moreover, if we define

$$\begin{aligned} m_q' &= \inf\{f(x) \mid x \in [t_{q-1},t']\}, \\ m_q'' &= \inf\{f(x) \mid x \in [t',t_q]\}, \\ M_q' &= \sup\{f(x) \mid x \in [t_{q-1},t']\}, \text{ and} \\ M_q'' &= \sup\{f(x) \mid x \in [t',t_q]\}, \end{aligned}$$

then we easily see that $m_q \leq m_q', m_q''$, that $M_q', M_q'' \leq M_q$, and that

$$\begin{aligned} L(f,\mathcal{Q}) &= m_q'(t' - t_{q-1}) + m_q''(t_q - t') + \sum_{\substack{k=1 \\ k\neq q}}^n m_k(t_k - t_{k-1}), \quad \text{and} \\ U(f,\mathcal{Q}) &= M_q'(t' - t_{q-1}) + M_q''(t_q - t') + \sum_{\substack{k=1 \\ k\neq q}}^n M_k(t_k - t_{k-1}). \end{aligned}$$

Therefore

$$\begin{aligned} L(f,\mathcal{Q}) - L(f,\mathcal{P}) &= m_q'(t' - t_{q-1}) + m_q''(t_q - t') - m_q(t_q - t_{q-1}) \\ &\geq m_q(t' - t_{q-1}) + m_q(t_q - t') - m_q(t_q - t_{q-1}) = 0 \end{aligned}$$

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

$$\begin{aligned} U(f,\mathcal{Q}) - U(f,\mathcal{P}) &= M_q'(t' - t_{q-1}) + M_q''(t_q - t') - M_q(t_q - t_{q-1}) \\ &\leq M_q(t' - t_{q-1}) + M_q(t_q - t') - M_q(t_q - t_{q-1}) = 0 \end{aligned}$$

so $U(f,\mathcal{Q}) \leq U(f,\mathcal{P})$. Hence the result follows when $\mathcal{Q} = \mathcal{P} \cup \{t'\}$.

To complete the proof, let $\mathcal{Q}$ be an arbitrary refinement of $\mathcal{P}$. Hence we can write $\mathcal{Q} = \mathcal{P} \cup \{t'_k\}_{k=1}^m$ for some $\{t'_k\}_{k=1}^m \subseteq (a,b)$. Thus, by adding a single point at a time, we obtain that

$$L(f,\mathcal{P}) \leq L(f,\mathcal{P} \cup \{t'_1\}) \leq L(f,\mathcal{P} \cup \{t'_1,t'_2\}) \leq \cdots \leq L(f,\mathcal{Q})$$

and

$$U(f,\mathcal{P}) \geq U(f,\mathcal{P} \cup \{t'_1\}) \geq U(f,\mathcal{P} \cup \{t'_1,t'_2\}) \geq \cdots \geq U(f,\mathcal{Q}),$$

which completes the proof. ∎

In order to answer our question of whether $L(f,\mathcal{Q}) \leq U(f,\mathcal{P})$ for all partitions $\mathcal{P}$ and $\mathcal{Q}$, we can use Lemma A.1.7 provided we have a partition that is a refinement of both $\mathcal{P}$ and $\mathcal{Q}$: that is, there is a least upper bound of $\mathcal{P}$ and $\mathcal{Q}$.

Definition A.1.8. 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 $[a,b]$.

Remark A.1.9. Clearly, given two partitions $\mathcal{P}$ and $\mathcal{Q}$, $\mathcal{P} \cup \mathcal{Q}$ is a partition that is a refinement of both $\mathcal{P}$ and $\mathcal{Q}$. Consequently, if $f : [a,b] \to \mathbb{R}$ is bounded, then Lemma A.1.7 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.

A.2 Definition of the Riemann Integral

In order to define the Riemann integral of a bounded function on a closed interval, we desire that the upper and lower Riemann sums both better and better approximate a single number. Using the above observations, we notice that if $f : [a,b] \to \mathbb{R}$ is bounded, then

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

Therefore, in order for there to be no reasonable discrepancy between our approximations, we will like an equality in the above inequality, in which case the value obtained should be the area under the graph. Unfortunately, this is not always the case.

Example A.2.1. 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 $x \in [0, 1]$. Since each open interval always contains at least one element from each of $\mathbb{Q}$ and $\mathbb{R} \setminus \mathbb{Q}$, we easily see that $L(f, \mathcal{P}) = 0$ and $U(f, \mathcal{P}) = 1$ for all partitions $\mathcal{P}$ of $[0, 1]$. Hence

$$\begin{aligned}\sup\{L(f, \mathcal{P}) \mid \mathcal{P} \text{ a partition of } [0,1]\} \\ \neq \inf\{U(f, \mathcal{P}) \mid \mathcal{P} \text{ a partition of } [0,1]\}.\end{aligned}$$

So what should be the area under the graph of this function?

Consequently we will just restrict our attention to the following type of functions.

Definition A.2.2. 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 defined to be

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

Remark A.2.3. Notice that if f is Riemann integrable on $[a, b]$, then

$$L(f, \mathcal{P}) \leq \int_b^a f(x)\,dx \leq U(f, \mathcal{P})$$

for every partition $\mathcal{P}$ of $[a, b]$ by the definition of the Riemann integral.

Clearly the function f in Example A.2.1 is not Riemann integrable. However, which types of function are Riemann integrable and how can we compute the value of the integral? To illustrate the definition, we note the following simple examples (note if the first example did not work out the way it does, we clearly would not have a well-defined notion of area under a graph using Riemann integrals).

Example A.2.4. Let $c \in \mathbb{R}$ and let $f : [a, b] \to \mathbb{R}$ be defined by $f(x) = c$ for all $x \in [a, b]$. If $\mathcal{P} = \{t_k\}_{k=0}^n$ is a partition of $[a, b]$, we see that

$$L(f, \mathcal{P}) = U(f, \mathcal{P}) = \sum_{k=1}^n c(t_k - t_{k-1}) = c\sum_{k=1}^n t_k - t_{k-1} = c(t_n - t_0) = c(b-a).$$

Hence f is Riemann integrable and $\int_a^b f(x)\,dx = c(b - a)$. (Was there any doubt?)

Example A.2.5. Let $f : [0,1] \to \mathbb{R}$ be defined by $f(x) = x$ for all $x \in [0,1]$. For each $n \in \mathbb{N}$, note Example A.1.3 demonstrates the existence of a partition $\mathcal{P}_n$ such that $L(f, \mathcal{P}_n) = \frac{1-\frac{1}{n}}{2}$. Hence

$$\sup\{L(f,\mathcal{P}) \mid \mathcal{P} \text{ a partition of } [a,b]\} \geq \limsup_{n\to\infty} \frac{1-\frac{1}{n}}{2} = \frac{1}{2}.$$

Similarly, for each $n \in \mathbb{N}$, Example A.1.5 demonstrates the existence of a partition $\mathcal{Q}_n$ such that $U(f, \mathcal{Q}_n) = \frac{1+\frac{1}{n}}{2}$. Hence

$$\inf\{U(f,\mathcal{P}) \mid \mathcal{P} \text{ a partition of } [a,b]\} \leq \liminf_{n\to\infty} \frac{1+\frac{1}{n}}{2} = \frac{1}{2}.$$

Therefore, since

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

the above computations show both the inf and sup must be $\frac{1}{2}$. Hence f is Riemann integrable on $[0,1]$ and $\int_0^1 x\,dx = \frac{1}{2}$.

Example A.2.6. Let $f : [0,1] \to \mathbb{R}$ be defined by $f(x) = x^2$ for all $x \in [0,1]$. We claim that f is Riemann integrable on $[0,1]$ and $\int_0^1 x^2\,dx = \frac{1}{3}$. To see this, let $n \in \mathbb{N}$ and let $\mathcal{P}_n = \{t_k\}_{k=1}^n$ be the partition of $[0,1]$ such that $t_k = \frac{k}{n}$ for all $n \in \mathbb{N}$. Then, by an induction argument to compute the value of the sums,

$$\begin{aligned}L(f,\mathcal{P}) &= \sum_{k=1}^n \frac{(k-1)^2}{n^2}\left(\frac{k}{n} - \frac{k-1}{n}\right) \\ &= \sum_{k=1}^n \frac{1}{n^3}(k-1)^2 \\ &= \frac{1}{n^3}\left(\sum_{j=1}^{n-1} j^2\right) \\ &= \frac{1}{n^3}\frac{(n-1)(n)(2(n-1)+1)}{6} = \frac{2n^3-3n^2+n}{6n^3}\end{aligned}$$

and

$$\begin{aligned}U(f,\mathcal{P}) &= \sum_{k=1}^n \frac{k^2}{n^2}\left(\frac{k}{n} - \frac{k-1}{n}\right) \\ &= \sum_{k=1}^n \frac{1}{n^3}k^2 \\ &= \frac{1}{n^3}\left(\sum_{k=1}^{n} k^2\right) \\ &= \frac{1}{n^3}\frac{n(n+1)(2n+1)}{2} = \frac{2n^3+3n^2+n}{6n^3}.\end{aligned}$$

Hence, since $\lim_{n\to\infty} \frac{2n^3-3n^2+1}{6n^3} = \lim_{n\to\infty} \frac{2n^3+3n^2+1}{6n^3} = \frac{1}{3}$, we see that

$$\begin{aligned}\frac{1}{3} &\leq \sup\{L(f,\mathcal{P}) \mid \mathcal{P} \text{ a partition of } [a,b]\} \\ &\leq \inf\{U(f,\mathcal{P}) \mid \mathcal{P} \text{ a partition of } [a,b]\} \leq \frac{1}{3}.\end{aligned}$$

Hence the inequalities must be equalities so f is Riemann integrable on $[0,1]$ by definition with $\int_0^1 x^2\,dx = \frac{1}{3}$

Note in the previous two examples, the functions were demonstrated to be Riemann integrable on $[0,1]$ via partitions $\mathcal{P}$ such that $L(f,\mathcal{P})$ and $U(f,\mathcal{P})$ were as closes as one would like. Coincidence, I think not!

Theorem A.2.7. *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*

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

Proof. Note we must have that $0 \leq U(f,\mathcal{P}) - L(f,\mathcal{P})$ for any partition $\mathcal{P}$ by earlier discussions.

First assume that f is Riemann integrable. Hence, with $I = \int_a^b f(x)\,dx$, we have by the definition of the integral that

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

Let $\epsilon > 0$ be arbitrary. By the definition of the supremum, there exists a partition $\mathcal{P}_1$ of $[a,b]$ such that

$$I - \frac{\epsilon}{2} < L(f,\mathcal{P}_1).$$

Similarly, by the definition of the infimum, there exists a partition $\mathcal{P}_2$ of $[a,b]$ such that

$$U(f,\mathcal{P}_2) < I + \frac{\epsilon}{2}.$$

Let $\mathcal{P} = \mathcal{P}_1 \cup \mathcal{P}_2$ which is a partition of $[a,b]$. Since $\mathcal{P}$ is a refinement of both $\mathcal{P}_1$ and $\mathcal{P}_2$, we obtain that

$$L(f,\mathcal{P}_1) \leq L(f,\mathcal{P}) \leq U(f,\mathcal{P}) \leq U(f,\mathcal{P}_2)$$

by Lemma A.1.7. Hence

$$\begin{aligned}U(f,\mathcal{P}) - L(f,\mathcal{P}) &\leq U(f,\mathcal{P}_2) - L(f,\mathcal{P}_1) \\ &= (U(f,\mathcal{P}_2) - I) + (I - L(f,\mathcal{P}_1)) \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.\end{aligned}$$

Therefore, since $\epsilon > 0$ was arbitrary, this direction of the proof is complete.

For the other direction, assume for every $\epsilon > 0$ there exists a partition $\mathcal{P}$ of $[a, b]$ such that

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

In particular, for each $n \in \mathbb{N}$ there exists a partition $\mathcal{P}_n$ of $[a, b]$ such that

$$0 \leq U(f, \mathcal{P}_n) - L(f, \mathcal{P}_n) < \frac{1}{n}.$$

Let

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

Then $L, U \in \mathbb{R}$ are such that $L \leq U$. Moreover, for each $n \in \mathbb{N}$

$$0 \leq U - L \leq U(f, \mathcal{P}_n) - L(f, \mathcal{P}_n) < \frac{1}{n}.$$

Therefore it follows that $U = L$. Hence f is Riemann integrable on $[a, b]$ by definition. ∎

Remark A.2.8. Using Theorem A.2.7, there is an easier method for approximating the Riemann integral of a Riemann integrable function. Indeed suppose $\mathcal{P} = \{t_k\}_{k=0}^n$ is a partition of $[a, b]$ with

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

and let $f : [a, b] \to \mathbb{R}$ be bounded. For each k, let $x_k \in [t_{k-1}, t_k]$ and let

$$R(f, \mathcal{P}, \{x_k\}_{k=1}^n) = \sum_{k=1}^n f(x_k)(t_k - t_{k-1}).$$

The sum $R(f, \mathcal{P}, \{x_k\}_{k=1}^n)$ is called a *Riemann sum*.

Clearly

$$L(f, \mathcal{P}) \leq R(f, \mathcal{P}, \{x_k\}_{k=1}^n) \leq U(f, \mathcal{P})$$

by definitions. Hence, if f is Riemann integrable, we obtain via Theorem A.2.7 that for any $\epsilon > 0$ there exists a partition $\mathcal{P}'$ of $[a, b]$ such that

$$L(f, \mathcal{P}') \leq \int_b^a f(x)\, dx \leq U(f, \mathcal{P}') \leq L(f, \mathcal{P})' + \epsilon$$

and thus

$$\left| \int_a^b f(x)\, dx - R(f, \mathcal{P}', \{x_k\}_{k=1}^n) \right| < \epsilon$$

for any choice of $\{x_k\}_{k=1}^n$. Consequently, if one knows that f is Riemann integrable, one may approximate $\int_a^b f(x)\, dx$ using Riemann sums oppose to lower/upper Riemann sums. This is occasionally useful as convenient choices of $\{x_n\}_{k=1}^n$ may make computing the sum much easier.

Of course, our next question is, “Which types of functions are Riemann integrable?”

A.3 Some Integrable Functions

If the theory of Riemann integration will be of use to us, we must have a wide variety of functions that are Riemann integrable. It is easy to show some functions are Riemann integrable.

Proposition A.3.1. *If $f : [a, b] \to \mathbb{R}$ is monotonic and bounded, then f is Riemann integrable on $[a, b]$.*

Proof. Assume $f : [a, b] \to \mathbb{R}$ is monotone and bounded. In addition, we will assume that f is non-decreasing as the proof when f is non-increasing is similar.

Let $\epsilon > 0$. Since

$$\lim_{n\to\infty} \frac{1}{n}(b-a)(f(b)-f(a)) = 0,$$

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

$$0 \leq \frac{1}{N}(b-a)(f(b)-f(a)) < \epsilon.$$

Let $\mathcal{P}_N = \{t_k\}_{k=0}^N$ be the partition such that

$$t_k = a + \frac{k}{N}(b-a)$$

for all $k \in \{0, \ldots, N\}$. Notice $t_k - t_{k-1} = \frac{1}{n}(b-a)$ for all k (and thus we call $\mathcal{P}_N$ the *uniform partition* of $[a, b]$ into N intervals). Since f is non-decreasing, if for all $k \in \{1, \ldots, N\}$

$$m_k = \inf\{f(x) \mid x \in [t_{k-1}, t_k]\} \quad \text{and} \quad M_k = \sup\{f(x) \mid x \in [t_{k-1}, t_k]\},$$

then

$$m_k = f(t_{k-1}) \qquad \text{and} \qquad M_k = f(t_k).$$

Hence

$$\begin{aligned}
0 &\leq U(f, \mathcal{P}_n) - L(f, \mathcal{P}_n) \\
&= \sum_{k=1}^N M_k(t_k - t_{k-1}) - \sum_{k=1}^N m_k(t_k - t_{k-1}) \\
&= \sum_{k=1}^N f(t_k)\frac{1}{N}(b-a) - \sum_{k=1}^N f(t_{k-1})\frac{1}{N}(b-a) \\
&= f(t_N)\frac{1}{N}(b-a) - f(t_0)\frac{1}{N}(b-a) \\
&= \frac{1}{N}(b-a)(f(b)-f(a)) < \epsilon.
\end{aligned}$$

Therefore, since $\epsilon > 0$ was arbitrary, Theorem A.2.7 implies that f is Riemann integrable on $[a, b]$. ∎

Of course, if continuous functions were not Riemann integrable, Riemann integration would be worthless to us. The fact that continuous functions on closed intervals are uniformly continuous is vital int he following proof.

Theorem A.3.2. *If $f : [a,b] \to \mathbb{R}$ is continuous, then f is Riemann integrable on $[a,b]$.*

Proof. Assume $f : [a,b] \to \mathbb{R}$ is continuous. Therefore f is bounded by the Extreme Value Theorem. Hence it makes sense to discuss whether f is Riemann integrable.

In order to invoke Theorem A.2.7 to show that f is Riemann integrable, let $\epsilon > 0$ be arbitrary. Since $f : [a,b] \to \mathbb{R}$ is continuous, f is uniformly continuous on $[a,b]$. Hence there exists a $\delta > 0$ such that if $x, y \in [a,b]$ and $|x - y| < \delta$ then $|f(x) - f(y)| < \frac{\epsilon}{b-a}$.

Choose $n \in \mathbb{N}$ such that $\frac{1}{n} < \delta$. Let $\mathcal{P}$ be the uniform partition of $[a,b]$ into n intervals; that is, let $\mathcal{P} = \{t_k\}_{k=0}^n$ be the partition such that

$$t_k = a + \frac{k}{n}(b-a)$$

for all $k \in \{0, \ldots, n\}$. For all $k \in \{0, \ldots, n\}$, let

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

Since $|t_k - t_{k-1}| = \frac{1}{n} < \delta$ so $|x - y| < \delta$ for all $x, y \in [t_{k-1}, t_k]$, it must be the case that $M_k - m_k = |M_k - m_k| \leq \frac{\epsilon}{b-a}$ for all $k \in \{1, \ldots, n\}$. Hence

$$\begin{aligned} 0 \leq U(f, \mathcal{P}) - L(f, \mathcal{P}) &= \sum_{k=1}^n (M_k - m_k)(t_k - t_{k-1}) \\ &\leq \sum_{k=1}^n \frac{\epsilon}{b-a}(t_k - t_{k-1}) \\ &= \frac{\epsilon}{b-a} \sum_{k=1}^n t_k - t_{k-1} = \frac{\epsilon}{b-a}(b-a) = \epsilon. \end{aligned}$$

Thus, since $\epsilon > 0$ was arbitrary, f is Riemann integrable on $[a,b]$ by Theorem A.2.7. ∎

Of course, not all functions we desire to integrate are continuous. However, many functions one sees and deals with in real-world applications are continuous at almost every point. In particular, the following shows that if our functions are piecewise continuous, then they are Riemann integrable.

Corollary A.3.3. *If $f : [a,b] \to \mathbb{R}$ is continuous on $[a,b]$ except at a finite number of points and f is bounded on $[a,b]$, then f is Riemann integrable on $[a,b]$.*

Proof. Assume $f : [a, b] \to \mathbb{R}$ is continuous except at a finite number of points and $f([a, b])$ is bounded. Let $\{a_k\}_{k=0}^{q}$ contain all of the points for which f is not continuous at and be such that

$$a = a_0 < a_1 < a_2 < \cdots < a_q = b.$$

The idea of the proof is to construct a partition such that each interval of the partition contains at most one a_k, and if an interval of the partition contains an a_k, then its length is really small.

Let $\epsilon > 0$ be arbitrary. Since $f([a, b])$ is bounded, there exists a $K > 0$ such that $|f(x)| \leq K$ for all $x \in [a, b]$. Therefore, if

$$L = \sup\{f(x) - f(y) \mid x, y \in [a, b]\},$$

then $0 \leq L \leq 2K < \infty$.

Let

$$\delta = \frac{\epsilon}{2(q+1)(L+1)} > 0.$$

By taking a and b together with endpoints of intervals centred at each a_k of radius less than $\frac{\delta}{2}$, there exists a partition $\mathcal{P}' = \{t_k\}_{k=0}^{2q+1}$ with

$$a = t_0 < t_1 < t_2 < \cdots < t_{2q+1} = b$$

such that $t_{2k+1} - t_{2k} < \delta$ for all $k \in \{0, \ldots, q\}$ and $t_{2k} < a_k < t_{2k+1}$ for all $k \in \{1, \ldots, q-1\}$. For all $k \in \{1, \ldots, 2q+1\}$, let

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

Thus $M_k - m_k \leq L$ for all $k \in \{1, \ldots, 2q+1\}$.

Since f is continuous on $[t_{2k-1}, t_{2k}]$ for all $k \in \{1, \ldots, q\}$, f is Riemann integrable on $[t_{2k-1}, t_{2k}]$ by Theorem A.3.2. Hence, by the definition of Riemann integration, there exist partitions $\mathcal{P}_k$ of $[t_{2k-1}, t_{2k}]$ such that

$$0 \leq U(f, \mathcal{P}_k) - L(f, \mathcal{P}_k) < \frac{\epsilon}{2q}.$$

Let $\mathcal{P} = \mathcal{P}' \cup \left(\bigcup_{k=1}^{q} \mathcal{P}_k\right)$. Then $\mathcal{P}$ is a partition of $[a, b]$ such that

$$\begin{aligned} 0 &\leq U(f, \mathcal{P}) - L(f, \mathcal{P}) \\ &= \sum_{k=1}^{q} (U(f, \mathcal{P}_k) - L(f, \mathcal{P}_k)) + \sum_{k=0}^{q} (M_{2k+1} - m_{2k+1})(t_{2k+1} - t_{2k}). \end{aligned}$$

(that is, on each $[t_{2k-1}, t_{2k}]$ the partition behaves like $\mathcal{P}_k$ and thus so do the sums, and the parts of the partition remaining are of the form $[t_{2k}, t_{2k+1}]$

each of which contains at most one a_j). Hence

$$\begin{aligned}
0 &\leq U(f,\mathcal{P}) - L(f,\mathcal{P}) \\
&\leq \sum_{k=1}^{q} \frac{\epsilon}{2q} + \sum_{k=0}^{q} L\delta \\
&\leq \frac{\epsilon}{2} + (q+1)L\delta \\
&\leq \frac{\epsilon}{2} + (q+1)L\frac{\epsilon}{2(q+1)(L+1)} \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.
\end{aligned}$$

Thus, as $\epsilon > 0$ was arbitrary, f is Riemann integrable on $[a,b]$ by Theorem A.2.7. ∎

Using the similar ideas to those used to prove Corollary A.3.3, it is possible to show that some truly bizarre functions are Riemann integrable.

Example A.3.4. Let $f : [0,1] \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 0 & \text{if } x \text{ is irrational} \\ 1 & \text{if } x = 0 \\ \frac{1}{b} & \text{if } x = \frac{a}{b} \text{ where } a \in \mathbb{Z} \setminus \{0\},\, b \in \mathbb{N}, \text{ and } \gcd(a,b) = 1 \end{cases}.$$

Clearly f is bounded.

We claim that f is Riemann integrable on $[0,1]$. To see this, let $\epsilon > 0$ be arbitrary. Choose $N \in \mathbb{N}$ such that $\frac{1}{N} < \frac{\epsilon}{2}$.

By the definition of f, let $\{a_k\}_{k=0}^{q}$ be the finite set of $x \in [0,1]$ such that $f(x) \leq \frac{1}{N}$ and

$$0 = a_0 < a_1 < a_2 < \cdots < a_q = 1.$$

Let

$$\delta = \frac{\epsilon}{2(q+1)} > 0.$$

By taking 0 and 1 together with endpoints of intervals centred at each a_k of radius less than $\frac{\delta}{2}$, there exists a partition $\mathcal{P} = \{t_k\}_{k=0}^{2q+1}$ with

$$0 = t_0 < t_1 < t_2 < \cdots < t_{2q+1} = 1$$

such that $t_{2k+1} - t_{2k} < \delta$ for all $k \in \{0,\ldots,q\}$ and $t_{2k} < a_k < t_{2k+1}$ for all $k \in \{1,\ldots,q-1\}$.

For all $k \in \{1,\ldots,2q+1\}$, let

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

Since $0 \leq f(x) \leq 1$ for all $x \in [0,1]$, we see that $M_k - m_k \leq 1$ for all $k \in \{1,\ldots,2q+1\}$. Moreover, since $t_{2k} < a_k < t_{2k+1}$ for all $k \in \{1,\ldots,q-1\}$, we have that

$$M_{2k} - m_{2k} \leq \frac{1}{N} - 0 < \frac{\epsilon}{2}$$

for all $k \in \{1, \ldots, q\}$. Therefore

$$\begin{aligned}
0 &\leq U(f,\mathcal{P}) - L(f,\mathcal{P}) \\
&= \sum_{k=1}^{q} (M_{2k} - m_{2k})(t_{2k} - t_{2k-1}) + \sum_{k=0}^{q} (M_{2k+1} - m_{2k+1})(t_{2k+1} - t_{2k}) \\
&\leq \sum_{k=1}^{q} \frac{\epsilon}{2}(t_{2k} - t_{2k-1}) + \sum_{k=0}^{q} 1\delta \\
&\leq \frac{\epsilon}{2}\left(\sum_{k=1}^{q}(t_{2k} - t_{2k-1})\right) + (q+1)\delta \\
&\leq \frac{\epsilon}{2}(1-0) + (q+1)\delta \\
&\leq \frac{\epsilon}{2} + (q+1)\frac{\epsilon}{2(q+1)} \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.
\end{aligned}$$

Thus, as $\epsilon > 0$ was arbitrary, f is Riemann integrable on $[0,1]$ by Theorem A.2.7.

A.4 Properties of the Riemann Integral

Now that we know several functions are Riemann integrable, we desire to derive the basic properties of the Riemann integral just as we did for limits of sequences and functions. We begin with the following that enables us to divide up a closed interval into a finite number of closed subintervals when considering Riemann integration.

Proposition A.4.1. *Let $f : [a,b] \to \mathbb{R}$ be bounded and let $c \in (a,b)$. Then f is Riemann integrable on $[a,b]$ if and only if f is Riemann integrable on $[a,c]$ and $[c,b]$. Moreover, when f is Riemann integrable on $[a,b]$, we have that*

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

Proof. To begin, assume that f is Riemann integrable on $[a,b]$. To see that f is Riemann integrable on $[a,c]$ and $[c,b]$, let $\epsilon > 0$ be arbitrary. Since f is Riemann integrable on $[a,b]$, Theorem A.2.7 implies that there exists a partition $\mathcal{P}$ of $[a,b]$ such that

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

Therefore, if $\mathcal{P}_0 = \mathcal{P} \cup \{c\}$, then $\mathcal{P}_0$ is a partition of $[a,b]$ containing c that is a refinement of $\mathcal{P}$. Therefore, by Remark A.2.3 and Lemma A.1.7

$$\begin{aligned}
L(f,\mathcal{P}_0) &\leq U(f,\mathcal{P}_0) \\
&\leq U(f,\mathcal{P}) \\
&\leq L(f,\mathcal{P}) + \epsilon \\
&\leq L(f,\mathcal{P}_0) + \epsilon.
\end{aligned}$$

Let

$$\mathcal{P}_1 = \mathcal{P}_0 \cap [a,c] \qquad \text{and} \qquad \mathcal{P}_2 = \mathcal{P}_0 \cap [c,b].$$

Then $\mathcal{P}_1$ is a partition of $[a,c]$ and $\mathcal{P}_2$ is a partition of $[c,b]$. Furthermore, due to the nature of these partitions and the definitions of the upper and lower Riemann sums, we easily see that

$$L(f,\mathcal{P}_0) = L(f,\mathcal{P}_1) + L(f,\mathcal{P}_2) \qquad \text{and} \qquad U(f,\mathcal{P}_0) = U(f,\mathcal{P}_1) + U(f,\mathcal{P}_2).$$

Hence

$$0 \leq (U(f,\mathcal{P}_1) - L(f,\mathcal{P}_1)) + (U(f,\mathcal{P}_2) - L(f,\mathcal{P}_2)) = U(f,\mathcal{P}_0) - L(f,\mathcal{P}_0) \leq \epsilon.$$

Therefore, since $0 \leq U(f,\mathcal{P}_1) - L(f,\mathcal{P}_1)$ and $0 \leq U(f,\mathcal{P}_2) - L(f,\mathcal{P}_2)$, it must be the case that

$$0 \leq U(f,\mathcal{P}_1) - L(f,\mathcal{P}_1) \leq \epsilon \qquad \text{and} \qquad 0 \leq U(f,\mathcal{P}_2) - L(f,\mathcal{P}_2) \leq \epsilon.$$

Hence f is integrable on both $[a,c]$ and $[c,b]$ by Theorem A.2.7.

To prove the converse and demonstrate the desired integral equation, assume that f is Riemann integrable on $[a,c]$ and $[c,b]$. To see that f is Riemann integrable on $[a,b]$, let $\epsilon > 0$ be arbitrary. Since f is Riemann integrable on $[a,c]$ and $[c,b]$, Remark A.2.3 together with Theorem A.2.7 imply that there exists partitions $\mathcal{P}_1$ and $\mathcal{P}_2$ of $[a,c]$ and $[c,b]$ respectively such that

$$\begin{aligned} L(f,\mathcal{P}_1) &\leq \int_a^c f(x)\,dx \leq U(f,\mathcal{P}_1) \leq L(f,\mathcal{P}_1) + \frac{\epsilon}{2} \text{ and} \\ L(f,\mathcal{P}_2) &\leq \int_c^b f(x)\,dx \leq U(f,\mathcal{P}_2) \leq L(f,\mathcal{P}_2) + \frac{\epsilon}{2}. \end{aligned}$$

Let $\mathcal{P} = \mathcal{P}_1 \cup \mathcal{P}_2$. It is elementary to see that $\mathcal{P}$ is a partition of $[a,b]$. Moreover, due to the nature of these partitions and the definitions of the upper and lower Riemann sums, we easily see that

$$L(f,\mathcal{P}) = L(f,\mathcal{P}_1) + L(f,\mathcal{P}_2) \qquad \text{and} \qquad U(f,\mathcal{P}) = U(f,\mathcal{P}_1) + U(f,\mathcal{P}_2).$$

Hence

$$\begin{aligned} 0 &\leq U(f,\mathcal{P}) - L(f,\mathcal{P}) \\ &= (U(f,\mathcal{P}_1) + U(f,\mathcal{P}_2)) + (L(f,\mathcal{P}_1) + L(f,\mathcal{P}_2)) \\ &= (U(f,\mathcal{P}_1) - L(f,\mathcal{P}_1)) + (U(f,\mathcal{P}_2) - L(f,\mathcal{P}_2)) \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \end{aligned}$$

Therefore, since $\epsilon > 0$ was arbitrary, f is Riemann integrable on $[a,b]$ by Theorem A.2.7. Moreover, we have for all $\epsilon > 0$ that

$$\begin{aligned}
\int_a^c f(x)\,dx + \int_c^b f(x)\,dx - \epsilon &\leq L(f,\mathcal{P}_1) + L(f,\mathcal{P}_2) \\
&= L(f,\mathcal{P}) \\
&\leq \int_a^b f(x)\,dx \\
&\leq U(f,\mathcal{P}) \\
&= U(f,\mathcal{P}_1) + U(f,\mathcal{P}_2) \\
&\leq \int_a^c f(x)\,dx + \int_c^b f(x)\,dx + \epsilon.
\end{aligned}$$

Hence

$$\left| \int_a^c f(x)\,dx + \int_c^b f(x)\,dx - \int_a^b f(x)\,dx \right| < \epsilon.$$

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

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

as desired. ∎

Of course, integrals behave well with respect to many of the same arithmetic properties that limits satisfy as the following result shows. Unfortunately, notice that multiplication is absent from this result.

Proposition A.4.2. *Let $f, g : [a,b] \to \mathbb{R}$ be Riemann integrable functions on $[a,b]$. The following are true:*

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

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

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

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

Proof. a) Assume $f : [a, b] \to \mathbb{R}$ is a Riemann integrable function and $\alpha \in \mathbb{R}$. To see that αf is Riemann integrable, consider an arbitrary partition $\mathcal{P}$ of $[a, b]$.

Notice if $\alpha \geq 0$ then $\sup(\alpha A) = \alpha \sup(A)$ and $\inf(\alpha A) = \alpha \inf(A)$ for all subsets $A \subseteq \mathbb{R}$. Therefore, if $\alpha > 0$, we have that

$$L(\alpha f, \mathcal{P}) = \alpha L(f, \mathcal{P}) \qquad \text{and} \qquad U(\alpha f, \mathcal{P}) = \alpha U(f, \mathcal{P})$$

Furthermore, since if A is a bounded subset of $\mathbb{R}$ then $\inf(-A) = -\sup(A)$, it follows that if $\alpha < 0$ then

$$L(\alpha f, \mathcal{P}) = \alpha U(f, \mathcal{P}) \qquad \text{and} \qquad U(\alpha f, \mathcal{P}) = \alpha L(f, \mathcal{P})$$

Since f is Riemann integrable on $[a, b]$, we obtain by the definition of the Riemann integral that

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

Therefore, the previous above computations we obtain that

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

Hence αf is Riemann integrable on $[a, b]$ with

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

b) Let $f, g : [a, b] \to \mathbb{R}$ be Riemann integrable. To begin the proof, consider an arbitrary partition $\mathcal{P}$ of $[a, b]$. Since

$$\sup\{f(x)+g(x) \mid x \in [c, d]\} \leq \sup\{f(x) \mid x \in [c, d]\}+\sup\{g(x) \mid x \in [c, d]\}$$

and

$$\inf\{f(x) + g(x) \mid x \in [c, d]\} \geq \inf\{f(x) \mid x \in [c, d]\} + \inf\{g(x) \mid x \in [c, d]\}$$

for all $c, d \in [a, b]$ with $c < d$, we obtain that

$$L(f, \mathcal{P}) + L(g, \mathcal{P}) \leq L(f + g, \mathcal{P}) \leq U(f + g, \mathcal{P}) \leq U(f, \mathcal{P}) + U(g, \mathcal{P})$$

by the definition of the Riemann sums.

To prove that $f + g$ is Riemann integrable and obtain the desired integral equation, let $\epsilon > 0$ be arbitrary. Since f is Riemann integrable on $[a, b]$, Remark A.2.3 together with Theorem A.2.7 imply that there exists a partition $\mathcal{P}_1$ of $[a, b]$ such that

$$L(f, \mathcal{P}_1) \leq \int_a^b f(x)\,dx \leq U(f, \mathcal{P}_1) \leq L(f, \mathcal{P}_1) + \frac{\epsilon}{2}.$$

Similarly, since g is Riemann integrable on $[a, b]$, Remark A.2.3 together with Theorem A.2.7 imply that there exists a partition $\mathcal{P}_2$ of $[a, b]$ such that

$$L(g, \mathcal{P}_2) \leq \int_a^b g(x)\,dx \leq U(g, \mathcal{P}_2) \leq L(g, \mathcal{P}_2) + \frac{\epsilon}{2}.$$

Let $\mathcal{P} = \mathcal{P}_1 \cup \mathcal{P}_2$. Then $\mathcal{P}$ is a partition of $[a, b]$ that is a refinement of both $\mathcal{P}_1$ and $\mathcal{P}_2$. Therefore, Remark A.2.3 together with Lemma A.1.7 imply that

$$\begin{aligned} L(f, \mathcal{P}) \leq \int_a^b f(x)\,dx &\leq U(f, \mathcal{P}) \\ &\leq U(f, \mathcal{P}_1) \\ &\leq L(f, \mathcal{P}_1) \\ &\leq L(f, \mathcal{P}) + \frac{\epsilon}{2} \end{aligned}$$

and similarly

$$L(g, \mathcal{P}) \leq \int_a^b g(x)\,dx \leq U(g, \mathcal{P}) \leq L(g, \mathcal{P}) + \frac{\epsilon}{2}.$$

Hence, since we know that

$$L(f, \mathcal{P}) + L(g, \mathcal{P}) \leq L(f + g, \mathcal{P}) \leq U(f + g, \mathcal{P}) \leq U(f, \mathcal{P}) + U(g, \mathcal{P})$$

we obtain that

$$L(f, \mathcal{P}) + L(g, \mathcal{P}) \leq L(f + g, \mathcal{P}) \leq U(f + g, \mathcal{P}) \leq L(f, \mathcal{P}) + L(g, \mathcal{P}) + \epsilon.$$

Hence $0 \leq U(f + g, \mathcal{P}) - L(f + g, \mathcal{P}) < \epsilon$. Therefore, since ϵ was arbitrary, Theorem A.2.7 implies that $f + g$ is Riemann integrable on $[a, b]$. Moreover, by repeating the above now knowing that $f + g$ is Riemann integrable on $[a, b]$, we obtain that for all $\epsilon > 0$ there exists a partition $\mathcal{P}$ such that

$$\begin{aligned} \int_a^b f(x)\,dx + \int_a^b g(x)\,dx - \epsilon &\leq L(f, \mathcal{P}) + L(g, \mathcal{P}) \\ &\leq L(f + g, \mathcal{P}) \\ &\int_a^b (f + g)(x)\,dx \\ &\leq U(f + g, \mathcal{P}) \\ &\leq U(f, \mathcal{P}) + U(g, \mathcal{P}) \\ &\leq \int_a^b f(x)\,dx + \int_a^b g(x)\,dx + \epsilon. \end{aligned}$$

Hence
$$\left|\int_a^b f(x)\,dx + \int_a^b g(x)\,dx - \int_a^b (f+g)(x)\,dx\right| \leq \epsilon.$$
Therefore, as $\epsilon > 0$ was arbitrary, we obtain that
$$\int_a^b (f+g)(x)\,dx = \int_a^b f(x)\,dx + \int_a^b g(x)\,dx$$
as desired.

c) Let $f, g : [a,b] \to \mathbb{R}$ be Riemann integrable and assume $f(x) \leq g(x)$ for all $x \in [a,b]$. To see the desired result, let $\epsilon > 0$ be arbitrary. Remark A.2.3 together with Theorem A.2.7 imply that there exists a partition $\mathcal{P}$ of $[a,b]$ such that
$$L(f,\mathcal{P}) \leq \int_a^b f(x)\,dx \leq U(f,\mathcal{P}) \leq L(f,\mathcal{P}) + \epsilon.$$
However, since $f(x) \leq g(x)$ for all $x \in [a,b]$, we know that
$$\inf\{f(x) \mid x \in [c,d]\} \leq \inf\{g(x) \mid x \in [c,d]\}$$
for all $c, d \in [a,b]$ with $c < d$. Therefore $L(f,\mathcal{P}) \leq L(g,\mathcal{P})$. Hence
$$\int_a^b f(x)\,dx - \epsilon \leq L(f,\mathcal{P}) \leq L(g,\mathcal{P}) \leq \int_a^b g(x)\,dx.$$
Hence, for all $\epsilon > 0$, we have that
$$\int_a^b f(x)\,dx \leq \int_a^b g(x)\,dx + \epsilon.$$
Therefore, we have ("by sending ϵ to 0") that
$$\int_a^b f(x)\,dx \leq \int_a^b g(x)\,dx$$
as desired.

d) By part c) and Example A.2.4, we have that
$$m(b-a) = \int_a^b m\,dx \leq \int_a^b f(x)\,dx \leq \int_a^b M\,dx = M(b-a)$$
as desired. ∎

Remark A.4.3. Note that Proposition A.4.2 does not produce a formula for the Riemann integral of the product of Riemann integrable functions. Indeed it is almost always the case that $\int_a^b (fg)(x)\,dx \neq \left(\int_a^b f(x)\,dx\right)\left(\int_a^b g(x)\,dx\right)$. For example, using Examples A.2.5 and A.2.6, we see that
$$\int_0^1 x^2\,dx = \frac{1}{3} \quad \text{whereas} \quad \left(\int_0^1 x\,dx\right)^2 = \frac{1}{4}.$$

In lieu of the above remark, it is still possible to show that if f and g are Riemann integrable on $[a, b]$, then fg is Riemann integrable on $[a, b]$. To begin this proof, we first must deal with the case that $f = g$.

Lemma A.4.4. *Let $f : [a, b] \to \mathbb{R}$ be a Riemann integrable function on $[a, b]$. The function $f^2 : [a, b] \to \mathbb{R}$ defined by $f^2(x) = (f(x))^2$ for all $x \in [a, b]$ is Riemann integrable on $[a, b]$.*

Proof. Since f is bounded by the definition of Riemann integrable,

$$K = \sup\{|f(x)| \mid x \in [a, b]\} < \infty.$$

To see that f^2 is Riemann integrable, let $\epsilon > 0$ be arbitrary. Since f is Riemann integrable on $[a, b]$, Theorem A.2.7 implies that there exists a partition $\mathcal{P}$ of $[a, b]$ such that

$$0 \leq U(f, \mathcal{P}) - L(f, \mathcal{P}) < \frac{1}{2(K+1)}\epsilon.$$

Write $\mathcal{P} = \{t_k\}_{k=0}^n$ where

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

For each $k \in \{1, \ldots, n\}$ let

$$\begin{aligned} m_k(f) &= \inf\{f(x) \mid x \in [t_{k-1}, t_k]\}, \\ M_k(f) &= \sup\{f(x) \mid x \in [t_{k-1}, t_k]\}, \\ m_k(f^2) &= \inf\{(f(x))^2 \mid x \in [t_{k-1}, t_k]\}, \text{ and} \\ M_k(f^2) &= \sup\{(f(x))^2 \mid x \in [t_{k-1}, t_k]\}. \end{aligned}$$

Notice for all $x, y \in [a, b]$ we have that

$$\begin{aligned} |(f(x))^2 - (f(y))^2| &= |f(x) + f(y)||f(x) - f(y)| \\ &\leq (|f(x)| + |f(y)|)|f(x) - f(y)| \\ &\leq (K + K)|f(x) - f(y)| = 2K|f(x) - f(y)|. \end{aligned}$$

Hence we obtain that

$$M_k(f^2) - m_k(f^2) \leq 2K(M_k(f) - m_k(f))$$

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

$$0 \leq U(f^2, \mathcal{P}) - L(f^2, \mathcal{P}) \leq 2K(U(f, \mathcal{P}) - L(f, \mathcal{P})) \leq 2K\frac{1}{2(K+1)}\epsilon < \epsilon.$$

Hence f^2 is Riemann integrable by Proposition A.4.6. ∎

Using the above and a clever decomposition of functions, we obtain the product of Riemann integrable functions is Riemann integrable.

Proposition A.4.5. *Let $f, g : [a, b] \to \mathbb{R}$ be Riemann integrable functions on $[a, b]$. Then $fg : [a, b] \to \mathbb{R}$ is Riemann integrable on $[a, b]$.*

Proof. Since

$$f(x)g(x) = \frac{1}{2}\left((f(x) + g(x))^2 - f(x)^2 - g(x)^2\right)$$

and since $f + g, f^2, g^2$, and $(f + g)^2$ are Riemann integrable by Proposition A.4.2 and Lemma A.4.4, it follows by Proposition A.4.2 that fg is Riemann integrable. ∎

To complete our section on the properties of the Riemann integral, we have one more useful result. The main reason why this result is useful in analysis is that it plays the same role for integrals as the triangle inequality plays for sums.

Proposition A.4.6. *Let $f : [a, b] \to \mathbb{R}$ a Riemann integrable function on $[a, b]$. Then the function $|f| : [a, b] \to \mathbb{R}$ defined by $|f|(x) = |f(x)|$ for all $x \in [a, b]$ is Riemann integrable on $[a, b]$ and*

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

Proof. Let $\epsilon > 0$ be arbitrary. By Theorem A.2.7, there exists a partition $\mathcal{P}$ of $[a, b]$ such that

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

Write $\mathcal{P} = \{t_k\}_{k=0}^n$ where

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

For each $k \in \{1, \ldots, n\}$ let

$$\begin{aligned} m_k(f) &= \inf\{f(x) \mid x \in [t_{k-1}, t_k]\}, \\ M_k(f) &= \sup\{f(x) \mid x \in [t_{k-1}, t_k]\}, \\ m_k(|f|) &= \inf\{|f(x)| \mid x \in [t_{k-1}, t_k]\}, \text{ and} \\ M_k(|f|) &= \sup\{|f(x)| \mid x \in [t_{k-1}, t_k]\}. \end{aligned}$$

We claim that

$$M_k(|f|) - m_k(|f|) \leq M_k(f) - m_k(f)$$

for all $k \in \{1, \ldots, n\}$. Indeed notice if $x, y \in [t_{k-1}, t_k]$ are such that:

- $f(x), f(y) \geq 0$, then

$$|f(x)| - |f(y)| = f(x) - f(y) \leq M_k(f) - m_k(f).$$

- $f(x) \geq 0 \geq f(y)$, then
$$|f(x)| - |f(y)| \leq f(x) - f(y) \leq M_k(f) - m_k(f).$$
- $f(y) \geq 0 \geq f(x)$, then
$$|f(x)| - |f(y)| \leq f(y) - f(x) \leq M_k(f) - m_k(f).$$
- $f(x), f(y) \leq 0$, then
$$|f(x)| - |f(y)| = f(y) - f(x) \leq M_k(f) - m_k(f).$$

By considering the supreme of the above equations over x followed by the infimum of the above equations over y, we obtain that
$$M_k(|f|) - m_k(|f|) \leq M_k(f) - m_k(f).$$
Hence
$$\begin{aligned} U(|f|, \mathcal{P}) - L(|f|, \mathcal{P}) &= \sum_{k=1}^{n} (M_k(|f|) - m_k(|f|))(t_k - t_{k-1}) \\ &\leq \sum_{k=1}^{n} (M_k(f) - m_k(f))(t_k - t_{k-1}) \\ &= U(f, \mathcal{P}) - L(f, \mathcal{P}) < \epsilon. \end{aligned}$$
Therefore, since $\epsilon > 0$ was arbitrary, $|f|$ is Riemann integrable on $[a, b]$ by Theorem A.2.7.

Since $|f|$ is Riemann integrable, Proposition A.4.2 implies that $-|f|$ is Riemann integrable. Moreover, since
$$-|f(x)| \leq f(x) \leq |f(x)|$$
for all $x \in [a, b]$, Proposition A.4.2 also implies that
$$-\int_a^b |f(x)|\, dx \leq \int_a^b f(x)\, dx \leq \int_a^b |f(x)|\, dx.$$
Hence
$$\left| \int_a^b f(x)\, dx \right| \leq \int_a^b |f(x)|\, dx.$$
which completes the proof. ∎

Appendix B

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?"

B.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 B.1.1. Given two non-empty sets X and Y, a *relation* between X and Y is a subset of the product $X \times Y$. Given a relation R, we write xRy if $(x, y) \in R$. In the case that $Y = X$, we call R a relation on X.

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

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

Given an $x \in X$, the set $\{a \in X \mid a \sim x\}$ 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 B.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 B.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 $a, b \in X$, $a \preceq b$, and $b \preceq a$, then $a = b$, 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 B.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 Y 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 B.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).

B.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.

Remark B.2.1. 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 B.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 g and f then 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 B.2.2. 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 B.2.3. 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 B.2.4. 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

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

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 writing 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 B.2.5. 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 B.2.2. 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 B.2.5 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 B.2.6. 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 B.2.7. 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 $\{1, 2, 3\}$. In order to show this, we require one of the 'optional' axioms of set theory.

Axiom B.2.8 (Axiom of Choice). *Let I be a non-empty set. For each $i \in I$ let A_i be a non-empty set. Then there exists a function $f : I \to \bigcup_{i\in I} A_i$ such that $f(i) \in A_i$ for all $i \in I$.*

Note the Axiom of Choice says that for any collection of non-empty sets, we can always choose an element from each set. This may seem natural, but it is not one of the necessary axioms of Zermelo-Fraenkel set theory and many mathematicians examine what happens when this axiom is removed. However, for the purposes of analysis, the Axiom of Choice should be included for otherwise arguments become substantially more complicated and some results actually fail. One example argument using the Axiom of Choice is the following that shows surjective functions give us information on the cardinality of sets.

Proposition B.2.9. *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 $y \in Y$. By the Axiom of Choice (Axiom B.2.8) there exists a function $g : Y \to \bigcup_{y\in Y} A_y \subseteq X$ is such that $g(y) \in A_y$ for all $y \in Y$.

We claim that g is injective. To see this, assume $y_1, y_2 \in Y$ 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. ∎

B.3 Finite and Infinite Sets

Before we attempt to determine whether the relation in Definition B.2.5 is a partial ordering, let us first formalize the notions of finite and infinite sets.

Definition B.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 B.3.2. *If X is an infinite set, there exists an injection $f : \mathbb{N} \to X$.*

Proof. Since X is non-empty, the power set of X is non-empty. By the Axiom of Choice (Axiom B.2.8) there exists a function $f : \mathcal{P}(X) \setminus \{\emptyset\} \to \mathcal{P}(X)$ such that $f(A) \in A$ for all $A \in \mathcal{P}(X) \setminus \{\emptyset\}$.

Let $a_1 = f(X)$. Since $|X| \neq 1$, $X \setminus \{a_1\}$ 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, since $|X| \neq 2$, 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. 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 B.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 B.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 $g : X \to Y$ by

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

for all $x \in X$. 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 B.3.4. *A non-empty set X is infinite if and only if X is equinumerous to a proper subset.*

B.4 Cantor-Schröder-Bernstein Theorem

To show that $\leq$ from Definition B.2.5 is a partial ordering, we must show that $\leq$ is antisymmetric. To begin, let us first consider the following. In Example B.2.7, 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, we see f is an injective function. Hence $|\mathbb{Q}| \leq |\mathbb{N}|$!

Since $|\mathbb{N}| \leq |\mathbb{Q}|$ 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 B.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 since 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 $n \in \mathbb{N}$.

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 since $A_n \subseteq A_{n-1}$ for all $n \in \mathbb{N}$.

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, since 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. Hence $|A| = |X|$ as claimed.

However $|A| = |Y|$ since $g : Y \to A$ 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 B.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 $|\mathbb{Q}| = |\mathbb{R}|$?

B.5 Countable Sets

One nice corollary about $|\mathbb{N}| = |\mathbb{Q}|$ 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 B.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 B.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 B.3.2. Since Y is infinite, X is infinite. Thus, since 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 B.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 B.5.2. Consequently, for each $n \in \mathbb{N}$, we may write

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

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 B.5.4. *If X and Y are countable sets, $X \bigcup Y$ is a countable set.*

Proof. Apply Theorem B.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 B.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 B.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 B.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

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

Since 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 B.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 B.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 B.5.2, which contradicts the fact that Y is uncountable. Hence X must be uncountable. ∎

Combining Theorem B.5.7 and Proposition B.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 B.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 B.5.3. Since $\mathbb{R}$ is uncountable by Proposition B.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.

Theorem B.5.10. *The Cantor set is uncountable.*

Proof. Recall by Lemma 1.6.9 that every element of the Cantor set $\mathcal{C}$ has a unique ternary representation using only 0s and 2s. Define $f : \mathcal{C} \to [0, 1]$ as follows: If $x \in \mathcal{C}$ has ternary representation $x = \sum_{n=1}^{\infty} \frac{a_n}{3^n}$ with $a_n \in \{0, 2\}$, for all $n \in \mathbb{N}$ let $b_n = \frac{a_n}{2} \in \{0, 1\}$ and define $f(x) = \sum_{n=1}^{\infty} \frac{b_n}{2^n}$. Clearly f is a surjective function so $|\mathcal{C}| \geq |[0, 1]|$ by Proposition B.2.9. Hence, since $\mathcal{C} \subseteq [0, 1]$ so $|\mathcal{C}| \leq |[0, 1]|$, we obtain that $|\mathcal{C}| = |[0, 1]|$ so $\mathcal{C}$ is uncountable. ∎

One question we may ask since $\mathbb{R}$ is whether $\mathbb{R}$ the 'smallest' set larger than $\mathbb{N}$? In particular:

Question B.5.11 (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.

B.6 Comparability of Cardinals

Using the Cantor-Schröder–Bernstein Theorem (Theorem B.4.1), we saw that cardinality gives a partial ordering on the size of sets. However, is it a total ordering (Definition B.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 B.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 B.1. Our main focus is a 'result' about totally ordered subsets of partially ordered sets:

Definition B.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 B.6.3. Recall from Example B.1.5 that the power set $\mathcal{P}(\mathbb{R})$ of $\mathbb{R}$ has a partial ordering $\preceq$ where

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

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 B.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 B.6.5. Recall from Example B.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 B.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. 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 $a \not\preceq b$ 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 B.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 (Axiom B.2.8). 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 B.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, assume $\vec{v}_1, \ldots, \vec{v}_n \in A$ 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, since 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 $A \in \mathcal{L}$. Since 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 for the sake of a contradiction 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, since $B \preceq B \cup \{\vec{v}\}$ 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 B.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 B.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, assume $x \in A_i$ 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 $f_j(x) = f_i(x)$. Since 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, assume $x_1, x_2 \in A$ 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 $x_1 = x_2$. Since 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 $(A, B, f) \in \mathcal{F}$.

Since $(A, B, f) \in \mathcal{F}$, 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, since $\mathcal{C}$ 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 for the sake of a contradiction 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 $(A_0, B_0, f_0) \neq (A', B', g)$. Since 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 B.2.9. ∎

B.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 B.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 B.7.1. *Let X be an infinite set and let Y be a finite subset of X. Then $|X \setminus Y| = |X|$.*

Proof. Assume X is an infinite set and Y is a finite subset of X. Then $Z = X \setminus Y$ is an infinite set. Since Z is infinite, there exists an infinite countable set $W \subseteq Z$ by Lemma B.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 B.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 B.7.2 to uncountable sets is more involved.

Lemma B.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 B.7.2. Thus suppose X is an uncountable set. Define

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

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 (Axiom B.6.7). To do this, first we must verify that $\mathcal{F}$ is non-empty. Since X is uncountable, by Lemma B.3.2 there exists a $W \subseteq X$ such that W is infinite and countable. By Lemma B.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 B.6.9, f and g are well-defined bijections. Furthermore, we claim that $A \cap B = \emptyset$. To see this, suppose for the sake of a contradiction 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. Since this contradicts the definition of $\mathcal{F}$, we obtain that $A \cap B = \emptyset$. Since it is clear that $W = A \cup B$, we see that $(W, A, B, f, g) \in \mathcal{F}$. Since $(W_\alpha, A_\alpha, B_\alpha, f_\alpha, g_\alpha) \preceq (W, A, B, f, g)$ 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 for the sake of a contradiction that $X \setminus W_0$ is infinite. Thus there exists a countable subset $Z \subseteq X \setminus W_0$. By Lemma B.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 B.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 B.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 B.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 B.7.4, we note the following result which improves upon Theorem B.5.3.

Corollary B.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 B.5.3, we may assume that the X_n are pairwise disjoint.

Since X is infinite, Lemma B.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 B.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 $n \in \mathbb{N}$.

Since $|X_n| \leq |X| = |Z_n|$ 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 appendix 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 B.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. To see that $|X| \leq |\mathcal{P}(X)|$, define $f : X \to \mathcal{P}(X)$ by $f(x) = \{x\}$. Clearly f is injective so $|X| \leq |\mathcal{P}(X)|$ by definition.

To see that $|X| \neq |\mathcal{P}(X)|$, we return to a Russell's Paradox-like argument. Suppose for the sake of a contradiction that there exists a function $f : X \to \mathcal{P}(X)$ that is bijective (in particular, f is surjective). Consider the set

$$\Psi = \{x \in X \mid x \notin f(x)\}.$$

Since f is surjective, there exists a $z \in X$ such that $f(z) = \Psi$.

If $z \in \Psi$ then, by the definition of Ψ, it must be the case that $z \notin f(z) = \Psi$, which is a contradiction. Hence it must be the case that $z \notin \Psi$. Therefore, by the definition of Ψ, it must be the case that $z \in f(z) = \Psi$, which is also a contradiction. Hence we have a contradiction to the existence of such an f and thus $|X| \neq |\mathcal{P}(X)|$. ∎

Example B.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 B.6.9 as desired.

Appendix C

Hausdorff Dimension

The goal of this appendix chapter is to modify the definition of the Lebesgue measure in order to obtain a definition for the dimension of subsets of $\mathbb{R}$. The definition of the Hausdorff dimension of a subset of $\mathbb{R}$ will be obtained via a limit of specific outer measures on $\mathbb{R}$.

C.1 Metric Outer Measures

To begin our discussion of dimension of subsets of $\mathbb{R}$, we desire to analyze specific outer measures on $\mathbb{R}$ similar to the Lebesgue outer measure. As we do so, we will obtain an alternative way to demonstrate that every Borel set of $\mathbb{R}$ is Lebesgue measurable. The key to constructing these outer measures is to use the distance function on $\mathbb{R}$ and consider the following pairs of sets.

Definition C.1.1. Two subsets $A, B \subseteq \mathbb{R}$ are said to have *positive separation* if

$$\operatorname{dist}(A, B) = \inf\{|a - b| \mid a \in A, b \in B\} > 0.$$

Example C.1.2. Using the Extreme Value Theorem along with the fact that the distance to a set is a continuous function, it is possible to show that any two disjoint compact subsets of a metric space have positive separation. However, two disjoint closed subsets of a metric space need not have positive separation. Indeed consider $A = \mathbb{N}$ and $B = \{n + \frac{1}{n} \mid n \in \mathbb{N}, n \geq 2\}$. Clearly A and B are disjoint closed subsets of $\mathbb{R}$ that do not have positive separation.

The special collection of outer measures we wish to study are as follows.

Definition C.1.3. An outer measure $\mu^* : \mathcal{P}(\mathbb{R}) \to [0, \infty]$ is said to be a *metric outer measure* if

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

for all $A, B \subseteq \mathbb{R}$ such that A and B have positive separation.

Remark C.1.4. It is not difficult to see that if $\mu^* : \mathcal{P}(\mathbb{R}) \to [0, \infty]$ is an outer measure on $\mathbb{R}$ such that every Borel set of $\mathcal{X}$ is μ^*-measurable, then μ^* must be a metric outer measure. Indeed, assume $A, B \subseteq \mathbb{R}$ have positive separation. It is elementary to see that if $\overline{A}$ and $\overline{B}$ denote the closures of A and B respectively, then $\overline{A}$ and $\overline{B}$ are Borel sets such that $\overline{A} \cap \overline{B} = \emptyset$. Hence

$$(A \cup B) \cap \overline{A} = A \qquad \text{and} \qquad (A \cup B) \cap \overline{A}^c = B.$$

Therefore, since $\overline{A}$ is then μ^*-measurable, we obtain that

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

as desired.

Of course, our desire is to prove the converse; that is, given a metric outer measure μ^* every Borel set is μ^*-measurable. To see this, we will make use of the following lemma.

Lemma C.1.5. *Let $\mu^* : \mathcal{P}(\mathbb{R}) \to [0, \infty]$ be a metric outer measure, let $\{A_n\}_{n=1}^\infty \subseteq \mathcal{P}(\mathbb{R})$ be such that $A_k \subseteq A_{k+1}$ for all $k \in \mathbb{N}$, and let $A = \bigcup_{n=1}^\infty A_n$. If*

$$\mathrm{dist}(A_k, A \setminus A_{k+1}) > 0$$

for all $k \in \mathbb{N}$, then

$$\mu^*(A) = \lim_{n\to\infty} \mu^*(A_k).$$

Proof. Due to the monotonicity of outer measures, $(\mu^*(A_n))_{n\geq 1}$ is a monotone sequence. Therefore, either $\lim_{n\to\infty} \mu^*(A_n)$ exists and is finite, or is infinity. Furthermore, since $\mu^*(A_k) \leq \mu^*(A)$ for all $k \in \mathbb{N}$ due to the monotonicity of outer measures, we obtain that $\lim_{n\to\infty} \mu^*(A_n) \leq \mu^*(A)$. Therefore, if $\lim_{n\to\infty} \mu^*(A_n) = \infty$ then clearly $\mu^*(A) = \infty$ and the result holds. Hence we may assume that $\lim_{n\to\infty} \mu^*(A_n) < \infty$.

Let $B_1 = A_1$ and for each $k \geq 2$ let $B_k = A_k \setminus A_{k-1}$. Clearly $\bigcup_{m=1}^k B_m \subseteq A_k$ and $B_k \subseteq A \setminus A_{k-1}$ for all $k \in \mathbb{N}$. Therefore, if $m \geq k+2$ and $B \subseteq \bigcup_{j=1}^k B_j$ then

$$\begin{aligned} \mathrm{dist}\,(B_m, B) &\geq \mathrm{dist}\left(B_m, \bigcup_{j=1}^k B_j\right) \\ &\geq \mathrm{dist}(A \setminus A_{m-1}, A_k) \\ &\geq \mathrm{dist}(A \setminus A_{m-1}, A_{m-2}) > 0 \end{aligned}$$

by assumption so $\mu^*(B_m \cup B) = \mu^*(B_m) + \mu^*(B)$ as μ^* is a metric outer measure. Hence

$$\begin{aligned}\mu^*\left(\bigcup_{k=1}^{n} B_{2k}\right) &= \mu^*\left(B_{2n} \cup \left(\bigcup_{k=1}^{n-1} B_{2k}\right)\right)\\ &= \mu^*(B_{2n}) + \mu^*\left(\bigcup_{k=1}^{n-1} B_{2k}\right)\\ &= \ldots = \sum_{k=1}^{n} \mu^*(B_{2k})\end{aligned}$$

and

$$\begin{aligned}\mu^*\left(\bigcup_{k=1}^{n} B_{2k-1}\right) &= \mu^*\left(B_{2n-1} \cup \left(\bigcup_{k=1}^{n-1} B_{2k-1}\right)\right)\\ &= \mu^*(B_{2n-1}) + \mu^*\left(\bigcup_{k=1}^{n-1} B_{2k-1}\right)\\ &= \ldots = \sum_{k=1}^{n} \mu^*(B_{2k-1})\end{aligned}$$

for all $n \in \mathbb{N}$. Therefore, since

$$\mu^*\left(\bigcup_{k=1}^{n} B_{2k}\right) \leq \mu^*(A_{2n}) \qquad \text{and} \qquad \mu^*\left(\bigcup_{k=1}^{n} B_{2k-1}\right) \leq \mu^*(A_{2n-1}),$$

we obtain that the infinite sums $\sum_{k=1}^{\infty} \mu^*(B_{2k})$ and $\sum_{k=1}^{\infty} \mu^*(B_{2k-1})$ converge as $\lim_{n\to\infty} \mu^*(A_n) < \infty$.

For each $m \in \mathbb{N}$ notice that

$$\begin{aligned}\mu^*(A) &= \mu^*\left(A_m \cup \left(\bigcup_{k=m+1}^{\infty} B_k\right)\right)\\ &\leq \mu^*(A_m) + \sum_{k=m+1}^{\infty} \mu^*(B_k)\end{aligned}$$

by the subadditivity of outer measures. However, since

$$\lim_{m\to\infty} \sum_{k=m+1}^{\infty} \mu^*(B_k) = 0$$

as $\sum_{k=1}^{\infty} \mu^*(B_{2k})$ and $\sum_{k=1}^{\infty} \mu^*(B_{2k-1})$ converge, and since $\lim_{n\to\infty} \mu^*(A_k)$ exists, we obtain that

$$\mu^*(A) \leq \lim_{n\to\infty} \mu^*(A_k)$$

which when combined with $\lim_{n\to\infty} \mu^*(A_n) \leq \mu^*(A)$ yields the desired result.

∎

Proposition C.1.6. *If $\mu^* : \mathcal{P}(\mathbb{R}) \to [0,\infty]$ is a metric outer measure, then every Borel subset of $\mathbb{R}$ is μ^*-measurable.*

Proof. Since the set of μ^*-measurable sets is a σ-algebra by Theorem 1.5.6 and since the set of closed subsets of $\mathbb{R}$ generate the Borel σ-algebra, it suffices to prove that every closed subset of $\mathbb{R}$ is μ^*-measurable.

Let F be an arbitrary closed subset of $\mathbb{R}$. To see that F is μ^*-measurable, let $A \subseteq \mathbb{R}$ be arbitrary. For each $n \in \mathbb{N}$ let

$$A_n = \left\{ a \in A \,\middle|\, \mathrm{dist}(\{a\}, F) \geq \frac{1}{n} \right\}.$$

Notice that $A_n \subseteq A_{n+1}$ for all $n \in \mathbb{N}$ and that

$$\bigcup_{n=1}^{\infty} A_n = \{a \in A \mid \mathrm{dist}(\{a\}, F) > 0\} = A \cap F^c$$

since F is closed (so $x \in F^c$ if and only if $\mathrm{dist}(\{x\}, F) > 0$).

We claim that

$$\mathrm{dist}(A_k, (A \cap F^c) \setminus A_{k+1}) \geq \frac{1}{k(k+1)}.$$

To see this, let $a \in A_k$ and $x \in (A \cap F^c) \setminus A_{k+1}$ be arbitrary. Clearly this implies $x \in A$, $x \notin F$, and $x \notin A_{k+1}$. Hence $0 < \mathrm{dist}(\{x\}, F) < \frac{1}{k+1}$. Furthermore, since $a \in A_k$, we obtain that $\mathrm{dist}(\{a\}, F) \geq \frac{1}{k}$. By the triangle inequality, we obtain by taking an infimum over all $y \in F$ that

$$\frac{1}{k} \leq \mathrm{dist}(\{a\}, F) \leq |a - x| + \mathrm{dist}(\{x\}, F) < |a - x| + \frac{1}{k+1}.$$

Hence

$$|a - x| \geq \frac{1}{k(k+1)}.$$

Therefore, since $a \in A_k$ and $x \in (A \cap F^c) \setminus A_{k+1}$ were arbitrary, the claim is complete.

By Lemma C.1.5 we obtain that

$$\lim_{n\to\infty} \mu^*(A_n) = \mu^*(A \cap F^c).$$

Since $A_n \cup (A \cap F) \subseteq A$, since

$$\mathrm{dist}(A_n, (A \cap F)) \geq \mathrm{dist}(A_n, F) \geq \frac{1}{n} > 0,$$

and since μ^* is a metric outer measure, we obtain that

$$\mu^*(A) \geq \mu^*(A_n \cup (A \cap F)) = \mu^*(A_n) + \mu^*(A \cap F)$$

for all $n \in \mathbb{N}$. Therefore, by taking a limit of the right-hand-side, we obtain that

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

Therefore, since $A \subseteq \mathbb{R}$ was arbitrary, F is μ^*-measurable. Therefore, since F was an arbitrary closed subset of $\mathbb{R}$, the proof is complete. ∎

To complete our alternative approach to demonstrating Borel subsets of $\mathbb{R}$ are Lebesgue measurable, we demonstrate that the Lebesgue outer measure is a metric outer measure.

Proposition C.1.7. *The Lebesgue outer measure is a metric outer measure.*

Proof. Let $A, B \subseteq \mathbb{R}$ have positive separation. Since λ^* is an outer measure, clearly $\lambda^*(A \cup B) \leq \lambda^*(A) + \lambda^*(B)$ by subadditivity.

To see the other inequality, let $\delta = \frac{1}{4}\mathrm{dist}(A, B) > 0$. For each $0 < \epsilon < \delta$ there exists a countable collection of open intervals $\{I_n\}_{n=1}^\infty$ such that $A \cup B \subseteq \bigcup_{n=1}^\infty I_n$ and

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

We desire to modify $\{I_n\}_{n=1}^\infty$ in order to control bound the lengths of each interval we use. To begin if $I_n = (a, b)$ where $a, b \in \mathbb{R}$, for each $k \in \mathbb{N}$ let

$$I_{n,k} = \left(a + k\delta, \min\left\{b, a + (k+1)\delta + \frac{\epsilon}{2^{nk}}\right\}\right).$$

Clearly each $I_{n,k}$ is an open interval with length

$$\ell(I_{n,k}) \leq \delta + \frac{\epsilon}{2^{nk}} < \frac{3}{2}\delta < \mathrm{dist}(A, B).$$

Furthermore $I_n \subseteq \bigcup_{k=1}^\infty I_{n,k}$ and

$$\sum_{k=1}^\infty \ell(I_{n,k}) \leq b - a + \sum_{k=1}^\infty \frac{\epsilon}{2^{nk}} = \ell(I_n) + \frac{\epsilon}{2^n}.$$

If $a = -\infty$ or $b = \infty$ then we can apply a similar process to construct a countable number of open intervals $I_{n,k}$ such that $\ell(I_{n,k}) < \mathrm{dist}(A, B)$ for all $k \in \mathbb{N}$, and $\sum_{k=1}^\infty \ell(I_{n,k}) \leq \ell(I_n) + \frac{\epsilon}{2^n}$. Therefore $\{I_{n,k} \mid n, k \in \mathbb{N}\}$ is a countable collection of open intervals such that $A \cup B \subseteq \bigcup_{n,k=1}^\infty I_{n,k}$ and

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

Since $\ell(I_{n,k}) < \mathrm{dist}(A, B)$, each $I_{n,k}$ can intersect at most one of A and B. Let

$$\begin{aligned} J_A &= \{(n,k) \in \mathbb{N}^2 \mid I_{n,k} \cap A \neq \emptyset\} \quad \text{and} \\ J_B &= \{(n,k) \in \mathbb{N}^2 \mid I_{n,k} \cap B \neq \emptyset\}. \end{aligned}$$

Then J_A and J_B are countable disjoint sets such that

$$A \subseteq \bigcup_{(n,k)\in J_A} I_{n,k} \qquad \text{and} \qquad B \subseteq \bigcup_{(n,k)\in J_B} I_{n,k}.$$

Hence

$$\begin{aligned} \lambda^*(A \cup B) + 2\epsilon &\geq \sum_{n,k=1}^{\infty} \ell(I_{n,k}) \\ &\geq \sum_{(n,k)\in J_A} \ell(I_{n,k}) + \sum_{(n,k)\in J_B} \ell(I_{n,k}) \\ &\geq \lambda^*(A) + \lambda^*(B). \end{aligned}$$

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

$$\lambda^*(A \cup B) = \lambda^*(A) + \lambda^*(B).$$

Therefore, since A and B were arbitrary subsets of $\mathbb{R}$ with positive separation, the result follows. ∎

C.2 Hausdorff Outer Measures

To define the Hausdorff dimension of subsets of $\mathbb{R}$, we will construct the Hausdorff outer measures. The process for modifying the definition of the Lebesgue outer measure to obtain the Hausdorff outer measures comes from both modifying the length function and the collection of open intervals permitted. In particular, we want restrict the lengths of the open intervals used.

Definition C.2.1. Let $\mathcal{I}$ denote the set of all open intervals in $\mathbb{R}$. For each $\epsilon > 0$ let

$$\mathcal{F}_\epsilon = \{I \subseteq \mathbb{R} \mid I \in \mathcal{I}, \ell(I) \leq \epsilon\}.$$

For each $s \in (0, \infty)$ let $\mu^*_{s,\epsilon}$ denote the outer measure defined by

$$\mu^*_{s,\epsilon}(A) = \inf\left\{ \sum_{n=1}^{\infty} \ell(I_n)^s \;\middle|\; \{I_n\}_{n=1}^{\infty} \subseteq \mathcal{F}_\epsilon, A \subseteq \bigcup_{n=1}^{\infty} I_n \right\}$$

for all $A \subseteq \mathbb{R}$. Notice trivially that if $0 < \epsilon' < \epsilon$ then $\mu^*_{s,\epsilon}(A) \leq \mu^*_{s,\epsilon'}(A)$ for all $A \subseteq \mathbb{R}$.

Definition C.2.2. For $s \in (0, \infty)$, the *s-dimensional outer Hausdorff measure on* $\mathbb{R}$ is the outer measure $H_s^* : \mathcal{P}(\mathbb{R}) \to [0, \infty]$ defined by

$$H_s^*(A) = \sup_{\epsilon > 0} \mu_{s,\epsilon}^*(A) = \lim_{\epsilon \to 0^+} \mu_{s,\epsilon}^*(A)$$

for all $A \subseteq \mathbb{R}$.

Unsurprisingly, the s-dimensional outer Hausdorff measure is a outer measure with the properties of the previous section.

Proposition C.2.3. *For all $s \in (0, \infty)$ the s-dimensional outer Hausdorff measure H_s^* is a metric outer measure.*

Proof. To see that H_s^* is an outer measure, recall that each $\mu_{s,\epsilon}^*$ is an outer measure. Since the defining properties of an outer measure from Definition 1.5.1 are easily seen to pass to limits, H_s^* is an outer measure.

To see that H_s^* is a metric outer measure, assume $A, B \subseteq \mathbb{R}$ have positive separation. Therefore $\mathrm{dist}(A, B) > 0$. Clearly

$$H_s^*(A \cup B) \leq H_s^*(A) + H_s^*(B)$$

as H_s^* is an outer measure, so it suffices to prove the other inequality.

Assume $\epsilon < \frac{1}{2}\mathrm{dist}(A, B)$. Let $\{I_n\}_{n=1}^\infty \subseteq \mathcal{F}_\epsilon$ be such that $A \cup B \subseteq \bigcup_{n=1}^\infty I_n$. Since

$$\ell(I_n) \leq \epsilon < \frac{1}{2}\mathrm{dist}(A, B),$$

every I_n intersects at most one of A and B. Let

$$J_A = \{n \in \mathbb{N} \mid I_n \cap A \neq \emptyset\} \qquad \text{and} \qquad J_B = \{n \in \mathbb{N} \mid I_n \cap B \neq \emptyset\}$$

Then J_A and J_B are countable disjoint sets such that

$$A \subseteq \bigcup_{n \in J_A} I_n \qquad \text{and} \qquad B \subseteq \bigcup_{n \in J_B} I_n.$$

Hence

$$\begin{aligned} \sum_{n=1}^\infty \ell(I_n)^s &\geq \sum_{n \in J_A} \ell(I_n)^s + \sum_{n \in J_B} \ell(I_n)^s \\ &\geq \mu_{s,\epsilon}^*(A) + \mu_{s,\epsilon}^*(B). \end{aligned}$$

Therefore, since $\{I_n\}_{n=1}^\infty \subseteq \mathcal{F}_\epsilon$ was an arbitrary collection such that $A \cup B \subseteq \bigcup_{n=1}^\infty I_n$, we obtain that

$$\mu_{s,\epsilon}^*(A \cup B) \geq \mu_{s,\epsilon}^*(A) + \mu_{s,\epsilon}^*(B).$$

As this holds for all $\epsilon < \frac{1}{2}\mathrm{dist}(A, B)$, we obtain by taking limits that

$$H_s^*(A \cup B) \geq H_s^*(A) + H_s^*(B).$$

Therefore, since A and B were arbitrary subsets of $\mathbb{R}$ with positive separation, the result follows. ∎

By applying the Carathéodory Method to the Hausdorff outer measures, we obtain the following collection of measures.

Definition C.2.4. For $s \in (0, \infty)$, the *s-dimensional Hausdorff measure* on $\mathbb{R}$, denoted H_s, is the measure H_s obtained by restricting H_s^* to the set of H_s^*-measurable sets.

Remark C.2.5. Note that Proposition C.1.6 implies that every Borel subset of $\mathbb{R}$ is H_s^*-measurable for all $s \in (0, \infty)$.

Example C.2.6. The 1-dimensional Hausdorff measure on $\mathbb{R}$ is the Lebesgue measure. To see this, first note for all $A \subseteq \mathbb{R}$ that $\lambda^*(A) \leq \mu_{1,\epsilon}^*(A)$ for all $\epsilon > 0$. Hence $\lambda^*(A) \leq H_1^*(A)$ for all $A \subseteq \mathbb{R}$. To see the other inclusion, notice by the proof of Proposition C.1.7 that for all $A \subseteq \mathbb{R}$ and all $0 < \epsilon < \delta$ there exists a collection $\{I_n\}_{n=1}^\infty$ of open intervals with $\ell(I_n) < \frac{3}{2}\delta$ such that

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

This implies $\mu_{1,\frac{3}{2}\delta}^*(A) \leq \lambda^*(A) + \epsilon$ for all $0 < \epsilon < \delta$ and thus $H_s^*(A) = \lambda^*(A)$. Therefore, by the definitions of H_1 and λ, we obtain that $H_1 = \lambda$.

Remark C.2.7. Notice that if $s, t \in (0, \infty)$ and $t < s$ then $x^s \leq x^t$ whenever $0 \leq x < 1$. Consequently, by the above definitions, we see that $\mu_{s,\epsilon}^*(A) \leq \mu_{t,\epsilon}^*(A)$ for all $A \subseteq \mathbb{R}$ and $\epsilon < 1$. Hence $H_s(A) \leq H_t(A)$ for all $A \in \mathfrak{B}(\mathbb{R})$ whenever $s, t \in (0, \infty)$ and $t < s$ (note we restrict to Borel sets as this is the largest common domain of H_s and H_t).

In fact, something rather spectacular occurs.

Theorem C.2.8. *If $s, t \in (0, \infty)$ are such that $t < s$ and $A \in \mathfrak{B}(\mathbb{R})$ is such $H_t(A) < \infty$, then $H_s(A) = 0$.*

Proof. Fix a Borel set $A \subseteq \mathbb{R}$ and assume $H_t(A) < \infty$. Let $0 < \epsilon < 1$. Then for any collection $\{I_n\}_{n=1}^\infty \in \mathcal{F}_\epsilon$ such that $A \subseteq \bigcup_{n=1}^\infty I_n$, observe that

$$\sum_{n=1}^\infty \ell(I_n)^s = \sum_{n=1}^\infty \ell(I_n)^{s-t}\ell(I_n)^t \leq \epsilon^{s-t} \sum_{n=1}^\infty \ell(I_n)^t.$$

Therefore, by taking the infimum over all such $\{I_n\}_{n=1}^\infty$, we obtain that

$$\mu_{s,\epsilon}^*(A) \leq \epsilon^{s-t}\mu_{t,\epsilon}^*(A) \leq \epsilon^{s-t}H_t(A).$$

Therefore, since $H_t(A) < \infty$, we obtain that $H_s(A) = 0$ by taking the limit as ϵ tends to zero. ∎

By Theorem C.2.8, we arrive at a definition of dimension for a Borel subset of $\mathbb{R}$.

Definition C.2.9. Let A be a Borel subset of $\mathbb{R}$. The *Hausdorff dimension of* A, denoted $\dim_H(A)$, is

$$\dim_H(A) = \inf\{s > 0 \mid H_s(A) = 0\} = \sup\{s > 0 \mid H_s(A) = \infty\}.$$

Remark C.2.10. Since $A \subseteq B \subseteq \mathbb{R}$ implies $H_s(A) \leq H_s(B)$ for all $s \in (0, \infty)$, we see that $\dim_H(A) \leq \dim_H(B)$ by construction. This is clearly a property we would expect for a good dimension function.

Remark C.2.11. We claim that if $A \subseteq \mathbb{R}$ then $\dim_H(A) \leq 1$. To see this, fix $s > 1$ and let $0 < \epsilon < 1$. Since $\sum_{n=1}^\infty \frac{\epsilon}{n} = \infty$, it is possible to cover $\mathbb{R}$ with a countable collection open intervals I_n such that $\ell(I_n) = \frac{\epsilon}{n}$ for all n (i.e. place a symmetric interval of length ϵ around 0 and alternate placing intervals at the left most endpoint of the last interval placed in the negative numbers and the right most endpoint of the last interval placed in the positive numbers). Thus

$$\mu^*_{s,\epsilon}(\mathbb{R}) \leq \sum_{n=1}^\infty \left(\frac{\epsilon}{n}\right)^s = \epsilon^s \sum_{n=1}^\infty \frac{1}{n^s}.$$

Since $s > 1$, we know that $\sum_{n=1}^\infty \frac{1}{n^s} < \infty$. Therefore, since $\lim_{\epsilon \to 0^+} \epsilon^s = 0$, we obtain that $H_s(\mathbb{R}) = H_s^*(\mathbb{R}) = 0$. Moreover, since the 1-dimensional Hausdorff measure is the Lebesgue measure and $\lambda(\mathbb{R}) = \infty$, we obtain that $\dim_H(\mathbb{R}) = 1$. Thus the claim follows from Remark C.2.10.

Example C.2.12. Let I be a non-singleton finite intervals. Hence $0 < \lambda(I) < \infty$. Since the 1-dimensional Hausdorff measure is the Lebesgue measure so $H_1(I) = \lambda(I) \in (0, \infty)$, Theorem C.2.8 implies that $H_s(I) = 0$ for all $s > 1$ and $H_s(I) = \infty$ for all $s < 1$. Therefore $\dim_H(I) = 1$ by definition.

Similarly, if I is an infinite interval, then $H_1(I) = \lambda(I) = \infty$. Thus $\dim_H(I) \geq 1$. Hence Remark C.2.11 implies $\dim_H(I) = 1$.

To finish our discussion of Hausdorff dimension, we compute the Hausdorff dimension of the most notorious set in Lebesgue measure theory.

Proposition C.2.13. *The Hausdorff dimension of the Cantor set is* $\frac{\ln(2)}{\ln(3)}$.

Proof. Let

$$s_0 = \frac{\ln(2)}{\ln(3)}.$$

To compute $H_{s_0}(\mathcal{C})$, let $0 < \epsilon < 1$. Choose n such that $\frac{1}{3^n} < \epsilon$. By taking P_n as in Definition 1.6.8, by replacing each closed interval I in P_n with an open interval J such that $I \subseteq J$ and $\ell(J) < \ell(I) + \delta$ for some δ such that $\frac{1}{3^n} + \delta < \epsilon$, and by sending δ to 0, we obtain that

$$\mu^*_{s_0,\epsilon}(\mathcal{C}) \leq \sum_{k=1}^{2^n} \left(\frac{1}{3^n}\right)^{s_0} = \frac{2^n}{3^{ns_0}}.$$

However

$$3^{ns_0} = 3^{\frac{\ln(2^n)}{\ln(3)}} = 3^{\log_3(2^n)} = 2^n$$

so $\mu^*_{s_0,\epsilon}(\mathcal{C}) \leq 1$. Therefore, by taking the limit as ϵ tends to 0, we obtain that $H_{s_0}(\mathcal{C}) \leq 1$. Hence $\dim_H(\mathcal{C}) \leq s_0$.

To see the other inequality, let $0 < \epsilon < 1$ and let $\{I_n\}_{n=1}^\infty \subseteq \mathcal{F}_\epsilon$ be such that $\mathcal{C} \subseteq \bigcup_{n=1}^\infty I_n$. Since $\mathcal{C}$ is compact, there exists an $M \in \mathbb{N}$ such that $\mathcal{C} \subseteq \bigcup_{n=1}^M I_n$.

Choose $N \in \mathbb{N}$ such that

$$\frac{1}{3^{N+1}} \leq \epsilon < \frac{1}{3^N}$$

and choose $k \in \mathbb{N}$ such that

$$\frac{1}{3^k} < \ell(I_n)$$

for all $1 \leq n \leq M$. Consider P_k as in Definition 1.6.8. If $1 \leq n \leq M$ and

$$\frac{1}{3^j} \leq \ell(I_n) < \frac{1}{3^{j-1}}$$

for some $j \leq k$, we see that I_n can intersect at most one closed interval in the definition of P_{j-1} since each such closed interval has length $\frac{1}{3^{j-1}}$ and is separated from each other closed interval by an open interval of length $\frac{1}{3^{j-1}}$. Therefore, since each closed interval in the definition of P_{j-1} contains 2^{k-j+1} of the closed intervals in the definition of P_k, we see that I_n can intersect at most 2^{k-j+1} of the closed intervals in the definition of P_k. Since

$$2^{k-j+1} = 2^{k+1}2^{-j} = 2^{k+1}3^{-js_0} = 2^{k+1}\left(\frac{1}{3^j}\right)^{s_0} \leq 2^{k+1}\ell(I_n)^{s_0},$$

we see that each I_n can intersect at most $2^{k+1}\ell(I_n)^{s_0}$ of the closed intervals in the definition of P_k. Thus, since $\mathcal{C} \subseteq \bigcup_{n=1}^M I_n$ and since P_k contains 2^k intervals, we obtain that

$$\sum_{n=1}^M 2^{k+1}\ell(I_n)^{s_0} \geq 2^k.$$

Thus

$$\sum_{n=1}^\infty \ell(I_n)^{s_0} \geq \sum_{n=1}^M \ell(I_n)^{s_0} \geq \frac{1}{2}.$$

Therefore, since $\{I_n\}_{n=1}^\infty \subseteq \mathcal{F}_\epsilon$ was arbitrary, we obtain that

$$\mu^*_{s_0,\epsilon}(\mathcal{C}) \geq \frac{1}{2}$$

for all $0 < \epsilon < 1$. Therefore $\frac{1}{2} \leq H_{s_0}(\mathcal{C}) \leq 1$. Hence Theorem C.2.8 implies that $\dim_H(\mathcal{C}) = s_0$ as desired. ∎

Appendix D

Fourier Series on $\mathbb{R}$

In this appendix chapter, we will explore using Lebesgue measure theory to study Fourier Analysis. Since students in this course have already taken MATH 3001 and studied Fourier series for functions on $[0, 2\pi)$, we will focus instead on the Fourier series for functions on $\mathbb{R}$.

D.1 Complex Lebesgue Measure Theory

In order to begin our study of Fourier series for functions on $\mathbb{R}$, we first need the notion of the Lebesgue integral for complex-valued functions as, like with the Fourier series for functions on $[0, 2\pi)$, we desire to use complex exponentials. Unsurprisingly, our solution to constructing a Lebesgue integral for complex-valued functions is to use the real and imaginary parts. Thus we make the following definitions.

Definition D.1.1. Let $f : \mathbb{R} \to \mathbb{C}$. The *real and imaginary parts of f* are the functions $\mathrm{Re}(f), \mathrm{Im}(f) : \mathbb{R} \to \mathbb{R}$ defined by

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

for all $x \in \mathbb{R}$ respectively.

Definition D.1.2. A function $f : \mathbb{R} \to \mathbb{C}$ is said to be *Lebesgue measurable* if $\mathrm{Re}(f), \mathrm{Im}(f) : \mathbb{R} \to \mathbb{R}$ are Lebesgue measurable as real-valued functions.

Remark D.1.3. Note that the theory of real-valued Lebesgue measurable functions from Section 2.2 immediately transfer to complex-valued Lebesgue measurable functions with minor work. For example, if $f : \mathbb{R} \to \mathbb{C}$ is Lebesgue measurable, then since

$$|f| = \sqrt{|\mathrm{Re}(f)(x)|^2 + |\mathrm{Im}(f)(x)|^2},$$

we see that $|f|$ is Lebesgue measurable since the square and sum of Lebesgue measurable functions is Lebesgue measurable and the square root of a Lebesgue measurable function is Lebesgue measurable by Proposition refprop:measurable-composed-with-continuous-is-measurable.

Moreover, it is not too difficult to verify that Egoroff's Theorem (Theorem 2.4.1) and Lusin's Theorem (Theorem 2.6.1) hold for complex-valued measurable functions.

Remark D.1.4. Let $f : \mathbb{R} \to \mathbb{C}$ and for a complex number z, let $\overline{z}$ denote the complex conjugation of z. Consider the function $\overline{f} : \mathbb{R} \to \mathbb{C}$ defined by $\overline{f}(x) = \overline{f(x)}$ for all $x \in \mathbb{R}$. Since

$$\mathrm{Re}(\overline{f}) = \mathrm{Re}(f) \qquad \text{and} \qquad \mathrm{Im}(\overline{f}) = -\mathrm{Im}(f),$$

we see that $\overline{f}$ is Lebesgue measurable if and only if f is Lebesgue measurable.

Onto integration!

Definition D.1.5. A function $f : \mathbb{R} \to \mathbb{C}$ is said to be *Lebesgue integrable* if $\mathrm{Re}(f), \mathrm{Im}(f) : \mathbb{R} \to \mathbb{R}$ are Lebesgue integrable as real-valued functions.

Remark D.1.6. Let $f : \mathbb{R} \to \mathbb{C}$ be Lebesgue measurable. Since

$$\begin{aligned} |f(x)| &= \sqrt{|\mathrm{Re}(f)(x)|^2 + |\mathrm{Im}(f)(x)|^2} \\ &\leq \sqrt{2} \max\{|\mathrm{Re}(f)(x)|, |\mathrm{Im}(f)(x)|\} \\ &\leq \sqrt{2}|\mathrm{Re}(f)(x)| + \sqrt{2}|\mathrm{Im}(f)(x)|, \end{aligned}$$

we see that if $\mathrm{Re}(f), \mathrm{Im}(f) : \mathbb{R} \to \mathbb{R}$ are Lebesgue integrable as real-valued functions, then $|f|$ is Lebesgue integrable as a real-valued function. Conversely, since

$$|\mathrm{Re}(f)(x)| \leq |f(x)| \qquad \text{and} \qquad |\mathrm{Im}(f)(x)| \leq |f(x)|,$$

we see that if $|f| : \mathbb{R} \to \mathbb{R}$ is Lebesgue integrable as a real-valued function, then $\mathrm{Re}(f), \mathrm{Im}(f) : \mathbb{R} \to \mathbb{R}$ are Lebesgue integrable as real-valued functions.

Definition D.1.7. Let $f : \mathbb{R} \to \mathbb{C}$ be Lebesgue integral. The *Lebesgue integral of* f is defined to be

$$\int_{\mathbb{R}} f \, d\lambda = \int_{\mathbb{R}} \mathrm{Re}(f) \, d\lambda + i \int_{\mathbb{R}} \mathrm{Im}(f) \, d\lambda.$$

Remark D.1.8. Note that the set of complex-valued Lebesgue integrable forms a vector space over $\mathbb{C}$ by the same arguments used in the proof of Theorem 3.4.9. Moreover, if $f : \mathbb{R} \to \mathbb{C}$ is Lebesgue integrable, we see that if is Lebesgue integrable with

$$\mathrm{Re}(if) = -\mathrm{Im}(f) \qquad \text{and} \qquad \mathrm{Im}(if) = \mathrm{Re}(f)$$

so

$$\begin{aligned}\int_{\mathbb{R}} if\,d\lambda &= \int_{\mathbb{R}} -\mathrm{Im}(f)\,d\lambda + i\int_{\mathbb{R}} \mathrm{Re}(f)\,d\lambda \\ &= -\int_{\mathbb{R}} \mathrm{Im}(f)\,d\lambda + i\int_{\mathbb{R}} \mathrm{Re}(f)\,d\lambda \\ &= i\left(\int_{\mathbb{R}} \mathrm{Re}(f)\,d\lambda + i\int_{\mathbb{R}} \mathrm{Im}(f)\,d\lambda\right) \\ &= i\int_{\mathbb{R}} f\,d\lambda.\end{aligned}$$

Thus the Lebesgue integral is complex-linear.

Remark D.1.9. Let $f : \mathbb{R} \to \mathbb{C}$ be Lebesgue integrable. Since

$$\mathrm{Re}(\overline{f}) = \mathrm{Re}(f) \qquad \text{and} \qquad \mathrm{Im}(\overline{f}) = -\mathrm{Im}(f),$$

we see that $\overline{f}$ is Lebesgue integrable and

$$\begin{aligned}\int_{\mathbb{R}} \overline{f}\,d\lambda &= \int_{\mathbb{R}} \mathrm{Re}(f)\,d\lambda + i\int_{\mathbb{R}} -\mathrm{Im}(f)\,d\lambda \\ &= \int_{\mathbb{R}} \mathrm{Re}(f)\,d\lambda - i\int_{\mathbb{R}} \mathrm{Im}(f)\,d\lambda \\ &= \overline{\int_{\mathbb{R}} f\,d\lambda}\end{aligned}$$

by the definition of the complex-valued Lebesgue integral.

One result for the complex-valued Lebesgue integral that does not immediately follow from the real-valued Lebesgue integral is the following.

Theorem D.1.10. *If $f : \mathbb{R} \to \mathbb{C}$ is Lebesgue integrable, then*

$$\left|\int_{\mathbb{R}} f\,d\lambda\right| \leq \int_{\mathbb{R}} |f|\,d\lambda.$$

Proof. By properties of complex numbers, there exists a $z \in \mathbb{C}$ such that $|z| = 1$ and

$$z\int_{\mathbb{R}} f\,d\lambda = \left|\int_{\mathbb{R}} f\,d\lambda\right| \geq 0$$

(i.e. rotate the complex number $\int_{\mathbb{R}} f\,d\lambda$ until it is positive). Hence zf is integrable and

$$0 \leq \left|\int_{\mathbb{R}} f\,d\lambda\right| = \int_{\mathbb{R}} zf\,d\lambda = \int_{\mathbb{R}} \mathrm{Re}(zf)\,d\lambda + i\int_{\mathbb{R}} \mathrm{Im}(zf)\,d\lambda.$$

However, since $\int_{\mathbb{R}} \mathrm{Re}(zf)\,d\lambda, \int_{\mathbb{R}} \mathrm{Im}(zf)\,d\lambda \in \mathbb{R}$, it must be the case that $\int_{\mathbb{R}} \mathrm{Im}(zf)\,d\lambda = 0$. Hence

$$\begin{aligned}\left|\int_{\mathbb{R}} f\,d\lambda\right| &= \int_{\mathbb{R}} \mathrm{Re}(zf)\,d\lambda \\ &\leq \int_{\mathbb{R}} |\mathrm{Re}(zf)|\,d\lambda \\ &\leq \int_{\mathbb{R}} |zf|\,d\lambda = \int_{\mathbb{R}} |f|\,d\lambda\end{aligned}$$

as desired. ∎

Remark D.1.11. Although the Monotone Convergence Theorem (Theorem 3.3.2) and Fatou's Lemma (Theorem 3.6.1) don't make sense for complex-valued functions, the Dominated Convergence Theorem (Theorem 3.7.1) does and holds by the same proof presented in the real-valued case. Moreover, Fubini's Theorem (Theorem 5.2.1) and Tonelli's Theorem (Theorem 5.2.2) transfer immediately to the complex setting.

D.2 Fourier Transform on $\mathbb{R}$

With the above construction of the complex Lebesgue integral, we can begin our study of Fourier series for functions on $\mathbb{R}$. Unlike with Fourier series for functions on $[0, 2\pi)$, in order to recover a function from its Fourier series (see Theorem D.6.2), we will need to expand the Fourier series from a function on $\mathbb{Z}$ to a function on $\mathbb{R}$.

Definition D.2.1. Let $f : \mathbb{R} \to \mathbb{C}$ be Lebesgue integrable. The *Fourier transform of f* is the function $\widehat{f} : \mathbb{R} \to \mathbb{C}$ defined by

$$\hat{f}(y) = \int_{\mathbb{R}} f(x)e^{-iyx}\,d\lambda(x)$$

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

Remark D.2.2. Note if $f : \mathbb{R} \to \mathbb{C}$ is Lebesgue integrable, then for each $y \in \mathbb{R}$ the function $g_y : \mathbb{R} \to \mathbb{C}$ defined by

$$g_y(x) = f(x)e^{-iyx} = f(x)\cos(yx) + if(x)\sin(yx)$$

for all $x \in \mathbb{R}$ is Lebesgue measurable. Moreover, since $|g_y| = |f|$, we see that g_y is Lebesgue integrable for all $y \in \mathbb{R}$. Hence $\widehat{f}$ is well-defined.

Not only is the Fourier transform well-defined, it produces nice functions on $\mathbb{R}$.

Theorem D.2.3. *Let $f : \mathbb{R} \to \mathbb{C}$ be Lebesgue integrable. Then $\widehat{f} : \mathbb{R} \to \mathbb{C}$ is a bounded, continuous function.*

Proof. To see that $\widehat{f}$ is continuous, let $(y_n)_{n\geq 1}$ be a sequence in $\mathbb{R}$ that converges to $y \in \mathbb{R}$. For each $n \in \mathbb{N}$, let $g_n : \mathbb{R} \to \mathbb{C}$ be defined by

$$g_n(x) = f(x)e^{-iy_n x}$$

for all $x \in \mathbb{R}$. Therefore, since

$$\lim_{n\to\infty} g_n(x) = f(x)e^{-iyx}$$

for all $x \in \mathbb{R}$, since $|g_n(x)| = |f(x)|$ for all $x \in \mathbb{R}$, and since f is Lebesgue integrable, the Dominated Convergence Theorem (Theorem 3.7.1) implies that

$$\begin{aligned}\lim_{n\to\infty} \widehat{f}(y_n) &= \lim_{n\to\infty} \int_{\mathbb{R}} f(x)e^{-iy_n x}\, d\lambda(x) \\ &= \lim_{n\to\infty} g_n\, d\lambda \\ &= \int_{n\to\infty} f(x)e^{-iyx} d\lambda(x) \\ &= \widehat{f}(y).\end{aligned}$$

Therefore, since $(y_n)_{n\geq 1}$ was arbitrary, $\widehat{f}$ is continuous on $\mathbb{R}$.

To see that $\widehat{f}$ is bounded, note since f is Lebesgue integrable that

$$M = \int_{\mathbb{R}} |f|\, d\lambda < \infty.$$

Therefore, since for all $y \in \mathbb{R}$ we have by Theorem D.1.10 that

$$|\widehat{f}(y)| = \left|\int_{\mathbb{R}} f(x)e^{-iyx}\, d\lambda(x)\right| \leq \int_{\mathbb{R}} |f(x)e^{-iyx}|\, d\lambda = M,$$

we see that $\widehat{f}$ is bounded by M. ∎

Unsurprisingly, the Fourier transform for functions on $\mathbb{R}$ shares similar properties to the Fourier transform studied in MATH 3001.

Proposition D.2.4. *Let $f, g : \mathbb{R} \to \mathbb{C}$ be Lebesgue integrable functions. Then*

a) $\widehat{\alpha f + g} = \alpha\widehat{f} + \widehat{g}$ *for all $\alpha \in \mathbb{C}$ (i.e. the Fourier transform is linear),*

b) $\widehat{\overline{f}}(y) = \overline{\widehat{f}(-y)}$ *for all $y \in \mathbb{R}$,*

c) if $t \in \mathbb{R}$ and $f_t : \mathbb{R} \to \mathbb{C}$ is defined by $f_t(x) = f(x-t)$ for all $x \in \mathbb{R}$, then $\widehat{f_t}(y) = e^{-iyt}\widehat{f}(y)$ for all $y \in \mathbb{R}$,

d) if $\check{f} : \mathbb{R} \to \mathbb{C}$ is defined by $\check{f}(x) = f(-x)$ for all $x \in \mathbb{R}$, then $\widehat{\check{f}}(y) = \widehat{f}(-y)$ for all $y \in \mathbb{R}$,

e) *if* $\delta > 0$ *and* $k_\delta : \mathbb{R} \to \mathbb{C}$ *is defined by* $k_\delta(x) = f(\delta x)$ *for all* $x \in \mathbb{R}$*, then* $\widehat{k_\delta}(y) = \frac{1}{\delta}\widehat{f}\left(\frac{1}{\delta}y\right)$ *for all* $y \in \mathbb{R}$*, and*

f) *if* $t \in \mathbb{R}$ *and* $h : \mathbb{R} \to \mathbb{C}$ *is define by* $h(x) = e^{-itx}f(x)$ *for all* $x \in \mathbb{R}$*, then* $\widehat{h}(y) = \widehat{f}(y+t)$ *for all* $y \in \mathbb{R}$*.*

Proof. To see a) is true, note for all $\alpha \in \mathbb{C}$ and $y \in \mathbb{R}$ that

$$\begin{aligned}
(\widehat{\alpha f + g})(y) &= \int_{\mathbb{R}} (\alpha f + g)(x) e^{-iyx} \, d\lambda(x) \\
&= \int_{\mathbb{R}} (\alpha f(x) + g(x)) e^{-iyx} \, d\lambda(x) \\
&= \alpha \int_{\mathbb{R}} f(x) e^{-iyx} \, d\lambda(x) + \int_{\mathbb{R}} g(x) e^{-iyx} \, d\lambda(x) \\
&= \alpha \widehat{f}(y) + \widehat{g}(y)
\end{aligned}$$

as desired.

To see that b) is true, note for all $y \in \mathbb{R}$ that

$$\begin{aligned}
\overline{\widehat{f}(-y)} &= \overline{\int_{\mathbb{R}} f(x) e^{-i(-y)x} \, d\lambda(x)} \\
&= \overline{\int_{\mathbb{R}} f(x) e^{iyx} \, d\lambda(x)} \\
&= \int_{\mathbb{R}} \overline{f(x)} e^{-iyx} \, d\lambda(x) \\
&= \widehat{\overline{f}}(y)
\end{aligned}$$

as desired.

To see that c) is true, note by Proposition 3.4.11 that f_t is Lebesgue integrable. Moreover, for all $y, t \in \mathbb{R}$ we have that

$$\begin{aligned}
\widehat{f_t}(y) &= \int_{\mathbb{R}} f_t(x) e^{-iyx} \, d\lambda(x) \\
&= \int_{\mathbb{R}} f(x-t) e^{-iyx} \, d\lambda(x) \\
&= \int_{\mathbb{R}} f(x) e^{-iy(x+t)} \, d\lambda(x) && \text{by Proposition 3.4.11} \\
&= e^{-iyt} \int_{\mathbb{R}} f(x) e^{-iyx} \, d\lambda(x) \\
&= e^{-iyt} \widehat{f}(y)
\end{aligned}$$

as desired.

To see that d) is true, note by Proposition 3.4.12 that $\check{f}$ is Lebesgue integrable. Moreover, for all $y \in \mathbb{R}$ we have that

$$\begin{aligned}
\widehat{\check{f}}(y) &= \int_{\mathbb{R}} \check{f}(x) e^{-iyx} \, d\lambda(x) \\
&= \int_{\mathbb{R}} f(-x) e^{-iyx} \, d\lambda(x) \\
&= \int_{\mathbb{R}} f(x) e^{-iy(-x)} \, d\lambda(x) && \text{by Proposition 3.4.12} \\
&= \int_{\mathbb{R}} f(x) e^{-i(-y)x} \, d\lambda(x) \\
&= \widehat{f}(-y)
\end{aligned}$$

as desired.

To see that e) is true, note by Proposition 3.4.13 that k_δ is Lebesgue integrable for all $\delta > 0$. Moreover, for all $y \in \mathbb{R}$ and $\delta > 0$ we have that

$$\begin{aligned}
\widehat{k_\delta}(y) &= \int_{\mathbb{R}} k_\delta(x) e^{-iyx} \, d\lambda(x) \\
&= \int_{\mathbb{R}} f(\delta x) e^{-iyx} \, d\lambda(x) \\
&= \frac{1}{\delta} \int_{\mathbb{R}} f(x) e^{-iy\frac{x}{\delta}} \, d\lambda(x) && \text{by Proposition 3.4.13} \\
&= \frac{1}{\delta} \int_{\mathbb{R}} f(x) e^{-i\left(\frac{y}{\delta}\right)x} \, d\lambda(x) \\
&= \frac{1}{\delta} \widehat{f}\left(\frac{1}{\delta} y\right)
\end{aligned}$$

as desired.

To see that f) is true, note that h is Lebesgue measurable and, since $|h| = |f|$, h is Lebesgue integrable. Moreover for all $y, t \in \mathbb{R}$ we have that

$$\begin{aligned}
\widehat{h}(y) &= \int_{\mathbb{R}} h(x) e^{-iyx} \, d\lambda(x) \\
&= \int_{\mathbb{R}} e^{-itx} f(x) e^{-iyx} \, d\lambda(x) \\
&= \int_{\mathbb{R}} f(x) e^{-i(y+t)x} \, d\lambda(x) \\
&= \widehat{f}(y+t)
\end{aligned}$$

as desired. ∎

D.3 The Riemann-Lebesgue Lemma

Using Proposition D.2.4 and other facts from Lebesgue measure theory, we can prove the following useful "lemma" that demonstrates a property of the Fourier transform and extends Theorem D.2.3 to show that not every continuous bounded function is obtain via the Fourier transform.

Theorem D.3.1 (Riemann-Lebesgue Lemma). *Let $f : \mathbb{R} \to \mathbb{C}$ be Lebesgue integrable. Then*

$$\lim_{y\to\pm\infty} \widehat{f}(y) = 0.$$

Proof. We will proceed as we often do when trying to demonstrate a result for Lebesgue integrable functions; first we will demonstrate the result for characteristic functions, then use linearity to demonstrate the result for simple functions, then use limits to demonstrate the result for non-negative functions, and finally use linearity to demonstrate the result for all Lebesgue measurable functions. It turns out we cannot jump straight to all characteristic functions. Luckily, all Lebesgue measurable sets are “almost” intervals.

First, assume $f = \chi_I$ for some open interval I. Since f is Lebesgue integrable, we have that $\lambda(I) < \infty$. Hence we can write $I = (a, b)$ for some $a, b \in \mathbb{R}$ with $a < b$. Thus for all $y \in \mathbb{R} \setminus \{0\}$,

$$\begin{aligned}
\widehat{f}(y) &= \int_{\mathbb{R}} \chi_I(x) e^{-iyx}\, d\lambda(x) \\
&= \int_{(a,b)} e^{-iyx}\, d\lambda(x) \\
&= \int_a^b e^{-iyx}\, dx && \text{by Theorem 3.5.5} \\
&= \frac{e^{-iyb} - e^{-iya}}{-iy}.
\end{aligned}$$

Therefore since

$$\left| \frac{e^{-iyb} - e^{-iya}}{-iy} \right| \le \frac{1}{|y|}$$

for all $y \in \mathbb{R} \setminus \{0\}$, we see that $\lim_{y\to\pm\infty} \widehat{f}(y) = 0$ when $f = \chi_I$ for an interval I.

Next, assume $f = \chi_A$ for some Lebesgue measurable set A. Since f is Lebesgue integrable, we have that $\lambda(A) < \infty$. Let $\epsilon > 0$. By Littlewood's First Principle (Theorem 2.5.1) there exists a finite number of disjoint open intervals $I_1, \ldots, I_n$ such that if $U = \bigcup_{k=1}^n I_k$ then

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

Thus $\lambda(U) \le \lambda(A) + \epsilon$ so $\lambda(I_k) < \infty$ for all $k \in \{1, \ldots, n\}$.

Let $g = \sum_{k=1}^n \chi_{I_k}$. Note since $I_1, \ldots, I_n$ are disjoint that

$$|f(x) - g(x)| = \chi_{(A\setminus U)\cup(U\setminus A)}(x)$$

for all $x \in \mathbb{R}$. Moreover, since $\lambda(I_k) < \infty$ implies χ_{I_k} is Lebesgue integrable so $\widehat{g} = \sum_{k=1}^n \widehat{\chi_{I_k}}$ by Proposition D.2.4, the above demonstrates that

$$\lim_{y \to \pm\infty} \widehat{g}(y) = 0.$$

Hence there exists an $M \in \mathbb{R}$ such that if $y \in \mathbb{R}$ and $|y| \geq M$ then $|\widehat{g}(y)| < \epsilon$. Hence if $y \in \mathbb{R}$ and $|y| \geq M$ then

$$\begin{aligned}
\left|\widehat{f}(y)\right| &\leq \left|\widehat{f}(y) - \widehat{g}(y)\right| + |\widehat{g}(y)| \\
&= \left|\widehat{(f-g)}(y)\right| + \epsilon \\
&= \left|\int_{\mathbb{R}} (f-g)(x) e^{-iyx} \, d\lambda(x)\right| + \epsilon \\
&\leq \int_{\mathbb{R}} |f(x) - g(x)| \, d\lambda(x) + \epsilon \\
&= \int_{\mathbb{R}} \chi_{(A \setminus U) \cup (U \setminus A)}(x) \, d\lambda(x) + \epsilon \\
&= \lambda((A \setminus U) \cup (U \setminus A)) + \epsilon \\
&= 2\epsilon.
\end{aligned}$$

Therefore, since $\epsilon > 0$ was arbitrary, $\lim_{y \to \pm\infty} \widehat{f}(y) = 0$ when $f = \chi_A$ for a Lebesgue measurable set A.

Next, assume $f = \sum_{k=1}^n a_k \chi_{A_k}$ for some Lebesgue measurable sets $\{A_k\}_{k=1}^n$ and $\{a_k\}_{k=1}^n \in (0, \infty)$. Since f is Lebesgue integrable, $\lambda(A_k) < \infty$ for all k and thus Proposition D.2.4 implies that

$$\widehat{f}(y) = \sum_{k=1}^n a_k \widehat{\chi_{A_k}}(y)$$

for all $y \in \mathbb{R}$. Therefore, the above implies that $\lim_{y \to \pm\infty} \widehat{f}(y) = 0$ when f is a Lebesgue integrable simple function.

Next assume f is non-negative. Let $\epsilon > 0$. By Theorem 2.3.5 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 $0 \leq \varphi_n \leq f = |f|$, we obtain that φ_n is Lebesgue integrable for all $n \in \mathbb{N}$. Moreover, since the Monotone Convergence Theorem (Theorem 3.3.2) implies that

$$\lim_{n \to \infty} \int_{\mathbb{R}} \varphi_n \, d\lambda = \int_{\mathbb{R}} f \, d\lambda,$$

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

$$\int_{\mathbb{R}} |f - \varphi_N| \, d\lambda = \int_{\mathbb{R}} f - \varphi_N \, d\lambda = \int_{\mathbb{R}} f \, d\lambda - \int_{\mathbb{R}} \varphi_N \, d\lambda < \epsilon.$$

Since the above demonstrated that

$$\lim_{y \to \pm\infty} \widehat{\varphi_N}(y) = 0,$$

there exists an $M \in \mathbb{R}$ such that if $y \in \mathbb{R}$ and $|y| \geq M$ then $|\widehat{\varphi_N}(y)| < \epsilon$. Hence if $y \in \mathbb{R}$ and $|y| \geq M$ then

$$\begin{aligned}
\left|\widehat{f}(y)\right| &\leq \left|\widehat{f}(y) - \widehat{\varphi_N}(y)\right| + |\widehat{\varphi_N}(y)| \\
&= \left|(\widehat{f - \varphi_N})(y)\right| + \epsilon \\
&= \left|\int_{\mathbb{R}} (f - \varphi_N)(x) e^{-iyx} \, d\lambda(x)\right| + \epsilon \\
&\leq \int_{\mathbb{R}} |f(x) - \varphi_N(x)||e^{-iyx}| \, d\lambda(x) + \epsilon \\
&= \int_{\mathbb{R}} |f - \varphi_N| \, d\lambda + \epsilon \\
&= \epsilon + \epsilon = 2\epsilon.
\end{aligned}$$

Therefore, since $\epsilon > 0$ was arbitrary, $\lim_{y \to \pm\infty} \widehat{f}(y) = 0$ when f is non-negative.

Finally, if f is Lebesgue integrable, then so too are the positive and negative parts of the real and imaginary parts of f. Since the above shows the result holds for the the positive and negative parts of the real and imaginary parts of f and since the Fourier transform is linear, the result follows. ∎

D.4 Convolution of Functions on ℝ

In order to prove some desirable results about the Fourier transform, we again turn our attention to convolutions as we did in MATH 3001.

Definition D.4.1. Let $f, g : \mathbb{R} \to \mathbb{C}$ be Lebesgue integrable functions. The *convolution of f and g* is the function $f * g : \mathbb{R} \to \mathbb{C}$ defined by

$$(f * g)(x) = \int_{\mathbb{R}} f(x - y) g(y) \, d\lambda(y).$$

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

However, unlike in MATH 3001 we run into an immediate issue; why is the convolution of Lebesgue integrable functions well-defined? In MATH 3001 the verification that the convolution was well-defined was a simple task since all the functions that were considered were continuous functions on finite intervals so the Riemann integrable made sense. However, since we are now considering Lebesgue integrable functions on $\mathbb{R}$, we must ensure the functions we are defining are Lebesgue integrable. In particular, how do we know for each $x \in \mathbb{R}$ that the function $h(y) = f(x - y)g(y)$ is Lebesgue integrable?

Note one way to avoid this issue is to assume that g is bounded in which case it would follow that the $f * g$ is well-defined since the translation and inversion of a Lebesgue integrable function is Lebesgue integrable and since the product of a bounded Lebesgue measurable function with a Lebesgue integrable function is Lebesgue integrable. By restricting to bounded g, we could proceed and develop a good portion of the theory we desire. However, to obtain a complete theory, we desire to resolve this problem.

The way we will resolve our issues is straightforward to comprehend: we desire to show $H(x, y) = f(x - y)g(y)$ is 2-dimensional Lebesgue integrable so that we may use Fubini's Theorem (Theorem 5.2.1) to obtain $h(y) = f(x - y)g(y)$ is Lebesgue integrable for all $x \in \mathbb{R}$. To show that $H(x, y)$ is 2-dimensional Lebesgue integrable is a simple application of Tonelli's Theorem (Theorem 5.2.2) provided H is 2-dimensional Lebesgue measurable. Since f and g are Lebesgue measurable, it is believable that H will be 2-dimensional Lebesgue measurable. However, an issue arises because it is not clear based on the definition of a Lebesgue measurable function that $F(x, y) = f(x - y)$ is 2-dimensional Lebesgue measurable since

$$F^{-1}((a, \infty)) = \{(x, y) \in \mathbb{R}^2 \mid x - y \in f^{-1}((a, \infty))\}.$$

In particular, if $A \in \mathcal{M}(\mathbb{R})$, why is

$$\{(x, y) \in \mathbb{R}^2 \mid x - y \in A\} \in \mathcal{M}(\mathbb{R}^2)?$$

This question can be resolved via the following two lemmata. Note the Borel subsets of $\mathbb{R}^2$, denoted $\mathfrak{B}(\mathbb{R}^2)$, are defined to be the smallest σ-algebra generated by the open subsets of $\mathbb{R}^2$. Moreover, since every measurable rectangle is 2-dimensional Lebesgue measurable, we obtain that $\mathfrak{B}(\mathbb{R}^2) \subseteq \mathcal{M}(\mathbb{R}^2)$.

Lemma D.4.2. *Let $B \subseteq \mathbb{R}$ be a Borel set. Then*

$$B' = \{(x, y) \in \mathbb{R}^2 \mid x - y \in B\} \in \mathcal{M}(\mathbb{R}^2).$$

Proof. Let $g : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by

$$g(x, y) = (x - y, y)$$

for all $(x, y) \in \mathbb{R}^2$. It is elementary to verify that g is a continuous bijection with continuous inverse (i.e. g is a homeomorphism) and thus g is a bijection on the open subsets of $\mathbb{R}^2$.

Let

$$\mathcal{A} = \{A \subseteq \mathbb{R}^2 \mid g^{-1}(A) \text{ is Borel in } \mathbb{R}^2\}.$$

By the proof of Proposition 2.1.14 we see that $\mathcal{A}$ is a σ-algebra on $\mathbb{R}^2$ that contains the open subsets of $\mathbb{R}^2$. Hence $\mathfrak{B}(\mathbb{R}^2) \subseteq \mathcal{A}$. Thus

$$\begin{aligned} g^{-1}(\mathfrak{B}(\mathbb{R}^2)) &\subseteq g^{-1}(\mathcal{A}) \\ &= \{g^{-1}(A) \mid A \subseteq \mathbb{R}^2, g^{-1}(A) \text{ is Borel in } \mathbb{R}^2\} \\ &\subseteq \mathfrak{B}(\mathbb{R}^2) \end{aligned}$$

so $\mathfrak{B}(\mathbb{R}^2) \subseteq g(\mathfrak{B}(\mathbb{R}^2))$. Similarly

$$\mathcal{A}' = \{A \subseteq \mathbb{R}^2 \mid g(A) = (g^{-1})^{-1}(A) \text{ is Borel in } \mathbb{R}^2\}$$

is a σ-algebra on $\mathbb{R}^2$ such that $\mathfrak{B}(\mathbb{R}^2) \subseteq \mathcal{A}'$. Thus

$$\begin{aligned} g(\mathfrak{B}(\mathbb{R}^2)) &\subseteq g(\mathcal{A}') \\ &= \{g(A) \subseteq \mathbb{R}^2 \mid A \subseteq \mathbb{R}^2, g(A) \text{ is Borel in } \mathbb{R}^2\} \\ &\subseteq \mathfrak{B}(\mathbb{R}^2). \end{aligned}$$

Hence g is a bijection between the Borel subsets of $\mathbb{R}^2$.

Let $B \in \mathfrak{B}(\mathbb{R})$. Then

$$B' = \{(x, y) \in \mathbb{R}^2 \mid x - y \in B\} = g^{-1}(B \times \mathbb{R}).$$

Since the Borel sets on $\mathbb{R}$ are the smallest σ-algebra containing the open sets, we see since B is a Borel set that $B \times \mathbb{R}$ is an element of the σ-algebra generated by

$$\{U \times \mathbb{R} \mid U \subseteq \mathbb{R} \text{ open}\}.$$

Since $U \times \mathbb{R}$ is open in $\mathbb{R}^2$ for all open $U \subseteq \mathbb{R}$ and since the Borel sets are the smallest σ-algebra generated by the open sets (both in $\mathbb{R}$ and $\mathbb{R}^2$), we see that $B \times \mathbb{R} \in \mathfrak{B}(\mathbb{R}^2)$. Hence

$$BB' = g^{-1}(B \times \mathbb{R}) \in \mathfrak{B}(\mathbb{R}^2) \subseteq \mathcal{M}(\mathbb{R}^2)$$

as desired. ∎

Lemma D.4.3. *Let $A \in \mathcal{M}(\mathbb{R})$ be such that $\lambda(A) = 0$. Then*

$$A' = \{(x, y) \in \mathbb{R}^2 \mid x - y \in A\} \in \mathcal{M}(\mathbb{R}^2)$$

and $\lambda_2(A') = 0$.

Proof. Since the 2-dimensional Lebesgue measure is complete being produced by the Carathéodory Method (see Proposition 1.5.8), it suffices to show that $\lambda_2^*(A') = 0$.

To see that $\lambda_2^*(A') = 0$, for each $n \in \mathbb{N}$ let

$$A_n = \{(x, y) \in \mathbb{R}^2 \mid x - y \in A, y \in (-n, n)\}.$$

Clearly $A' = \bigcup_{n=1}^\infty A_n$ so, since λ_2^* is an outer measure, it suffices to show that $\lambda_2^*(A_n) = 0$ for all $n \in \mathbb{N}$.

Fix $n \in \mathbb{N}$ and let $\epsilon > 0$ be arbitrary. Since $\lambda(A) = 0$, there exists a countable collection $\{I_k\}_{k=1}^\infty$ of open intervals such that $A \subseteq \bigcup_{k=1}^\infty I_k$ and

$$\sum_{k=1}^\infty \lambda(I_k) < \epsilon.$$

Since $A \subseteq \bigcup_{k=1}^\infty I_k$, we see that

$$A_n \subseteq \bigcup_{k=1}^\infty \{(x,y) \in \mathbb{R}^2 \mid x - y \in I_k, y \in (-n,n)\}.$$

For each $k \in \mathbb{N}$ let

$$P_{n,k} = \{(x,y) \in \mathbb{R}^2 \mid x - y \in I_k, y \in (-n,n)\}.$$

Clearly $P_{n,k}$ is an open set as I_k is an open interval for all k. Hence $P_{n,k}$ is Borel and thus 2-dimensional Lebesgue measurable. Therefore, Tonelli's Theorem (Theorem 5.2.2) implies that

$$\begin{aligned}
\lambda_2(P_{n,k}) &= \int_{\mathbb{R}^2} \chi_{P_{n,k}} \, d\lambda_2 \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} \chi_{P_{n,k}}(x,y) \, d\lambda(x) \, d\lambda(y) \\
&= \int_{(-n,n)} \int_{\mathbb{R}} \chi_{P_{n,k}}(x,y) \, d\lambda(x) \, d\lambda(y) \\
&= \int_{(-n,n)} \lambda(I_k) \, d\lambda(y) \\
&= 2n\lambda(I_k).
\end{aligned}$$

Therefore, we obtain that

$$\lambda_2^*(A_n) \leq \sum_{k=1}^\infty \lambda_2(P_{n,k}) = 2n \sum_{k=1}^\infty \lambda(I_k) \leq 2n\epsilon.$$

Therefore, since $\epsilon > 0$ was arbitrary, $\lambda_2^*(A_n) = 0$ for all $n \in \mathbb{N}$ and thus $\lambda_2^*(A') = 0$. ∎

Lemma D.4.4. *Let $f : \mathbb{R} \to \mathbb{R}$ be a Lebesgue measurable function and define $h : \mathbb{R}^2 \to \mathbb{R}$ by*

$$h(x,y) = f(x-y)$$

for all $(x,y) \in \mathbb{R}^2$. Then h is 2-dimensional Lebesgue measurable.

Proof. To see that h is 2-dimensional Lebesgue measurable, let $a \in \mathbb{R}$ be arbitrary. Our goal is to show that $h^{-1}((a,\infty)) \in \mathcal{M}(\mathbb{R}^2)$.

Let $A = f^{-1}((a,\infty))$ and note $A \in \mathcal{M}(\mathbb{R})$. Thus

$$h^{-1}((a,\infty)) = \{(x,y) \in \mathbb{R}^2 \mid x - y \in A\}.$$

By Proposition 1.6.13 there exists a Borel set F (i.e. F is a countable union of closed sets) such that $F \subseteq A$ and $\lambda(A \setminus F) = 0$. Hence

$$h^{-1}((a,\infty)) = \{(x,y) \in \mathbb{R}^2 \mid x - y \in F\} \cup \{(x,y) \in \mathbb{R}^2 \mid x - y \in A \setminus F\}.$$

Since F is Borel, Lemma D.4.2 implies that

$$\{(x, y) \in \mathbb{R}^2 \mid x - y \in F\} \in \mathcal{M}(\mathbb{R}^2).$$

Moreover, since $\lambda(A \setminus F) = 0$, Lemma D.4.3 implies that

$$\{(x, y) \in \mathbb{R}^2 \mid x - y \in A \setminus F\} \in \mathcal{M}(\mathbb{R}^2).$$

Hence $h^{-1}((a, \infty)) \in \mathcal{M}(\mathbb{R}^2)$ completing the proof. ∎

With the above, we can now prove the convolution is a well-defined function. In fact, the convolution will be Lebesgue integrable!

Theorem D.4.5. *If $f : \mathbb{R} \to \mathbb{C}$ is a Lebesgue integrable function and $g : \mathbb{R} \to \mathbb{C}$ is a bounded Lebesgue measurable function, then $f * g$ is Lebesgue integrable such that*

$$\int_{\mathbb{R}} |f * g| \, d\lambda \leq \left(\int_{\mathbb{R}} |f| \, d\lambda \right) \left(\int_{\mathbb{R}} |g| \, d\lambda \right).$$

Proof. Define $h : \mathbb{R}^2 \to \mathbb{R}$ by $h(x, y) = f(x - y)g(y)$ for all $(x, y) \in \mathbb{R}^2$. Clearly h is well-defined and 2-dimensional Lebesgue measurable by Lemma D.4.4.

We claim that h is 2-dimensional Lebesgue integrable. To see this, note by Tonelli's Theorem (Theorem 5.2.2) that

$$\begin{aligned}
&\int_{\mathbb{R}^2} |h| \, d\lambda_2 \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} |h(x, y)| \, d\lambda(x) \, d\lambda(y) \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} |f(x - y)||g(y)| \, d\lambda(x) \, d\lambda(y) \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} |f(x)||g(y)| \, d\lambda(x) \, d\lambda(y) \quad \text{by Proposition 3.4.11} \\
&= \left(\int_{\mathbb{R}} f(x) \, d\lambda(x) \right) \left(\int_{\mathbb{R}} g(y) \, d\lambda(y) \right) < \infty.
\end{aligned}$$

Hence h is 2-dimensional Lebesgue integrable.

By Fubini's Theorem (Theorem 5.2.1) that the function

$$\Phi(x) = \int_{\mathbb{R}} h(x, y) \, d\lambda(y) = (f * g)(x)$$

is a well-defined Lebesgue integrable function. Moreover, again by Tonelli's

Theorem, we have that that

$$\begin{aligned}
\int_{\mathbb{R}} |f * g|(x)\, d\lambda(x) &= \int_{\mathbb{R}} \left| \int_{\mathbb{R}} f(x-y)g(y)\, d\lambda(y) \right| d\lambda(x) \\
&\leq \int_{\mathbb{R}} \int_{\mathbb{R}} |f(x-y)g(y)|\, d\lambda(x)\, d\lambda(y) \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} |f(x-y)g(y)|\, d\lambda(y)\, d\lambda(x) \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} |f(x)||g(y)|\, d\lambda(x)\, d\lambda(y) \quad \text{by Proposition 3.4.11} \\
&= \left(\int_{\mathbb{R}} f(x)\, d\lambda(x) \right) \left(\int_{\mathbb{R}} g(y)\, d\lambda(y) \right)
\end{aligned}$$

as desired. ∎

Theorem D.4.6. *Let $f, g, h : \mathbb{R} \to \mathbb{R}$ be Lebesgue integrable functions. Then*

a) $f * (g + h) = (f * g) + (f * h)$,

b) $(f + g) * h) = (f * h) + (g * h)$,

c) $(zf) * g) = z(f * g) = f * (zg)$ *for all* $z \in \mathbb{C}$,

d) $f * g = g * f$,

e) $(f * g) * h = f * (g * h)$, *and*

f) $\widehat{(f * g)}(y) = \hat{f}(y)\hat{g}(y)$ *for all* $y \in \mathbb{R}$.

Proof. To see that a) is true, note for all $x \in \mathbb{R}$ that

$$\begin{aligned}
(f * (g + h))(x) &= \int_{\mathbb{R}} f(x-y)(g+h)(y)\, d\lambda(y) \\
&= \int_{\mathbb{R}} f(x-y)(g(y) + h(y))\, d\lambda(y) \\
&= \int_{\mathbb{R}} f(x-y)g(y)\, d\lambda(y) + \int_{\mathbb{R}} f(x-y)h(y)\, d\lambda(y) \\
&= (f * g)(y) + (f * h)(y)
\end{aligned}$$

as desired.

To see that b) is true, we can either repeat the proof of part a) or use part a) along with part d). Thus we omit the proof.

Next, note c) is clearly true by similar arguments used to prove part a) as the integral is linear.

To see that d) is true, note for all $x \in \mathbb{R}$ that

$$\begin{aligned}(f * g)(x) &= \int_{\mathbb{R}} f(x-y)g(y)\, d\lambda(y) \\ &= \int_{\mathbb{R}} f(x+y)g(-y)\, d\lambda(y) && \text{by Proposition 3.4.12} \\ &= \int_{\mathbb{R}} f(y)g(-(y-x))\, d\lambda(y) && \text{by Proposition 3.4.11} \\ &= \int_{\mathbb{R}} g(x-y)f(y)\, d\lambda(y) \\ &= (g * f)(x)\end{aligned}$$

as desired.

To see that e) is true, first note since f, g, and h are Lebesgue integrable, that $f * g$ and $g * h$ are Lebesgue integrable by Theorem D.4.5 so $(f * g) * h$ and $f * (g * h)$ make sense.

Let $H : \mathbb{R}^3 \to \mathbb{C}$ be defined by

$$H(x, y, z) = f(x - y - z)g(z)h(y)$$

for all $(x, y, z) \in \mathbb{R}^3$. By a similar argument to that used in Lemma D.4.4, we obtain that H is 3-dimensional Lebesgue measurable. Moreover, by the 3-dimensional Tonelli's Theorem (Theorem 5.2.2), we have that

$$\begin{aligned}&\int_{\mathbb{R}^3} |H|\, d\lambda_3 \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \int_{\mathbb{R}} |f(x-y-z)||g(z)||h(y)|\, d\lambda(x)\, d\lambda(z)\, d\lambda(y) \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \int_{\mathbb{R}} |f(x)||g(z)||h(y)|\, d\lambda(x)\, d\lambda(z)\, d\lambda(y) && \text{by Proposition 3.4.11} \\ &= \left(\int_{\mathbb{R}} |f|\, d\lambda\right)\left(\int_{\mathbb{R}} |g|\, d\lambda\right)\left(\int_{\mathbb{R}} |h|\, d\lambda\right) \\ &< \infty.\end{aligned}$$

Hence, by the 3-dimensional Fubini's Theorem (Theorem 5.2.1), we have that for each $x \in \mathbb{R}$ the function $\Psi : \mathbb{R}^2 \to \mathbb{C}$ defined by

$$\Psi(y, z) = f(x - y - z)g(z)h(y)$$

is 2-dimensional Lebesgue integrable. Therefore, by Fubini's Theorem,

$$\begin{aligned}
((f*g)*h)(x) &= \int_{\mathbb{R}} (f*g)(x-y)h(y)\, d\lambda(y) \\
&= \int_{\mathbb{R}} (g*f)(x-y)h(y)\, d\lambda(y) && \text{by part d)} \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} g(x-y-z)f(z)\, d\lambda(z) \right) h(y)\, d\lambda(y) \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} f(z)g(x-y-z)h(y)\, d\lambda(z)\, d\lambda(y) \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} f(z)g(x-y-z)h(y)\, d\lambda(y)\, d\lambda(z) \\
&= \int_{\mathbb{R}} f(z)(g*h)(x-z)\, d\lambda(z) \\
&= ((g*h)*f)(x) \\
&= (f*(g*h))(x) && \text{by part d).}
\end{aligned}$$

Hence the proof of part e) is complete.

To see that f) is true, note for all $y \in \mathbb{R}$ that the function $K(x,z) = f(x-z)g(z)e^{-iyx}$ is 2-dimensional Lebesgue measurable by Lemma D.4.4. Moreover, by Tonelli's Theorem (Theorem 5.2.2), we have that

$$\begin{aligned}
\int_{\mathbb{R}^2} |K|\, d\lambda_2 &= \int_{\mathbb{R}} \int_{\mathbb{R}} |f(x-z)g(z)e^{-iyx}|\, d\lambda(x)\, d\lambda(z) \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} |f(x-z)||g(z)|\, d\lambda(x)\, d\lambda(z) \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} |f(x)||g(z)|\, d\lambda(x)\, d\lambda(z) && \text{by Proposition 3.4.11} \\
&= \left(\int_{\mathbb{R}} |f|\, d\lambda \right) \left(\int_{\mathbb{R}} |g|\, d\lambda \right) < \infty.
\end{aligned}$$

Therefore K is 2-dimensional Lebesgue integrable. Hence we have by Fubini's

Theorem (Theorem 5.2.1) that

$$\begin{aligned}
\widehat{(f*g)}(y) &= \int_{\mathbb{R}} (f*g)(x)e^{-iyx}\, d\lambda(x) \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x-z)g(z)\, d\lambda(z) \right) e^{-iyx}\, d\lambda(x) \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} f(x-z)g(z)e^{-iyx}\, d\lambda(z)\, d\lambda(x) \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} f(x-z)g(z)e^{-iyx}\, d\lambda(x)\, d\lambda(z) \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} f(x)g(z)e^{-iy(x+z)}\, d\lambda(x)\, d\lambda(z) && \text{by Proposition 3.4.11} \\
&= \int_{\mathbb{R}} g(z)e^{-iyz} \left(\int_{\mathbb{R}} f(x)e^{-iyx}\, d\lambda(x) \right) d\lambda(z) \\
&= \int_{\mathbb{R}} g(z)e^{-iyz}\widehat{f}(y)\, d\lambda(z) \\
&= \widehat{f}(y)\widehat{g}(y)
\end{aligned}$$

as desired. ∎

To complete our discussion of properties of the convolution, we desire to show that if g is bounded then $f * g$ is continuous. To do so, we require two preliminary results.

Lemma D.4.7. *Let $f : \mathbb{R} \to \mathbb{C}$ be a Lebesgue integrable function. For all $\epsilon > 0$ there exists a continuous function $g : \mathbb{R} \to \mathbb{C}$ and a compact set K such that $g(x) = 0$ for all $x \notin K$ and*

$$\int_{\mathbb{R}} |f - g|\, d\lambda < \epsilon.$$

Proof. Let $f : \mathbb{R} \to \mathbb{C}$ be Lebesgue integrable and let $\epsilon > 0$. For each $m \in \mathbb{N}$, let

$$A_m = \{x \in \mathbb{R} \mid |f(x)| \leq m\}.$$

Note $A_m \in \mathcal{M}(\mathbb{R})$ since f is Lebesgue measurable. Moreover, $A_m \subseteq A_{m+1}$ for all $m \in \mathbb{N}$ and $\bigcup_{m=1}^{\infty} A_m = \mathbb{R}$ by construction.

For each $m \in \mathbb{N}$, let $f_m = f\chi_{A_m}$. Note f_m is Lebesgue measurable for all $m \in \mathbb{N}$ since f is Lebesgue measurable. Moreover, f_m is Lebesgue integrable for all $m \in \mathbb{N}$ since $|f_m| \leq |f|$ and $(f_m)_{m\geq 1}$ converges pointwise to f on $\mathbb{R}$ since $A_m \subseteq A_{m+1}$ for all $m \in \mathbb{N}$ and $\bigcup_{m=1}^{\infty} A_m = \mathbb{R}$. Therefore, since $|f_m| \leq |f|$ for all $m \in \mathbb{N}$, and since f is Lebesgue integrable, the proof of the Dominated Convergence Theorem (Theorem 3.7.1) implies that

$$\lim_{m\to\infty} \int_{\mathbb{R}} |f - f_m|\, d\lambda = 0$$

(see Remark 3.7.2). Therefore, there exists an $M \in \mathbb{N}$ such that

$$\int_{\mathbb{R}} |f - f_M| \, d\lambda < \epsilon.$$

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

$$f_{M,n} = f_M \chi_{[-n,n]}.$$

Note $f_{M,n}$ is Lebesgue measurable for all $n \in \mathbb{N}$ since f_M is Lebesgue measurable. Furthermore $f_{n,M}$ is Lebesgue integrable for all $n \in \mathbb{N}$ since $|f_{M,n}| \leq |f_M|$ for all $n \in \mathbb{N}$ and f_M is Lebesgue integrable. Therefore, since $(f_{M,n})_{n\geq 1}$ converges pointwise to f_M on $\mathbb{R}$, since $|f_{M,n}| \leq |f_M|$ for all $n \in \mathbb{N}$, and since f_M is Lebesgue integrable, the proof of the Dominated Convergence Theorem (Theorem 3.7.1) implies that

$$\lim_{n\to\infty} \int_{\mathbb{R}} |f_M - f_{M,n}| \, d\lambda = 0$$

(see Remark 3.7.2). Therefore, there exists an $N \in \mathbb{N}$ such that

$$\int_{\mathbb{R}} |f_M - f_{M,N}| \, d\lambda < \epsilon.$$

Considering $f_{M,N}$ as a function on $[-N, N]$, Lusin's Theorem (Theorem 2.6.1) implies there exists a continuous function $h : [-N, N] \to \mathbb{C}$ such that

$$\sup\{|h(x)| \mid x \in [-N, N]\} \leq \sup\{|f_{M,N}(x)| \mid x \in [-N, N]\} \leq M$$

and, if

$$B = \{x \in [-N, N] \mid f_{M,N}(x) \neq h(x)\},$$

then $\lambda(B) \leq \frac{\epsilon}{2M+1}$. Therefore

$$\begin{aligned}
\int_{\mathbb{R}} |f_{M,N} - h\chi_{[-N,N]}| \, d\lambda &= \int_B |f_{M,N} - h| \, d\lambda \\
&\leq \int_B |f_{M,N}| + |h| \, d\lambda \\
&\leq \int_B M + M \, d\lambda \\
&\leq 2M\lambda(B) < \epsilon.
\end{aligned}$$

Finally, choose δ such that $0 < \delta < \frac{\epsilon}{2M+1}$ and define $g : \mathbb{R} \to \mathbb{C}$ by

$$g(x) = \begin{cases} h(x) & \text{if } x \in [-N, N] \\ h(N) - \frac{1}{\delta}h(N)(x - N) & \text{if } x \in (N, N + \delta) \\ 0 & \text{if } x \geq N + \delta \\ h(-N) + \frac{1}{\delta}h(N)(x + N) & \text{if } x \in (-N - \delta, -N) \\ 0 & \text{if } x \leq -N - \delta \end{cases}.$$

Then g is a continuous function such that $g(x) = 0$ for all $x \notin [-N-\delta, N+\delta]$, $|g(x)| \leq M$ for all $x \in \mathbb{R}$, and

$$\begin{aligned}
&\int_{\mathbb{R}} |f-g| \, d\lambda \\
&\leq \int_{\mathbb{R}} |f-f_M| + |f_M - f_{M,N}| + |f_{M,N} - h\chi_{[-N,N]}| + |h\chi_{[-N,N]} - g| \, d\lambda \\
&\leq \epsilon + \epsilon + \epsilon + \int_{\mathbb{R}} |h\chi_{[-N,N]} - g| \, d\lambda \\
&= 3\epsilon + \int_{[-N-\delta,-N]\cup[N,N+\delta]} |g| \, d\lambda \\
&\leq 3\epsilon + \int_{[-N-\delta,-N]\cup[N,N+\delta]} M \, d\lambda \\
&= 3\epsilon + 2\delta M \\
&< 3\epsilon + 2\left(\frac{\epsilon}{2M+1}\right) M < 4\epsilon.
\end{aligned}$$

Therefore, since $\epsilon > 0$ was arbitrary, the proof is complete. ∎

Lemma D.4.8. *Let $f : \mathbb{R} \to \mathbb{C}$ be a Lebesgue integrable function and for each $y \in \mathbb{R}$, let $f_y : \mathbb{R} \to \mathbb{C}$ be the function as defined in Proposition 3.4.11; that is $f_y(x) = f(x-y)$ for all $x \in \mathbb{R}$. For all $\epsilon > 0$ there exists a $\delta > 0$ such that if $|y| < \delta$ then*

$$\int_{\mathbb{R}} |f - f_y| \, d\lambda < \epsilon.$$

Proof. Let $\epsilon > 0$ be arbitrary. Since f is Lebesgue integrable, Lemma D.4.7 implies there exists a continuous function $g : \mathbb{R} \to \mathbb{C}$ and a compact set K such that $g(x) = 0$ for all $x \notin K$ and

$$\int_{\mathbb{R}} |f-g| \, d\lambda < \frac{\epsilon}{3}.$$

Hence

$$\int_{\mathbb{R}} |f_y - g_y| \, d\lambda < \frac{\epsilon}{3}$$

for all $y \in \mathbb{R}$ by the translation invariance of the Lebesgue integral.

Since K is compact, K is bounded. Hence there exists an $M > 0$ such that $K \subseteq [-M, M]$. Moreover, since $g(x) = 0$ for all $x \notin K$, it is elementary to verify that g is uniformly continuous on $\mathbb{R}$. Hence there exists a $\delta > 0$ such that if $|y| < \delta$ then

$$|g(x) - g_y(x)| = |g(x) - g(x-y)| < \frac{\epsilon}{3(4M+1)}$$

for all $x \in \mathbb{R}$. Therefore, if $|y| < \delta$ then

$$\begin{aligned}
\int_{\mathbb{R}} |f - f_y| \, d\lambda &\leq \int_{\mathbb{R}} |f - g| + |g - g_y| + |g_y - f_y| \, d\lambda \\
&\leq \frac{\epsilon}{3} + \int_{\mathbb{R}} |g(x) - g(x-y)| \, d\lambda(x) + \frac{\epsilon}{3} \\
&= \frac{2\epsilon}{3} + \int_{[-M,M]\cup(y+[-M,M])} |g(x) - g(x-y)| \, d\lambda(x) \\
&\leq \frac{2\epsilon}{3} + \int_{[-M,M]\cup(y+[-M,M])} \frac{\epsilon}{3(4M+1)} \, d\lambda \\
&\leq \frac{2\epsilon}{3} + \frac{\epsilon}{3(4M+1)} \lambda([-M,M] \cup (y + [-M,M])) \\
&\leq \frac{2\epsilon}{3} + \frac{\epsilon}{3(4M+1)}(4M) < \epsilon.
\end{aligned}$$

Hence the proof is complete. ∎

Theorem D.4.9. *If $f : \mathbb{R} \to \mathbb{C}$ is a Lebesgue integrable function and $g : \mathbb{R} \to \mathbb{C}$ is a bounded Lebesgue measurable function, then $f * g$ is a uniformly continuous function.*

Proof. To see that $f * g$ is continuous, let $\epsilon > 0$. Since g is bounded, there exists an $M > 0$ such that $|g(x)| \leq M$ for all $x \in \mathbb{R}$. By Lemma D.4.8 there exists a $\delta > 0$ such that if $|t| < \delta$ then

$$\int_{\mathbb{R}} |f(y) - f(y-t)| \, d\lambda(y) < \frac{\epsilon}{M}.$$

Therefore, if $x, x_0 \in \mathbb{R}$ are such that $|x - x_0| < \delta$, then

$$\begin{aligned}
&|(f * g)(x) - (f * g)(x_0)| \\
&= \left| \int_{\mathbb{R}} (f(x-y) - f(x_0 - y)) g(y) \, d\lambda(y) \right| \\
&\leq \int_{\mathbb{R}} |f(x-y) - f(x_0-y)||g(y)| \, d\lambda(y) \\
&\leq \int_{\mathbb{R}} |f(x-y) - f(x_0-y)| M \, d\lambda(y) \\
&= M \int_{\mathbb{R}} |f(x+y) - f(x_0+y)| \, d\lambda(y) \qquad \text{by Propositon 3.4.12} \\
&= M \int_{\mathbb{R}} |f(y) - f((x_0 - x) + y)| \, d\lambda(y) \\
&\leq M \frac{\epsilon}{M} = \epsilon.
\end{aligned}$$

Therefore, since $\epsilon > 0$ was arbitrary, $f * g$ is uniformly continuous. ∎

D.5 The Gaussian Approximate Identity

With the above convolution, we can proceed like we did in MATH 3001; we can construct a summability kernel and obtain results about the Fourier transform and convergence. The summability kernel we will use is drastically different than the one from MATH 3001 and actually is related to the following nice function used in statistics.

Definition D.5.1. The function $G : \mathbb{R} \to [0, \infty)$ defined by

$$G(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

for all $x \in \mathbb{R}$ is called the *normalized Gaussian function*.

To begin, we note that the normalized Gaussian function has some nice properties.

Lemma D.5.2. *Let $G : \mathbb{R} \to [0, \infty)$ the normalized Gaussian function. Then*

a) $\int_{\mathbb{R}} G \, d\lambda = 1$, and

b) $\widehat{G}(y) = e^{-\frac{y^2}{2}} = \sqrt{2\pi} G(y)$ for all $y \in \mathbb{R}$.

Proof. The fact that

$$\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, d\lambda(x) = 1$$

can either be proved using polar coordinates and results from multivariate calculus (MATH 2015) or via integrals of holomorphic functions in from complex analysis (MATH 3410). As such, we leave the proof to those courses.

To see that b) is true, recall that

$$\widehat{G}(y) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} e^{-iyx} \, d\lambda(x).$$

Consider the function $F : \mathbb{R}^2 \to \mathbb{C}$ defined by

$$F(x, y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} e^{-iyx}$$

for all $(x, y) \in \mathbb{R}^2$. Since

- for each $y \in \mathbb{R}$, we have $F(x, y)$ is Lebesgue integrable in x,
- fir each $x \in \mathbb{R}$, we we have $F(x, y)$ is differentiable in y with $\frac{\partial F}{\partial y}(x, y) = -ixF(x, y)$, and
- since $\frac{1}{\sqrt{2\pi}}|x|e^{-\frac{x^2}{2}}$ is Lebesgue integrable with $\left|\frac{\partial F}{\partial y}(x, y)\right| \leq \frac{1}{\sqrt{2\pi}}|x|e^{-\frac{x^2}{2}}$ for all $(x, y) \in \mathbb{R}^2$,

Leibniz's Integration Rule (Theorem 4.6.1) implies that $\widehat{G}$ is differentiable with

$$\begin{aligned}\widehat{G}'(y) &= \int_{\mathbb{R}} -ix\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}e^{-iyx}\,d\lambda(x)\\ &= \frac{i}{\sqrt{2\pi}}\int_{\mathbb{R}} -xe^{-\frac{x^2}{2}}e^{-iyx}\,d\lambda(x)\\ &= \frac{i}{\sqrt{2\pi}}\int_{\mathbb{R}} iye^{-\frac{x^2}{2}}e^{-iyx} + \frac{\partial}{\partial x}e^{-\frac{x^2}{2}}e^{-iyx}\,d\lambda(x)\\ &= -y\widehat{G}(y) + \int_{\mathbb{R}}\frac{\partial}{\partial x}e^{-\frac{x^2}{2}}e^{-iyx}\,d\lambda(x)\\ &= -y\widehat{G}(y) + 0\end{aligned}$$

since

$$\begin{aligned}&\int_{\mathbb{R}}\frac{\partial}{\partial x}e^{-\frac{x^2}{2}}e^{-iyx}\,d\lambda(x)\\ &= \lim_{N\to\infty}\int_{[-N,N]}\frac{\partial}{\partial x}e^{-\frac{x^2}{2}}e^{-iyx}\,d\lambda(x) \qquad \text{by the DCT (Theorem 3.7.1)}\\ &= \lim_{N\to\infty}\int_{-N}^{N}\frac{\partial}{\partial x}e^{-\frac{x^2}{2}}e^{-iyx}\,d\lambda(x)\\ &= \lim_{N\to\infty} e^{-\frac{-N^2}{2}}e^{-iyN} - e^{-\frac{-(-N)^2}{2}}e^{-iy(-N)}\\ &= 0.\end{aligned}$$

Hence $\widehat{G}$ satisfies the differential equation $\widehat{G}'(y) = -y\widehat{G}(y)$ with the initial condition

$$\widehat{G}(0) = \int_{\mathbb{R}} G(x)e^{-i(0)x}\,d\lambda(x) = \int_{\mathbb{R}} G\,d\lambda = 1.$$

Thus it follows that $\widehat{G}(y) = e^{-\frac{y^2}{2}}$ for all $y \in \mathbb{R}$. ∎

To obtain our desired summability kernel from the normalized Gaussian function, we observe the following.

Lemma D.5.3. *Let G be normalized Gaussian function. For each $\epsilon > 0$, let $G_\epsilon : \mathbb{R} \to [0,\infty)$ be defined by*

$$G_\epsilon(x) = \frac{1}{\epsilon}G\left(\frac{x}{\epsilon}\right)$$

for all $x \in \mathbb{R}$. Then

a) $\widehat{G_\epsilon}(y) = e^{-\frac{\epsilon^2 y^2}{2}}$ *for all $y \in \mathbb{R}$,*

b) $\int_{\mathbb{R}} G_\epsilon\,d\lambda = 1$, *c) for all* $\epsilon_0 > 0$ *and* $\delta > 0$ *there exists an* $\epsilon' > 0$ *such that* $|G_\epsilon(x)| < \epsilon_0$ *for all* $|x| \geq \delta$ *and all* $0 < \epsilon \leq \epsilon'$*, and*

d) $\lim_{\epsilon\to 0^+} \int_{\mathbb{R}\setminus[-\delta,\delta]} G_\epsilon \, d\lambda = 0$ *for all* $\delta > 0$*.*

Proof. Note part a) follows from Proposition D.2.4 since $\widehat{G}(y) = e^{-\frac{y^2}{2}}$ for all $y \in \mathbb{R}$ by Lemma D.5.2.

To see that b) is true, we note by Proposition 3.4.13 that

$$\begin{aligned}\int_{\mathbb{R}} G_\epsilon \, d\lambda &= \frac{1}{\epsilon}\int_{\mathbb{R}} G\left(\frac{x}{\epsilon}\right) \, d\lambda(x) \\ &= \frac{1}{\epsilon}\epsilon \int_{\mathbb{R}} G\left(x\right) \, d\lambda(x) \\ &= 1.\end{aligned}$$

To see that c) is true, let $\epsilon_0 > 0$ and $\delta > 0$. Since each G_ϵ is non-negative, decreasing on $(0, \infty)$, and $G(x) = G(-x)$ for all $x \in \mathbb{R}$, it suffices to prove that there exists an $\epsilon' > 0$ such that $|G_\epsilon(\delta)| < \epsilon_0$ for all $0 < \epsilon \leq \epsilon'$. Since

$$G_\epsilon(\delta) = \frac{1}{\sqrt{2\pi}\epsilon} e^{-\frac{\delta^2}{2\epsilon^2}}$$

for all $\epsilon \in (0, \infty)$, a standard L'Höpital's rule argument shows

$$\lim_{x\to\infty} \frac{x}{\sqrt{2\pi} e^{\frac{\delta^2 x^2}{2}}} = 0$$

so

$$\lim_{\epsilon\to 0^+} G_\epsilon(\delta) = \lim_{\epsilon\to 0^+} \frac{1}{\sqrt{2\pi}\epsilon} e^{-\frac{\delta^2}{2\epsilon^2}} = 0.$$

Hence the result follows.

To see that d) is true, fix $\delta > 0$. To see that

$$\lim_{\epsilon\to 0^+} \int_{\mathbb{R}\setminus[-\delta,\delta]} G_\epsilon \, d\lambda = 0,$$

let $\epsilon_0 > 0$.

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

$$f_n = G\chi_{[-n,n]}.$$

Note f_n is Lebesgue measurable for all $n \in \mathbb{N}$ since G is Lebesgue measurable. Furthermore f_n is Lebesgue integrable for all $n \in \mathbb{N}$ since $|f_n| \leq |G|$ for all $n \in \mathbb{N}$ and G is Lebesgue integrable. Therefore, since $(f_n)_{n\geq 1}$ converges pointwise to G on $\mathbb{R}$, since $|f_n| \leq |G|$ for all $n \in \mathbb{N}$, and since G is Lebesgue integrable, the proof of the Dominated Convergence Theorem (Theorem 3.7.1) implies that

$$\lim_{n\to\infty} \int_{\mathbb{R}} |G - f_n| \, d\lambda = 0$$

(see Remark 3.7.2). Therefore, there exists an $N \in \mathbb{N}$ such that

$$\int_{(-\infty,-N]\cup[N,\infty)} G\,d\lambda = \int_{\mathbb{R}} |G - f_n|\,d\lambda < \epsilon_0.$$

Note for all $0 < \epsilon < \frac{\delta}{N}$ that $\frac{\delta}{\epsilon} > N$ so

$$\begin{aligned} 0 \leq \int_{\mathbb{R}\setminus[-\delta,\delta]} G_\epsilon\,d\lambda &= \int_{\mathbb{R}} \frac{1}{\epsilon} G\left(\frac{x}{\epsilon}\right) \chi_{\mathbb{R}\setminus[-\delta,\delta]}(x)\,d\lambda(x) \\ &= \int_{\mathbb{R}} G(x)\,\chi_{\mathbb{R}\setminus[-\delta,\delta]}(\epsilon x)\,d\lambda(x) \qquad \text{by Proposition 3.4.13} \\ &= \int_{\mathbb{R}} G(x)\,\chi_{\mathbb{R}\setminus[-\frac{\delta}{\epsilon},\frac{\delta}{\epsilon}]}(x)\,d\lambda(x) \\ &= \int_{\mathbb{R}\setminus[-\frac{\delta}{\epsilon},\frac{\delta}{\epsilon}]} G(x)\,d\lambda(x) \\ &\leq \int_{\mathbb{R}\setminus[-N,N]} G(x)\,d\lambda(x) \qquad \text{as } G \geq 0 \\ &< \epsilon_0. \end{aligned}$$

Therefore, since $\epsilon > 0$ was arbitrary, the proof is complete. ∎

Consequently, by a very similar proof to one used in MATH 3001, we can recover functions by the convolution of the function against the functions from Lemma D.5.3. We begin with the following which says we get close 'in the integral' for any Lebesgue integrable function.

Theorem D.5.4. *With G_ϵ as in Lemma D.5.3, if $f : \mathbb{R} \to \mathbb{C}$ is Lebesgue integrable, then*

$$\lim_{\epsilon\to 0^+} \int_{\mathbb{R}} |f - f * G_\epsilon|\,d\lambda = 0.$$

Proof. Let $f : \mathbb{R} \to \mathbb{C}$ be Lebesgue integrable and fix $\epsilon_0 > 0$. Since f is Lebesgue integrable,

$$M = \int_{\mathbb{R}} |f|\,d\lambda < \infty.$$

Moreover Lemma D.4.8 implies there exists a $\delta > 0$ such that if $|y| \leq \delta$ then

$$\int_{\mathbb{R}} |f - f_y|\,d\lambda < \epsilon_0$$

where $f_y : \mathbb{R} \to \mathbb{C}$ is defined by $f_y(x) = f(x - y)$ for all $x \in \mathbb{R}$.

By Lemma D.5.3, there exists an $\epsilon' > 0$ such that if $0 < \epsilon < \epsilon'$, then

$$\int_{\mathbb{R}\setminus[-\delta,\delta]} G_\epsilon\,d\lambda < \frac{\epsilon_0}{M+1}.$$

Hence for all ϵ such that $0 < \epsilon < \epsilon'$, we have that

$$\begin{aligned}
&\int_{\mathbb{R}} |f - f * G_\epsilon| \, d\lambda \\
&= \int_{\mathbb{R}} \left| f(x) - \int_{\mathbb{R}} f(x-y) G_\epsilon(y) \, d\lambda(y) \right| \, d\lambda(x) \\
&= \int_{\mathbb{R}} \left| \int_{\mathbb{R}} f(x) G_\epsilon(y) \, d\lambda(y) - \int_{\mathbb{R}} f(x-y) G_\epsilon(y) \, d\lambda(y) \right| \, d\lambda(x) \\
&= \int_{\mathbb{R}} \left| \int_{\mathbb{R}} (f(x) - f_y(x)) \, G_\epsilon(y) \, d\lambda(y) \right| \, d\lambda(x) \\
&\leq \int_{\mathbb{R}} \int_{\mathbb{R}} |f(x) - f_y(x)| \, G_\epsilon(y) \, d\lambda(y) \, d\lambda(x) \\
&= \int_{\mathbb{R}} \int_{[-\delta,\delta]} |f(x) - f_y(x)| \, G_\epsilon(y) \, d\lambda(y) \, d\lambda(x) \\
&\qquad + \int_{\mathbb{R}} \int_{\mathbb{R} \setminus [-\delta,\delta]} |f(x) - f_y(x)| \, G_\epsilon(y) \, d\lambda(y) \, d\lambda(x) \\
&= \int_{[-\delta,\delta]} \int_{\mathbb{R}} |f(x) - f_y(x)| \, G_\epsilon(y) \, d\lambda(x) \, d\lambda(y) \\
&\qquad + \int_{\mathbb{R} \setminus [-\delta,\delta]} \int_{\mathbb{R}} |f(x) - f_y(x)| \, G_\epsilon(y) \, d\lambda(x) \, d\lambda(y) \\
&\leq \int_{[-\delta,\delta]} \epsilon_0 G_\epsilon(y) \, d\lambda(y) + \int_{\mathbb{R} \setminus [-\delta,\delta]} 2M G_\epsilon(y) \, d\lambda(y) \\
&= \int_{\mathbb{R}} \epsilon_0 G_\epsilon(y) \, d\lambda(y) + \int_{\mathbb{R} \setminus [-\delta,\delta]} 2M G_\epsilon(y) \, d\lambda(y) \\
&\leq \epsilon_0 + 2M \left(\frac{\epsilon_0}{M+1} \right) \leq 3\epsilon_0.
\end{aligned}$$

Therefore, since ϵ_0 was arbitrary, the result follows. ∎

To prove we can recover continuous Lebesgue integrable functions by convolution against the functions from Lemma D.5.3, we first restrict our attention to bounded functions.

Lemma D.5.5. *With G_ϵ as in Lemma D.5.3, if $f : \mathbb{R} \to \mathbb{C}$ is Lebesgue integrable, bounded, and continuous at a point $x_0 \in \mathbb{R}$, then*

$$\lim_{\epsilon \to 0^+} (f * G_\epsilon)(x_0) = f(x_0).$$

Moreover, if f is uniformly continuous, then $(f * G_\epsilon)_{\epsilon > 0}$ converges to f uniformly on $\mathbb{R}$.

Proof. Since f is bounded, there exists a $K > 0$ such that

$$|f(x)| \leq K$$

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

First assume f is continuous at x_0. To see that

$$\lim_{\epsilon\to 0^+}(f * G_\epsilon)(x_0) = f(x_0),$$

let $\epsilon_0 > 0$ be arbitrary. Since f is continuous at x_0, there exists a $\delta > 0$ such that if $y \in [-\delta, +\delta]$, then

$$|f(x_0) - f(x_0 - y)| < \frac{\epsilon}{2}.$$

By Lemma D.5.3, there exists an $\epsilon' > 0$ such that if $0 < \epsilon < \epsilon'$, then

$$\int_{\mathbb{R}\setminus[-\delta,\delta]} G_\epsilon \, d\lambda < \frac{\epsilon_0}{4K+1}.$$

Hence for all ϵ such that $0 < \epsilon < \epsilon'$, we have that

$$\begin{aligned}
&|(f * G_\epsilon)(x_0) - f(x_0)| \\
&= \left| \int_{\mathbb{R}} f(x_0 - y) G_\epsilon(y) \, d\lambda(y) - f(x_0) \right| \\
&= \left| \int_{\mathbb{R}} f(x_0 - y) G_\epsilon(y) \, d\lambda(y) - f(x_0) \int_{\mathbb{R}} G_\epsilon(y) \, d\lambda(y) \right| \\
&= \left| \int_{\mathbb{R}} (f(x_0 - y) - f(x_0)) G_\epsilon(y) \, d\lambda(y) \right| \\
&\leq \int_{\mathbb{R}} |f(x_0 - y) - f(x_0)| G_\epsilon(y) \, d\lambda(y) \\
&= \int_{[-\delta,\delta]} |f(x_0 - y) - f(x_0)| G_\epsilon(y) \, d\lambda(y) \\
&\qquad + \int_{\mathbb{R}\setminus[-\delta,\delta]} |f(x_0 - y) - f(x_0)| G_\epsilon(y) \, d\lambda(y) \\
&\leq \int_{[-\delta,\delta]} \frac{\epsilon_0}{2} G_\epsilon(y) \, d\lambda(y) + \int_{\mathbb{R}\setminus[-\delta,\delta]} 2K G_\epsilon(y) \, d\lambda(y) \\
&\leq \frac{\epsilon_0}{2} \int_{\mathbb{R}} G_\epsilon(y) \, d\lambda(y) + 2K \int_{\mathbb{R}\setminus[-\delta,\delta]} G_\epsilon(y) \, d\lambda(y) \\
&\leq \frac{\epsilon_0}{2} + 2K \left(\frac{\epsilon_0}{4K+1} \right) < \epsilon_0.
\end{aligned}$$

Therefore, since ϵ_0 was arbitrary, the proof of the first part of the theorem is complete.

To see the second part of the proof, we simply note that one can choose δ to work simultaneously for all $x_0 \in \mathbb{R}$ and thus the proof is complete. ∎

Theorem D.5.6. *With G_ϵ as in Lemma D.5.3, if $f : \mathbb{R} \to \mathbb{C}$ is Lebesgue integrable and continuous at a point $x_0 \in \mathbb{R}$, then*

$$\lim_{\epsilon\to 0^+}(f * G_\epsilon)(x_0) = f(x_0).$$

Proof. Let $f : \mathbb{R} \to \mathbb{C}$ be Lebesgue integrable and continuous at a point $x_0 \in \mathbb{R}$. To see the desired limit, let $\epsilon_0 > 0$. For each $m \in \mathbb{N}$, let

$$A_m = \{x \in \mathbb{R} \mid |f(x)| \leq m\}.$$

Note $A_m \in \mathcal{M}(\mathbb{R})$ since f is Lebesgue measurable. Moreover, $A_m \subseteq A_{m+1}$ for all $m \in \mathbb{N}$ and $\bigcup_{m=1}^{\infty} A_m = \mathbb{R}$ by construction.

For each $m \in \mathbb{N}$, let $f_m = f\chi_{A_m}$. Note f_m is Lebesgue measurable for all $m \in \mathbb{N}$ since f is Lebesgue measurable. Moreover, f_m is Lebesgue integrable for all $m \in \mathbb{N}$ since $|f_m| \leq |f|$ and $(f_m)_{m\geq 1}$ converges pointwise to f on $\mathbb{R}$ since $A_m \subseteq A_{m+1}$ for all $m \in \mathbb{N}$ and $\bigcup_{m=1}^{\infty} A_m = \mathbb{R}$. Therefore, since $|f_m| \leq |f|$ for all $m \in \mathbb{N}$, and since f is Lebesgue integrable, the proof of the Dominated Convergence Theorem (Theorem 3.7.1) implies that

$$\lim_{m\to\infty} \int_{\mathbb{R}} |f - f_m| \, d\lambda = 0$$

(see Remark 3.7.2). Therefore, there exists an $M \in \mathbb{N}$ such that $|f(x_0)| \leq M$ and

$$\int_{\mathbb{R}} |f - f_M| \, d\lambda < \epsilon_0.$$

Since f is continuous at x_0 and since $f_M(x) = \max\{f(x), M\}$ for all $x \in \mathbb{R}$, we see that f_M is continuous at x_0. Moreover, since $|f(x_0)| \leq M$, we obtain that $x_0 \in A_M$ so $f_M(x_0) = f(x_0)$. Therefore, since $f_M(x_0) = f(x_0)$ and since f_M and f are continuous at x_0, there exists a $\delta > 0$ such that if $|y| \leq \delta$ then

$$|f(x_0 - y) - f_M(x_0 - y)| < \epsilon_0.$$

By Lemma D.5.3 there exists a $\epsilon_1 > 0$ such that $|G_\epsilon(x)| < \epsilon_0$ for all $|x| \geq \delta$ and $0 < \epsilon < \epsilon_1$. Moreover, Lemma D.5.5 implies that

$$\lim_{\epsilon\to 0^+} (f_M * G_\epsilon)(x_0) = f_M(x_0) = f(x_0)$$

so there exists an $\epsilon_2 > 0$ such that if $0 < \epsilon < \epsilon_2$ then

$$|(f_M * G_\epsilon)(x_0) - f(x_0)| < \epsilon_0.$$

Finally, notice for all $0 < \epsilon < \min\{\epsilon_1, \epsilon_2\}$ that

$$\begin{aligned}
&|(f * G_\epsilon)(x_0) - (f_M * G_\epsilon)(x_0)| \\
&= \left| \int_{\mathbb{R}} f(x_0 - y) G_\epsilon(y)\, d\lambda(y) - \int_{\mathbb{R}} f_M(x_0 - y) G_\epsilon(y)\, d\lambda(y) \right| \\
&= \left| \int_{\mathbb{R}} \left(f(x_0 - y) - f_M(x_0 - y) \right) G_\epsilon(y)\, d\lambda(y) \right| \\
&\leq \int_{\mathbb{R}} |f(x_0 - y) - f_M(x_0 - y)| G_\epsilon(y)\, d\lambda(y) \\
&= \int_{[-\delta,\delta]} |f(x_0 - y) - f_M(x_0 - y)| G_\epsilon(y)\, d\lambda(y) \\
&\qquad + \int_{\mathbb{R}\setminus[-\delta,\delta]} |f(x_0 - y) - f_M(x_0 - y)| G_\epsilon(y)\, d\lambda(y) \\
&\leq \int_{[-\delta,\delta]} \epsilon_0 G_\epsilon(y)\, d\lambda(y) \\
&\qquad + \int_{\mathbb{R}\setminus[-\delta,\delta]} |f(x_0 - y) - f_M(x_0 - y)| \epsilon_0\, d\lambda(y) \\
&\leq \int_{\mathbb{R}} \epsilon_0 G_\epsilon(y)\, d\lambda(y) + \int_{\mathbb{R}} |f(x_0 - y) - f_M(x_0 - y)| \epsilon_0\, d\lambda(y) \\
&\leq \epsilon_0 + \int_{\mathbb{R}} |f(y) - f_M(y)| \epsilon_0\, d\lambda(y) \\
&\leq \epsilon_0 + \epsilon_0^2.
\end{aligned}$$

Therefore, since ϵ_0 was arbitrary, the result follows. ■

D.6 Inversion of the Fourier Transform

Using the Gaussian approximate identity, we can demonstrate that continuous Lebesgue integrable functions can be recovered from their Fourier transforms. To begin, we require the following lemma.

Lemma D.6.1. *Let $f, g : \mathbb{R} \to \mathbb{C}$ be Lebesgue integrable. Then $\widehat{f} g$ and $f \widehat{g}$ are Lebesgue integrable and*

$$\int_{\mathbb{R}} \widehat{f} g \, d\lambda = \int_{\mathbb{R}} f \widehat{g} \, d\lambda.$$

Proof. Recall $\widehat{f}$ is continuous and bounded by Theorem D.2.3. Thus $\widehat{f} g$ is Lebesgue measurable and there exists an $M > 0$ such that $|\widehat{f}(y)| \leq M$ for all $y \in \mathbb{R}$. Therefore, since g is Lebesgue integrable,

$$\int_{\mathbb{R}} |\widehat{f} g| \, d\lambda \leq \int_{\mathbb{R}} M |g| \, d\lambda < \infty.$$

Hence $\widehat{f} g$ is Lebesgue integrable. Similarly, $f \widehat{g}$ is Lebesgue integrable.

To see the second claim, note if $h : \mathbb{R}^2 \to \mathbb{C}$ is defined by $h(x, y) = g(x)f(y)$, then h is 2-dimensional Lebesgue measurable. Moreover, by Tonelli's Theorem (Theorem 5.2.2)

$$\begin{aligned}\int_{\mathbb{R}^2} |h|\, d\lambda_2 &= \int_{\mathbb{R}}\int_{\mathbb{R}} |g(x)||f(y)|\, d\lambda(x)\, d\lambda(y) \\ &= \left(\int_{\mathbb{R}} |g(x)|\, d\lambda(x)\right)\left(\int_{\mathbb{R}} |f(y)|\, d\lambda(y)\right) < \infty\end{aligned}$$

Thus h is 2-dimensional Lebesgue integrable. Therefore, by Fubini's Theorem (Theorem 5.2.1), we have that

$$\begin{aligned}\int_{\mathbb{R}} \widehat{f}(x)g(x)\, d\lambda(x) &= \int_{\mathbb{R}}\left(\int_{\mathbb{R}} f(y)e^{-iyx}\, d\lambda(y)\right) g(x)\, d\lambda(x) \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} f(y)g(x)e^{-iyx}\, d\lambda(y)\, d\lambda(x) \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} f(y)g(x)e^{-iyx}\, d\lambda(x)\, d\lambda(y) \\ &= \int_{\mathbb{R}} f(y)\left(\int_{\mathbb{R}} g(x)e^{-iyx}\, d\lambda(x)\right) d\lambda(y) \\ &= \int_{\mathbb{R}} f(y)\widehat{g}(y)\, d\lambda(y)\end{aligned}$$

as desired. ∎

Theorem D.6.2. *If $f : \mathbb{R} \to \mathbb{C}$ is continuous and Lebesgue integrable, and $\widehat{f}$ is Lebesgue integrable, then*

$$f(x) = \frac{1}{2\pi}\int_{\mathbb{R}} \widehat{f}(y)e^{iyx}\, d\lambda(y)$$

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

Proof. Fix $x_0 \in \mathbb{R}$. For all $\epsilon > 0$, let G_ϵ be as in Lemma D.5.3. Hence Theorem D.5.6 implies that

$$f(x_0) = \lim_{\epsilon \to 0^+} (f * G_\epsilon)(x_0).$$

For all $\epsilon > 0$ define $h_\epsilon : \mathbb{R} \to \mathbb{C}$ by

$$h_\epsilon(x) = \frac{1}{\sqrt{2\pi}} e^{ix_0x} G(\epsilon x)$$

for all $x \in \mathbb{R}$ where G is the normalized Gaussian function. Therefore h is Lebesgue integrable. Moreover, by Proposition D.2.4 and Lemma D.5.2, we have that

$$\widehat{h_\epsilon}(y) = \frac{1}{\sqrt{2\pi}\epsilon}\widehat{G}\left(\frac{y - x_0}{\epsilon}\right) = \frac{1}{\epsilon}G\left(\frac{x_0 - y}{\epsilon}\right) = G_\epsilon(x_0 - y).$$

Therefore

$$\begin{aligned}(f * G_\epsilon)(x_0) &= (G_\epsilon * f)(x_0) \\ &= \int_{\mathbb{R}} G_\epsilon(x_0 - y) f(y)\, d\lambda(y) \\ &= \int_{\mathbb{R}} \widehat{h_\epsilon}(y) f(y)\, d\lambda(y) \\ &= \int_{\mathbb{R}} h_\epsilon(y) \widehat{f}(y)\, d\lambda(y) \qquad \text{by Lemma D.6.1.}\end{aligned}$$

Thus

$$f(x_0) = \lim_{\epsilon \to 0^+} \int_{\mathbb{R}} h_\epsilon(y) \widehat{f}(y)\, d\lambda(y).$$

Notice for all $y \in \mathbb{R}$ that

$$\lim_{\epsilon \to 0^+} h_\epsilon(y) = \frac{1}{\sqrt{2\pi}} e^{ix_0 y} G(0) = \frac{1}{2\pi} e^{ix_0 y}.$$

Moreover

$$\left| h_\epsilon(y) \widehat{f}(y) \right| \leq \left| \widehat{f}(y) \right|$$

for all $y \in \mathbb{R}$. Therefore, since $\widehat{f}$ is Lebesgue integrable, we have by the Dominated Convergence Theorem (Theorem 3.7.1) (see the proof of Leibniz's Rule (Theorem 4.6.1) for how the Dominated Convergence Theorem generalizes from sequences to a continuum) that

$$\begin{aligned}f(x_0) &= \lim_{\epsilon \to 0^+} \int_{\mathbb{R}} h_\epsilon(y) \widehat{f}(y)\, d\lambda(y) \\ &= \int_{\mathbb{R}} \frac{1}{2\pi} e^{ix_0 y} \widehat{f}(y)\, d\lambda(y) \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} \widehat{f}(y) e^{iyx_0}\, d\lambda(y)\end{aligned}$$

as desired. ∎

Using Theorem D.6.2, we obtain one more useful fact relating the Fourier transform and the original function.

Corollary D.6.3 (Parseval's Theorem). *If $f : \mathbb{R} \to \mathbb{C}$ is continuous and Lebesgue integrable, and $\widehat{f}$ is Lebesgue integrable, then*

$$\frac{1}{2\pi} \int_{\mathbb{R}} |\widehat{f}|^2\, d\lambda = \int_{\mathbb{R}} |f|^2\, d\lambda.$$

Proof. Let $h : \mathbb{R}^2 \to \mathbb{C}$ be defined by $h(x, y) = \overline{f(x)} \widehat{f}(y) e^{-iyx}$. Thus h is 2-dimensional Lebesgue measurable. Moreover, by Tonelli's Theorem (Theorem 5.2.2)

$$\begin{aligned}\int_{\mathbb{R}^2} |h|\, d\lambda_2 &= \int_{\mathbb{R}} \int_{\mathbb{R}} \left| \overline{f(x)} \widehat{f}(y) e^{-iyx} \right| d\lambda(x)\, d\lambda(y) \\ &= \left(\int_{\mathbb{R}} |f(x)|\, d\lambda(x) \right) \left(\int_{\mathbb{R}} \left| \widehat{f}(y) \right| d\lambda(y) \right) < \infty\end{aligned}$$

Thus h is 2-dimensional Lebesgue integrable. Therefore, by Fubini's Theorem (Theorem 5.2.1), we have that

$$\begin{aligned}
\frac{1}{2\pi}\int_{\mathbb{R}} |\widehat{f}(y)|^2 \, d\lambda(y) &= \frac{1}{2\pi}\int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x)e^{-iyx}\, d\lambda(x)\right) \overline{\widehat{f}(y)}\, d\lambda(y) \\
&= \frac{1}{2\pi}\int_{\mathbb{R}}\int_{\mathbb{R}} f(x)\overline{\widehat{f}(y)}e^{-iyx}\, d\lambda(x)\, d\lambda(y) \\
&= \overline{\frac{1}{2\pi}\int_{\mathbb{R}}\int_{\mathbb{R}} \overline{f(x)}\widehat{f}(y)e^{iyx}\, d\lambda(x)\, d\lambda(y)} \\
&= \overline{\frac{1}{2\pi}\int_{\mathbb{R}}\int_{\mathbb{R}} \overline{f(x)}\widehat{f}(y)e^{iyx}\, d\lambda(y)\, d\lambda(x)} \\
&= \overline{\int_{\mathbb{R}} \overline{f(x)} \left(\frac{1}{2\pi}\int_{\mathbb{R}} \widehat{f}(y)e^{iyx}\, d\lambda(y)\right) d\lambda(x)} \\
&= \overline{\int_{\mathbb{R}} \overline{f(x)}f(x)\, d\lambda(x)} \\
&= \int_{\mathbb{R}} f(x)\overline{f(x)}\, d\lambda(x) \\
&= \int_{\mathbb{R}} |f|^2 \, d\lambda
\end{aligned}$$

as desired. ∎

Index