MATH 6280
Measure Theory

Paul Skoufranis

April 28, 2025

Preface:

If you come across any typos, errors, omissions, or unclear explanations in these lecture notes for MATH 6280 (Graduate Measure Theory), please feel free to contact me so that these notes may continually be improved.

Contents

Chapter 1

Measure Spaces

As per its title, this course is dedicated to the study of the theory of measures. So what sort of course would this be if we did not define the main object of study in the first chapter? After defining the notion of a measure, we will examine several examples and properties of measures that immediately follow from definitions. We will then turn to constructing measures from various notions of length and extending measure-like functions to actual measures. This will lead us to looking at measures with important analytical properties including the Lebesgue-Stieltjes measures and metric outer measures. By using metric outer measures, we will obtain the notion of Hausdorff dimension for subsets of metric spaces.

1.1 Measure Spaces

Before we can define the notion of a measure, we must describe collections of sets which are valid for the domain of a measure. One may think of these collections as all sets which have a valid length (or measure) or as all events which have a well-defined probability. After all, it is not always true that we can assign every event in the universe a probability at once.

Definition 1.1.1. Let X be a non-empty set. A *σ-algebra on X* is a subset $\mathcal{A} \subseteq \mathcal{P}(X)$ (the power set of X) such that

1. $\emptyset \in \mathcal{A}$ and $X \in \mathcal{A}$ (that is, we can measure the empty event and the full event),
2. if $A \in \mathcal{A}$ then $A^c = X \setminus A \in \mathcal{A}$ (that is, we can measure the complement of an event), and
3. if $\{A_n\}_{n=1}^\infty \subseteq \mathcal{A}$, then $\bigcup_{n=1}^\infty A_n \in \mathcal{A}$ (that is, we can measure the union of a countable collection of events).

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

Remark 1.1.2. One may ask why we only ask for countable unions of measurable sets to be measurable. One answer for this comes with the definition of a measure in that we want to have additivity over disjoint unions and adding over an uncountable set only works if only a countable number of elements are non-zero. Another reason is that restricting to countable collections is quite powerful as we will see in this course.

Remark 1.1.3. One may also ask why we have not required that the intersection of a countable collection of measurable sets is measurable. The reason for this is that countable intersections come for free. Indeed if $(X, \mathcal{A})$ is a measurable space and $\{A_n\}_{n=1}^\infty \subseteq \mathcal{A}$ then

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

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

Of course, there are some trivial examples of measurable spaces.

Example 1.1.4. Let X be a non-empty set. Then $(X, \mathcal{P}(X))$ is a measurable space and $(X, \{\emptyset, X\})$ is a measurable space.

Of course there are some more complicated examples of measurable spaces.

Example 1.1.5. Let X be a non-empty set and let

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

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

Moreover, if one has a collection of σ-algebras on a set X, there are ways of constructing new σ-algebras. In particular, it is elementary to verify the following using set properties and Definition 1.1.1.

Lemma 1.1.6. *Let X be a non-empty set and let $\{\mathcal{A}_\alpha \mid \alpha \in I\}$ be a collection of σ-algebras of X. Then*

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

is a σ-algebra of X.

Remark 1.1.7. Using Lemma 1.1.6 we can construct the smallest σ-algebra containing a collection of subsets. Indeed let X be a non-empty set and let $A \subseteq \mathcal{P}(X)$. Define

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

Clearly $\mathcal{P}(X) \in I$ so I is non-empty. Hence Lemma 1.1.6 implies that

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

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

Definition 1.1.8. Let $(\mathcal{X}, d)$ be a metric space. The σ-algebra generated by the open subsets of $\mathcal{X}$ is called the *Borel σ-algebra* and is denoted $\mathfrak{B}(\mathcal{X})$. In particular, $\mathfrak{B}(\mathcal{X})$ is also the σ-algebra generated by the closed subsets of $\mathcal{X}$ since open and closed sets are complements of each other and as σ-algebras are closed under complements. Elements of $\mathfrak{B}(\mathcal{X})$ are called *Borel sets*.

Remark 1.1.9. In terms of the Borel subsets of $\mathbb{R}$, the sets

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

all can be shown to generate $\mathfrak{B}(\mathbb{R})$ via unions, intersections, and complements (show that $\mathfrak{B}(\mathbb{R})$ contains each of these sets and any σ-algebra containing one of these sets contains all open intervals and thus all open sets by the fact that every open set is a countable union of open intervals; see Proposition C.2.11). We note it is possible to show that $|\mathfrak{B}(\mathbb{R})| = |\mathbb{R}| < |\mathcal{P}(\mathbb{R})|$ (that is, the cardinality of the Borel subsets of $\mathbb{R}$ is strictly less than the cardinality of the power set of $\mathbb{R}$ so not every subset of $\mathbb{R}$ is Borel). Said proof requires the use of transfinite induction.

With σ-algebras, we may now define the central object of study in this course.

Definition 1.1.10. Let $(X, \mathcal{A})$ be a measurable space. A *(countably additive, positive) measure* on $(X, \mathcal{A})$ is a function $\mu : \mathcal{A} \to [0, \infty]$ such that

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

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

(where the sum is infinite if one of the elements is ∞ or if the sum diverges).

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*.

Remark 1.1.11. 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 countable additivity with $A_k = \emptyset$ for all $k > n$.

Before we get too deep into the study of properties of measures, let's examine some common measures which are easy to define.

Example 1.1.12. Let X be a non-empty set and let $x \in X$. The *point-mass measure at x* is the measure δ_x on $(X, \mathcal{P}(X))$ defined by

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

It is elementary to verify that δ_x is a measure.

Example 1.1.13. Let X be a non-empty set. The *counting measure on X* is the measure μ on $(X, \mathcal{P}(X))$ defined by

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

It is elementary to verify that μ is a measure.

Example 1.1.14. A function $\mu : \mathcal{P}(\mathbb{N}) \to [0, \infty]$ is a measure on $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ if and only if there exists a sequence $(a_n)_{n\geq 1}$ of elements of $[0, \infty]$ such that

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

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

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

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

We claim that

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

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

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

we obtain by the properties of a measure that

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

as desired.

Note the measures in Example 1.1.14 can be constructed using Example 1.1.12 and the following technique (which will be of use to us later).

Example 1.1.15. Let $(X, \mathcal{A}, \mu)$ be a measure space, let $\{A_k\}_{k=1}^\infty \subseteq \mathcal{A}$, and let $\{a_k\}_{k=1}^\infty \in [0, \infty]$. Define $\nu : \mathcal{A} \to [0, \infty]$ by

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

for all $A \in \mathcal{A}$ where the sum equates to ∞ if the sum diverges or one of the terms is ∞, and

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

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

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

Hence ν is a measure as desired.

Although we can define many more measures, we turn our attention to properties of measures immediately implied by Definition 1.1.10 and set manipulations. We begin with the following.

Remark 1.1.16. Let $(X, \mathcal{A}, \mu)$ be a measure space and let $E, F \in \mathcal{A}$. Assume $E \subseteq F$. Since $F \setminus E = F \cap E^c \in \mathcal{A}$ 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. 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 1.1.17. 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 1.1.16 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.

Of course if a measure is going to represent the probability of an event occurring, we must dictate the probability of all possible events is one. As such, when discussing probability, we use the following terminology.

Definition 1.1.18. 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, X is called the sample space, elements of $\mathcal{A}$ are called *events*, and, given $A \in \mathcal{A}$, $\mu(A)$ denotes the probability that the event A occurs.

Remark 1.1.19. 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 1.1.17 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.

Of course, when studying probability, one may only have finite additivity instead of countable additivity. As will be seen in Section 1.3, it is not difficult to extend finitely additive measures to countably additive measures, which is far more desirable in our analytic realm.

Of course requiring the measure of the entire space to be one is a specific property of a measure we may wish to study. The following generalizations of probability measures are vital for this course.

Definition 1.1.20. A measure μ on a measurable space $(X, \mathcal{A})$ is said to be

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

In most cases, if one can prove a property for any finite measure, one can extend the result to all σ-finite measures using analytical techniques. This is often done using the following additional partition decompositions of a σ-finite measure space.

Remark 1.1.21. Assume μ is a σ-finite measure on $(X, \mathcal{A})$. Thus there exists a collection $\{A_n\}_{n=1}^\infty \subseteq \mathcal{A}$ such that $X = \bigcup_{n=1}^\infty A_n$ and $\mu(A_n) < \infty$ for all $n \in \mathbb{N}$. Let $B_1 = C_1 = A_1$ and for each $n \geq 2$ let

$$B_n = A_n \setminus \left(\bigcup_{k=1}^{n-1} B_k \right) \qquad \text{and} \qquad C_n = \bigcup_{k=1}^{n} A_k.$$

Then $\{B_n\}_{n=1}^\infty$ are pairwise disjoint elements of $\mathcal{A}$ are such that $X = \bigcup_{n=1}^\infty B_n$ and $\mu(B_n) \leq \mu(A_n) < \infty$ for all $n \in \mathbb{N}$. Similarly $\{C_n\}_{n=1}^\infty$ are elements of $\mathcal{A}$ are such that $X = \bigcup_{n=1}^\infty C_n$, $C_n \subseteq C_{n+1}$ for all $n \in \mathbb{N}$, and $\mu(C_n) < \infty$ for all $n \in \mathbb{N}$. The reason $\mu(C_n) < \infty$ can be seen via the following result as $\sum_{k=1}^n \mu(A_k) < \infty$.

Proposition 1.1.22 (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 1.1.16), 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. ∎

As seen above, being able to replace our measurable sets with disjoint measurable is a very useful technique. In particular, the same idea is helpful in proving the following.

Theorem 1.1.23 (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 1.1.16 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. ∎

Remark 1.1.24. Note that part b) of the Monotone Convergence Theorem (Theorem 1.1.23) fails if the condition $\mu(A_1) < \infty$ is removed. Indeed if μ is the counting measure on $\mathbb{N}$ and $A_n = \mathbb{N} \setminus \{1, 2, \ldots, n\}$ for all $n \in \mathbb{N}$, then $A_{n+1} \subseteq A_n$ for all $n \in \mathbb{N}$, $\mu(A_n) = \infty$ for all $n \in \mathbb{N}$, yet $\bigcap_{n=1}^{\infty} A_n = \emptyset$ so $\mu\left(\bigcap_{n=1}^{\infty} A_n\right) = 0 \neq \infty = \lim_{n\to\infty} \mu(A_n)$.

1.2 The Carathéodory Method

Based on the above notions, it is very natural to ask whether there exists a measure λ on $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$ that emulates the length of a set. In particular, we desire such a measure to have some very natural properties, such as:

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

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

Example 1.2.1. Suppose for the sake of a contradiction that λ is a measure on $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$ with the above two properties. 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.

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, suppose $x \in A_n \cap A_m$ 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 we have $x - y = -1 \in \mathbb{Q}$ so $x \neq y$ and $x \sim y$. Therefore, as A contains exactly one element from each equivalence class, we have obtained a contradiction. Hence $B_{n,1} \cap B_{n,2} = \emptyset$.

To obtain our contradiction, note that

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

This yields our contradiction since $\lambda(A) \in [0,\infty]$ yet no number in $[0,\infty]$ when summed an infinite number of times produces 1. Thus we have obtained a contradiction to the existence of such a λ defined on $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$.

The above example illustrates that $\mathcal{P}(\mathbb{R})$ is too large; that is, there are too many sets in $\mathcal{P}(\mathbb{R})$ to define such a measure in a consistent way. The set A in Example 1.2.1 is one of these sets.

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

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

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

Definition 1.2.2. Let X be a non-empty set. A function $\mu^* : \mathcal{P}(X) \to [0, \infty]$ is said to be an *outer measure* if

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

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

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

Notice that every measure is an outer measure by the results of Section 1.1 whereas an outer measure need not be a measure as it is not necessary that equality occur in the third property of Definition 1.2.2 when the collection $\{A_n\}_{n=1}^\infty$ are pairwise disjoint. Of course, it is a priori possible that every outer measure is automatically a measure. For an example to show this is not the case, we will need to construct some outer measures. The most natural way to do so is the following which attempts to assign certain sets a specific value.

Definition 1.2.3. Let X be a non-empty set, let $\mathcal{F} \subseteq \mathcal{P}(X)$ be a family of subsets of X such that $\emptyset, X \in \mathcal{F}$, and let $\ell : \mathcal{F} \to [0, \infty]$ be any function such that $\ell(\emptyset) = 0$. The *outer measure associated to* ℓ is the function $\mu_\ell^* : \mathcal{P}(X) \to [0, \infty]$ defined by

$$\mu_\ell^*(A) = \inf\left\{ \sum_{n=1}^\infty \ell(A_n) \,\middle|\, \{A_n\}_{n=1}^\infty \subseteq \mathcal{F} \text{ such that } A \subseteq \bigcup_{n=1}^\infty A_n \right\}$$

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

Of course, we should prove that the outer measure associated to ℓ is actually an outer measure!

Proposition 1.2.4. *Let X be a non-empty set, let $\mathcal{F} \subseteq \mathcal{P}(X)$ be a family of subsets of X such that $\emptyset, X \in \mathcal{F}$, and let $\ell : \mathcal{F} \to [0, \infty]$ be any function such that $\ell(\emptyset) = 0$. The outer measure associated to ℓ is an outer measure μ_ℓ^* such that $\mu_\ell^*(A) \leq \ell(A)$ for all $A \subseteq X$.*

Furthermore, if $\nu^* : \mathcal{P}(X) \to [0, \infty]$ is an outer measure such that $\nu^*(A) \leq \ell(A)$ for all $A \subseteq X$, then $\nu^*(A) \leq \mu_\ell^*(A)$ for all $A \in \mathcal{F}$. Hence μ_ℓ^* is the largest outer measure bounded above by ℓ.

Proof. First notice that since $X \in \mathcal{F}$ that the set whose infimum defines $\mu_\ell^*(A)$ is non-empty for all $A \subseteq X$. This fact will be used throughout the proof.

Clearly $\mu_\ell^* : \mathcal{P}(X) \to [0, \infty]$. Furthermore, since $\emptyset \in \mathcal{F}$ and $\ell(\emptyset) = 0$, we clearly see that $\mu_\ell^*(\emptyset) = 0$ as $\{\emptyset\}_{n=1}^\infty$ is a cover of $\emptyset$. Moreover, if $A \subseteq B \subseteq X$, it is easy to see that $\mu_\ell^*(A) \leq \mu_\ell^*(B)$ since the infimum in the definition of $\mu_\ell^*(A)$ is taken over a larger collection of sets than the infimum in the definition of $\mu_\ell^*(B)$.

To verify the final property from Definition 1.2.2 for μ_ℓ^*, let $\{A_n\}_{n=1}^\infty \subseteq \mathcal{P}(X)$ and let $A = \bigcup_{n=1}^\infty A_n$. Fix $\epsilon > 0$. By the definition of μ_ℓ^*, for each $n \in \mathbb{N}$ there exists a collection $\{A_{n,k} \mid k \in \mathbb{N}\} \subseteq \mathcal{F}$ such that $A_n \subseteq \bigcup_{k=1}^\infty A_{n,k}$ and

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

Clearly $\{A_{n,k} \mid n, k \in \mathbb{N}\}$ is a countable subset of $\mathcal{F}$ is such that

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

Hence by the definition of μ_ℓ^*

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

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

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

Hence μ_ℓ^* is an outer measure.

To complete the proof, assume $\nu^* : \mathcal{P}(X) \to [0, \infty]$ is an outer measure such that $\nu^*(A) \leq \ell(A)$ for all $A \subseteq X$. Notice for each $A \subseteq X$ and each collection $\{A_n\}_{n=1}^\infty \subseteq \mathcal{F}$ such that $A \subseteq \bigcup_{n=1}^\infty A_n$ that

$$\nu^*(A) \leq \nu^*\left(\bigcup_{n=1}^\infty A_n\right) \leq \sum_{n=1}^\infty \nu^*(A_n) \leq \sum_{n=1}^\infty \ell(A_n)$$

by the properties of an outer measure and the assumptions on ν^*. Therefore, since $\mu_\ell^*(A)$ is the infimum of $\sum_{n=1}^\infty \ell(A_n)$ over all collections $\{A_n\}_{n=1}^\infty \subseteq \mathcal{F}$ such that $A \subseteq \bigcup_{n=1}^\infty A_n$, we easily see that $\nu^*(A) \leq \mu^*(A)$ for all $A \subseteq X$ as desired. ∎

The outer measure one uses on $\mathbb{R}$ to define length is the following.

Definition 1.2.5. 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*, denoted λ^*, is the outer measure associated to ℓ restricted to the open intervals. In particular $\lambda^* : \mathcal{P}(\mathbb{R}) \to [0, \infty]$ is defined by

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

for all $A \subseteq \mathbb{R}$.

Clearly we can extend the Lebesgue outer measure on $\mathbb{R}$ to measures on $\mathbb{R}^n$ to measure areas and volumes.

Definition 1.2.6. Let $n \in \mathbb{N}$, let

$$\mathcal{F} = \{(a_1, b_1) \times \cdots \times (a_n, b_n) \subseteq \mathbb{R}^n \mid a_k, b_k \in \mathbb{R} \cup \{\pm\infty\}, a_k < b_k\},$$

and define $\ell : \mathcal{F} \to [0, \infty]$ by

$$\ell((a_1, b_1) \times \cdots \times (a_n, b_n)) = \prod_{k=1}^{n} b_k - a_k$$

(where the product is zero if $b_k = a_k$ for some k, and otherwise if $b_k = \infty$ or $a_k = -\infty$ for some k then the product is infinite). The *n-dimensional Lebesgue outer measure*, denoted λ_n^*, is the outer measure on $\mathbb{R}^n$ associated to ℓ.

With the above notion of outer measures, we desire to construct measures from outer measures. To do so, we need to define a σ-algebra of sets for which the restriction of our outer measure produces a measure. These sets are described as follows.

Definition 1.2.7. Let X be a non-empty set and let $\mu^* : \mathcal{P}(X) \to [0, \infty]$ be an outer measure on X. A subset $A \subseteq X$ is said to be *μ^*-measurable* or *outer measurable* if for every $B \in \mathcal{P}(X)$

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

Remark 1.2.8. 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 1.2.9. 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 1.2.10. *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 1.1.3.

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 1.1.23) 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 1.1.10, 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. ∎

Let λ^* be the Lebesgue outer measure from Definition 1.2.5. By Theorem 1.2.10 the collection $\mathcal{M}(\mathbb{R})$ of λ^*-measurable sets is a σ-algebra and $\lambda^*|_{\mathcal{M}(\mathbb{R})}$ is a measure. Since these objects will be the focus for the remainder of our course, we make the following definition.

Definition 1.2.11. The *Lebesgue measure* on $\mathbb{R}$ is the measure $\lambda = \lambda^*|_{\mathcal{M}(\mathbb{R})}$. The elements of $\mathcal{M}(\mathbb{R})$ are called *Lebesgue measurable sets*.

Similarly, we have the n-dimensional analogue of the Lebesgue measure.

Definition 1.2.12. For each $n \in \mathbb{N}$, the *n-dimensional Lebesgue measure* on $\mathbb{R}^n$ is the measure λ_n obtained by restricting λ_n^* to the λ_n^*-measurable subsets of $\mathbb{R}^n$.

One by-product of the Carathéodory Method is that the measures constructed have a specific additional property that we now describe.

Definition 1.2.13. 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 1.2.14. *Let X be a non-empty set, let $\mu^* : \mathcal{P}(X) \to [0, \infty]$ be an outer measure on X, and let $\mathcal{A}$ be the σ-algebra of all outer measurable sets. If $A \in \mathcal{P}(X)$ and $\mu^*(A) = 0$, then $A \in \mathcal{A}$. Hence $(X, \mathcal{A}, \mu^*|_{\mathcal{A}})$ is complete by the monotonicity of μ^*.*

Proof. 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. ∎

Remark 1.2.15. Note by Proposition 1.2.14 that λ is a complete measure.

One may think the Carathéodory Method may not be that useful as it can only construct measures that are complete and thereby might be limited. However, this is not the case as it is always possible to 'complete' a measure rather simply.

Proposition 1.2.16. *Let $(X, \mathcal{A}, \mu)$ be a measure space. Then there exists a complete measure space $(X, \overline{\mathcal{A}}, \overline{\mu})$ such that $\mathcal{A} \subseteq \overline{\mathcal{A}}$ and $\overline{\mu}(A) = \mu(A)$ for all $A \in \mathcal{A}$.*

Proof. Exercise. Alternatively, see Theorem 1.3.7. ∎

1.3 Extending Measures

Although the Carathéodory Method has enabled us to construct the Lebesgue measure and other measures, the process produces startlingly little information about the properties the measure inherits from the length function ℓ used to define the outer measure μ^*_ℓ. In particular, does the Lebesgue measure have the properties described at the beginning of Section 1.2 and is every Borel set Lebesgue measurable? Verifying the desired properties for the Lebesgue measure is not difficult to do directly (and will be demonstrated in Section 1.4). However, we will take a more indirect approach to produce further results and obtain the properties of the Lebesgue measure for free.

In this section, we will analyze how properties of specific length functions are immediately inherited by the measures produced via the Carathéodory Method. In particular, we will see how a ‘finitely additive measure’ can be extended uniquely to an actual measures. Consequently, the results of this section will immediately give almost all properties of the Lebesgue measure we desire in the next section. Alternatively, Section 1.5 analyzes measure similar to the Lebesgue measure and can also be used to produce another indirect approach to obtaining the desired properties.

To begin, we desire to describe functions that are similar to measures on a more general notion than a σ-algebra.

Definition 1.3.1. Let X be a non-empty set. An *algebra on X* is a subset $\mathcal{A} \subseteq \mathcal{P}(X)$ such that

1. $\emptyset \in \mathcal{A}$ and $X \in \mathcal{A}$,
2. if $A \in \mathcal{A}$ then $A^c = X \setminus A \in \mathcal{A}$, and
3. if $A_1, A_2 \subseteq \mathcal{A}$, then $A_1 \cup A_2 \in \mathcal{A}$.

Remark 1.3.2. Notice if $\mathcal{A}$ is an algebra then $\mathcal{A}$ is closed under finite unions by iterating the third property in Definition 1.3.1. Furthermore, if $A_1, A_2 \in \mathcal{A}$ then clearly

$$A_1 \cap A_2 = (A_1^c \cup A_2^c)^c \in \mathcal{A}$$

so $\mathcal{A}$ is also closed under finite intersections.

Example 1.3.3. Clearly if $\mathcal{A}$ is a σ-algebra, then $\mathcal{A}$ is an algebra. However, there are algebras that are not σ-algebras. Indeed for $X = \mathbb{R}$ let

$$\mathcal{F} = \{(a, b] \mid a, b \in [-\infty, \infty), a < b\} \cup \{(a, \infty) \mid a \in \mathbb{R} \cup \{-\infty\}\}$$

and let $\mathcal{A}$ denote the collection of all sets obtained by taking all finite unions of elements of $\mathcal{F}$ (including the empty set). It is not difficult to see that $\mathcal{A}$ is an algebra as the complement of each element of $\mathcal{F}$ is a finite union of elements of $\mathcal{F}$. However $\mathcal{A}$ is not a σ-algebra since $\bigcup_{n=1}^\infty (2n, 2n+1] \notin \mathcal{A}$ yet $(2n, 2n+1] \in \mathcal{A}$ for all $n \in \mathbb{N}$.

Using algebras in place of σ-algebras, we obtain the beginnings of a measure.

Definition 1.3.4. Let X be a non-empty set and let $\mathcal{A}$ be an algebra on X. A *pre-measure* on $\mathcal{A}$ is a function $\mu : \mathcal{A} \to [0, \infty]$ such that

- $\mu(\emptyset) = 0$, and

- if $\{A_n\}_{n=1}^\infty \subseteq \mathcal{A}$ are pairwise disjoint and $\bigcup_{n=1}^\infty A_n \in \mathcal{A}$, then

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

Remark 1.3.5. Notice the difference between a pre-measure on an algebra and a measure on a σ-algebra stems from the fact that if $\mathcal{A}$ is an algebra and $\{A_n\}_{n=1}^\infty \subseteq \mathcal{A}$ then it need not be the case that $\bigcup_{n=1}^\infty A_n \in \mathcal{A}$. Hence, if $\mathcal{A}$ is a σ-algebra, there is no difference between Definitions 1.1.10 and 1.3.4. It is also not difficult to see that a pre-measure shares several properties with measures by repeating some of the proofs demonstrated in Section 1.1. Indeed every pre-measure is finitely additive (by taking the empty set for an infinite number of times) and is monotone. To see the monotonicity of pre-measures, assume $A, B \in \mathcal{A}$ are such that $A \subseteq B$. Then $B \cap A^c \in \mathcal{A}$ as $\mathcal{A}$ is an algebra. Therefore, since A and $B \cap A^c$ are disjoint subsets, we notice for every pre-measure μ on $\mathcal{A}$ that

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

as claimed.

Instead of repeating the theory to show that pre-measures have properties similar to measures, we will demonstrate that every pre-measure can be extended to a measure. We begin with the following lemma.

Lemma 1.3.6. *Let X be a non-empty set, let $\mathcal{A}$ be an algebra on X, and let $\mu : \mathcal{A} \to [0, \infty]$ be a pre-measure on $\mathcal{A}$. Let μ^* be the outer measure associated to μ; that is, $\mu^* : \mathcal{P}(X) \to [0, \infty]$ is defined by*

$$\mu^*(A) = \inf\left\{ \sum_{k=1}^\infty \mu(A_n) \,\middle|\, \{A_n\}_{n=1}^\infty \subseteq \mathcal{A}, A \subseteq \bigcup_{n=1}^\infty A_n \right\}$$

for all $A \subseteq X$. Then μ^* is a an outer measure on X such that $\mu^*(A) = \mu(A)$ for all $A \in \mathcal{A}$.

Proof. Clearly μ^* is an outer measure by Proposition 1.2.4.

For the other claim, let $A \in \mathcal{A}$ be arbitrary. To see that $\mu^*(A) = \mu(A)$, first notice that trivially $\mu^*(A) \leq \mu(A)$ by definition. To see the other inequality, assume $\{A_n\}_{n=1}^\infty \subseteq \mathcal{A}$ is an arbitrary countable collection such that $A \subseteq \bigcup_{n=1}^\infty A_n$. Let $B_1 = A_1 \cap A$ and for each $n \geq 2$ let

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

Since $\mathcal{A}$ is an algebra, we see that $B_n \in \mathcal{A}$ for all $n \in \mathbb{N}$. Furthermore $\{B_n\}_{n=1}^\infty$ are disjoint subsets such that $B_n \subseteq A_n$ for all $n \in \mathbb{N}$ and

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

Therefore, since μ is a pre-measure, we obtain that

$$\mu(A) = \sum_{n=1}^{\infty} \mu(B_n) \leq \sum_{n=1}^{\infty} \mu(A_n)$$

where the inequality follows from the fact that μ is monotone as demonstrated in Remark 1.3.5. Therefore, since $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{A}$ was arbitrary, it follows that $\mu^*(A) = \mu(A)$ for all $A \in \mathcal{A}$. ∎

Theorem 1.3.7 (Carathéodory-Hahn Extension Theorem). *Let X be a non-empty set, let $\mathcal{A}$ be an algebra on X, and let $\mu : \mathcal{A} \to [0, \infty]$ be a pre-measure on $\mathcal{A}$. Let μ^* be the outer measure associated to μ and let $\mathcal{A}^*$ denote the set of all μ^*-measurable sets. Recall from Theorem 1.2.10 that $\mathcal{A}^*$ is a σ-algebra on X and $\overline{\mu} = \mu^*|_{\mathcal{A}^*}$ is a measure on $\mathcal{A}^*$. Then $\mathcal{A} \subseteq \mathcal{A}^*$ and $\overline{\mu}(A) = \mu(A)$ for all $A \in \mathcal{A}$.*

Moreover assume μ is σ-finite in the sense that there exists $\{X_n\}_{n=1}^{\infty} \subseteq \mathcal{A}$ such that $X = \bigcup_{n=1}^{\infty} X_n$ and $\mu(X_n) < \infty$. If $\nu : \mathcal{A}^* \to [0, \infty]$ is a measure such that $\nu(A) = \mu(A)$ for all $A \in \mathcal{A}$, then $\nu = \overline{\mu}$.

Proof. Recall from Lemma 1.3.6 that $\mu^*(A) = \mu(A)$ for all $A \in \mathcal{A}$. Therefore, to complete the first claim, it suffices to show that $\mathcal{A} \subseteq \mathcal{A}^*$.

Let $A \in \mathcal{A}$ be arbitrary. To see that A is μ^*-measurable, let $B \subseteq X$ and let $\epsilon > 0$ be arbitrary. By the definition of μ^* there exists $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{A}$ such that $B \subseteq \bigcup_{n=1}^{\infty} A_n$ and

$$\sum_{n=1}^{\infty} \mu(A_n) \leq \mu^*(B) + \epsilon.$$

Notice that

$$B \cap A \subseteq \bigcup_{n=1}^{\infty} A_n \cap A \qquad \text{and} \qquad B \cap A^c \subseteq \bigcup_{n=1}^{\infty} A_n \cap A^c.$$

Since $\mathcal{A}$ is an algebra, $A_n \cap A, A_n \cap A^c \in \mathcal{A}$ for all $n \in \mathbb{N}$ and therefore, since μ^* is an outer measure,

$$\mu^*(B \cap A) \leq \mu^*\left(\bigcup_{n=1}^{\infty} A_n \cap A\right) \leq \sum_{n=1}^{\infty} \mu^*(A_n \cap A) = \sum_{n=1}^{\infty} \mu(A_n \cap A) \text{ and}$$
$$\mu^*(B \cap A^c) \leq \mu^*\left(\bigcup_{n=1}^{\infty} A_n \cap A^c\right) \leq \sum_{n=1}^{\infty} \mu^*(A_n \cap A^c) = \sum_{n=1}^{\infty} \mu(A_n \cap A^c)$$

Hence

$$\begin{aligned}\mu^*(B\cap A)+\mu^*(B\cap A^c) &\leq \sum_{n=1}^\infty \mu(A_n\cap A)+\sum_{n=1}^\infty \mu(A_n\cap A^c)\\ &= \sum_{n=1}^\infty \mu(A_n\cap A)+\mu(A_n\cap A^c)\\ &= \sum_{n=1}^\infty \mu(A_n)\\ &\leq \mu^*(B)+\epsilon.\end{aligned}$$

where $\mu(A_n\cap A)+\mu(A_n\cap A^c)=\mu(A_n)$ follows from the fact that μ is a pre-measure and $A_n\cap A$ and $A_n\cap A^c$ are disjoint sets. Therefore, since $\epsilon>0$ was arbitrary, we obtain that

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

Hence, since $B\subseteq X$ was arbitrary, A is μ^*-measurable as desired.

For the uniqueness, assume there exists $\{X_n\}_{n=1}^\infty\subseteq \mathcal{A}$ such that $X=\bigcup_{n=1}^\infty X_n$ and $\mu(X_n)<\infty$, and $\nu:\mathcal{A}^*\to[0,\infty]$ is a measure such that $\nu(A)=\mu(A)$ for all $A\in\mathcal{A}$. Notice if $Y_n=\bigcup_{k=1}^n X_k$ for all $n\in\mathbb{N}$, then $\bigcup_{n=1}^\infty Y_n=X$, $Y_n\in\mathcal{A}$ for all $n\in\mathbb{N}$, $Y_n\subseteq Y_{n+1}$ for all $n\in\mathbb{N}$, and

$$\mu(Y_n)=\overline{\mu}(Y_n)=\overline{\mu}\left(\bigcup_{k=1}^n X_k\right)\leq\sum_{k=1}^n\overline{\mu}(X_k)=\sum_{k=1}^n\mu(X_k)<\infty.$$

To see that $\nu(B)=\overline{\mu}(B)$ for all $B\in\mathcal{A}^*$, let $B\in\mathcal{A}^*$ be arbitrary. Then for every k and collection $\{A_n\}_{n=1}^\infty\subseteq\mathcal{A}$ such that $B\cap Y_k\subseteq\bigcup_{n=1}^\infty A_n$, we see by the properties of measures that

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

Hence $\nu(B\cap Y_k)\leq\mu^*(B\cap Y_k)=\overline{\mu}(B\cap Y_k)$ as $B\cap Y_k\in\mathcal{A}^*$. By repeating with B^c in place of B, we obtain that $\nu(B^c\cap Y_k)\leq\overline{\mu}(B^c\cap Y_k)$. However

$$\begin{aligned}\mu(Y_k)=\nu(Y_k) &= \nu(B\cap Y_k)+\nu(B^c\cap Y_k)\\ &\leq \overline{\mu}(B\cap Y_k)+\overline{\mu}(B^c\cap Y_k)\\ &= \overline{\mu}(Y_k)=\mu(Y_k).\end{aligned}$$

Since $\mu(Y_k)<\infty$, we obtain that $\nu(B\cap Y_k)=\overline{\mu}(B\cap Y_k)$ for all k. Therefore, since $Y_n\subseteq Y_{n+1}$ for all $n\in\mathbb{N}$ and $\bigcup_{n=1}^\infty Y_n=X$, we obtain by the Monotone Convergence Theorem for measures that

$$\nu(B)=\lim_{n\to\infty}\nu(B\cap Y_k)=\lim_{n\to\infty}\overline{\mu}(B\cap Y_k)=\overline{\mu}(B).$$

Hence $\nu=\overline{\mu}$ as desired. ∎

Due to the Carathéodory-Hahn Extension Theorem (Theorem 1.3.7), we make the following definition.

Definition 1.3.8. Let X be a non-empty set, let $\mathcal{A}$ be an algebra on X, and let $\mu : \mathcal{A} \to [0, \infty]$ be a pre-measure. The measure $\overline{\mu}$ constructed in Theorem 1.3.7 is called the *Carathéodory extension of* μ.

Before seeing what the Carathéodory-Hahn Extension Theorem (Theorem 1.3.7) produces in regards to the Lebesgue measure, we note the following example demonstrating why the σ-finite condition is necessary in order to prove uniqueness.

Example 1.3.9. If μ is a pre-measure on an algebra $\mathcal{A}$ that is not σ-finite, the Carathéodory extension of μ need not be the unique extension of μ to the set of all μ^*-measurable sets. Indeed consider $X = \mathbb{Q} \cap (0, 1]$ and let $\mathcal{A}$ be the collection of all finite unions of sets of the form $\mathbb{Q} \cap (a, b]$. It is not difficult to see that $\mathcal{A}$ is an algebra on X.

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

$$\mu(A) = \begin{cases} 0 & \text{if } A = \emptyset \\ \infty & \text{otherwise} \end{cases}.$$

Clearly μ is a pre-measure on $\mathcal{A}$. Let μ^* be the outer measure associated to μ and let $\mathcal{A}^*$ denote the σ-algebra of all μ^*-measurable sets. Clearly we see that

$$\mu^*(A) = \begin{cases} 0 & \text{if } A = \emptyset \\ \infty & \text{otherwise} \end{cases}.$$

for all $A \subseteq X$.

We claim that μ has multiple extensions to $\mathcal{A}^*$. First, we claim that $\mathcal{A}^* = \mathcal{P}(X)$. Indeed since

$$\{q\} = \bigcap_{n=1}^{\infty} \left(q - \frac{1}{n}, q\right] \cap \mathbb{Q}$$

for all $q \in \mathbb{Q}$, we see that $\{q\} \in \mathcal{A}^*$ for all $q \in X$. Therefore, since X is countable so every subset of X is countable, and since σ-algebras are closed under countable unions, the claim follows.

Since $\mathcal{A}^* = \mathcal{P}(X)$, we see that the Carathéodory extension of μ is μ^*. However, since the counting measure on X is a measure that extends μ but does not equal μ^*, the claim is complete.

To see the full power of the Carathéodory-Hahn Extension Theorem (Theorem 1.3.7), we note the following generalization of the Lebesgue measure.

Example 1.3.10. Recall from Example 1.3.3 that if

$$\mathcal{F} = \{(a,b] \mid a,b \in [-\infty,\infty), a < b\} \cup \{(a,\infty) \mid a \in \mathbb{R} \cup \{-\infty\}\}$$

then the set $\mathcal{A}$ consisting of all finite unions of elements of $\mathcal{F}$ (including the empty set) is algebra. Notice that if $A, B \in \mathcal{F}$ and

$$\operatorname{dist}(A,B) = \inf\{|a-b| \mid a \in A, b \in B\} = 0$$

then $A \cup B \in \mathcal{F}$. Hence it is easy to see that if $A \in \mathcal{A}$ then there exists a unique $n \in \mathbb{N}$ and a unique collection $A_1, \ldots, A_n \in \mathcal{F}$ such that $A = \bigcup_{k=1}^n A_k$ and $\operatorname{dist}(A_k, A_m) > 0$ for all $k \neq m$.

Let $F : \mathbb{R} \to \mathbb{R}$ be a non-decreasing function such that

$$F(c) = \lim_{x \to c^+} F(x)$$

for all $c \in \mathbb{R}$ (that is, F is right continuous). Since F is non-decreasing, $\lim_{x\to\infty} F(x)$ either exists or equals ∞ and $\lim_{x\to-\infty} F(x)$ either exists or equals $-\infty$. Define $\lambda_F : \mathcal{F} \to [0,\infty]$ by

$$\begin{aligned}
\lambda_F(\emptyset) &= 0 \\
\lambda_F((a,b]) &= F(b) - F(a) \\
\lambda_F((a,\infty)) &= \lim_{x\to\infty} F(x) - F(a) \\
\lambda_F((-\infty,b]) &= F(b) - \lim_{x\to-\infty} F(x) \\
\lambda_F((-\infty,\infty)) &= \lim_{x\to\infty} F(x) - \lim_{x\to-\infty} F(x)
\end{aligned}$$

for all $a, b \in \mathbb{R}$ with $a < b$ (where $\infty + c = \infty$ for all $c \in [-\infty,\infty)$ and $d - (-\infty) = \infty$ for all $d \in (-\infty,\infty]$).

Notice we can extend λ_F to a function on $\mathcal{A}$ which we will also denoted by λ_F as follows: If $A \in \mathcal{A}$ define

$$\lambda_F(A) = \sum_{k=1}^n \lambda_F(A_k)$$

where $A_1, \ldots, A_n \in \mathcal{F}$ are the unique elements such that $A = \bigcup_{k=1}^n A_n$ and $\operatorname{dist}(A_k, A_m) > 0$ for all $k \neq m$.

We claim that λ_F is a pre-measure on $\mathcal{A}$. To see this, first notice that clearly $\lambda_F(\emptyset) = 0$ and $\lambda_F(A) \geq 0$ for all $A \in \mathcal{A}$ as F is non-decreasing.

Before we demonstrated that λ_F is countably additive, first notice that if $A, B \in \mathcal{F}$ are such that $A \cap B = \emptyset$ and $\operatorname{dist}(A,B) = 0$, then $\lambda_F(A \cup B) = \lambda_F(A) + \lambda_F(B)$ trivially. Hence it is easy to see that if $\{A_k\}_{k=1}^n \subseteq \mathcal{A}$ are pairwise disjoint, then

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

(that is, λ_F is finitely additive). Moreover if $A, B \in \mathcal{A}$ are such that $A \subseteq B$, we see that $B \cap A^c \in \mathcal{A}$ and

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

Hence λ_F is monotone. This further implies that λ_F is finitely subadditive. Indeed if $\{A_k\}_{k=1}^n \subseteq \mathcal{A}$ are arbitrary subsets of $\mathcal{A}$, define $B_1 = A_1$ and

$$B_k = A_k \setminus \left(\bigcup_{j=1}^{k-1} A_j \right)$$

for all $k \in \{1, \ldots, n\}$. Then $\{B_k\}_{k=1}^n$ are pairwise disjoint elements of $\mathcal{A}$ such that $B_k \subseteq A_k$ for all $k \in \{1, \ldots, n\}$ and $\bigcup_{k=1}^n B_k = \bigcup_{k=1}^n A_k$. Consequently

$$\lambda_F\left(\bigcup_{k=1}^n A_k \right) = \lambda_F\left(\bigcup_{k=1}^n B_k \right) = \sum_{k=1}^n \lambda_F(B_k) \leq \sum_{k=1}^n \lambda_F(A_k).$$

Hence λ_F is finitely subadditive.

To see that λ_F is countably additive, assume $\{A_n\}_{n=1}^\infty \subseteq \mathcal{A}$ are pairwise disjoint sets such that $A = \bigcup_{n=1}^\infty A_n \in \mathcal{A}$. To see that $\lambda_F(A) = \sum_{n=1}^\infty \lambda_F(A_n)$, we notice we may assume that $A \in \mathcal{F}$ and $A_n \in \mathcal{F}$ for all $n \in \mathbb{N}$ as every element of $\mathcal{A}$ is a finite union of elements of $\mathcal{F}$ and λ_F is finitely additive.

Notice since λ_F is monotone that for all $m \in \mathbb{N}$

$$\lambda_F(A) \geq \lambda_F\left(\bigcup_{n=1}^m A_n \right) = \sum_{n=1}^m \lambda_F(A_n).$$

Hence, by taking the limit as m tends to infinity, we see that $\sum_{n=1}^\infty \lambda_F(A_n) \leq \lambda_F(A)$.

To see the reverse inequality, first assume that $A = (a, b]$ for some $a, b \in \mathbb{R}$. Therefore, since $A_n \in \mathcal{F}$ for all $n \in \mathbb{N}$, we have for each $n \in \mathbb{N}$ we have that $A_n = (a_n, b_n]$ for some $a_n, b_n \in \mathbb{R}$. Fix $\epsilon > 0$ and notice since $F(b_n) = \lim_{x \to b_n^+} F(x)$ for each $n \in \mathbb{N}$ that there exists a $c_n > b_n$ such that

$$F(c_n) < F(b_n) + \frac{\epsilon}{2^n}.$$

Furthermore, there exists a $\delta > 0$ such that

$$F(a + \delta) < F(a) + \epsilon.$$

Since $(a, b] = \bigcup_{n=1}^\infty (a_n, b_n]$, we see that

$$[a + \delta, b] \subseteq \bigcup_{n=1}^\infty (a_n, c_n).$$

Hence $\{(a_n, c_n)\}_{n=1}^{\infty}$ is an open cover of the compact set $[a+\delta, b]$ and thus has a finite subcover, say $\{(a_{n_k}, c_{n_k})\}_{k=1}^{m}$. Thus

$$(a+\delta, b] \subseteq \bigcup_{k=1}^{m} (a_{n_k}, c_{n_k}].$$

Therefore, by the monotonicity and finite subadditivity of λ_F, we see that

$$\begin{aligned}
\lambda_F(A) = F(b) - F(a) &< \epsilon + F(b) - F(a+\delta) \\
&= \epsilon + \lambda_F((a+\delta, b]) \\
&\leq \epsilon + \lambda_F\left(\bigcup_{k=1}^{m} (a_{n_k}, c_{n_k}]\right) \\
&\leq \epsilon + \sum_{k=1}^{m} \lambda_F((a_{n_k}, c_{n_k}]) \\
&= \epsilon + \sum_{k=1}^{m} F(c_{n_k}) - F(a_{n_k}) \\
&\leq \epsilon + \sum_{k=1}^{m} \frac{\epsilon}{2^{n_k}} + F(b_{n_k}) - F(a_{n_k}) \\
&\leq 2\epsilon + \sum_{k=1}^{m} \lambda_F(A_{n_k}) \\
&\leq 2\epsilon + \sum_{n=1}^{\infty} \lambda_F(A_n).
\end{aligned}$$

Therefore, since $\epsilon > 0$ was arbitrary, the claim follows in the case that $A = (a, b]$.

To see that $\lambda_F(A) \leq \sum_{n=1}^{\infty} \lambda_F(A_n)$ for arbitrary $A \in \mathcal{F}$, notice that $A \cap (-m, m]$ is of the form $(a, b]$ for some $a, b \in \mathbb{N}$ for all $m \in \mathbb{N}$. Therefore, the previous case along with monotonicity implies that

$$\lambda_F(A \cap (-m, m]) = \sum_{n=1}^{\infty} \lambda_F(A_n \cap (-m, m]) \leq \sum_{n=1}^{\infty} \lambda_F(A_n).$$

Therefore, as it is easily seen by the definition of λ_F that

$$\lim_{m\to\infty} \lambda_F(A \cap (-m, m]) = \lambda_F(A),$$

the claim follows.

Hence λ_F is a pre-measure on the algebra $\mathcal{A}$. Furthermore, since

$$\lambda_F((-n, n]) = F(n) - F(-n) < \infty$$

for all $n \in \mathbb{N}$, λ_F is σ-finite. Hence the Carathéodory-Hahn Extension Theorem (Theorem 1.3.7) implies that λ_F has a unique extension, which will also be denoted by λ_F, to the set of all λ_F^*-measurable sets. This Carathéodory extension is called the *Lebesgue-Stieljtes measure associated to* F. Notice since every element of $\mathcal{F}$ is λ_F^*-measurable and since $\sigma(\mathcal{F}) = \mathfrak{B}(\mathbb{R})$, all Borel sets are Lebesgue-Stieljtes measurable. Consequently, $\lambda_F|_{\mathfrak{B}(\mathbb{R})}$ is often called the *Borel-Stieljtes measure associated to* F.

Remark 1.3.11. We claim that the Lebesgue measure λ is a specific instance of a Lebesgue-Stieljtes measure. Indeed if $F(x) = x$ for all $x \in \mathbb{R}$, we claim that $\lambda = \lambda_F$. To see this, it suffices to show that $\lambda^* = \lambda_F^*$. Recall for all $A \in \mathbb{R}$ that

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

$$\lambda_F^*(A) = \inf\left\{ \sum_{n=1}^\infty \ell(J_n) \,\middle|\, \substack{\{J_n\}_{n=1}^\infty \text{ are open on the left and closed on the right} \\ \text{such that } A \subseteq \bigcup_{n=1}^\infty J_n} \right\}$$

where ℓ denotes the length of an interval. Hence clearly $\lambda_F^*(A) \leq \lambda^*(A)$ for all $A \subseteq \mathbb{R}$. For the reverse inequality, let $\epsilon > 0$. By the definition of λ_F^* there exists a collection $\{J_n\}_{n=1}^\infty$ of intervals that are open on the left and closed on the right such that $A \subseteq \bigcup_{n=1}^\infty J_n$ and

$$\sum_{n=1}^\infty \ell(J_n) < \lambda_F^*(A) + \epsilon.$$

For each $n \in \mathbb{N}$ choose an open interval I_n such that $J_n \subseteq I_n$ and $\ell(I_n) \leq \ell(J_n) + \frac{\epsilon}{2^n}$. Therefore $\{I_n\}_{n=1}^\infty$ are open intervals such that $A \subseteq \bigcup_{n=1}^\infty I_n$. Thus

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

Therefore, since $\epsilon > 0$ and $A \subseteq \mathbb{R}$ were arbitrary, $\lambda_F^* = \lambda^*$.

1.4 Properties of the Lebesgue Measure

Since the Lebesgue measure is a specific instance of the Lebesgue-Stieljtes measure which was constructed using the Carathéodory-Hahn Extension Theorem (Theorem 1.3.7), we immediately obtain several properties.

Corollary 1.4.1. *Every Borel subset of* $\mathbb{R}$ *is Lebesgue measurable. Furthermore, if* $I \subseteq \mathbb{R}$ *is an interval, then* $\lambda(I) = \ell(I)$.

Proof. Since the Lebesgue measure is a specific example of the Lebesgue-Stieltjes measure by Remark 1.3.11, Example 1.3.10 shows all Borel subsets of $\mathbb{R}$ are Lebesgue measurable. Moreover, by the Carathéodory-Hahn Extension

Theorem (Theorem 1.3.7), we obtain that $\lambda(I) = \ell(I)$ for all intervals I of the form $(a, b]$, (a, ∞), and $(-\infty, b]$ for $a, b \in \mathbb{R}$. Therefore, since

$$\{c\} = \bigcap_{n=1}^{\infty} \left(c - \frac{1}{n}, c\right]$$

for all $c \in \mathbb{R}$, we obtain by the Monotone Convergence Theorem (Theorem 1.1.23) that

$$\lambda(\{c\}) = \lim_{n\to\infty} \lambda\left(\left(c - \frac{1}{n}, c\right]\right) = \lim_{n\to\infty} c - c + \frac{1}{n} = 0.$$

Therefore, since every interval of $\mathbb{R}$ is differs from an interval of the form $(a, b]$, (a, ∞), and $(-\infty, b]$ for $a, b \in \mathbb{R}$ by at most two points, the result follow by properties of measures (i.e. to change "$(a$" to "$[a$", union the above intervals with the set $\{a\}$ which does not change the length nor measure, and to change "$b]$" to "$b)$", remove the set $\{b\}$ which does not change the length nor measure). ∎

If one does not want to prove the above via the Lebesgue-Stieltjes measures, one can use the following proof.

Another proof of Corollary 1.4.1. 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

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

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.

Since $\mathcal{M}(\mathbb{R})$ is a σ-algebra, since $(a, \infty) \in \mathcal{M}(\mathbb{R})$ 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 1.1.9, it follows that $\mathfrak{B}(\mathbb{R}) \subseteq \mathcal{M}(\mathbb{R})$.

To see that $\lambda(I) = \ell(I)$ for all intervals I, first assume that $I = [a, b]$. 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, 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 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 3) 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. ∎

Using the above, we easily obtain the following important property of the Lebesgue measure.

Corollary 1.4.2. *The Lebesgue measure is σ-finite.*

Proof. For each $n \in \mathbb{N}$ let $X_n = [-n, n]$. Then $\{X_n\}_{n=1}^\infty$ are Borel sets (and hence Lebesgue measurable sets) such that $\mathbb{R} = \bigcup_{n=1}^\infty X_n$ and $\lambda(X_n) = 2n < \infty$. Hence λ is σ-finite by definition. ∎

Unsurprisingly, it is easy to compute the Lebesgue measure of any countable set.

Proposition 1.4.3. *Let $A \subseteq \mathbb{R}$ be countable. Then $A \in \mathcal{M}(\mathbb{R})$ and $\lambda(A) = 0$.*

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

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

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

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

$$A \subseteq \bigcup_{n=1}^\infty I_n.$$

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. ∎

In terms of uncountable subsets of $\mathbb{R}$, one of the most interesting sets when studying the Lebesgue measure is the following.

Definition 1.4.4. Let $P_0 = [0, 1]$. Construct P_1 from P_0 by removing the open interval of length $\frac{1}{3}$ from the middle of P_0 (i.e. $P_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$). Then construct P_2 from P_1 by removing the open intervals of length $\frac{1}{3^2}$ from the middle of each closed subinterval of P_1. Subsequently, having constructed P_n, construct P_{n+1} by removing the open intervals of length $\frac{1}{3^{n+1}}$ from the middle of each of the 2^n closed subintervals of P_n. Specifically, P_n is the union of the 2^n closed intervals of the form

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

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

The set

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

is known as the *Cantor set*.

Remark 1.4.5. The Cantor set has many interesting properties. In particular, the Cantor set is a compact set with no interior.

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

Lemma 1.4.6. *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. Suppose $x \in \mathcal{C}$. Hence $x \in P_n$ for all $n \in \mathbb{N}$. Thus, by the recursive construction of the P_n, there exists numbers $a_1, a_2, a_3, \ldots \in \{0, 2\}$ such that

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

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 that $x \in \mathbb{R}$ 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 $x \in \bigcap_{n\geq 1} P_n = \mathcal{C}$. ∎

Using the above description of the Cantor set, it is not difficult to see that the Cantor set has the same cardinality as $\mathcal{P}(\mathbb{N})$ and thus the Cantor set is uncountable. However, since we have demonstrated that the Lebesgue measure of an interval is its length, we can easily compute the Lebesgue measure of the Cantor set to be zero.

Example 1.4.7. We claim that the Cantor set $\mathcal{C}$ has Lebesgue measure zero. To begin, note C is a closed set, hence C is a Borel set, and thus Lebesgue measurable. To see that $\lambda(C) = 0$, recall from the definition of the Cantor set that

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

where $P_n \subseteq [0,1]$ (as described in Definition 1.4.4) is the union of 2^n closed intervals each of length $\frac{1}{3^n}$. Therefore, we obtain for each $n \in \mathbb{N}$ that

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

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

One important property of the Lebesgue measure is its invariance under translation and multiplicative under scaling.

Proposition 1.4.8. *If $A \in \mathcal{M}(\mathbb{R})$ and $x \in \mathbb{R}$, then $x + A \in \mathcal{M}(\mathbb{R})$ and $\lambda(x + A) = \lambda(A)$.*

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

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

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

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

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

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

Proposition 1.4.9. *If $A \in \mathcal{M}(\mathbb{R})$, $r \in \mathbb{R} \setminus \{0\}$, and $rA = \{ar \mid a \in A\}$, then $rA \in \mathcal{M}(\mathbb{R})$ and $\lambda(rA) = |r|\lambda(A)$.*

Proof. Fix $A \in \mathcal{M}(\mathbb{R})$. Since $r \neq 0$ it is easy to see that if I is an open interval then rI is an open interval with $\ell(rI) = |r|\ell(I)$. Therefore it is elementary to see that if $B \subseteq \mathbb{R}$ then

$$\lambda^*(rB) = |r|\lambda^*(B).$$

Thus it suffices to show that rA is measurable.

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

$$\begin{aligned}
\lambda^*(B) &= |r|\lambda^*(r^{-1}B) \\
&= |r|\lambda^*((r^{-1}B) \cap A) + |r|\lambda^*((rB) \cap A^c) \qquad A \in \mathcal{M}(\mathbb{R}) \\
&= \lambda^*(B \cap (rA)) + \lambda^*(B \cap (rA^c)) \\
&= \lambda^*(B \cap (rA)) + \lambda^*(B \cap (rA)^c).
\end{aligned}$$

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

Remark 1.4.10. Note Corollary 1.4.1 shows us that $\mathfrak{B}(\mathbb{R}) \subseteq \mathcal{M}(\mathbb{R}) \subseteq \mathcal{P}(\mathbb{R})$. However, we have seen (claimed really) that $|\mathfrak{B}(\mathbb{R})| = |\mathbb{R}|$ whereas Cantor's Theorem implies that $|\mathbb{R}| < |\mathcal{P}(\mathbb{R})|$. 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 1.4.5 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 $|\mathcal{C}| = |\mathbb{R}|$, 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 Example 1.2.1 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.

In fact, by modifying the proof used in Example 1.2.1, one may prove the following.

Proposition 1.4.11. *If $A \subseteq \mathbb{R}$ is such that $\lambda^*(A) > 0$, then there exists a subset $B \subseteq A$ such that B is not Lebesgue measurable.*

Proof. Let $A \subseteq \mathbb{R}$ be such that $\lambda^*(A) > 0$. For each $n \in \mathbb{Z}$ let $A_n = A \cap [n, n + 1)$. Therefore, since $A = \bigcup_{n=1}^{\infty} A_n$, we obtain by the subadditivity of the Lebesgue outer measure that

$$0 < \lambda^*(A) \leq \sum_{n=1}^{\infty} \lambda^*(A_n).$$

Hence there exists an $N \in \mathbb{N}$ such that $\lambda^*(A_N) > 0$.

We claim there exists a subset $B \subseteq A_N$ such that B is not Lebesgue measurable. To see this, note since the notion of Lebesgue measurability is invariant under translation, we may assume that $N = 0$.

Define an equivalence relation $\sim$ on $\mathbb{R}$ by $x \sim y$ if and only if $x - y \in \mathbb{Q}$. Clearly every equivalence class under $\sim$ has an element in $[0, 1)$ and by the Axiom of Choice there exists a subset B of A_0 that contains precisely one element from each equivalence class with a representative from A_0. We claim that B is not Lebesgue measurable. To see this, suppose for the sake of a contradiction that B is Lebesgue measurable.

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

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

(that is, B_n is $r_n + B$ modulo 1). Since $B_n \subseteq A_n$ where $\{A_n\}_{n=1}^{\infty}$ are as in Example 1.2.1, we see that $\{B_n\}_{n=1}^{\infty}$ is a collection of pairwise disjoint subsets of $[0, 1)$.

Moreover, we claim that

$$A_0 \subseteq \bigcup_{n=1}^{\infty} B_n.$$

To see this, note if $x \in A_0$ then there exists a unique $y \in B$ 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 $x = r_n + y \in B_n$. 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 $x = r_n + y - 1 \in B_n$. Thus the claim is complete.

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

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

Clearly $B_n = B_{n,1} \cup B_{n,2}$ since $r_n + B \subseteq [0, 2)$ 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 $x, y \in B$ such that $r_n + x \in [0, 1)$, $r_n + y \in [1, 2)$, and $b = r_n + x = -1 + r_n + y$. Clearly $r_n + x \in [0, 1)$ and $r_n + y \in [1, 2)$ imply that $x \neq y$ whereas we have $x - y = -1 \in \mathbb{Q}$ so $x \sim y$. Therefore, since B 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

$$\begin{aligned} 0 &< \lambda(A_0)) \\ &\leq \lambda\left(\bigcup_{n=1}^{\infty} B_n\right) && \text{monotonicity} \\ &= \sum_{n=1}^{\infty} \lambda(B_n) && \{A_n\}_{n=1}^{\infty} \text{ are disjoint} \\ &= \sum_{n=1}^{\infty} \lambda(B_{n,1} \cup B_{n,2}) \\ &= \sum_{n=1}^{\infty} \lambda(B_{n,1}) + \lambda(B_{n,2}) && B_{n,1} \text{ and } B_{n,2} \text{ are disjoint} \\ &= \sum_{n=1}^{\infty} \lambda((r_n + B) \cap [0, 1)) + \lambda(((r_n + B) \cap [1, 2)) \\ &= \sum_{n=1}^{\infty} \lambda((r_n + B) \cap [0, 2)) \\ &= \sum_{n=1}^{\infty} \lambda(r_n + B) && r_n + B \subseteq [0, 2) \\ &= \sum_{n=1}^{\infty} \lambda(B). \end{aligned}$$

This yields our contradiction since $\lambda(B) \in [0, \infty]$ yet no number in $[0, \infty]$ when summed an infinite number of times produces a number in $(0, \infty)$. Hence we have obtained our contradiction so B is not Lebesgue measurable. ∎

The Lebesgue measure has many additional important properties. The most important properties are described in the following two results and are used later in these notes.

Proposition 1.4.12. *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 1.1.23) that

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

For each $n \in \mathbb{N}$, let $B = A_n^c \cap [-n, n]$. 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 1.4.13. *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.

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 1.4.12 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. ∎

1.5 Metric Outer Measures

In this section we will analyze an alternative way to demonstrate that every Borel subset of $\mathbb{R}$ is Lebesgue measurable. The idea is to develop a property for outer measures on metric spaces that will guarantee that Borel sets are measurable. It turns out that the metric structure makes specific outer measures more tractable. In particular, the special outer measures on metric spaces we wish to examine are related to the following property of subsets of metric spaces.

Definition 1.5.1. Let $(\mathcal{X}, d)$ be a metric space. Two subsets $A, B \subseteq \mathcal{X}$ are said to have *positive separation* if

$$\operatorname{dist}(A, B) = \inf\{d(a, b) \mid a \in A, b \in B\} > 0.$$

Example 1.5.2. 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 on metric spaces we wish to study are as follows.

Definition 1.5.3. Let $(\mathcal{X}, d)$ be a metric space. An outer measure $\mu^* : \mathcal{P}(\mathcal{X}) \to [0, \infty]$ is said to be a *metric outer measure* if

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

for all $A, B \subseteq \mathcal{X}$ such that A and B have positive separation.

Remark 1.5.4. It is not difficult to see that if $(\mathcal{X}, d)$ is a metric space and $\mu^* : \mathcal{P}(\mathcal{X}) \to [0, \infty]$ is an outer measure on $\mathcal{X}$ such that every Borel set of $\mathcal{X}$ is μ^*-measurable, then μ^* must be a metric outer measure. Indeed, assume $A, B \subseteq \mathcal{X}$ have positive separation. By metric space properties, 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 1.5.5. *Let $(\mathcal{X}, d)$ be a metric space, let $\mu^* : \mathcal{P}(\mathcal{X}) \to [0, \infty]$ be a metric outer measure, let $\{A_n\}_{n=1}^\infty \subseteq \mathcal{P}(\mathcal{X})$ 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

$$\begin{aligned}\operatorname{dist}(B_m, B) &\geq \operatorname{dist}\left(B_m, \bigcup_{j=1}^k B_j\right) \\ &\geq \operatorname{dist}(A \setminus A_{m-1}, A_k) \\ &\geq \operatorname{dist}(A \setminus A_{m-1}, A_{m-2}) > 0\end{aligned}$$

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 1.5.6. *If $(\mathcal{X}, d)$ be a metric space and $\mu^* : \mathcal{P}(\mathcal{X}) \to [0, \infty]$ is a metric outer measure, then every Borel subset of X is μ^*-measurable.*

Proof. Since the set of μ^*-measurable sets is a σ-algebra by Theorem 1.2.10 and since the set of closed subsets of $\mathcal{X}$ generate the Borel σ-algebra, it suffices to prove that every closed subset of $\mathcal{X}$ is μ^*-measurable.

Let F be an arbitrary closed subset of $\mathcal{X}$. To see that F is μ^*-measurable, let $A \subseteq \mathcal{X}$ 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 $\mathrm{dist}(\{a\}, F) \geq \frac{1}{k}$. Since for all $y \in F$

$$d(a, y) \leq d(a, x) + d(x, y)$$

by the triangle inequality, we obtain by taking an infimum over all $y \in F$ that

$$\frac{1}{k} \leq \mathrm{dist}(\{a\}, F) \leq d(a, x) + \mathrm{dist}(\{x\}, F) < d(a, x) + \frac{1}{k+1}.$$

Hence

$$d(a, x) \geq \frac{1}{k(k+1)}.$$

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 1.5.5 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, as $A \subseteq \mathcal{X}$ was arbitrary, F is μ^*-measurable. Therefore, as F was an arbitrary closed subset of $\mathcal{X}$, 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 1.5.7. *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

$$\ell(I_{n,k}) \leq \delta + \frac{\epsilon}{2^{nk}} < \frac{3}{2}\delta < \operatorname{dist}(A, B).$$

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 $\ell(I_{n,k}) < \operatorname{dist}(A, B)$ 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 $\ell(I_{n,k}) < \operatorname{dist}(A, B)$, each $I_{n,k}$ can intersect at most one of A and B. Let

$$J_A = \{(n, k) \in \mathbb{N}^2 \mid I_{n,k} \cap A \neq \emptyset\} \quad \text{and}$$
$$J_B = \{(n, k) \in \mathbb{N}^2 \mid I_{n,k} \cap B \neq \emptyset\}.$$

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. ∎

1.6 Hausdorff Measures

One important use of metric outer measures is that it enables us to construct Hausdorff measures on metric spaces. In addition to being interesting measures trying to measure the diameter of sets, Hausdorff measures enable us to define the Hausdorff dimension thereby generalizing our notion of dimension to fractal-like sets. In order to define these objects, we recall the following metric space definition.

Definition 1.6.1. Let $(\mathcal{X}, d)$ be a metric space and let $A \subseteq \mathcal{X}$. The *diameter* of A is

$$\mathrm{diam}(A) = \sup\{d(x,y) \mid x, y \in A\}.$$

In order to construct new metric outer measures, fix a metric space $(\mathcal{X}, d)$. For each $\epsilon > 0$ let

$$\mathcal{F}_\epsilon = \{A \subseteq \mathcal{X} \mid \mathrm{diam}(A) \leq \epsilon\}$$

and for each $s \in (0, \infty)$ let $\ell_s : \mathcal{P}(\mathcal{X}) \to [0, \infty]$ be defined by

$$\ell_s(A) = \mathrm{diam}(A)^s$$

for all $A \subseteq \mathcal{X}$. Let $\mu^*_{s,\epsilon}$ denote the outer measure associated to $\ell_s|_{\mathcal{F}_\epsilon}$; that is

$$\mu^*_{s,\epsilon}(A) = \inf\left\{ \sum_{n=1}^\infty \mathrm{diam}(A_n)^s \,\middle|\, \substack{\{A_n\}_{n=1}^\infty \text{ are subsets of } \mathcal{X} \\ \text{such that } A \subseteq \bigcup_{n=1}^\infty A_n \text{ and } \mathrm{diam}(A_n) \leq \epsilon} \right\}.$$

Notice trivially that if $0 < \epsilon' < \epsilon$ then $\mu^*_{s,\epsilon}(A) \leq \mu^*_{s,\epsilon'}(A)$ for all $A \subseteq \mathcal{X}$.

Definition 1.6.2. For $s \in (0, \infty)$, the *s-dimensional outer Hausdorff measure* on the metric space $(\mathcal{X}, d)$ is the outer measure $H^*_s : \mathcal{P}(\mathcal{X}) \to [0, \infty]$ defined by

$$H^*_s(A) = \sup_{\epsilon > 0} \mu^*_{s,\epsilon}(A) = \lim_{\epsilon \to 0^+} \mu^*_{s,\epsilon}(A)$$

for all $A \subseteq \mathcal{X}$.

Unsurprisingly, the s-dimensional outer Hausdorff measure is a outer measure with the properties of the previous section.

Proposition 1.6.3. *For every metric space $(\mathcal{X}, d)$ and $s \in (0, \infty)$, 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 1.2.2 are easily seen to pass to limits, H^*_s is an outer measure.

To see that H^*_s is a metric outer measure, suppose $A, B \subseteq \mathcal{X}$ have positive separation. Therefore $\mathrm{dist}(A, B) > 0$. Clearly

$$H^*_s(A \cup B) \leq H^*_s(A) + H^*_s(B)$$

since H_s^* is an outer measure. Thus it suffices to prove the other inequality.

Assume $\epsilon < \frac{1}{2}\mathrm{dist}(A,B)$. Let $\{A_n\}_{n=1}^\infty$ be an arbitrary collection of subsets of $\mathcal{F}_\epsilon$ such that $A \cup B \subseteq \bigcup_{n=1}^\infty A_n$. Since

$$\mathrm{diam}(A_n) \leq \epsilon < \frac{1}{2}\mathrm{dist}(A,B),$$

every A_n intersects at most one of A and B. Let

$$J_A = \{n \in \mathbb{N} \mid A_n \cap A \neq \emptyset\} \qquad \text{and} \qquad J_B = \{n \in \mathbb{N} \mid A_n \cap B \neq \emptyset\}$$

Then J_A and J_B are countable disjoint sets such that

$$A \subseteq \bigcup_{n \in J_A} A_n \qquad \text{and} \qquad B \subseteq \bigcup_{n \in J_B} A_n.$$

Hence

$$\begin{aligned}\sum_{n=1}^\infty \mathrm{diam}(A_n)^s &\geq \sum_{n \in J_A} \mathrm{diam}(A_n)^s + \sum_{n \in J_B} \mathrm{diam}(A_n)^s \\ &\geq \mu_{s,\epsilon}^*(A) + \mu_{s,\epsilon}^*(B).\end{aligned}$$

Therefore, since $\{A_n\}_{n=1}^\infty$ was an arbitrary collection of subsets of $\mathcal{F}_\epsilon$ such that $A \cup B \subseteq \bigcup_{n=1}^\infty A_n$, we obtain that

$$\mu_{s,\epsilon}^*(A \cup B) \geq \mu_{s,\epsilon}^*(A) + \mu_{s,\epsilon}^*(B).$$

Since 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 $\mathcal{X}$ with positive separation, the result follows. ∎

Note that Proposition 1.5.6 implies that every Borel subset of $(\mathcal{X}, d)$ is H_s^*-measurable for all $s \in (0,\infty)$.

Definition 1.6.4. For $s \in (0,\infty)$, the *s-dimensional Hausdorff measure* on $(\mathcal{X}, d)$, denoted H_s, is the measure H_s obtained by restricting H_s^* to the set of H_s^*-measurable sets.

Example 1.6.5. The 1-dimensional Hausdorff measure on $\mathbb{R}$ is the Lebesgue measure. Indeed for all $A \subseteq \mathbb{R}$ it is clear that $\lambda^*(A) \leq \mu_{1,\epsilon}^*(A)$ for all $\epsilon > 0$ so clearly $\lambda^*(A) \leq H_s^*(A)$. For the other inclusion, we notice that the proof of Proposition 1.5.7 implies that for all $A \subseteq \mathbb{R}$ and all $\delta, \epsilon > 0$ there exists a collection $\{I_n\}_{n=1}^\infty$ of open intervals with $\ell(I_n) = \mathrm{diam}(I_n) < \frac{3}{2}\delta$ such that

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

This implies $\mu_{1,\frac{3}{2}\delta}^*(A) \leq \lambda^*(A) + \epsilon$ for all $\delta, \epsilon > 0$ and thus $H_s^*(A) = \lambda^*(A)$. Therefore, due to the definitions of H_s and λ, we obtain that $H_s = \lambda$.

Remark 1.6.6. 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 $A \subseteq \mathcal{X}$ and $\epsilon < 1$. Hence $H_s(A) \leq H_t(A)$ for all $A \in \mathfrak{B}(\mathcal{X})$ 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 1.6.7. *Let $(\mathcal{X}, d)$ be a metric space. If $s,t \in (0,\infty)$ are such that $t < s$ and $A \in \mathfrak{B}(\mathcal{X})$ is such $H_t(A) < \infty$, then $H_s(A) = 0$.*

Proof. Fix a Borel set $A \subseteq \mathcal{X}$ and assume $H_t(A) < \infty$. Let $0 < \epsilon < 1$. Then for any collection $\{A_n\}_{n=1}^\infty \in \mathcal{F}_\epsilon$ such that $A \subseteq \bigcup_{n=1}^\infty A_n$, observe that

$$\sum_{n=1}^\infty \operatorname{diam}(A_n)^s = \sum_{n=1}^\infty diam(A_n)^{s-t} diam(A_n)^t \leq \epsilon^{s-t} \sum_{n=1}^\infty diam(A_n)^t.$$

Therefore, by taking the infimum over all such $\{A_n\}_{n=1}^\infty$, 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 1.6.7, we arrive at a definition of dimension for a Borel subset of $\mathbb{R}$.

Definition 1.6.8. Let $(\mathcal{X}, d)$ be a metric space and let A be a Borel subset of $\mathcal{X}$. 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 1.6.9. Since $A \subseteq B \subseteq \mathcal{X}$ 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 1.6.10. 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 1.6.9.

Example 1.6.11. Let $I \subseteq \mathbb{R}$ 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 1.6.7 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 1.6.10 implies $\dim_H(I) = 1$.

To motivate lower-dimensional subsets of $\mathbb{R}$, we leave the following to the reader.

Proposition 1.6.12. *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 1.4.4, 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 1.4.4. 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 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 1.6.7 implies that $\dim_H(\mathcal{C}) = s_0$ as desired. ∎

Chapter 2

Measurable Functions

As with everything in mathematics, once one has defined the main objects one desires to study, one then defines the morphisms or functions related to ones' central object. These so called measurable functions will be the focus of this chapter. After developing the basic properties of real- and complex-valued measurable functions, we will demonstrate that every measurable function can be 'approximated' by 'simple' functions. We will also demonstrate that convergence of measurable functions occurs 'uniformly almost everywhere' and that measurable functions on the reals are 'almost everywhere continuous'. The theory of measurable function is vital for a theory of integration as we will see in the next chapter.

2.1 Measurable Functions

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

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

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

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

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

Definition 2.1.3. Let X be a non-empty set and let $A \subseteq X$. The *characteristic function of* A (or *indicator function*) is the function $\chi_A : X \to \mathbb{R}$ defined by

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

for all $x \in X$.

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

Example 2.1.4. Let $(X, \mathcal{A})$ be a measurable space and let $A \subseteq X$. The characteristic function χ_A is measurable as a function to $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$ if and only if $A \in \mathcal{A}$. Indeed, notice for all $B \subseteq \mathbb{R}$ that

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

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

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

However we have a notion of a measurable function for each σ-algebra on $\mathbb{K}$. One might think we could use the σ-algebra $\{\emptyset, \mathbb{K}\}$ to force every function to be measurable. However would imply the characteristic functions of non-measurable sets are measurable, which is undesirable if we want to construct an integral for measurable functions. Since we desire any continuous function to be measurable, the σ-algebra on $\mathbb{K}$ should at least contain every open set, and thus must contain the Borel σ-algebra. Thus we define the following notion.

Definition 2.1.5. Let $(X, \mathcal{A})$ be a measurable space. A function $f : X \to \mathbb{K}$ is said to be *measurable* if f is measurable as a function from $(X, \mathcal{A})$ to $(\mathbb{K}, \mathfrak{B}(\mathbb{K}))$. The set of all measurable functions from $(X, \mathcal{A})$ to $(\mathbb{K}, \mathfrak{B}(\mathbb{K}))$ is denoted $\mathcal{M}(X, \mathbb{K})$.

Of course, one natural question to ask when $\mathbb{K} = \mathbb{R}$ is why we did not use the Lebesgue measurable functions. To see the reason why, we require the following peculiar function.

Definition 2.1.6 (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 2.1.7. *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,

$$\vec{a}_1 = (a_1, a_2, \dots, a_{m-1}, 0, 2, 2, 2, \dots)$$
$$\vec{a}_2 = (a_1, a_2, \dots, a_{m-1}, 1, 0, 0, 0, \dots).$$

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 $n \in \{1, \dots, m-1\}$ 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 $n \in \{1, \dots, m-1\}$. Hence

$$\vec{b}_{\vec{a}_1} = \left(\frac{a_1}{2}, \frac{a_2}{2}, \dots, \frac{a_{m-1}}{2}, 0, 1, 1, 1, \dots\right)$$
$$\vec{b}_{\vec{a}_2} = \left(\frac{a_1}{2}, \frac{a_2}{2}, \dots, \frac{a_{m-1}}{2}, 1, 0, 0, 0, \dots\right)$$

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,

$$\vec{a}_1 = (a_1, a_2, \dots, a_{m-1}, 1, 2, 2, 2, \dots)$$
$$\vec{a}_2 = (a_1, a_2, \dots, a_{m-1}, 2, 0, 0, 0, \dots).$$

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 $n \in \{1, \dots, m-1\}$ 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 $n \in \{1, \dots, m-1\}$. Hence

$$\vec{b}_{\vec{a}_1} = \left(\frac{a_1}{2}, \frac{a_2}{2}, \dots, \frac{a_{m-1}}{2}, 1, 0, 0, 0, \dots\right)$$
$$\vec{b}_{\vec{a}_2} = \left(\frac{a_1}{2}, \frac{a_2}{2}, \dots, \frac{a_{m-1}}{2}, 1, 0, 0, 0, \dots\right)$$

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 2.1.8. *Let $\mathcal{C}$ denote the Cantor set and let f denote the Cantor ternary function. Then f is a non-decreasing continuous function which is constant on each interval of $\mathcal{C}^c$. Furthermore $f(C) = [0,1]$.*

Proof. By Lemma 2.1.7 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 1.4.4) 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 1.4.4 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; the inverse image under a continuous function of a Lebesgue measurable set need not be Lebesgue measurable. Consequently, if we defined a function $f : X \to \mathbb{R}$ to be Lebesgue measurable if and only if the inverse image of a Lebesgue measurable set is Lebesgue measurable, there would be continuous functions that are not Lebesgue measurable.

Example 2.1.9. 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.

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 Proposition 1.4.11 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, $\varphi = \psi^{-1} : [0, 2] \to [0, 1]$ 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.

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

Proposition 2.1.10. *Let $(X, \mathcal{A}_X)$ and $(Y, \mathcal{A}_Y)$ be measurable spaces and let $f : X \to Y$. If $A \subseteq \mathcal{A}_Y$ and $\mathcal{A}_Y = \sigma(A)$ (that is, $\mathcal{A}_Y$ is the smallest σ-algebra containing A), then f is measurable if and only if*

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

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

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

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

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

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

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

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

Corollary 2.1.11. *Let $(X, \mathcal{A})$ be a measurable space, let Y be a metric space, and let $f : X \to Y$. Then f is measurable as a function from $(X, \mathcal{A})$ to $(Y, \mathfrak{B}(Y))$ if and only if $f^{-1}(U) \in \mathcal{A}$ for all open subsets $U \subseteq Y$; that is, a function to a metric space equipped with the Borel σ-algebra is measurable if and only if the inverse image of every open set is measurable.*

As the inverse image of an open set under a continuous function is open, Corollary 2.1.11 trivially implies the following.

Corollary 2.1.12. *Let $(\mathcal{X}, d)$ be a metric space and let $\mathcal{A}$ be a σ-algebra of X containing $\mathfrak{B}(\mathcal{X})$. If $f : \mathcal{X} \to \mathbb{K}$ is continuous, then f is measurable.*

Corollary 2.1.13. *Let $(X, \mathcal{A})$ be a measurable space and let $f : X \to \mathbb{R}$. The following are equivalent:*

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

Proof. The result follows easily from Proposition 2.1.10 as Remark 1.1.9 implies each of the sets used in the inverse images generate $\mathfrak{B}(\mathbb{R})$. ∎

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

Proposition 2.1.14. *Let $(X, \mathcal{A}_X)$ be a measurable space and let $(\mathcal{Y}, d_\mathcal{Y})$, and $(\mathcal{Z}, d_\mathcal{Z})$ be metric spaces. If $\mathcal{Y}$ and $\mathcal{Z}$ are equipped with their respective Borel σ-algebras, if $f : \mathcal{X} \to \mathcal{Y}$ is measurable, and if $g : \mathcal{Y} \to \mathcal{Z}$ is measurable (e.g. when g is continuous), then $g \circ f : \mathcal{X} \to \mathcal{Z}$ is measurable.*

Proof. Let $U \subseteq \mathcal{Z}$ be an arbitrary open set. Since g is measurable, $g^{-1}(U)$ is measurable in $\mathcal{Y}$. Hence, since f is measurable,

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

Therefore, since $U \subseteq \mathcal{Z}$ was an arbitrary open set and as open sets generated the Borel σ-algebra on $\mathcal{Z}$, the result follows from Proposition 2.1.10. ∎

Remark 2.1.15. Clearly the proof of Proposition 2.1.14 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 2.1.9.

To exhibit an example where Proposition 2.1.14 fails when g is only Lebesgue measurable, let φ and A be as in Example 2.1.9 and let $f = \varphi$ and $g = \chi_A$. Since A is Lebesgue measurable, g is Lebesgue measurable by

Example 2.1.4. Moreover, since φ is continuous, f is Lebesgue measurable by Corollary 2.1.11. 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 Corollary 2.1.11 and Proposition 2.1.14, we can easily use continuous operations to show specific operations on measurable functions preserve measurability. To get all of the operations we want, sometimes we need to double up.

Proposition 2.1.16. *Let $(X, \mathcal{A})$ be a measurable space and let $f, g \in \mathcal{M}(X, \mathbb{K})$. If $h : X \to \mathbb{K}^2$ is defined by*

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

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

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

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

Therefore Proposition 2.1.10 implies that h is measurable. ∎

Corollary 2.1.17. *Let $(X, \mathcal{A})$ be a measurable space and let $f, g \in \mathcal{M}(X, \mathbb{K})$. Then*

a) $cf \in \mathcal{M}(X, \mathbb{K})$ *for all* $c \in \mathbb{K}$.

b) $f + g \in \mathcal{M}(X, \mathbb{K})$.

c) $fg \in \mathcal{M}(X, \mathbb{K})$.

d) $|f| \in \mathcal{M}(X, \mathbb{K})$.

e) $\frac{1}{f} \in \mathcal{M}(X, \mathbb{K})$ *if* $f(x) \neq 0$ *for all* $x \in X$.

f) $\overline{f} \in \mathcal{M}(X, \mathbb{K})$ *where* $\overline{f}(z) = \overline{f(z)}$.

Proof. As constant functions are measurable, a) will follow from c).

Let $f, g \in \mathcal{M}(X, \mathbb{K})$. By Proposition 2.1.16, the function $h : X \to \mathbb{K}^2$ defined by $h(x) = (f(x), g(x))$ is measurable. Since the functions $h_+, h_\times : \mathbb{K}^2 \to \mathbb{K}$ defined by $h_+(x, y) = x + y$ and $h_\times(x, y) = xy$ are continuous functions, Propositions 2.1.14 implies that $f + g = h_+ \circ h$ and $fg = h_\times \circ h$ are measurable. Hence b) and c) follow.

Since the functions $a, C : \mathbb{K} \to \mathbb{R}$ defined by $a(z) = |z|$ and $C(z) = \overline{z}$ are continuous, Proposition 2.1.14 implies that $|f| = a \circ f$ and $\overline{f} = C \circ f$ are measurable. Hence d) and f) follow.

Finally define the function $q : \mathbb{K} \setminus \{0\} \to \mathbb{K} \setminus \{0\}$ by $q(z) = \frac{1}{z}$. Clearly q is continuous with respect to the metric on $\mathbb{K} \setminus \{0\}$ induced by $|\cdot|$. Since $\frac{1}{f} = q \circ f$ is well-defined as $f(x) \neq 0$ for all $x \in X$, Proposition 2.1.14 implies that $\frac{1}{f}$ is measurable. Hence e) follows. ∎

Remark 2.1.18. Using Corollary 2.1.17 we may reduce the study of complex-valued measurable functions to real-valued measurable functions. Indeed let $(X, \mathcal{A})$ be a measurable space and let $f : X \to \mathbb{C}$. Define $\mathrm{Re}(f), \mathrm{Im}(f) : X \to \mathbb{R}$ by

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

for all $x \in X$. Hence $f(x) = \mathrm{Re}(f)(x) + i\mathrm{Im}(f)(x)$ for all $x \in X$. Note by by Corollary 2.1.17 that f is measurable if and only if $\mathrm{Re}(f)$ and $\mathrm{Im}(f)$ are measurable. The functions $\mathrm{Re}(f)$ and $\mathrm{Im}(f)$ are called the *real and imaginary parts of f* respectively.

Remark 2.1.19. In fact, the theory of measurable functions can be reduced to non-negative measurable functions. Indeed let $(X, \mathcal{A})$ be a measurable space and let $f : X \to \mathbb{R}$. Define $f_+, f_- : X \to [0, \infty)$ by

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

and

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

for all $x \in X$. Hence $|f|(x) = f_+(x) + f_-(x)$ and $f(x) = f_+(x) - f_-(x)$ for all $x \in X$. Thus, by Corollary 2.1.17, f is measurable if and only if f_+ and f_- are measurable. The functions f_+ and f_- are called the *positive and negative parts of f* respectively.

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

Definition 2.1.20. Let $(X, \mathcal{A})$ be a measurable space. An extended real-valued function $f : X \to [-\infty, \infty]$ is said to be *measurable* if

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

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

Remark 2.1.21. It is not difficult to see that the characterization of measurable real-valued functions from Corollary 2.1.13 extends to extended real-valued functions. Indeed the second characterization of Corollary 2.1.13 will extend since

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

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 measurable functions are preserved under limits.

Proposition 2.1.22. *Let $(X, \mathcal{A})$ be a measurable space. For each $n \in \mathbb{N}$, let $f_n : X \to [-\infty, \infty]$ be a measurable function. Then the functions*

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

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

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

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

Hence $\sup_{n \geq 1} f_n$ is measurable by Corollary 2.1.13. Similarly, to see that $\inf_{n \geq 1} f_n$ is measurable, notice for all $a \in \mathbb{R}$ that

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

Hence $\inf_{n \geq 1} f_n$ is measurable by Corollary 2.1.13.

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

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

Since each g_k and h_k is measurable from above and since

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

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

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

Corollary 2.1.23. *Let $(X, \mathcal{A})$ be a measurable space. For each $n \in \mathbb{N}$, let $f_n : X \to \mathbb{C}$ be a measurable function. If $f : X \to \mathbb{C}$ is such that $f(x) = \lim_{n\to\infty} f_n(x)$ (that is, f_n converge to f pointwise), then f is measurable.*

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

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

Since $\mathrm{Re}(f_n)$ and $\mathrm{Im}(f_n)$ are measurable by Remark 2.1.18, we obtain that $\mathrm{Re}(f)$ and $\mathrm{Im}(f)$ are measurable by Proposition 2.1.22. Hence f is measurable by Remark 2.1.18. ■

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

Definition 2.1.24. Let $(X, \mathcal{A}, \mu)$ be a measure space and let P be a property that at each point in X is either true or false. It is said that P holds *μ-almost everywhere* (abbreviated μ-a.e. or simply a.e. if μ is clear) if there exists a set $A \subseteq \mathcal{A}$ such that $P(x)$ is true for all $x \in A$ and $\mu(A^c) = 0$.

Remark 2.1.25. For example, given a measure space $(X, \mathcal{A}, \mu)$, two functions $f, g : X \to \mathbb{K}$ are equal almost everywhere if there exists a set $A \subseteq \mathcal{A}$ such that $f(x) = g(x)$ for all $x \in A$ and $\mu(A^c) = 0$. Note this is not necessarily the same as saying

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

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

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

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

Example 2.1.26. It is elementary to see that $\chi_{\mathbb{Q}} = 0$ almost everywhere with respect to the Lebesgue measure. Similarly, if A is any measurable set with zero μ-measure, then $\chi_A = 0$ μ-almost everywhere.

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

Proposition 2.1.27. *Let $(X, \mathcal{A}_X, \mu)$ be a complete measure space, let $(Y, \mathcal{A}_Y)$ be a measurable space, and let $f, g : X \to Y$ be such that $f = g$ μ-almost everywhere. If f is measurable, then g is measurable.*

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

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

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

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

Corollary 2.1.28. *Let $(X, \mathcal{A}_X, \mu)$ be a complete measure space. For each $n \in \mathbb{N}$, let $f_n : X \to \mathbb{K}$ be a measurable function. If $f : X \to \mathbb{K}$ is such that $f(x) = \lim_{n\to\infty} f_n(x)$ a.e. (that is, f_n converge to f pointwise except on a set of measure zero), then f is measurable.*

Proof. Since $f(x) = \lim_{n\to\infty} f_n(x)$ for a.e. $x \in X$, there exists a set $A \in \mathcal{A}$ such that $f(x) = \lim_{n\to\infty} f_n(x)$ for all $x \in A$ and $\mu(A^c) = 0$. Consider the sequence of functions $(f_n\chi_A)_{n\geq 1}$. Clearly $f_n\chi_A$ is measurable for all $n \in \mathbb{N}$ by Corollary 2.1.17 since f_n is measurable and χ_A is measurable as $A \in \mathcal{A}$. Therefore, since $f(x)\chi_A(x) = \lim_{n\to\infty} f_n(x)\chi_A(x)$ for all $x \in X$, $f\chi_A$ is measurable by Corollary 2.1.23. Therefore, since $\mu(A^c) = 0$ and $f(x)\chi_A(x) = f(x)$ for all $x \in A$, we see that $f = f\chi_A$ almost everywhere. Hence Proposition 2.1.27 implies that f is measurable. ∎

Remark 2.1.29. Note the trick used in Corollary 2.1.28 of multiplying by a characteristic function of a set with measure zero complement is incredibly useful in proving properties hold almost everywhere or when using a condition that only holds almost everywhere. We will see several instances of this technique in subsequent sections and chapters.

2.2 Simple Functions

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

Definition 2.2.1. Let $(X, \mathcal{A})$ be a measurable space. A function $\varphi : X \to [0, \infty)$ is said to be *simple* if there exists an $n \in \mathbb{N}$, non-empty, pairwise disjoint sets $\{A_k\}_{k=1}^n \subseteq \mathcal{A}$ such that $X = \bigcup_{k=1}^n A_k$, and $\{a_k\}_{k=1}^n \subseteq [0, \infty)$ distinct (i.e. $a_i \neq a_j$ whenever $i \neq j$) such that

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

Note clearly all simple functions are measurable by Example 2.1.4 and Corollary 2.1.17. In particular, students have already encountered specific types of simple functions in previous courses.

Example 2.2.2. Recall that $\varphi : [a, b] \to [0, \infty)$ is said to be a *step function* if $\varphi = \sum_{k=1}^n a_k \chi_{A_k}$ where $\{A_k\}_{k=1}^\infty$ are disjoint intervals whose union is $[a, b]$.

Remark 2.2.3. Suppose $(X, \mathcal{A})$ is a measurable space and $\varphi : X \to [0, \infty)$ is measurable with finite range. We claim that φ is a simple function. Indeed write $\varphi(X) = \{b_1, \ldots, b_m\}$. Since φ is measurable, $A_k = \varphi^{-1}(\{b_k\}) \in \mathcal{A}$ for all $k \in \{1, \ldots, m\}$. It is then easy to see that $\varphi = \sum_{k=1}^m b_k \chi_{A_k}$ and $\{A_k\}_{k=1}^n \subseteq \mathcal{A}$ pairwise disjoint non-empty with $X = \bigcup_{k=1}^n A_k$.

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

Consequently, if $g : X \to [0, \infty]$ is such that $g = \sum_{k=1}^n a_k \chi_{A_k}$ where $\{A_k\}_{k=1}^n \subseteq \mathcal{A}$ and $\{a_k\}_{k=1}^n \subseteq [0, \infty)$, then g has finite range and thus is a simple function. Note the description of g differs from that in Definition 2.2.1 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 2.2.1 is called the *canonical representation of a simple function*.

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

Theorem 2.2.4. *Let $(X, \mathcal{A})$ be a measurable space and let $f : X \to [0, \infty]$. Then f is measurable if and only if there exists a sequence $(\varphi_n)_{n \geq 1}$ of simple functions on X such that $\varphi_n \leq \varphi_{n+1}$ for all $n \in \mathbb{N}$ and $(\varphi_n)_{n \geq 1}$ converges to f pointwise (that is, $\lim_{n \to \infty} \varphi_n(x) = f(x)$ for all $x \in X$).*

Proof. 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 measurable, we obtain that f is measurable by Proposition 2.1.22.

Conversely assume f is 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 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 $\varphi_n : X \to [0, \infty)$ 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 2.2.4 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.

2.3 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 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 measure, pointwise convergence implies uniform convergence.

Theorem 2.3.1 (Egoroff's Theorem). *Let $(X, \mathcal{A}, \mu)$ be a finite measure space. For each $n \in \mathbb{N}$ let $f_n : X \to \mathbb{C}$ be a measurable function. If $f : X \to \mathbb{C}$ is a measurable function such that $f(x) = \lim_{n\to\infty} f_n(x)$ for all $x \in X$, then for all $\delta > 0$ there exists an $B \in \mathcal{A}$ such that $\mu(B) < \delta$ and $f = \lim_{n\to\infty} f_n$ uniformly on B^c.*

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

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

Therefore, since f and $(f_n)_{n\geq 1}$ are measurable functions, we see that $B_{m,k} \in \mathcal{A}$ for all $m, k \in \mathbb{N}$. Notice that $B_{m+1,k} \subseteq B_{m,k}$ for all $m \in \mathbb{N}$. Moreover, since $f(x) = \lim_{n\to\infty} f_n(x)$ for all $x \in X$, we see that

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

for all $k \in \mathbb{N}$. Therefore, since $\mu(\emptyset) = 0$ and $\mu(X) < \infty$, the Monotone Convergence Theorem (Theorem 1.1.23) implies that

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

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

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

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

Hence, to complete the proof, it suffices to show that $(f_n)_{n\geq 1}$ converges uniformly to f on B^c.

To see that $(f_n)_{n\geq 1}$ converges uniformly to f on B^c, let $\epsilon > 0$ be arbitrary. Choose $k \in \mathbb{N}$ be such that $\frac{1}{k} < \epsilon$. Notice that if $x \in B^c$ then $x \notin B$ so $x \notin B_{n_k,k}$. Hence for all $x \in B^c$ 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 B^c as desired. ∎

Remark 2.3.2. If in the statement of Egoroff's Theorem (Theorem 2.3.1) 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 an $A \in \mathcal{A}$ such that $\mu^c(A) = 0$ 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 2.3.1) as stated, there exists a $B \in \mathcal{A}$ such that $\mu(B) < \delta$ and $f\chi_A = \lim_{n\to\infty} f_n\chi_A$ uniformly on B^c. Hence, if $C = B \cup A^c$, then $C \in \mathcal{A}$, $\mu(C) < \delta$, and $f = \lim_{n\to\infty} f_n$ uniformly on C^c as desired.

Example 2.3.3. The conclusions of Egoroff's Theorem (Theorem 2.3.1) fail if the assumption that μ is finite is removed (or replaced with σ-finite). Indeed consider $(\mathbb{R}, \mathcal{M}(\mathbb{R}), \lambda)$ and for each $n \in \mathbb{N}$ let $f_n = \chi_{[n,\infty)}$. Clearly $(f_n)_{n\geq 1}$ converges pointwise to the constant function 0. However there does not exists a set $B \in \mathcal{M}(\mathbb{R})$ such that $(f_n)_{n\geq 1}$ converges uniformly to 0 on B^c and $\mu(B)$ is finite. To see this, assume $(f_n)_{n\geq 1}$ converged uniformly to 0 on B^c for some $B \in \mathcal{M}(\mathbb{R})$. Thus if $\epsilon = 1$ there exists an $N \in \mathbb{N}$ such that

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

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

2.4 Littlewood's First Principle

Our next goal in this course is to proof Lusin's Theorem (Theorem 2.5.1), which is also know as Littlewood's second principle. One proof of Lusin's Theorem can be constructed using Littlewood's first principle. However, we will present a different proof of Lusin's Theorem that is shorter and bypasses 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. Consequently, we begin with the following notions.

Definition 2.4.1. Let $(\mathcal{X}, d)$ be a metric space and let $\mathcal{A}$ be a σ-algebra on $\mathcal{X}$ containing the Borel sets. A measure $\mu : \mathcal{A} \to [0, \infty]$ is said to be:

- *outer regular* if $\mu(A) = \inf\{\mu(U) \mid A \subseteq U, U \text{ open}\}$ for all $A \in \mathcal{A}$.
- *inner regular* $\mu(A) = \sup\{\mu(K) \mid K \subseteq A, K \text{ compact}\}$ for all $A \in \mathcal{A}$.

A measure that both inner and outer regular is said to be *regular*.

Example 2.4.2. Note the Lebesgue measure is regular by Proposition 1.4.12.

For Littlewood's first principle, outer regularity is key.

Theorem 2.4.3 (Littlewood's First Principle). *Let $\mathcal{A} \subseteq \mathcal{P}(\mathbb{R})$ be a σ-algebra containing $\mathfrak{B}(\mathbb{R})$ and let $\mu : \mathcal{A} \to [0, \infty]$ be an outer regular measure. If $A \in \mathcal{A}$ is such that $\mu(A) < \infty$, 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*

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

Proof. Let $\epsilon > 0$. Since μ is outer regular, there exists an open set V such that $A \subseteq V$ and

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

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

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

Since every open subset of $\mathbb{R}$ is a countable union of disjoint open intervals (see Proposition C.2.11), we can write $V = \bigcup_{k=1}^{\infty} I_k$ where $\{I_k\}_{k=1}^{\infty}$ are disjoint open intervals. By the Monotone Convergence Theorem for measures (Theorem 1.1.23), we know that

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

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

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

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

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

and

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

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

2.5 Lusin's Theorem

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

Theorem 2.5.1 (Lusin's Theorem). *Let $a, b \in \mathbb{R}$ with $a < b$, let μ be a finite, regular measure on a σ-algebra containing the Borel subsets of $[a, b]$ (e.g. the Lebesgue measure), and let $f : [a, b] \to \mathbb{C}$ be μ-measurable. For all $\epsilon > 0$ there exists a closed subset $F \subseteq [a, b]$ such that $\mu([a, b] \setminus F) < \epsilon$ and $f|_F$ is continuous.*

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

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

and

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

To see why the first part of Lusin's Theorem implies the second, we note the following that will also be of use in the proof of the first part of Lusin's Theorem.

Theorem 2.5.2 (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 proceed with the proof of Lusin's Theorem (Theorem 2.5.1), we begin with the simplest case.

Lemma 2.5.3. *Lusin's Theorem (Theorem 2.5.1) holds under the additional assumption that the function f is simple.*

Proof. Let

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

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

Fix $\epsilon > 0$. Since μ is inner regular, for every k there exists a compact subset $F_k \subseteq A_k$ such that

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

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

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

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

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 2.5.2) 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 2.5.1).

Proof of Lusin's Theorem (Theorem 2.5.1). Let $f : [a,b] \to \mathbb{C}$ be an arbitrary measurable function and fix $\epsilon > 0$. By applying Theorem 2.2.4 to the positive and negative parts of the real and imaginary parts of f, we can construct a sequence $(f_n)_{n\geq 1}$ of functions that are linear combinations of simple functions that converge to f pointwise. By applying Lemma 2.5.3 to each of the four simple functions in the linear combination of f_n and by taking the intersection of four closed sets (each whose measure in $[a,b]$ is at least $(b-a) - \frac{\epsilon}{2^{n+3}}$), there exists a closed subset $F_n \subseteq [a,b]$ and a continuous function $g_n : [a,b] \to \mathbb{C}$ such that $f_n(x) = g_n(x)$ for all $x \in F_n$ and $\mu([a,b] \setminus F_n) < \frac{\epsilon}{2^{n+1}}$.

Since $(f_n)_{n\geq 1}$ converges pointwise to f, Egoroff's Theorem (Theorem 2.3.1) implies there exists a measurable set B such that $\mu(B) < \frac{\epsilon}{4}$ and $(f_n)_{n\geq 1}$ converges uniformly to f on $[a,b] \setminus B$. Since μ is outer regular, there exists an open set U (by this we really mean an open subset $U \subseteq [a,b]$, but by the relative topology we can view U as an open subset of $\mathbb{R}$) such that $B \subseteq U$ and

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

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

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

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

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

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

Although Lusin's Theorem (Theorem 2.5.1) appears to rely on the finiteness of the measure used, this is not necessarily required as the following result demonstrates.

Theorem 2.5.4 (Lusin's Theorem, Lebesgue measure on $\mathbb{R}$). *Let $f : \mathbb{R} \to \mathbb{C}$ be Lebesgue measurable. For all $\epsilon > 0$ there exists a closed subset $F \subseteq \mathbb{R}$ such that $\lambda(F^c) < \epsilon$ and $f|_F$ is continuous.*

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

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

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 2.5.1) to each A_n and stitch together the results.

Let $\epsilon > 0$. Since Lusin's Theorem (Theorem 2.5.1) 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. ∎

To conclude this section, we note Lusin's Theorem (Theorem 2.5.1) extends to a far more general context (including the n-dimensional Lebesgue measure) as described below. Note the following is not as nice as Theorem 2.5.4 as $\mathbb{R}$ has a nice ordering to it whereas we will not have a nice ordering in general.

Theorem 2.5.5 (Lusin's Theorem, Locally Compact version). *Let $(\mathcal{X}, \mathcal{T})$ be a locally compact Hausdorff space, let μ be a regular measure on the Borel subsets of $(\mathcal{X}, \mathcal{T})$ such that $\mu(K) < \infty$ for all compact subsets $K \subseteq \mathcal{X}$, and let $f : \mathcal{X} \to \mathbb{C}$ be a measurable function that vanishes outside a Borel set of finite μ-measure. For all $\epsilon > 0$ there exists a compact set $K \subseteq \mathcal{X}$ and a continuous function $g : \mathcal{X} \to \mathbb{C}$ with compact support such that $\mu(K^c) < \epsilon$, $g(x) = f(x)$ for all $x \in K$, and*

$$\sup(\{|g(x)| \mid x \in \mathcal{X}\}) \leq \sup(\{|f(x)| \mid x \in \mathcal{X}\}).$$

To prove the locally compact version of Lusin's Theorem (Theorem 2.5.5), first one verifies the theorem for simple functions (or, more simply, characteristic functions and then uses linearity) by using the regularity of the measure and the fact that if $(\mathcal{X}, \mathcal{T})$ is a locally compact Hausdorff space,

$K \subseteq \mathcal{X}$ is compact, and $U \in \mathcal{T}$ is such that $K \subseteq U$, then there exists a continuous function f with compact support such that $f(x) = 1$ for all $x \in K$, $f(x) = 0$ for all $x \in U^c$, and $0 \leq f(x) \leq 1$ for all $x \in X$. Indeed if $f = \chi_A$ where A is a Borel set with finite μ-measure, then by the regularity properties there exists a $U \in \mathcal{T}$ and a compact $K \subseteq \mathcal{X}$ such that $A \subseteq U$, $K \subseteq U$, $\mu(U) < \infty$, and $\mu(U \setminus K) < \epsilon$. Applying the locally compact Hausdorff property described above then gives the desired approximation.

To extend Theorem 2.5.5 from simple functions to arbitrary functions, one uses Theorem 2.2.4 together with Egoroff's Theorem 2.3.1, the finite intersection property for compact sets, and the fact that the uniform limit of continuous functions is continuous to prove that there exists a compact subset K and a continuous function g on K such that $\mu(K^c) < \epsilon$ and $f(x) = g(x)$ for all $x \in K$. One then uses a version of the Tietze Extension Theorem to extend g to a continuous function on $\mathcal{X}$ with compact support.

Chapter 3

Integration over Measure Spaces

Lusin's Theorem showed us that every measurable function on a closed interval in $\mathbb{R}$ is 'almost continuous'. Consequently, since continuous functions are Riemann integrable, it is natural to ask "Can we integrate Lebesgue measurable functions?" A review of the Riemann integral can be found in Appendix A.

It is elementary to see that we cannot integrate every Lebesgue measurable function. Indeed $\chi_{\mathbb{Q}}$ is Lebesgue measurable since $\mathbb{Q}$ is countable and thus Lebesgue measurable, but $\chi_{\mathbb{Q}}$ is not Riemann integrable (see Example A.2.1). This seems like a fundamental flaw in the Riemann integrable since $\mathbb{Q}$ has zero Lebesgue measure (i.e. zero length) so we would believe "the area under the curve" should be defined to be 0.

Another flaw of the Riemann integral occurs with respect to limits; the concept at the heart analysis. For one example define $f_n : [0, 1] \to \mathbb{R}$ by

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

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

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

For 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!

In this chapter, we will define a notion of integral with respect to any measure. In particular, if we choose the Lebesgue measure, we obtain a generalization of the Riemann integral that is much less rigid and much easier to deal with mathematically. However, unlike the Riemann integral, our integral will need to be developed in stages.

3.1 The Integral of Non-Negative Functions

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

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

Definition 3.1.1. Let $(X, \mathcal{A}, \mu)$ be a measure space and let $\varphi : X \to [0, \infty)$ be a simple function with canonical representation $\varphi = \sum_{k=1}^{n} a_k \chi_{A_k}$. For every $A \in \mathcal{A}$, we define the *integral of φ over A against μ* to be

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

where

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

In fact, we have see the quantity in Definition 3.1.1 before.

Remark 3.1.2. Let $(X, \mathcal{A}, \mu)$ be a measure space and let $\varphi : X \to [0, \infty)$ be a simple function. If one defines $\nu : \mathcal{A} \to [0, \infty]$ by

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

then ν is a measure as shown in Example 1.1.15.

Example 3.1.3. Let $(X, \mathcal{A}, \mu)$ be a measure space and let $A \in \mathcal{A}$. For $B \in \mathcal{A}$, consider the simple function $\varphi = \chi_B$. If $B = X$, then the canonical representation of φ is $\varphi = 1\chi_B = 1\chi_X$ so

$$\int_A \varphi \, d\mu = 1\mu(A) = \mu(A \cap B)$$

If $B = \emptyset$, then the canonical representation of φ is $\varphi = 0\chi_{B^c} = 0\chi_X$ so

$$\int_A \varphi \, d\mu = 0\mu(X) = 0 = \mu(A \cap B).$$

Finally, if $B \neq X$ and $B \neq \emptyset$, then the canonical representation of φ is $\varphi = 1\chi_B + 0\chi_{B^c}$ so

$$\int_A \chi_B \, d\mu = 1\mu(A \cap B) + 0\mu(A \cap B^c) = \mu(A \cap B).$$

Note all of these cases produces the expected result.

Note that Definition 3.1.1 is a bit cumbersome to use in Example 3.1.3 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 2.2.3 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 3.1.4. Let $(X, \mathcal{A}, \mu)$ be a measure space and let $g : X \to [0, \infty)$ be such that $g = \sum_{k=1}^n a_k \chi_{A_k}$ where $\{A_k\}_{k=1}^n \subseteq \mathcal{A}$ are pairwise disjoint possibly empty sets with union X, and $\{a_k\}_{k=1}^n \subseteq [0, \infty)$. By Remark 2.2.3 we know that g is a simple function. In particular, Remark 2.2.3 shows that if $g(X) = \{b_1, \ldots, b_m\}$ and $B_j = g^{-1}(\{b_j\})$ then $g = \sum_{j=1}^m b_j \chi_{B_j}$ is the canonical representation of g. 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 $A \in \mathcal{A}$, then

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

Hence in Definition 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 3.1.4, we can verify the integral of simple functions has the desired properties of an integral.

Theorem 3.1.5. *Let $(X,\mathcal{A},\mu)$ be a measure space, let $A \in \mathcal{A}$, and let $\varphi,\psi : X \to [0,\infty)$ be simple functions. Then:*

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

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

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

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

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

Proof. Let

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

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

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

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

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

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

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

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

is a collection of pairwise disjoint measurable sets with union X 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 3.1.4,

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

Hence b) is true.

Next note c) follows from the monotonicity of measures since $A \mapsto \int_A \varphi \, d\mu$ is a measure by Remark 3.1.2.

To see that d) is true, we notice that

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

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

To see that e) is true, note that $\psi\chi_A - \varphi\chi_A$ is a simple function by Remark 2.2.3. Hence, by part b),

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

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

Using Theorem 3.1.5, we can conclude the representation of a simple function does not effect the integral.

Corollary 3.1.6. *Suppose* $(X, \mathcal{A}, \mu)$ *is a measure space and* $\varphi : X \to [0, \infty)$ *is such that* $\varphi = \sum_{k=1}^{n} a_k \chi_{A_k}$ *where* $\{A_k\}_{k=1}^{n} \subseteq \mathcal{A}$ *and* $\{a_k\}_{k=1}^{n} \subseteq [0, \infty)$ *(that is,* $\{A_k\}_{k=1}^{n}$ *are not necessarily disjoint with union* X *and* $\{a_k\}_{k=1}^{n}$ *need not be distinct). Then for all* $A \in \mathcal{A}$*,*

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

Our next goal is to extend the integral of simple functions to non-negative 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 2.2.4 and approximate from below.

Definition 3.1.7. Let $(X, \mathcal{A}, \mu)$ be a measure space, let $A \in \mathcal{A}$, and let $f : X \to [0, \infty]$ be measurable. The *integral of* f *over* A *against* μ is defined to be

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

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

Remark 3.1.8. One incredibly subtlety that we need to be careful of is that every simple function is a non-negative measurable function and thus we have two definitions for the integral of a simple function: Definition 3.1.1 and Definition 3.1.7. We better make sure these definitions agree.

Let $(X, \mathcal{A}, \mu)$ be a measure space, let $A \in \mathcal{A}$ and let $\psi : X \to [0, \infty)$ be a simple function. Let $\alpha = \int_A \psi \, d\mu$ when we evaluate the integral viewing ψ as a simple function and let $\beta = \int_A \psi \, d\mu$ when we evaluate the integral viewing ψ as a non-negative measurable function. By Definition 3.1.7, we see using $\varphi = \psi$ that $\alpha \leq \beta$. However, if $\varphi : X \to [0, \infty)$ is a simple function such that $\varphi \leq \psi$, we obtain by part e) of Theorem 3.1.5 that

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

Hence taking the supremum in Definition 3.1.7 yields $\beta \leq \alpha$ so $\alpha = \beta$. Thus the two definitions for the integral of a simple function are equal.

Example 3.1.9. Let X be a non-empty set, let $x \in X$, and let δ_x denote the point-mass measure at x. Then it is easy to see via definitions that

$$\int_X f \, d\delta_x = f(x)$$

for all $f : X \to [0, \infty]$.

Example 3.1.10. Let μ be the counting measure on $\mathbb{N}$. If $f : \mathbb{N} \to [0, \infty]$, then it is not difficult to see by the definition of the integral that

$$\int_{\mathbb{N}} f \, d\mu = \sum_{n=1}^{\infty} f(n).$$

Hence integrals are truly generalizations of sums! To see the above formula, notice for all $N \in \mathbb{N}$ that $\varphi_N = \sum_{k=1}^{N} f(k)\chi_{\{k\}}$ is a characteristic function such that $\chi_N \leq f$. Hence, for all $N \in \mathbb{N}$,

$$\int_{\mathbb{N}} f \, d\mu \geq \int_{\mathbb{N}} \varphi_N \, d\mu = \sum_{k=1}^{N} f(k)\mu(\{k\}) = \sum_{k=1}^{N} f(k).$$

Therefore $\int_{\mathbb{N}} f \, d\mu \geq \sum_{n=1}^{\infty} f(n)$. Hence the inequality holds if the series diverges. Otherwise, assume the series converges. If φ is a simple function such that $\varphi \leq f$, then, since φ has finite range and the series converges, there exists an $N \in \mathbb{N}$ such that $\varphi(k) = 0$ for all $k \geq N$. From this it is elementary to see that $\varphi \leq \varphi_N$ and hence

$$\int_{\mathbb{N}} f \, d\mu = \sup\left(\left\{ \int_{\mathbb{N}} \varphi_N \, d\mu \,\middle|\, N \in \mathbb{N} \right\}\right) = \sum_{k=1}^{\infty} f(k)$$

as desired.

Using Theorem 3.1.5, several properties of integrating simple functions transfer to integrating non-negative measurable functions.

Theorem 3.1.11. *Let $(X, \mathcal{A}, \mu)$ be a measure space, let $A \in \mathcal{A}$, and let $f, g : X \to [0, \infty]$ be measurable functions. Then:*

a) If $c \geq 0$, then $\int_A cf \, d\mu = c \int_A f \, d\mu$.

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

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

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

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

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

Proof. Clearly a) holds if $c = 0$. Otherwise if $c > 0$, it is clear that if $\varphi : X \to [0, \infty)$ is a simple function and $\varphi \leq f$ then $c\varphi$ is a simple function and $c\varphi \leq cf$. Hence, since Theorem 3.1.5 implies

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

we obtain that $c\int_A f\,d\mu \leq \int_A cf\,d\mu$. Similarly, if $\varphi : X \to [0,\infty)$ is a simple function and $\varphi \leq cf$ then $\frac{1}{c}\varphi$ is a simple function and $\frac{1}{c}\varphi \leq f$. Hence, since Theorem 3.1.5 implies

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

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

Note b) clearly follows from Theorem 3.1.5 and d) clearly follows by Definition 3.1.7 once c) is complete. Similarly, f) follows easily from e).

To see that c) is true, notice by Theorem 3.1.5

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

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 $B = \{x \in X \mid f(x) > 0\} \cap A$. Note that $B \in \mathcal{A}$ since $A \in \mathcal{A}$ and f is measurable.

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

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

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

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

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

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

as $f : X \to [0,\infty]$, we obtain by the subadditivity of measures that $\mu(B) = 0$.

Conversely, assume $\mu(B) = 0$. Let $\varphi : X \to [0,\infty)$ be 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

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

Hence the monotonicity of measures implies that

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

Thus

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

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

Immediately we have some observations based on Theorem 3.1.11

Remark 3.1.12. As Theorem 3.1.11 implies that $\int_X \chi_A f \, d\mu = \int_A f \, d\mu$, when developing the theory of integrals, it suffices to consider only integrals over all of X when developing our theory of integrals since the results for integrating over an arbitrary 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 3.1.11 is the additivity of integrals:

$$\int_X f + g \, d\mu = \int_X f \, d\mu + \int_X g \, d\mu.$$

Clearly if φ and ψ are simple functions with $\varphi \leq f$ and $\psi \leq g$, then $\varphi + \psi$ is a simple function with $\varphi + \psi \leq f + g$. Thus Theorem 3.1.5 clearly implies

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

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

3.2 The Monotone Convergence Theorem

In order to try and demonstrate the additivity of the integral of non-negative functions, we turn our attention to Theorem 2.2.4. We know every non-negative measurable function is the pointwise limit of an increasing sequence of simple functions. If we knew that the integral preserved these limits, then we would obtain

$$\int_X f \, d\mu + \int_X g \, d\mu = \int_X f + g \, d\mu$$

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

To prove our Monotone Convergence Theorem, we will make use of the Monotone Convergence Theorem for measures (Theorem 1.1.23) since, by Remark 3.1.2, the integral against a simple function produces a measure.

Theorem 3.2.1 (Monotone Convergence Theorem). *Let $(X, \mathcal{A}, \mu)$ be a measure space. For each $n \in \mathbb{N}$ let $f_n : X \to [0, \infty]$ be a measurable function such that $f_n \leq f_{n+1}$ for all $n \in \mathbb{N}$. If $f : X \to [0, \infty]$ is a measurable function and the pointwise limit of $(f_n)_{n\geq 1}$, then for all $A \in \mathcal{A}$*

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

Proof. First note since f is the pointwise limit of measurable functions that f is measurable by Proposition 2.1.22. Next note Remark 3.1.12 implies we may assume that $A = X$ since multiplying by a characteristic function will preserve measurability, pointwise limits, and the value of the integral by Theorem 3.1.11.

Since $f_n \leq f$ for all $n \in \mathbb{N}$, Theorem 3.1.11 implies that

$$\int_X f_n \, d\mu \leq \int_X f \, d\mu$$

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

$$\limsup_{n\to\infty} \int_X f_n \, d\mu \leq \int_X f \, d\mu.$$

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

$$\int_X f \, d\mu \leq \liminf_{n\to\infty} \int_X f_n \, d\mu.$$

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

$$\alpha \int_X f \, d\mu \leq \liminf_{n\to\infty} \int_X f_n \, d\mu$$

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 $\varphi : X \to [0, \infty)$ be an arbitrary simple function such that $\varphi \leq f$. Thus, if we can prove that

$$\alpha \int_X \varphi \, d\mu \leq \liminf_{n\to\infty} \int_X f_n \, d\mu,$$

the proof will be complete by the definition of the integral of f (Definition 3.1.7).

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

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

Since each $f_n - \alpha\varphi$ is a measurable function, A_n is measurable for all $n \in \mathbb{N}$. Moreover, by Theorem 3.1.11, we have for all $n \in \mathbb{N}$ that

$$\begin{aligned}
\alpha \int_{A_n} \varphi \, d\mu &= \int_{A_n} \alpha\varphi \, d\mu && \text{by Theorem 3.1.11, part a)} \\
&\leq \int_{A_n} f_n \, d\mu && \text{since } \alpha\varphi\chi_{A_n} \leq f_n\chi_{A_n} \\
&\leq \int_X f_n \, d\mu && \text{since } A_n \subseteq X \\
&\leq \liminf_{k\to\infty} \int_X f_k \, d\mu && \text{since } f_k \leq f_{k+1} \text{ so } \left(\textstyle\int_X f_k \, d\mu\right)_{k\geq 1} \text{ is an increasing sequence.}
\end{aligned}$$

Thus, to complete the proof, it suffices to replace A_n with X 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

$$X = \bigcup_{n\geq 1} A_n.$$

To see this, let $x \in X$ be arbitrary. If $f(x) = 0$ then $f_n \leq f$ and $\varphi \leq f$ implies that $f_n(x) = 0 = \alpha\varphi(x)$ and thus $x \in A_n$ for all $n \in \mathbb{N}$. Otherwise, if $f(x) > 0$, then we notice $\varphi \leq f$ implies that $f(x) > \alpha\varphi(x)$ 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 $X = \bigcup_{n\geq 1} A_n$.

Let $\nu : X \to [0, \infty]$ be defined by

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

for all $A \in \mathcal{A}$. Since φ is a simple function, Remark 3.1.2 implies that ν is a measure on $(X, \mathcal{A})$. Therefore, since $\{A_n\}_{n=1}^{\infty}$ is an increasing sequence of measurable sets with $X = \bigcup_{n\geq 1} A_n$, the Monotone Convergence Theorem for measures (Theorem 1.1.23) implies that

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

Hence the proof is complete. ∎

Using the Monotone Convergence Theorem (Theorem 3.2.1), we easily obtain the following final properties of integrals of positive functions we desire.

Theorem 3.2.2. *Let $(X, \mathcal{A}, \mu)$ be a measure space, let $A \in \mathcal{A}$, and let $f, g : X \to [0, \infty]$ be measurable functions. Then:*

a) $\int_A f + g \, d\mu = \int_A f \, d\mu + \int_A g \, d\mu$.

b) If $f = g$ *a.e., then* $\int_X f \, d\mu = \int_X g \, d\mu$.

Proof. To see that a) is true, note by Theorem 2.2.4 there exists increasing sequences of simple functions $(\varphi_n)_{n\geq 1}$ and $(\psi_n)_{n\geq 1}$ on X that converge pointwise to f and g respectively such that $\varphi_n \leq f$ and $\psi_n \leq g$ for all $n \in \mathbb{N}$. Therefore $(\varphi_n + \psi_n)_{n\geq 1}$ is an increasing sequence of simple functions that converges to $f + g$ pointwise such that $\varphi_n + \psi_n \leq f + g$ for all $n \in \mathbb{N}$. Therefore, by applying the Monotone Convergence Theorem (Theorem 3.2.1) twice and the additivity of integrals of simple functions from Theorem 3.1.5, we obtain that

$$\begin{aligned}
\int_A f + g \, d\mu &= \lim_{n\to\infty} \int_A \varphi_n + \psi_n \, d\mu \\
&= \lim_{n\to\infty} \int_A \varphi_n \, d\mu + \int_A \psi_n \, d\mu \\
&= \int_A f \, d\mu + \int_A g \, d\mu.
\end{aligned}$$

To see that b) is true, let $B \in \mathcal{A}$ be such that $f(x) = g(x)$ for all $x \in B$ and $\mu(B^c) = 0$. Thus $f\chi_B = g\chi_B$. Since $\mu(B^c) = 0$, Theorem 3.1.5 implies that

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

Hence we see that

$$\begin{aligned}
\int_X f \, d\mu &= \int_X f\chi_B \, d\mu + \int_X f\chi_{B^c} \, d\mu \\
&= \int_X f\chi_B \, d\mu + \int_{B^c} f \, d\mu \\
&= \int_X g\chi_B \, d\mu + \int_{B^c} g \, d\mu \\
&= \int_X g\chi_B \, d\mu + \int_X g\chi_{B^c} \, d\mu = \int_X g \, d\mu
\end{aligned}$$

as desired. ∎

Remark 3.2.3. Using part b) of Theorem 3.2.2 and the fact that the integral of any non-negative measurable function against a set of measure zero is zero, the Monotone Convergence Theorem (Theorem 3.2.1) also holds if the condition that "$f : X \to [0, \infty]$ is the pointwise limit of $(f_n)_{n\geq 1}$" is replaced with the condition that "$f(x) = \lim_{n\to\infty} f_n(x)$ almost everywhere" provided we know f is measurable (which is automatically implied when μ is complete as $(f_n)_{n\geq 1}$ converges to f pointwise almost everywhere).

Moreover, under an relatively mild assumption on the measure, part b) of Theorem 3.2.2 has a converse.

Proposition 3.2.4. *Let $(X, \mathcal{A}, \mu)$ be a measure space, and let $f, g : X \to [0, \infty]$ be measurable functions. If μ is σ-finite and*

$$\int_A f \, d\mu = \int_A g \, d\mu$$

for every $A \in \mathcal{A}$, then $f = g$ almost everywhere.

Proof. Let $B = \{x \in X \mid f(x) > g(x)\} \in \mathcal{A}$. We desire to show that $\mu(B) = 0$.

Consider $\int_B g \, d\mu$. Assume $\int_B g \, d\mu < \infty$. Thus, since $(f - g)\chi_B$ is a non-negative measurable function on X and since

$$\begin{aligned} \int_B f \, d\mu &= \int_X f\chi_B \, d\mu \\ &= \int_X (f - g)\chi_B + g\chi_B \, d\mu \\ &= \int_X (f - g)\chi_B \, d\mu + \int_X g\chi_B \, d\mu \\ &= \int_X (f - g)\chi_B \, d\mu + \int_B g \, d\mu, \end{aligned}$$

we obtain by cancelling $\int_B g \, d\mu = \int_B f \, d\mu$ from both sides that

$$\int_X (f - g)\chi_B \, d\mu = 0.$$

Hence part e) of Theorem 3.1.11 implies that

$$\{x \in X \mid (f(x) - g(x))\chi_B(x) > 0\} = B$$

has μ-measure zero. Hence $\mu(B) = 0$ as desired.

Otherwise, assume $\int_B g \, d\mu = \infty$. Since μ is σ-finite, there exists a collection $\{X_n\}_{n=1}^\infty \subseteq \mathcal{A}$ such that $X = \bigcup_{n=1}^\infty X_n$ and $\mu(X_n) < \infty$ for all $n \in \mathbb{N}$. For each $n, m \in \mathbb{N}$ let

$$B_{n,m} = \{x \in B \cap X_m \mid g(x) \leq n\}.$$

Since $g(x) < \infty$ for all $x \in B$ by definition, we see that $\{B_{n,m}\}_{n=1}^\infty$ are measurable subsets such that $\mu(B_{n,m}) \leq \mu(X_m) < \infty$, $B_{n,m} \subseteq B_{n+1,m}$ for all $n \in \mathbb{N}$ and $m \in \mathbb{N}$, and $\bigcup_{n=1}^\infty B_{n,m} = B \cap X_m$. However, since

$$\int_{B_{n,m}} f \, d\mu = \int_{B_{n,m}} g \, d\mu \leq n\mu(B_{n,m}) < \infty,$$

we can repeat the first part of the proof to obtain that $\mu(B_{n,m}) = 0$. Hence, by the Monotone Convergence Theorem for Measures (Theorem 1.1.23) we obtain that $\mu(B \cap X_m) = 0$ for all $m \in \mathbb{N}$. Hence

$$0 \leq \mu(B) = \mu\left(\bigcup_{m=1}^{\infty} B \cap X_m\right) \leq \sum_{m=1}^{\infty} \mu(B \cap X_m) = 0$$

so $\mu(B) = 0$ as desired.

Similarly $\mu(\{x \in X \mid f(x) < g(x)\}) = 0$ so $f = g$ μ-almost everywhere. ∎

The Monotone Convergence Theorem (Theorem 3.2.1) can also be used to prove several interesting properties of integrals of non-negative measurable functions.

Corollary 3.2.5. *Let $(X, \mathcal{A}, \mu)$ be a measure space. For each $n \in \mathbb{N}$ let $f_n : X \to [0, \infty]$ be a measurable function. If $f : X \to [0, \infty]$ is a measurable function such that $f(x) = \sum_{n=1}^{\infty} f_n(x)$ for almost every $x \in X$ (note f is automatically measurable if μ is complete), then for all $A \subseteq \mathcal{A}$*

$$\int_A f \, d\mu = \sum_{n=1}^{\infty} \int_A f_n \, d\mu.$$

Proof. For each $m \in \mathbb{N}$, let $g_m : X \to [0, \infty]$ be defined by $g_m = \sum_{n=1}^{m} f_n$. Clearly $(g_m)_{m \geq 1}$ is an increasing sequence of non-negative measurable functions that converges to f pointwise almost everywhere. Hence the Monotone Convergence Theorem (Theorem 3.2.1) implies

$$\int_A f \, d\mu = \lim_{m \to \infty} \int_A g_m \, d\mu = \lim_{m \to \infty} \sum_{n=1}^{m} \int_A f_n \, d\mu = \sum_{n=1}^{\infty} \int_A f_n \, d\mu$$

as desired. ∎

As Remark 3.1.2 (which demonstrated that integrating against a simple function gave rise to a measure) was instrumental in the proof of the Monotone Convergence Theorem (Theorem 3.2.1), we note the following extension to integrating against non-negative measurable functions.

Corollary 3.2.6. *Let $(X, \mathcal{A}, \mu)$ be a measure space and let $f : X \to [0, \infty]$ be measurable. Define $\nu : \mathcal{A} \to [0, \infty]$ by*

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

for all $A \in \mathcal{A}$. Then ν is a measure on $(X, \mathcal{A})$. Furthermore, if $A \in \mathcal{A}$ and $\mu(A) = 0$, then $\nu(A) = 0$.

Proof. Recall Theorem 3.1.5 that if $A \in \mathcal{A}$ and $\mu(A) = 0$, then

$$\nu(A) = \int_A f \, d\mu = 0.$$

Hence clearly $\nu(\emptyset) = 0$. To see that ν is countably additive, notice if $\{A_n\}_{n=1}^\infty$ is a collection of pairwise disjoint measurable sets in $(X, \mathcal{A})$, then

$$\begin{aligned}
\nu\left(\bigcup_{n=1}^\infty A_n\right) &= \int_{\bigcup_{n=1}^\infty A_n} f \, d\mu \\
&= \int_X f \chi_{\bigcup_{n=1}^\infty A_n} \, d\mu \\
&= \int_X \sum_{n=1}^\infty \chi_{A_n} f \, d\mu \qquad \text{since } \{A_n\}_{n=1}^\infty \text{ are pairwise disjoint} \\
&= \sum_{n=1}^\infty \int_X \chi_{A_n} f \, d\mu \qquad \text{by Corollary 3.2.5} \\
&= \sum_{n=1}^\infty \int_{A_n} f \, d\mu \\
&= \sum_{n=1}^\infty \nu(A_n).
\end{aligned}$$

Hence ν is a measure as desired. ∎

3.3 The Integral of Complex Functions

As the above notion of the integral for non-negative measurable functions has all of our desired properties, we now turn to extended this notion to all measurable functions. We have see that if f is a real-valued measurable function, then we can write $f = f_+ - f_-$ where f_+ and f_- are non-negative Lebesgue measurable function. If we want the 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 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 measurable functions so that we never end up in the "$\infty - \infty$" setting.

Definition 3.3.1. Let $(X, \mathcal{A}, \mu)$ be a measure space. A measurable function $f : X \to \mathbb{C}$ is said to be *integrable* if

$$\int_X |f| \, d\mu < \infty.$$

In the case that $(X, \mathcal{A}, \mu) = (\mathbb{R}, \mathcal{M}(\mathbb{R}), \lambda)$, it is said that f is *Lebesgue integrable*.

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

Remark 3.3.2. Notice if $(X, \mathcal{A}, \mu)$ and $f : X \to \mathbb{C}$ is integrable, then for all $A \in \mathcal{A}$

$$\int_A |f| \, d\mu = \int_X |f\chi_A| \, d\mu \leq \int_X |f| \, d\mu < \infty.$$

Hence the integral of $|f|$ with respect to μ against any measurable set is finite.

Remark 3.3.3. Let $(X, \mathcal{A}, \mu)$ be a measure space and let $f : X \to \mathbb{C}$ be a measurable function. Notice

$$f = \mathrm{Re}(f) + i\mathrm{Im}(f) = (\mathrm{Re}(f)_+ - \mathrm{Re}(f)_-) + i\,(\mathrm{Im}(f)_+ - \mathrm{Im}(f)_-)$$

(see Remarks 2.1.18 and 2.1.19 for definitions) where $\mathrm{Re}(f)_+$, $\mathrm{Re}(f)_-$, $\mathrm{Im}(f)_+$, and $\mathrm{Im}(f)_-$ are all measurable by Remarks 2.1.18 and 2.1.19. Since

$$\mathrm{Re}(f)_+, \mathrm{Re}(f)_-, \mathrm{Im}(f)_+, \mathrm{Im}(f)_- \leq |f|,$$

clearly if f is integrable then $\mathrm{Re}(f)_+$, $\mathrm{Re}(f)_-$, $\mathrm{Im}(f)_+$, and $\mathrm{Im}(f)_-$ are integrable. Conversely, since

$$\begin{aligned} |f| = \sqrt{\mathrm{Re}(f)^2 + \mathrm{Im}(f)^2} &\leq |\mathrm{Re}(f)| + |\mathrm{Im}(f)| \\ &= \mathrm{Re}(f)_+ + \mathrm{Re}(f)_- + \mathrm{Im}(f)_+ + \mathrm{Im}(f)_-, \end{aligned}$$

we see that f is integrable if and only if $\mathrm{Re}(f)_+$, $\mathrm{Re}(f)_-$, $\mathrm{Im}(f)_+$, and $\mathrm{Im}(f)_-$ are all integrable. More specifically, if $f : X \to \mathbb{R}$, then f is integrable if and only if f_+ and f_- are integrable.

Based on Remark 3.3.3, we make the following definition of the integral of an integrable function.

Definition 3.3.4. Let $(X, \mathcal{A}, \mu)$ be a measure space, let $A \in \mathcal{A}$, and let $f : X \to \mathbb{C}$ be integrable. The *integral of f over A against μ* is defined to be

$$\int_A f \, d\mu = \int_A \mathrm{Re}(f)_+ \, d\mu - \int_A \mathrm{Re}(f)_- \, d\mu + i \int_A \mathrm{Im}(f)_+ \, d\mu - i \int_A \mathrm{Im}(f)_- \, d\mu$$

where the four integrals on the right-hand-side are computed as integrals of non-negative measurable functions (i.e. via Definition 3.1.7). In the case that $(X, \mathcal{A}, \mu) = (\mathbb{R}, \mathcal{M}(\mathbb{R}), \lambda)$, the above is called the *Lebesgue integral of f*.

Remark 3.3.5. Due to Definition 3.3.4, given a measure space $(X, \mathcal{A}, \mu)$, a set $A \in \mathcal{A}$, and measurable function $f : X \to [0, \infty)$ we have two ways to compute $\int_A f \, d\mu$; one as a non-negative measurable function, and one as a complex-valued measurable function. However, the two notions agree as $\mathrm{Re}(f)_- = \mathrm{Im}(f)_+ = \mathrm{Im}(f)_- = 0$ when $f : X \to [0, \infty)$.

Remark 3.3.6. If $(X, \mathcal{A}, \mu)$ is a measure space is a measure space, $f : X \to \mathbb{C}$ is integrable, and $A \in \mathcal{A}$, then it is elementary to verify that

$$\mathrm{Re}(f\chi_A)_\pm = \mathrm{Re}(f)_\pm \chi_A \qquad \text{and} \qquad \mathrm{Im}(f\chi_A)_\pm = \mathrm{Im}(f)_\pm \chi_A.$$

Moreover, by Remark 3.3.2, we know that $f\chi_A$ is integrable. Hence

$$\begin{aligned}
&\int_A f\, d\mu \\
&= \int_A \mathrm{Re}(f)_+\, d\mu - \int_A \mathrm{Re}(f)_-\, d\mu + i\int_A \mathrm{Im}(f)_+\, d\mu - i\int_A \mathrm{Im}(f)_-\, d\mu \\
&= \int_X \chi_A \mathrm{Re}(f)_+\, d\mu - \int_X \chi_A \mathrm{Re}(f)_-\, d\mu + i\int_X \chi_A \mathrm{Im}(f)_+\, d\mu - i\int_X \chi_A \mathrm{Im}(f)_-\, d\mu \\
&= \int_X \mathrm{Re}(f\chi_A)_+\, d\mu - \int_X \mathrm{Re}(f\chi_A)_-\, d\mu + i\int_X \mathrm{Im}(f\chi_A)_+\, d\mu - i\int_X \mathrm{Im}(f\chi_A)_-\, d\mu \\
&= \int_X f\chi_A\, d\mu.
\end{aligned}$$

Therefore, when working with the integral and integrable functions, it suffices to just consider the Lebesgue integral over the entire space.

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

Proposition 3.3.7. *Let $(X, \mathcal{A}, \mu)$ be a measure space, let $f : X \to [-\infty, \infty]$ be integrable, and let $g : X \to [-\infty, \infty]$ be measurable. If $f = g$ a.e., then g is integrable and $\int_X g\, d\mu = \int_X f\, d\mu$.*

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 3.2.2, we obtain that

$$\int_X g_+\, d\mu = \int_X f_+\, d\mu < \infty \qquad \text{and} \qquad \int_X g_-\, d\mu = \int_X f_-\, d\mu < \infty.$$

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

$$\int_X f\, d\mu = \int_X f_+\, d\mu - \int_X f_-\, d\mu = \int_X g_+\, d\mu - \int_X g_-\, d\mu = \int_X g\, d\mu. \qquad \blacksquare$$

Remark 3.3.8. Let $(X, \mathcal{A}, \mu)$ be a measure space and let $f : X \to [-\infty, \infty]$ be integrable. Since f is measurable, the set $B = \{x \in X \mid |f(x)| = \infty\}$ is measurable. However, if $\mu(B) > 0$, it is elementary to see by the definition of the integral that $\int_X |f|\, d\mu = \infty$, which would contradict the fact that

$$\int_X |f|\, d\mu < \infty.$$

Hence $\mu(B) = 0$ so $f = \chi_{B^c} f$ almost everywhere. Since $\chi_{B^c} f : X \to \mathbb{R}$, and since

$$\int_X |\chi_{B^c} f| \, d\mu = \int_X |f| \, d\mu < \infty$$

it suffices to only consider real-valued measurable functions when discussing integrable functions.

By the way we constructed Definition 3.3.4, given a complex-valued measurable function f, we first check that $|f|$ has finite integral so that the integrals of $\mathrm{Re}(f)_+$, $\mathrm{Re}(f)_-$, $\mathrm{Im}(f)_+$, and $\mathrm{Im}(f)_-$ are finite, and then we define the integral of f to be the appropriate linear combination of the integrals of $\mathrm{Re}(f)_+$, $\mathrm{Re}(f)_-$, $\mathrm{Im}(f)_+$, and $\mathrm{Im}(f)_-$. Thus we have constructed our integral by demonstrating the theory of integration can be completely reduced to non-negative measurable functions!

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

Theorem 3.3.9. *Let $(X, \mathcal{A}, \mu)$ be a measure space. The set of integrable functions from X to $\mathbb{K}$ is a vector space over $\mathbb{K}$. In particular, if $f, g : X \to \mathbb{K}$ are integrable and $\alpha, \beta \in \mathbb{K}$, then*

$$\int_A \alpha f + \beta g \, d\mu = \alpha \int_A f \, d\mu + \beta \int_A g \, d\mu$$

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

Proof. We will focus on the case $\mathbb{K} = \mathbb{C}$ as the case $\mathbb{K} = \mathbb{R}$ follows as a subcase.

First if $f, g : X \to \mathbb{C}$ are integrable and $\alpha, \beta \in \mathbb{C}$, then

$$\int_X |\alpha f + \beta g| \, d\mu \leq \int_X |\alpha||f| + |\beta||g| \, d\mu = |\alpha| \int_X |f| \, d\mu + |\beta| \int_X |g| \, d\mu < \infty$$

since $\int_X |f| \, d\mu, \int_X |g| \, d\mu < \infty$. Hence $\alpha f + \beta g$ is integrable. Thus the set of integrable functions from X to $\mathbb{C}$ is a vector space over $\mathbb{C}$.

In order to show the linearity of the integral, it suffices to consider $A = X$ by multiplying the functions by χ_A if necessary (see Remark 3.3.2).

Next we claim that if $h_1, h_2, h_3, h_4 : X \to [0, \infty)$ are integrable functions, then

$$\int_X h_1 - h_2 + ih_3 - ih_4 \, d\mu = \int_X h_1 \, d\mu - \int_A h_2 \, d\mu + i \int_A h_3 \, d\mu - i \int_A h_4 \, d\mu.$$

To see this, let $h = h_1 - h_2 + ih_3 - ih_4$. Hence

$$h_1 - h_2 + ih_3 - ih_4 = \mathrm{Re}(h)_+ - \mathrm{Re}(h)_- + i\mathrm{Im}(h)_+ - i\mathrm{Im}(h)_-.$$

Thus

$$(h_1 + \mathrm{Re}(h)_-) + i(h_3 + \mathrm{Im}(h)_-) = (\mathrm{Re}(h)_+ + h_2) + i(\mathrm{Im}(h)_+ + h_4).$$

Therefore, by equating real and imaginary parts, we see that

$$h_1 + \mathrm{Re}(h)_- = \mathrm{Re}(h)_+ + h_2 \qquad \text{and} \qquad h_3 + \mathrm{Im}(h)_- = \mathrm{Im}(h)_+ + h_4.$$

Since $h_1, h_2, \mathrm{Re}(h)_+$, and $\mathrm{Re}(h)_-$ are non-negative measurable functions, we see that

$$\begin{aligned}\int_X h_1\,d\mu + \int_X \mathrm{Re}(h)_-\,d\mu &= \int_X h_1 + \mathrm{Re}(h)_-\,d\mu \\ &= \int_X \mathrm{Re}(h)_+ + h_2\,d\mu \\ &= \int_X h_2\,d\mu + \int_X \mathrm{Re}(h)_+\,d\mu.\end{aligned}$$

Therefore, since $h_1, h_2, \mathrm{Re}(h)_+$, and $\mathrm{Re}(h)_-$ are integrable so each integral is finite, we see that

$$\int_X h_1\,d\mu - \int_X h_2\,d\mu = \int_X \mathrm{Re}(h)_+\,d\mu - \int_X \mathrm{Re}(h)_-\,d\mu.$$

Similarly

$$\int_X h_3\,d\mu - \int_X h_4\,d\mu = \int_X \mathrm{Im}(h)_+\,d\mu - \int_X \mathrm{Im}(h)_-\,d\mu.$$

Hence

$$\begin{aligned}\int_X h\,d\mu &= \int_X \mathrm{Re}(h)_+\,d\mu - \int_X \mathrm{Re}(h)_-\,d\mu + i\int_X \mathrm{Im}(h)_+\,d\mu - i\int_X \mathrm{Im}(h)_-\,d\mu \\ &= \int_X h_1\,d\mu - \int_X h_2\,d\mu + i\int_X h_3\,d\mu - i\int_X h_4\,d\mu\end{aligned}$$

as claimed.

To proceed with the proof, for notational simplicity let

$$\begin{aligned}&f_1 = \mathrm{Re}(f)_+, \quad f_2 = \mathrm{Re}(f)_- \quad f_3 = \mathrm{Im}(f)_+ \quad f_4 = \mathrm{Im}(f)_- \\ &g_1 = \mathrm{Re}(g)_+, \quad g_2 = \mathrm{Re}(g)_- \quad g_3 = \mathrm{Im}(g)_+ \quad g_4 = \mathrm{Im}(g)_-.\end{aligned}$$

Since all f_i and g_j are positive integrable functions by Remark 3.3.3, we obtain by our above technical result that

$$\begin{aligned}\int_X f + g\,d\mu &= \int_X (f_1+g_1) - (f_2+g_2) + i(f_3+g_3) - i(f_4+g_4)\,d\mu \\ &= \int_X f_1\,d\mu + \int_X g_1\,d\mu - \int_X f_2\,d\mu - \int_X g_2\,d\mu \\ &\quad + i\int_X f_3\,d\mu + i\int_X g_3\,d\mu - i\int_X f_4\,d\mu - i\int_X fg_4\,d\mu \\ &= \int_X f\,d\mu + \int_X g\,d\mu.\end{aligned}$$

Hence the integral is additive. Thus, due to the properties of linear maps, it suffices to prove that the integral preserves scalar multiplication.

To begin, let $a \in \mathbb{R}$ be arbitrary. If $a \geq 0$, then af_1, af_2, af_3 and af_4 are positive integrable functions such that

$$\begin{aligned}\int_X af\,d\mu &= \int_X (af_1) - (af_2) + i(af_3) - i(af_4)\,d\mu \\ &= \int_X af_1\,d\mu - \int_X af_2\,d\mu + i\int_X af_3\,d\mu - i\int_X af_4\,d\mu \\ &= a\int_X f_1\,d\mu - a\int_X f_2\,d\mu + ai\int_X f_3\,d\mu - ai\int_X f_4\,d\mu \\ &= a\int_X f\,d\mu.\end{aligned}$$

Similarly, if $a < 0$, then $-a > 0$ and $(-a)f_1, (-a)f_2, (-a)f_3$ and $(-a)f_4$ are positive integral functions so

$$\begin{aligned}&\int_X af\,d\mu \\ &= \int_X ((-a)f_2) - ((-a)f_1) + i((-a)f_4) - i((-a)f_3)\,d\mu \\ &= \int_X (-a)f_2\,d\mu - \int_X (-a)f_1\,d\mu + i\int_X (-a)f_4\,d\mu - i\int_X (-a)f_3\,d\mu \\ &= (-a)\int_X f_2\,d\mu - (-a)\int_X f_1\,d\mu + (-a)i\int_X f_4\,d\mu - (-a)i\int_X f_2\,d\mu \\ &= a\int_X f\,d\mu.\end{aligned}$$

Furthermore, since f_1, f_2, f_3, and f_4 are positive integrable functions, we know that

$$\begin{aligned}\int_X if\,d\mu &= \int_X f_4 - f_3 + if_1 - if_2\,d\mu \\ &= \int_X f_4\,d\mu - \int_X f_3\,d\mu + i\int_X f_1\,d\mu - i\int_X f_2\,d\mu \\ &= i\left(-i\int_X f_4\,d\mu\right) + i\left(i\int_X f_3\,d\mu\right) + i\int_X f_1\,d\mu - i\int_X f_2\,d\mu \\ &= i\int_X f\,d\mu.\end{aligned}$$

Combining all of the above, we see that if $\alpha = a + bi$ where $a, b \in \mathbb{R}$, then

$$\begin{aligned}\int_X (a+bi)f\,d\mu &= \int_X af + b(if)\,d\mu \\ &= \int_X af\,d\mu + \int_X b(if)\,d\mu \\ &= a\int_X f\,d\mu + b\int_X if\,d\mu \\ &= a\int_X f\,d\mu + bi\int_X f\,d\mu = \alpha\int_X f\,d\mu.\end{aligned}$$

Hence the result follows. ∎

Remark 3.3.10. If $(X, \mathcal{A}, \mu)$ is a measure space and $f : X \to \mathbb{C}$ is integrable, then clearly the function $\overline{f} : X \to \mathbb{C}$ defined by $\overline{f}(x) = \overline{f(x)}$ is integrable since

$$\mathrm{Re}(\overline{f}) = \mathrm{Re}(f), \quad \mathrm{Im}\left(\overline{f}\right)_+ = \mathrm{Im}(f)_-, \quad \text{and} \quad \mathrm{Im}\left(\overline{f}\right)_- = \mathrm{Im}(f)_+.$$

Furthermore, from this we easily obtain that

$$\int_X \overline{f}\, d\mu = \overline{\int_X f\, d\mu}.$$

Thus our notion of integration plays well with respect to complex conjugation.

As we now have our integral of complex-valued integrable functions, we begin our study of the uses of this integral. To begin, we note three simple yet essential results, the first of which is reminiscent of a property of the Riemann integral.

Theorem 3.3.11. *Let $(X, \mathcal{A}, \mu)$ be a measure space. If $f : X \to \mathbb{C}$ is integrable, then*

$$\left|\int_A f\, d\mu\right| \leq \int_A |f|\, d\mu$$

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

Proof. By properties of complex numbers, there exists a $z \in \mathbb{C}$ such that $|z| = 1$ and

$$z\int_A f\, d\mu = \left|\int_A f\, d\mu\right| \geq 0$$

(i.e. rotate the complex number $\int_A f\, d\mu$ until it is positive). Hence zf is integrable and

$$0 \leq \left|\int_A f\, d\mu\right| = \int_A zf\, d\mu = \int_A \mathrm{Re}(zf)\, d\mu + i\int_A \mathrm{Im}(zf)\, d\mu.$$

However, since $\int_A \mathrm{Re}(zf)\, d\mu, \int_A \mathrm{Im}(zf)\, d\mu \in \mathbb{R}$, it must be the case that $\int_A \mathrm{Im}(zf)\, d\mu = 0$. Hence

$$\begin{aligned}
\left|\int_A f\, d\mu\right| &= \int_A \mathrm{Re}(zf)\, d\mu \\
&= \int_A \mathrm{Re}(zf)_+\, d\mu - \int_A \mathrm{Re}(zf)_-\, d\mu \\
&\leq \int_A \mathrm{Re}(zf)_+\, d\mu + \int_A \mathrm{Re}(zf)_-\, d\mu \\
&= \int_A \mathrm{Re}(zf)_+ + \mathrm{Re}(zf)_-\, d\mu \\
&= \int_A |\mathrm{Re}(zf)|\, d\mu \\
&\leq \int_A |zf|\, d\mu = \int_A |f|\, d\mu
\end{aligned}$$

as desired. ∎

Similar to Proposition 3.3.7, the integral does not distinguish under almost everywhere equivalence.

Theorem 3.3.12. *Let $(X, \mathcal{A}, \mu)$ be a measure space, let $f : X \to \mathbb{C}$ be integrable, and let $g : X \to \mathbb{C}$ be measurable. If $f = g$ a.e., then g is integrable and $\int_X f\, d\mu = \int_X g\, d\mu$.*

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

$$\begin{aligned} \mathrm{Re}(f)_+ &= \mathrm{Re}(g)_+, & \mathrm{Re}(f)_- &= \mathrm{Re}(g)_-, \\ \mathrm{Im}(f)_+ &= \mathrm{Im}(f)_+, & \mathrm{Im}(f)_- &= \mathrm{Im}(g)_- \end{aligned}$$

almost everywhere. Hence g is integrable. Thus we trivially obtain that $\int_X f\, d\mu = \int_X g\, d\mu$. ∎

To complete this section we note that the Lebesgue integral has some additional properties which follows from Proposition 1.4.8. In particular, the following shows that the Lebesgue integral behaves well with respect to the properties of $\mathbb{R}$.

Proposition 3.3.13 (Translation Invariance). *Let $f : \mathbb{R} \to \mathbb{C}$ be Lebesgue integrable. For each $y \in \mathbb{R}$ let $f_y : \mathbb{R} \to \mathbb{C}$ be defined by $f_y(x) = f(x - y)$. Then f_y is Lebesgue integrable and*

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

Proof. Since every Lebesgue integrable function can be written as a linear combination of four Lebesgue integrable non-negative functions, we may assume that $f : \mathbb{R} \to [0, \infty)$. Hence $f_y : \mathbb{R} \to [0, \infty)$. Since $f_y^{-1}([\alpha, \infty))$ is a translation of $f^{-1}([\alpha, \infty))$ for all $\alpha \in \mathbb{R}$ and thus measurable by Proposition 1.4.8 and the fact that f is measurable, we obtain that f_y is measurable.

To see that

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

(and thus f_y is integrable), first consider $A, B \subseteq \mathbb{R}$ and $y \in \mathbb{R}$ such that $B = y + A$. Hence Proposition 1.4.8 implies B is measurable if and only if A is measurable and

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

By the above, we obtain for all measurable $A \subseteq \mathbb{R}$ that

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

Therefore, since the above also shows us that φ is a simple function such that $\varphi \leq f$ if and only if φ_y is a simple function with $\varphi_y \leq f_y$ and as by the linearity of the integral we know that that

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

the result follows using the definition of the integral of a non-negative measurable function. ∎

By replacing Proposition 1.4.8 with Proposition 1.4.9 and repeating the above proof, we easily obtain the following.

Proposition 3.3.14 (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 3.3.15 (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*

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

3.4 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 3.4.1. *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 an undergraduate program.

Proposition 3.4.2. *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 3.4.1 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. 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 3.4.3. *If $f : [a,b] \to \mathbb{R}$ is Riemann integrable, then f is Lebesgue measurable.*

Proof. By Proposition 3.4.2, 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 2.1.13. 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 3.4.4. *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 3.4.3, 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 3.1.11 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

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

Theorem 3.4.5. *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 3.4.3, 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 3.4.4, 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 3.4.6. Of course, one may ask why in Definition 3.1.7 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 3.2.2 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 3.4.7. 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 3.4.4 together with the Monotone Convergence Theorem (Theorem 3.2.1), 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 3.4.4 together with the Monotone Convergence Theorem (Theorem 3.2.1), 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 3.4.8. Note the computations in Remark 3.4.7 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.

3.5 Fatou's Lemma

Due to the use of the Monotone Convergence Theorem (Theorem 3.2.1) in the theory of the integral, we desire two more limit theorems to demonstrate how well-behaved the integral is with respect to limits. The first is another limit theorem for non-negative 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 3.5.1 (Fatou's Lemma). *Let $(X, \mathcal{A}, \mu)$ be a measure space. For each $n \in \mathbb{N}$ let $f_n : X \to [0,\infty]$ be a measurable function. Then for each $A \in \mathcal{A}$*

$$\int_A \liminf_{n\to\infty} f_n \, d\mu \leq \liminf_{n\to\infty} \int_A f_n \, d\mu.$$

Proof. For each $k \in \mathbb{N}$, define $g_k : X \to [0,\infty]$ by

$$g_k(x) = \inf\{f_n(x) \mid n \geq k\}$$

for all $x \in X$. By Proposition 2.1.22 each g_k is a measurable function. Furthermore, for all $k \in \mathbb{N}$ and for all $n \geq k$ we see that $g_k \leq f_n$ so

$$\int_A g_k \, d\mu \leq \int_A f_n \, d\mu$$

for all $n \geq k$ and thus

$$\int_A g_k \, d\mu \leq \liminf_{n\to\infty} \int_A f_n \, d\mu$$

for all $k \in \mathbb{N}$. However, it is elementary to see that $(g_k)_{k\geq 1}$ is an increasing sequence of measurable functions that converges to $\liminf_{n\to\infty} f_n$ pointwise. Thus the Monotone Convergence Theorem (Theorem 3.2.1) implies that

$$\int_A \liminf_{n\to\infty} f_n \, d\mu = \lim_{n\to\infty} \int_A g_k \, d\mu \leq \liminf_{n\to\infty} \int_A f_n \, d\mu$$

as desired. ∎

Remark 3.5.2. It is not difficult to see that the inequality in Fatou's Lemma (Theorem 3.5.1) may be strict. Indeed if $f_n = \frac{1}{n}\chi_{[0,n]}$ for all $n \in \mathbb{N}$, it is easy to see that $\int_{\mathbb{R}} f_n \, d\lambda = 1$ for all $n \in \mathbb{N}$ yet $(f_n)_{n\geq 1}$ converges to zero pointwise almost everywhere so $\int_{\mathbb{R}} \liminf_{n\to\infty} f_n \, d\lambda = 0$.

3.6 The Dominated Convergence Theorem

Finally, we arrive at the most powerful notion of limit theorem for integrals of arbitrary integrable functions.

Theorem 3.6.1 (Dominated Convergence Theorem). *Let $(X, \mathcal{A}, \mu)$ be a measure space and let $g : X \to [0, \infty)$ be an integrable function. For each $n \in \mathbb{N}$ let $f_n : X \to \mathbb{C}$ be a measurable function such that $|f_n| \leq g$ almost everywhere. If $f : X \to \mathbb{C}$ is such that $(f_n)_{n\geq 1}$ converges to f pointwise almost everywhere and f is measurable (e.g. if μ is complete), then f is integrable with*

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

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

Proof. Since for each $n \in \mathbb{N}$ we have $|f_n| \leq g$ almost everywhere and since $(f_n)_{n\geq 1}$ converges to f pointwise almost everywhere, we see that $|f| \leq g$ almost everywhere. Hence, since g is integrable and since f and each f_n is measurable, f and each f_n is integrable by Theorem 3.1.11. Furthermore, as $|f - f_n|$ is measurable and since $|f - f_n| \leq |f| + |f_n| \leq 2g$, we also see that $|f - f_n|$ is integrable for all $n \in \mathbb{N}$.

Notice that for each $n \in \mathbb{N}$ that $2g - |f - f_n| \geq 0$ and that $(2g - |f - f_n|)_{n\geq 1}$ converges to $2g$ pointwise almost everywhere. Therefore Fatou's Lemma (Theorem 3.5.1) implies that

$$\begin{aligned}
\int_A 2g \, d\mu &= \int_A \liminf_{n\to\infty} 2g - |f - f_n| \, d\mu \\
&\leq \liminf_{n\to\infty} \int_A 2g - |f - f_n| \, d\mu \\
&= \liminf_{n\to\infty} \int_A 2g \, d\mu - \int_A |f - f_n| \, d\mu \\
&= \int_A 2g \, d\mu - \limsup_{n\to\infty} \int_A |f - f_n| \, d\mu.
\end{aligned}$$

Hence, since $0 \leq \int_{\mathbb{R}} 2g \, d\lambda < \infty$, we have that

$$\limsup_{n\to\infty} \int_A |f - f_n| \, d\mu = 0.$$

Therefore, by Theorem 3.3.11, we see that

$$\limsup_{n\to\infty} \left| \int_A f - f_n \, d\mu \right| \leq \limsup_{n\to\infty} \int_A |f - f_n| \, d\mu = 0$$

Hence

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

so the result follows. ∎

Remark 3.6.2. Notice that the proof of the Dominated Convergence Theorem (Theorem 3.6.1) actually produced that

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

This stronger claim will prove useful later.

Remark 3.6.3. The necessity of the existence of an integrable function $g : X \to [0, \infty)$ such that $|f_n| \leq g$ in the Dominated Convergence Theorem (Theorem 3.6.1) can be seen via the same example as in Remark 3.5.2.

To conclude, we note a result similar to part of the proof of Corollary 3.2.6 extends.

Corollary 3.6.4. *Let $(X, \mathcal{A}, \mu)$ be a measure space and let $f : X \to \mathbb{C}$ be an integrable function. If $\{A_n\}_{n=1}^\infty \subseteq \mathcal{A}$ are pairwise disjoint with $A = \bigcup_{n=1}^\infty A_n$, then*

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

Furthermore, the sum converges absolutely.

Proof. Let $\{A_n\}_{n=1}^\infty$ be a collection of pairwise disjoint measurable sets in $(X, \mathcal{A})$ with $A = \bigcup_{n=1}^\infty A_n$. To see that the series converges absolutely, notice by Theorem 3.3.11 and Corollary 3.2.6 that

$$\sum_{n=1}^\infty \left| \int_{A_n} f \, d\mu \right| \leq \sum_{n=1}^\infty \int_{A_n} |f| \, d\mu = \int_A |f| \, d\mu < \infty.$$

Hence the sum converges absolutely.

To see the series converges to the integral, for each $m \in \mathbb{N}$ let $g_m = f\chi_{\bigcup_{n=1}^m A_n}$. Clearly g_m is measurable since f is measurable and $\{A_n\}_{n=1}^\infty \subseteq \mathcal{A}$. Furthermore, since $|g_m| \leq |f|$, f is integrable, and $(g_m)_{m\geq 1}$ converge to $f\chi_A$ pointwise, we obtain by the Dominated Convergence Theorem (Theorem

3.6.1) that

$$\begin{aligned}\int_A f\,d\mu &= \int_X f\chi_A\,d\mu \\ &= \lim_{m\to\infty}\int_X g_m\,d\mu \\ &= \lim_{m\to\infty}\int_X f\chi_{\bigcup_{n=1}^m A_n}\,d\mu \\ &= \lim_{m\to\infty}\int_X \sum_{n=1}^m f\chi_{A_n}\,d\mu \\ &= \lim_{m\to\infty}\sum_{n=1}^m \int_X f\chi_{A_n}\,d\mu \\ &= \lim_{m\to\infty}\sum_{n=1}^m \int_{A_n} f\,d\mu \\ &= \sum_{n=1}^\infty \int_{A_n} f\,d\mu\end{aligned}$$

as claimed. ∎

3.7 L_p-Spaces

Before moving onto to analyzing uses of our new integrals, it is useful to describe specific collections of integrable functions in order to develop various norms on these functions. In particular, the following is a generalization of the notion of an integrable function.

Definition 3.7.1. Let $(X, \mathcal{A}, \mu)$ be a measure space and let $p \in [1, \infty)$. A measurable function $f : X \to \mathbb{C}$ is said to be *p-integrable* if

$$\int_X |f|^p\,d\mu < \infty.$$

The set of p-integrable functions on $(X, \mathcal{A}, \mu)$ is denoted $\mathcal{L}_p(X, \mu)$.

Example 3.7.2. Let $X = \mathbb{N}$, let $\mathcal{A} = \mathcal{P}(X)$, and let $\mu : \mathbb{N} \to [0, \infty)$ be the counting measure from Example 1.1.13. By our choice of σ-algebra, we obtain that every function on $\mathbb{N}$ is measurable. Furthermore, recall that there is a bijection from the set of all functions to all sequences of complex numbers by mapping a function $f : \mathbb{N} \to \mathbb{C}$ to the sequences $(f(n))_{n\geq 1}$. Finally, due to the choice of measure, we see that

$$\int_{\mathbb{N}} |f|^p\,d\mu = \sum_{n=1}^\infty |f(n)|^p.$$

Thus it is not difficult to see that $\mathcal{L}_p(\mathbb{N}, \mu) = \ell_p(\mathbb{N}, \mathbb{K})$ (as sets and, as we will see later, as normed linear spaces). In particular, p-integrable functions generalize the notion of ℓ_p-sequences spaces.

We know from undergraduate real analysis that $\ell_p(\mathbb{N}, \mathbb{K})$ are nice normed linear spaces. Thus we desire to prove the same for $\mathcal{L}_p(X, \mu)$. To begin, we note the following.

Lemma 3.7.3. *Let $(X, \mathcal{A}, \mu)$ be a measure space and let $p \in [1, \infty)$. Then $\mathcal{L}_p(X, \mu)$ is a vector space is a vector space over $\mathbb{C}$ (and thus, restricting to real-valued functions produces a vector space over $\mathbb{R}$).*

Proof. Let $f, g \in \mathcal{L}_p(X, \mu)$ and let $\alpha \in \mathbb{C}$. Then, since

$$\int_X |\alpha f|^p \, d\mu = |\alpha|^p \int_X |f|^p \, d\mu < \infty,$$

we see that $\alpha f \in \mathcal{L}_p(X, \mu)$. Moreover, since

$$\begin{aligned} |f+g|^p \leq (|f| + |g|)^p &\leq (2 \max\{|f|, |g|\})^p \\ &= 2^p \max\{|f|^p, |g|^p\} \leq 2^p \left(|f|^p + |g|^p\right), \end{aligned}$$

we see that

$$\int_X |f+g|^p \, d\mu \leq 2^p \int_X |f|^p \, d\mu + 2^p \int_X |g|^p \, d\mu < \infty$$

so $f + g \in \mathcal{L}_p(X, \mu)$. Hence $\mathcal{L}_p(X, \mu)$ is a vector space. ∎

Remark 3.7.4. Of course, given $p \in [1, \infty)$, we would like to define a norm on $\mathcal{L}_p(X, \mu)$ so that we can perform analysis. In particular, using our previous knowledge of the norm on $\ell_p(\mathbb{N}, \mathbb{K})$ from Example 3.7.2, given $f \in \mathcal{L}_p(X, \mu)$ we would like to define

$$\|f\|_p = \left(\int_X |f|^p \, d\mu\right)^{\frac{1}{p}}$$

to be the p-norm of f. It is elementary to see that if $f \in \mathcal{L}_p(X, \mu)$, then $\|f\|_p \in [0, \infty)$. Moreover, for all $\alpha \in \mathbb{C}$ we see that

$$\begin{aligned} \|\alpha f\|_p = \left(\int_X |\alpha f|^p \, d\mu\right)^{\frac{1}{p}} &= \left(|\alpha|^p \int_X |f|^p \, d\mu\right)^{\frac{1}{p}} \\ &= |\alpha| \left(\int_X |f|^p \, d\mu\right)^{\frac{1}{p}} = |\alpha| \, \|f\|_p \, . \end{aligned}$$

Furthermore, we will be able to verify the triangle inequality below. However, one problem remains. In the definition of a norm, the only vector that can have zero norm is the zero vector. However $\|f\|_p = 0$ if and only if f is zero almost everywhere. Thus it is possible there is a function f that is not the zero function (but zero almost everywhere) such that $\|f\|_p = 0$. How can we rectify this situation?

Well, since the problem is that functions that are equal almost everywhere are not equal, let's define a new notion of equality to make then equal. To begin, recall that $\mathcal{M}(X, \mathbb{C})$, the set of measurable functions from X to $\mathbb{C}$, is a vector space. Since it is elementary to verify that

$$W = \{f \in \mathcal{M}(X, \mathbb{C}) \mid f = 0 \ \mu\text{-almost everywhere}\}$$

is a subspace of $\mathcal{M}(X, \mathbb{C})$, we can form the quotient space $\mathcal{M}(X, \mathbb{C})/W$. Given a function $f \in \mathcal{M}(X, \mathbb{C})$, we will use $[f]$ to denote the equivalence class $f + W$ in $\mathcal{M}(X, \mathbb{C})/W$. Clearly if $f, g \in \mathcal{M}(X, \mathbb{C})$, then $[f] = [g]$ if and only if $f = g$ almost everywhere. In particular, if $[f] = [g]$ then

$$\int_X |f|^p \, d\mu = \int_X |g|^p \, d\mu$$

as $|f|^p = |g|^p$ almost everywhere so $f \in \mathcal{L}_p(X, \mu)$ if and only if $g \in \mathcal{L}_p(X, \mu)$. Furthermore, since W is clearly a subspace of $\mathcal{L}_p(X, \mu)$, we can consider $\mathcal{L}_p(X, \mu)/W$

Definition 3.7.5. Given a measure space $(X, \mathcal{A}, \mu)$ and a $p \in [1, \infty)$, the *L_p-space of* $(X, \mathcal{A}, \mu)$, denote $L_p(X, \mu)$, is the vector space over $\mathbb{C}$ defined by

$$L_p(X, \mu) = \{[f] \mid f \in \mathcal{L}_p(X, \mu)\}.$$

Furthermore, the *p-norm* is the function $\|\cdot\|_p : L_p(X, \mu) \to [0, \infty)$ defined by

$$\|[f]\|_p = \left(\int_X |f|^p \, d\mu\right)^{\frac{1}{p}}$$

for all $[f] \in L_p(X, \mu)$.

Remark 3.7.6. First, note that the p-norm is well-defined on $L_p(X, \mu)$. Indeed if $[f] = [g]$ then

$$\int_X |f|^p \, d\mu = \int_X |g|^p \, d\mu$$

so the value of $\|[f]\|_p$ does not depend on the representative of the equivalence class.

Due to the definition of $L_p(X, \mu)$ and Remark 3.7.4, we will often not distinguish elements of $L_p(X, \mu)$ and $\mathcal{L}_p(X, \mu)$. In actuality, elements of $\mathcal{L}_p(X, \mu)$ are functions whereas elements of $L_p(X, \mu)$ are equivalence classes of functions in $\mathcal{L}_p(X, \mu)$. However, each element $\vec{v}$ of $L_p(X, \mu)$ can be represented by a function $f \in \mathcal{L}_p(X, \mu)$ and if $g \in \mathcal{L}_p(X, \mu)$ is such that $g = f$ a.e., then $\vec{v}$ can also be represented by g. Consequently, we will treat elements of $L_p(X, \mu)$ as functions that are p-integrable where we are allowed to modify the functions on a set of μ-measure zero. Thus we will often omit the notation of an equivalence class. One thing to keep in mind is that we must verify that any function defined on $L_p(X, \mu)$ respects almost everywhere equivalence.

To prove that the p-norm is a norm on $L_p(X,\mu)$, we require some inequalities.

Lemma 3.7.7 (Young's Inequality). *Let $a,b \geq 0$ and let $p,q \in (1,\infty)$ be such that $\frac{1}{p}+\frac{1}{q}=1$. Then $ab \leq \frac{1}{p}a^p + \frac{1}{q}b^q$.*

Proof. Notice $1 = \frac{1}{p}+\frac{1}{q} = \frac{p+q}{pq}$ implies $p+q-pq=0$. Hence $q = \frac{p}{p-1}$.

Fix $b \geq 0$. Notice if $b=0$, the inequality easily holds. Thus we will assume $b>0$.

Define $f:[0,\infty)\to\mathbb{R}$ by $f(x) = \frac{1}{p}x^p + \frac{1}{q}b^q - bx$. Clearly $f(0)>0$ and $\lim_{x\to\infty} f(x) = \infty$ as $p>1$ so x^p grows faster than x. We claim that $f(x) \geq 0$ for all $x \in [0,\infty)$ thereby proving the inequality. Notice f is differentiable on $[0,\infty)$ with

$$f'(x) = x^{p-1} - b.$$

Therefore $f'(x)=0$ if and only if $x = b^{\frac{1}{p-1}}$. Moreover, it is elementary to see from the derivative that f has a local minimum at $b^{\frac{1}{p-1}}$ and thus f has a global minimum at $b^{\frac{1}{p-1}}$ due to the boundary conditions. Therefore, since

$$f\left(b^{\frac{1}{p-1}}\right) = \frac{1}{p}b^{\frac{p}{p-1}} + \frac{1}{q}b^q - b^{1+\frac{1}{p-1}} = \frac{1}{p}b^q + \frac{1}{q}b^q - b^q = 0,$$

we obtain that $f(x)\geq 0$ for all $x\in[0,\infty)$ as desired. ∎

Theorem 3.7.8 (Hölder's Inequality). *Let $(X,\mathcal{A},\mu)$ be a measure space and let $p,q\in(1,\infty)$ be such that $\frac{1}{p}+\frac{1}{q}=1$. If $f\in\mathcal{L}_p(X,\mu)$ and $g\in\mathcal{L}_q(X,\mu)$, then $fg\in\mathcal{L}_1(X,\mu)$ and*

$$\int_X |fg|\,d\mu \leq \left(\int_X |f|^p\,d\mu\right)^{\frac{1}{p}} \left(\int_X |g|^q\,d\mu\right)^{\frac{1}{q}}.$$

Proof. Let

$$\alpha = \left(\int_X |f|^p\,d\mu\right)^{\frac{1}{p}} \qquad \text{and} \qquad \beta = \left(\int_X |g|^q\,d\mu\right)^{\frac{1}{q}}.$$

If $\alpha=0$, then $|f|^p=0$ almost everywhere by Theorem 3.2.2. Hence $|f|=0$ almost everywhere so $|fg|=0$ almost everywhere and hence the inequality holds. Similarly, if $\beta=0$ then the inequality holds. Hence we may assume that $\alpha,\beta>0$.

Since $\alpha, \beta > 0$, we obtain that

$$\begin{aligned}\int_X |fg|\, d\mu &= \alpha\beta \int_X \frac{|f|}{\alpha}\frac{|g|}{\beta}\, d\mu \\ &\leq \alpha\beta \int_X \frac{|f|^p}{p\alpha^p} + \frac{|g|^q}{q\beta^q}\, d\mu \qquad \text{by Lemma 3.7.7} \\ &= \alpha\beta\left(\frac{1}{p\alpha^p}\int_X |f|^p\, d\mu + \frac{1}{q\beta^q}\int_X |g|^q\, d\mu\right) \\ &= \alpha\beta\left(\frac{1}{p} + \frac{1}{q}\right) = \alpha\beta\end{aligned}$$

as desired. ∎

In addition to being used to prove that $\mathcal{L}_p(X, \mu)$ is a vector space, Hölder's inequality (Theorem 3.7.8) also has following important corollary.

Corollary 3.7.9. *Let $(X, \mathcal{A}, \mu)$ be a measure space with $\mu(X) < \infty$ and let $p \in (1, \infty)$. If $f \in \mathcal{L}_p(X, \mu)$, then $f \in \mathcal{L}_1(X, \mu)$ with*

$$\int_X |f|\, d\mu \leq \mu(X)^{\frac{1}{q}} \left(\int_X |f|^p\, d\mu\right)^{\frac{1}{p}}.$$

where $q \in (1, \infty)$ is such that $\frac{1}{p} + \frac{1}{q} = 1$.

Proof. Since $\mu(X) < \infty$, it is elementary to see that $1 \in \mathcal{L}_q(X, \mu)$; that is, the function that is one everywhere is q-integrable as

$$\int_X 1^q\, d\mu = \mu(X) < \infty.$$

Hence, by Hölder's inequality (Theorem 3.7.8) $f = f1 \in \mathcal{L}_1(X, \mu)$ and

$$\int_X |f|\, d\mu \leq \mu(X)^{\frac{1}{q}} \left(\int_X |f|^p\, d\mu\right)^{\frac{1}{p}}.$$ ∎

Hölder's inequality (Theorem 3.7.8) also enables us to show that the p-norm satisfies the triangle inequality modulo one technicality.

Theorem 3.7.10 (Minkowski's Inequality). *Let $(X, \mathcal{A}, \mu)$ be a measure space and let $p \in [1, \infty)$. If $f, g \in \mathcal{L}_p(X, \mu)$, then*

$$\left(\int_X |f+g|^p\, d\mu\right)^{\frac{1}{p}} \leq \left(\int_X |f|^p\, d\mu\right)^{\frac{1}{p}} + \left(\int_X |g|^p\, d\mu\right)^{\frac{1}{p}}.$$

Proof. Let $f, g \in \mathcal{L}_p(X, \mu)$. Recall from Lemma 3.7.3 that $f + g \in \mathcal{L}_p(X, \mu)$. Moreover, if $p = 1$ then

$$\int_X |f+g|\, d\mu \leq \int_X |f| + |g|\, d\mu \leq \int_X |f|\, d\mu + \int_X |g|\, d\mu$$

so the inequality holds.

Now assume $p \in (1, \infty)$. Choose $q \in (1, \infty)$ so that $\frac{1}{p} + \frac{1}{q} = 1$. Thus $q = \frac{p}{p-1}$. Since $(|f+g|^{p-1})^q = |f+g|^p$, we see that $|f+g|^{p-1} \in \mathcal{L}_q(X, \mu)$. Hence Hölder's inequality (Theorem 3.7.8) implies that

$$\begin{aligned}
&\int_X |f+g|^p \, d\mu \\
&= \int_X |f+g||f+g|^{p-1} \, d\mu \\
&\leq \int_X (|f| + |g|)|f+g|^{p-1} \, d\mu \\
&= \int_X |f||f+g|^{p-1} \, d\mu + \int_X |g||f+g|^{p-1} \, d\mu \\
&\leq \left(\int_X |f|^p \, d\mu\right)^{\frac{1}{p}} \left(\int_X (|f+g|^{p-1})^q \, d\mu\right)^{\frac{1}{q}} \\
&\qquad + \left(\int_X |g|^p \, d\mu\right)^{\frac{1}{p}} \left(\int_X (|f+g|^{p-1})^q \, d\mu\right)^{\frac{1}{q}} \\
&= \left(\left(\int_X |f|^p \, d\mu\right)^{\frac{1}{p}} + \left(\int_X |g|^p \, d\mu\right)^{\frac{1}{p}}\right) \left(\int_X |f+g|^p \, d\mu\right)^{\frac{1}{q}}.
\end{aligned}$$

If $\int_X |f+g|^p \, d\mu = 0$, the result follows trivially. Otherwise, we may divide both sides of the equation by $(\int_X |f+g|^p \, d\mu)^{\frac{1}{q}}$ to obtain that

$$\left(\int_X |f+g|^p \, d\mu\right)^{\frac{1}{p}} = \left(\int_X |f+g|^p \, d\mu\right)^{1-\frac{1}{q}} \leq \left(\int_X |f|^p \, d\mu\right)^{\frac{1}{p}} + \left(\int_X |g|^p \, d\mu\right)^{\frac{1}{p}}$$

as desired. ∎

Corollary 3.7.11. *Let $(X, \mathcal{A}, \mu)$ be a measure space and let $p \in [1, \infty)$. The p-norm is a norm on $L_p(X, \mu)$.*

Proof. To see that $\|\cdot\|_p$ is indeed a norm on $L_p(X, \mu)$, we first note by Remark 3.7.6 that $\|\cdot\|_p$ is well-defined (i.e. its value does not depend on the representative of the equivalence class) and finite by the definition of $\mathcal{L}_p(X, \mu)$. Furthermore, notice that $\|f\|_p = 0$ if and only if $f = 0$ almost everywhere if and only if $[f] = 0$. Furthermore, as clearly $\|\alpha f\|_p = |\alpha| \, \|f\|_p$ for all $\alpha \in \mathbb{C}$ and $f \in L_p(X, \mu)$, and as the triangle inequality holds by Minkowski's Inequality (Theorem 3.7.10), we obtain that $\|\cdot\|_p$ is a norm on $L_p(X, \mu)$ as desired. ∎

Perhaps unsurprising for those with knowledge of undergraduate real analysis, each L_p-space is a Banach space. Of course, the proofs used in Section C.5 will not be of aid to us as how could we deduce 'pointwise Cauchy' knowing 'L_p-Cauchy'?

Theorem 3.7.12 (Riesz-Fisher Theorem). *Let $(X, \mathcal{A}, \mu)$ be a measure space and let $p \in [1, \infty)$. Then $L_p(X, \mu)$ is a Banach space.*

Proof. To see that $L_p(X, \mu)$ is a Banach space, let $(f_n)_{n\geq 1}$ be an arbitrary Cauchy sequence in $L_p(X, \mu)$ (of course this really means an Cauchy sequence of equivalence classes, each of which is represented by a function $f_n \in \mathcal{L}_p(X, \mu)$). Since $(f_n)_{n\geq 1}$ is Cauchy, it is not difficult to see that there exists a subsequence $(f_{k_n})_{n\geq 1}$ such that

$$\left\| f_{k_{n+1}} - f_{k_n} \right\|_p < \frac{1}{2^n}$$

for all $n \in \mathbb{N}$ (i.e. choose $k_n \in \mathbb{N}$ to be a natural number greater than k_{n-1} that works in the definition of a Cauchy sequence for $\epsilon = \frac{1}{2^n}$). Since $(f_n)_{n\geq 1}$ is Cauchy, it suffices to show that $(f_{k_n})_{n\geq 1}$ converges to some element in $L_p(X, \mu)$.

Define a function $g : X \to [0, \infty]$ by

$$g(x) = |f_{k_1}(x)| + \sum_{n=1}^{\infty} |f_{k_{n+1}}(x) - f_{k_n}(x)|$$

for all $x \in X$. Since the sum of measurable functions is measurable, the absolute value of measurable functions is measurable, and the pointwise limit of measurable functions is measurable by Proposition 2.1.22, we obtain that g is a measurable function. Furthermore, since g is the pointwise limit of

$$\left(|f_{k_1}| + \sum_{n=1}^{m} |f_{k_{n+1}} - f_{k_n}| \right)_{m\geq 1},$$

we obtain by Fatou's Lemma (Theorem 3.5.1) and Minkowski's inequality (Theorem 3.7.10) that

$$\begin{aligned}
\left(\int_X |g|^p \right)^{\frac{1}{p}} d\mu &\leq \liminf_{m\to\infty} \left(\int_X \left(|f_{k_1}(x)| + \sum_{n=1}^{m} |f_{k_{n+1}}(x) - f_{k_n}(x)| \right)^p d\mu \right)^{\frac{1}{p}} \\
&\leq \liminf_{m\to\infty} \left\| f_{k_1} \right\|_p + \sum_{n=1}^{m} \left\| f_{k_{n+1}} - f_{k_n} \right\|_p \\
&= \left\| f_{k_1} \right\|_p + 1 < \infty.
\end{aligned}$$

Hence $g \in L_p(X, \mu)$.

By Remark 3.3.8 we see that if $A = \{x \in X \mid g(x) = \infty\}$, then $A \in \mathcal{A}$ and $\mu(A) = 0$. By replacing each f_n with $f_n \chi_{A^c}$ (which does not affect the equivalence classes as $f_n = f_n \chi_{A^c}$ almost everywhere), we may assume that $g(x) < \infty$ for all $x \in X$.

Since $g(x) < \infty$ for all $x \in X$ and since $\mathbb{C}$ is complete so every absolutely summable sequence is summable by Theorem C.6.2, we obtain that the function $f : X \to \mathbb{C}$ defined by

$$f(x) = f_{k_1}(x) + \sum_{n=1}^{\infty} f_{k_{n+1}}(x) - f_{k_n}(x)$$

for all $x \in X$ is well-defined. Notice for all $m \in \mathbb{N}$ that

$$f_{k_1} + \sum_{n=1}^{m} f_{k_{n+1}} - f_{k_n} = f_{k_m}.$$

Hence $|f_{k_m}| \leq g$ for all $m \in \mathbb{N}$ and

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

for all $x \in X$. Hence f is measurable by Corollary 2.1.23 being the pointwise limit of measurable functions. Furthermore, since clearly $|f| \leq g$, we obtain that $f \in L_p(X, \mu)$.

We claim that $(f_{k_n})_{n\geq1}$ converges to f in the p-norm. To see this, notice since $|f|^p, |f_{k_m}|^p \leq g^p$ for all $m \in \mathbb{N}$ that

$$|f - f_{k_m}|^p \leq (|f| + |f_{k_m}|)^p \leq (2|g|)^p = 2^p|g|^p.$$

Therefore, since $g \in L_p(X, \mu)$ and since $(|f - f_{k_m}|^p)_{m\geq1}$ converges pointwise to zero, the Dominated Convergence Theorem (Theorem 3.6.1) implies that that

$$\lim_{m\to\infty} \int_X |f - f_{k_m}|^p \, d\mu = 0.$$

Hence $(f_{k_n})_{n\geq1}$ converges to f with respect to $\|\cdot\|_p$. Therefore, as $(f_n)_{n\geq1}$ was Cauchy, we obtain that $(f_n)_{n\geq1}$ converges to f in $L_p(X, \mu)$. Thus, since $(f_n)_{n\geq1}$ was an arbitrary Cauchy sequence in $L_p(X, \mu)$, we obtain that $L_p(X, \mu)$ is complete. ∎

Notice the proof of the Riesz-Markov Theorem (Theorem 7.3.2) immediately implies the following.

Corollary 3.7.13. *Let $(X, \mathcal{A}, \mu)$ be a measure space, let $p \in [1, \infty)$, and let $f \in L_p(X, \mu)$. If $(f_n)_{n\geq1}$ is a sequence of elements of $L_p(X, \mu)$ that converge to f in $L_p(X, \mu)$, then there exists a subsequence $(f_{k_n})_{n\geq1}$ of $(f_n)_{n\geq1}$ that converges to f pointwise almost everywhere.*

Proof. Since $(f_n)_{n\geq1}$ converges to f in $L_p(X, \mu)$, $(f_n)_{n\geq1}$ is Cauchy in $L_p(X, \mu)$. Therefore the proof of the Riesz-Markov Theorem (Theorem 7.3.2) implies there exists a subsequence $(f_{k_n})_{n\geq1}$ of $(f_n)_{n\geq1}$ that converges both pointwise almost everywhere and in $L_p(X, \mu)$ to some function h (i.e. $h(x) = f_{k_1}(x) + \sum_{n=1}^{\infty} f_{k_{n+1}}(x) - f_{k_n}(x)$ for all $x \in X$). Therefore, since limits in normed linear spaces are unique, we obtain that $h = f$ almost everywhere thereby completing the proof. ∎

For those familiar with undergraduate real analysis, it should not be surprising that the $p = 2$ case is special.

Corollary 3.7.14. *Let $(X, \mathcal{A}, \mu)$ be a measure space. Then $L_2(X, \mu)$ is a Hilbert space with inner product $\langle \cdot, \cdot \rangle : L_2(X, \mu) \times L_2(X, \mu) \to \mathbb{C}$ defined by*

$$\langle f, g\rangle = \int_X f\overline{g}\, d\mu.$$

Proof. First, clearly if $g \in L_2(X, \mu)$ then $\overline{g} \in L_2(X, \mu)$ and $\|g\|_2 = \|\overline{g}\|_2$. Hence, by Hölder's inequality (Theorem 3.7.8), we see that if $f, g \in L_2(X, \mu)$ then $f\overline{g} \in L_1(X, \mu)$ so

$$\langle f, g\rangle = \int_X f\overline{g}\, d\mu$$

is a well-defined element of $\mathbb{C}$. In addition, the definition of $\langle f, g\rangle$ does not depend on the representative of the equivalence classes of f and g selected, $\langle \cdot, \cdot \rangle$ is well-defined.

It is not difficult to see that $\langle f, f\rangle \geq 0$ for all $f \in L_2(X, \mu)$ with equality if and only if $f = 0$ almost everywhere, that $\langle \cdot, \cdot \rangle$ is linear in the first entry by the linearity of the integral, and that

$$\overline{\langle f, g\rangle} = \langle g, f\rangle$$

by Remark 3.3.10. Hence $\langle \cdot, \cdot \rangle$ is an inner product on $L_2(X, \mu)$. Since $\|f\|_2 = \sqrt{\langle f, f\rangle}$ for all $f \in L_2(X, \mu)$, we obtain that $L_2(X, \mu)$ is a Hilbert space by Theorem 3.7.12. ∎

Of course, the above did not deal with the case that $p = \infty$ as the formula for the norms does not make sense in this situation. To develop a notion of an ∞-norm for measurable functions, we define the following concept which is motivated by the fact that we don't need our functions to be bounded everywhere, just almost everywhere.

Definition 3.7.15. Let $(X, \mathcal{A}, \mu)$ be a measure space. A function $f : X \to \mathbb{K}$ is said to be *essentially bounded* if there exists an $M \geq 0$ such that

$$\mu(\{x \in X \mid |f(x)| > M\}) = 0.$$

The set of essentially bounded functions on $(X, \mathcal{A}, \mu)$ is denoted $\mathcal{L}_\infty(X, \mu)$.

Of course, $\mathcal{L}_\infty(X, \mu)$ will not have a well-defined norm for the same reason that $\mathcal{L}_p(X, \mu)$ did not have a well-defined norm; we have to deal with functions that are equal almost everywhere. Notice if $f, g : X \to \mathbb{C}$ are such that f is essentially bounded and $f = g$ almost everywhere, then g is essentially bounded as the union of μ-measure zero sets has μ-measure zero. Hence we may define the following.

Definition 3.7.16. Given a measure space $(X, \mathcal{A}, \mu)$, the *L_∞-space of* $(X, \mathcal{A}, \mu)$, denote $L_\infty(X, \mu)$, is

$$L_\infty(X, \mu) = \{[f] \mid f : X \to \mathbb{C} \text{ essentially bounded}\}.$$

Remark 3.7.17. Given a measure space $(X, \mathcal{A}, \mu)$ and $f, g \in \mathcal{M}(X, \mathbb{C})$ such that $[f] = [g]$, we have seen that $f \in \mathcal{L}_\infty(X, \mu)$ if and only if $g \in \mathcal{L}_\infty(X, \mu)$. In particular, every representative of an equivalence class in $L_\infty(X, \mu)$ is an element of $\mathcal{L}_\infty(X, \mu)$. Due to this and to abuse notation, we will consider elements of $\mathcal{L}_\infty(X, \mu)$ as elements of $L_\infty(X, \mu)$ and drop the explicit reminder that we are dealing with an equivalence class in most (if not all) arguments.

Theorem 3.7.18. *Let $(X, \mathcal{A}, \mu)$ be a measure space. Then $L_\infty(X, \mu)$ is a normed linear space with respect to the norm*

$$\|f\|_\infty = \inf\{M \geq 0 \mid \mu(\{x \mid |f(x)| > M\}) = 0\}.$$

Proof. First we claim that $L_\infty(X, \mu)$ is a subspace of $\mathcal{M}(X, \mathbb{C})/\sim$ and thus a vector space over $\mathbb{C}$. To see this, let $f, g \in L_\infty(X, \mu)$ be arbitrary. Then there exists $M_1, M_2 \geq 0$ such that

$$\mu(\{x \mid |f(x)| > M_1\}) = 0 \qquad \text{and} \qquad \mu(\{x \mid |g(x)| > M_2\}) = 0.$$

Hence since

$$\begin{aligned} &\{x \mid |f(x) + g(x)| > M_1 + M_2\} \\ &\subseteq \{x \mid |f(x)| + |g(x)| > M_1 + M_2\} \\ &\subseteq \{x \mid |f(x)| > M_1\} \cup \{x \mid |g(x)| > M_2\} \end{aligned}$$

we see that

$$\begin{aligned} &\mu(\{x \mid |f(x) + g(x)| > M_1 + M_2\}) \\ &\leq \mu(\{x \mid |f(x)| > M_1\}) + \mu(\{x \mid |g(x)| > M_2\}) = 0. \end{aligned}$$

Hence $f + g \in \mathcal{L}_\infty(X, \mu)$. Further for all $\alpha \in \mathbb{C}$

$$\mu(\{x \mid |\alpha f(x)| > |\alpha| M\}) = 0$$

so $\alpha f \in \mathcal{L}_\infty(X, \mu)$. Hence, since $0 \in \mathcal{L}_\infty(X, \mu)$, we have shown that $L_\infty(X, \mu)$ is a subspace of $\mathcal{M}(X, \mathbb{C})/\sim$ and thus a vector space over $\mathbb{C}$.

To see that $\|\cdot\|_\infty$ is a well-defined norm on $L_\infty(X, \mu)$, first notice that if $f = g$ almost everywhere and $M \geq 0$ then

$$\mu(\{x \mid |f(x)| > M\}) = 0 \quad \text{if and only if} \quad \mu(\{x \mid |g(x)| > M\}) = 0.$$

Hence $\|\cdot\|_\infty$ is well-defined. Furthermore, notice that $\|f\|_\infty < \infty$ for all $f \in \mathcal{L}_\infty(X, \mu)$ by the definition of an essentially bounded function. Next notice that $\|f\|_\infty \geq 0$ with equality if and only if

$$\mu\left(\left\{x \,\middle|\, |f(x)| > \frac{1}{n}\right\}\right) = 0$$

for all $n \in \mathbb{N}$ if and only if

$$\mu(\{x \mid |f(x)| > 0\}) = \mu\left(\bigcup_{n\geq 1}\left\{x \,\middle|\, |f(x)| > \frac{1}{n}\right\}\right) = 0$$

if and only if $f = 0$ almost everywhere if and only if $f = 0$ in $L_\infty(X, \mu)$.

Next let $\alpha \in \mathbb{C}$ and $f \in L_\infty(X, \mu)$ be arbitrary. If $\alpha = 0$, then clearly $\|\alpha f\|_\infty = 0 = |\alpha| \, \|f\|_\infty$. Otherwise, if $\alpha \neq 0$, we see that

$$\begin{aligned}
\|\alpha f\|_\infty &= \inf\{M \geq 0 \mid \mu(\{x \mid |\alpha f(x)| > M\}) = 0\} \\
&= \inf\left\{M \geq 0 \,\middle|\, \mu\left(\left\{x \,\middle|\, |f(x)| > \frac{M}{|\alpha|}\right\}\right) = 0\right\} \\
&= \inf\{|\alpha|M' \geq 0 \mid \mu(\{x \mid |f(x)| > M'\}) = 0\} \\
&= |\alpha| \, \|f\|_\infty
\end{aligned}$$

as desired.

Finally, to verify that $\|\cdot\|_\infty$ satisfies the triangle inequality, let $f, g \in L_\infty(X, \mu)$ be arbitrary. If $M_1, M_2 \geq 0$ are such that

$$\mu(\{x \mid |f(x)| > M_1\}) = 0 \qquad \text{and} \qquad \mu(\{x \mid |g(x)| > M_2\}) = 0,$$

the above shows that

$$\mu(\{x \mid |f(x) + g(x)| > M_1 + M_2\}) = 0.$$

Hence

$$\|f + g\|_\infty \leq M_1 + M_2.$$

Therefore, since this holds for all such M_1 and M_2, we obtain that

$$\|f + g\|_\infty \leq \|f\|_\infty + \|g\|_\infty$$

as desired. ∎

Remark 3.7.19. If $f \in L_\infty(X, \mu)$, then $\mu(\{x \in X \mid |f(x)| > \|f\|_\infty\}) = 0$. To see this, for each $n \in \mathbb{N}$ let

$$A_n = \left\{x \in X \,\middle|\, |f(x)| > \|f\|_\infty + \frac{1}{n}\right\}.$$

Note each A_n is measurable. Furthermore, by the definition of $\|f\|_\infty$, we obtain that $\mu(A_n) = 0$. Therefore, since

$$\{x \in X \mid |f(x)| > \|f\|_\infty\} = \bigcup_{n=1}^\infty A_n,$$

the claim follows by the Monotone Convergence Theorem for measures (Theorem 1.1.23) or simply the subadditivity of measures. Hence $|f(x)| \leq \|f\|_\infty$ almost everywhere.

Remark 3.7.20. If $f \in C[a,b]$, then the Extreme Value Theorem implies f is bounded. Thus, since f is Lebesgue measurable, $f \in L_\infty([a,b],\lambda)$. In addition, it is not difficult to verify that two notions of the ∞-norm (the one from Example C.1.15 and the one from Theorem 3.7.18) agree. Indeed if

$$M_0 = \sup(\{|f(x)| \mid x \in [a,b]\}) \geq 0$$

then clearly

$$\lambda(\{x \in [a,b] \mid |f(x)| > M_0\}) = 0.$$

Hence

$$\sup(\{|f(x)| \mid x \in [a,b]\}) \geq \inf\{M \geq 0 \mid \lambda(\{x \mid |f(x)| > M\}) = 0\}.$$

For the reverse inequality, assume

$$0 \leq M < \sup(\{|f(x)| \mid x \in [a,b]\}).$$

By the Extreme Value Theorem, there exists an $x_0 \in [a,b]$ such that

$$|f(x_0)| = \sup(\{|f(x)| \mid x \in [a,b]\}) > M.$$

However, if $\epsilon = \frac{1}{2}(|f(x_0)| - M) > 0$, there exists a $\delta > 0$ such that if $x \in [a,b]$ and $|x - x_0| < \delta$ then $|f(x) - f(x_0)| < \epsilon$. Hence, since $(x_0 - \delta, x_0 + \delta) \cap [a,b]$ has non-zero λ-measure and

$$|f(x)| > |f(x_0)| - \epsilon = \frac{1}{2}(|f(x_0)| + M) > M$$

for all $x \in (x_0 - \delta, x_0 + \delta) \cap [a,b]$, we see that

$$\lambda(\{x \in [a,b] \mid |f(x)| > M\}) > 0.$$

Thus it follows that

$$\sup(\{|f(x)| \mid x \in [a,b]\}) = \inf\{M \geq 0 \mid \lambda(\{x \mid |f(x)| > M\}) = 0\}.$$

as desired.

Unsurprisingly, we also have the following.

Theorem 3.7.21 (Riesz-Fisher Theorem). *Let $(X, \mathcal{A}, \mu)$ be a measure space. Then $L_\infty(X,\mu)$ is a Banach space.*

Proof. To see that $L_\infty(X,\mu)$ is a Banach space, let $(f_n)_{n\geq 1}$ be an arbitrary Cauchy sequence in $L_\infty(X,\mu)$. For each $n \in \mathbb{N}$ let

$$A_n = \{x \in X \mid |f_n(x)| > \|f_n\|_\infty\}$$

and for each $n, m \in \mathbb{N}$ let

$$B_{n,m} = \{x \in X \mid |f_n(x) - f_m(x)| > \|f_n - f_m\|_\infty\}.$$

Hence, by Remark 3.7.19, each A_n and $B_{n,m}$ are measurable for all $n, m \in \mathbb{N}$, $\mu(A_n) = 0$ for all $n \in \mathbb{N}$, and $\mu(B_{n,m}) = 0$ for all $n, m \in \mathbb{N}$. Let

$$B = \left(\bigcup_{n=1}^{\infty} A_n\right) \bigcup \left(\bigcup_{n,m=1}^{\infty} B_{n,m}\right).$$

Then B is a measurable set and $\mu(B) = 0$ since B is a countable union of μ-measure zero sets.

By replacing each f_n with $f_n \chi_{B^c}$ (which doesn't affect the equivalence classes), we may assume that $|f_n(x)| \leq \|f_n\|_\infty$ for all $x \in X$ and $n \in \mathbb{N}$, and that $|f_n(x) - f_m(x)| \leq \|f_n - f_m\|_\infty$ for all $x \in X$ and $n, m \in \mathbb{N}$. By this assumption, for each $x \in X$ we see that $(f_n(x))_{n\geq 1}$ is a Cauchy sequence in $\mathbb{C}$ and thus converges. Hence the function $f : X \to \mathbb{C}$ defined by

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

for all $x \in X$ is well-defined and measurable by Corollary 2.1.23.

We claim that $f \in \mathcal{L}_\infty(X, \mu)$ and that $(f_n)_{n\geq 1}$ converges to f in $L_\infty(X, \mu)$. To see this, notice for all $x \in X$ and $n \in \mathbb{N}$ that

$$|f(x) - f_n(x)| = \lim_{m\to\infty} |f_m(x) - f_n(x)| \leq \limsup_{m\to\infty} \|f_m - f_n\|_\infty.$$

Hence

$$\sup\{|f(x) - f_n(x)| \mid x \in X\} \leq \limsup_{m\to\infty} \|f_m - f_n\|_\infty$$

for all $n \in \mathbb{N}$. In particular

$$\begin{aligned} \sup\{|f(x) - f_1(x)| \mid x \in X\} &\leq \limsup_{m\to\infty} \|f_m - f_1\|_\infty \\ &\leq \limsup_{m\to\infty} \|f_m\|_\infty + \|f_1\|_\infty < \infty \end{aligned}$$

since Cauchy sequences are bounded. Hence by the definition of essentially bounded functions we see that $f - f_1 \in \mathcal{L}_\infty(X, \mu)$. Hence, since $f_1 \in \mathcal{L}_\infty(X, \mu)$ and $\mathcal{L}_\infty(X, \mu)$ is closed under addition, we see that $f \in \mathcal{L}_\infty(X, \mu)$. Thus the above shows that

$$\|f - f_n\|_\infty \leq \limsup_{m\to\infty} \|f_m - f_n\|_\infty$$

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

$$\lim_{n\to\infty} \limsup_{m\to\infty} \|f_m - f_n\|_\infty = 0$$

since $(f_n)_{n\geq 1}$ is Cauchy, we obtain that $(f_n)_{n\geq 1}$ converges to f in $L_\infty(X, \mu)$. Hence, as $(f_n)_{n\geq 1}$ was an arbitrary Cauchy sequence, we obtain that $L_\infty(X, \mu)$ is complete. ∎

Clearly essentially bounded functions behave like bounded functions when it comes to integration.

Theorem 3.7.22 (Hölder's Inequality). *Let $(X, \mathcal{A}, \mu)$ be a measure space. If $f \in \mathcal{L}_1(X, \mu)$ and $g \in \mathcal{L}_\infty(X, \mu)$, then $fg \in \mathcal{L}_1(X, \mu)$ and*

$$\|fg\|_1 \leq \|f\|_1 \|g\|_\infty .$$

Proof. Since $|g| \leq \|g\|_\infty$ almost everywhere by Remark 3.7.19, we obtain that

$$\|fg\|_1 = \int_X |f||g| \, d\mu \leq \int_X |f| \, \|g\|_\infty \, d\mu = \|f\|_1 \|g\|_\infty$$

as desired. ∎

Corollary 3.7.23. *Let $(X, \mathcal{A}, \mu)$ be a measure space with $\mu(X) < \infty$ and let $p \in [1, \infty)$. If $f \in L_\infty(X, \mu)$, then $f \in L_p(X, \mu)$ and*

$$\|f\|_p \leq \|f\|_\infty \mu(X)^{\frac{1}{p}}$$

Proof. Since $\mu(X) < \infty$, it is elementary to see that

$$\begin{aligned} \left(\int_X |f|^p \, d\mu \right)^{\frac{1}{p}} &\leq \left(\int_X \|f\|_\infty^p \, d\mu \right)^{\frac{1}{p}} \\ &= \left(\|f\|_\infty^p \mu(X) \right)^{\frac{1}{p}} = \|f\|_\infty \mu(X)^{\frac{1}{p}} < \infty. \end{aligned}$$

Hence $f \in L_p(X, \mu)$. ∎

To conclude this section, we note specific types of functions are dense in the L_p-spaces (well, when $p \neq \infty$).

Theorem 3.7.24. *Let $(X, \mathcal{A}, \mu)$ be a measure space. The set*

$$\mathcal{F} = \mathrm{span} \left\{ \varphi : X \to [0, \infty) \;\middle|\; {\textstyle \varphi \text{ is simple and there exists a } A \in \mathcal{A} \atop \text{such that } \mu(A) < \infty \text{ and } \varphi|_{A^c} = 0} \right\}$$

is dense in $L_p(X, \mu)$ for all $p \in [1, \infty)$.

Proof. Fix $p \in [1, \infty)$. To begin, we claim that if $\varphi : X \to [0, \infty)$ is a simple function, then $\varphi \in L_p(X, \mu)$ if and only if there exists an $A \in \mathcal{A}$ such that $\mu(A) < \infty$ and $\varphi|_{A^c} = 0$. Indeed, assume there exists an $A \in \mathcal{A}$ such that $\mu(A) < \infty$ and $\varphi|_{A^c} = 0$. Since φ is a simple function, φ is essentially bounded. Hence the proof of Corollary 3.7.23 yields

$$\|\varphi\|_p \leq \|\varphi\|_\infty^{\frac{1}{p}} \mu(A)^{\frac{1}{p}} < \infty$$

so $\varphi \in L_p(X, \mu)$.

Conversely, assume $\varphi \in L_p(X, \mu)$. Clearly if $\varphi = 0$ the result is true. Hence assume $\varphi \neq 0$. By the definition of a simple function, there exists pairwise disjoint sets $\{A_k\}_{k=1}^n \subseteq \mathcal{A}$ and elements $\{a_k\}_{k=1}^n \subseteq (0, \infty)$ such that $\varphi = \sum_{k=1}^n a_k \chi_{A_k}$ (where we have removed the characteristic function on which φ is zero). If $c = \min\{a_1, \ldots, a_n\} > 0$, then we see that

$$c^p \mu\left(\bigcup_{k=1}^n A_k\right) = c^p \sum_{k=1}^n \mu(A_k) \leq \sum_{k=1}^n a_k^p \mu(A_k) = \int_X \varphi^p \, d\mu < \infty.$$

Therefore, if $A = \bigcup_{k=1}^n A_k \in \mathcal{A}$, then $\mu(A) < \infty$ and $\varphi|_{A^c} = 0$ as desired.

To demonstrate the theorem, we must first show that $\mathcal{F} \subseteq L_p(X, \mu)$. However, this follows from the above claim as $\mathcal{F}$ is a span of elements of $L_p(X, \mu)$ and thus is a subspace of $L_p(X, \mu)$.

Finally, to show that $\mathcal{F}$ is dense in $L_p(X, \mu)$, it suffices (since $\mathcal{F}$ and $L_p(X, \mu)$ are vector spaces) to show that if $f \in L_p(X, \mu)$ and $f \geq 0$ then there exists a sequence $(\varphi_n)_{n\geq 1}$ of elements of $\mathcal{F}$ such that $\lim_{n\to\infty} \|f - \varphi_n\|_p = 0$. Indeed notice it is easy to see that the positive and negative parts of the real and imaginary parts of f are smaller than $|f|$ and thus elements of $L_p(X, \mu)$. If we can approximate each of these non-negative functions in $L_p(X, \mu)$ via elements of $\mathcal{F}$, then the triangle inequality will yield the result.

Fix $f \in L_p(X, \mu)$ such that $f \geq 0$. By Theorem 2.2.4 there exists an increasing sequence of simple functions $(\varphi_n)_{n\geq 1}$ that converge to f pointwise. Hence $0 \leq \varphi_n \leq f$ so

$$\int_X |\varphi_n|^p \, d\mu \leq \int_X |f|^p \, d\mu < \infty.$$

Hence $\varphi_n \in L_p(X, \mu)$ so $\varphi_n \in \mathcal{F}$ by the result at the beginning of the proof. Moreover, since $(|f - \varphi_n \chi_{A_n}|^p)_{n\geq 1}$ converges to zero pointwise and since

$$|f - \varphi_n \chi_{A_n}|^p \leq |f|^p \in L_1(X, \mu),$$

we obtain by the Dominated Convergence Theorem (Theorem 3.6.1) that

$$\lim_{n\to\infty} \int_X |f - \varphi_n \chi_{A_n}|^p \, d\mu = 0.$$

Hence $\lim_{n\to\infty} \|f - \varphi_n \chi_{A_n}\|_p = 0$ as desired. ∎

Theorem 3.7.25. *For all* $p \in [1, \infty)$,

$$\mathcal{C}_c(\mathbb{R}, \mathbb{C}) = \left\{ f : \mathbb{R} \to \mathbb{C} \,\middle|\, \substack{f \text{ is continuous and there exists a compact set} \\ K \subseteq \mathbb{C} \text{ such that } f|_{K^c} = 0} \right\}$$

is dense in $L_p(\mathbb{R}, \lambda)$.

Proof. By Theorem 3.7.24 we know that

$$\mathcal{F} = \operatorname{span}\left\{ \varphi : \mathbb{R} \to [0, \infty) \,\middle|\, \substack{\varphi \text{ is simple and there exists a } A \in \mathcal{M}(R) \\ \text{such that } \mu(A) < \infty \text{ and } \varphi|_{A^c} = 0} \right\}$$

is dense in $L_p(\mathbb{R}, \lambda)$. However, if $\varphi_n : \mathbb{R} \to [0, \infty)$ is simple and $\varphi \in L_p(\mathbb{R}, \lambda)$, then the end of the proof of Theorem 3.7.24 can be used to show that $\varphi\chi_{[-n,n]}$ converges to φ in $L_p(\mathbb{R}, \lambda)$. Therefore, since Corollary 3.7.23 implies that $\mathcal{C}_c(\mathbb{R}, \mathbb{C}) \subseteq L_p(\mathbb{R}, \lambda)$, to show that $\mathcal{C}_c(\mathbb{R}, \mathbb{C})$ is dense in $L_p(\mathbb{R}, \lambda)$, it suffices by the triangle inequality to show that each simple function φ such that $\varphi|_{[-n,n]^c} = 0$ for some $n \in \mathbb{N}$ can be approximated in $\|\cdot\|_p$ by an element of $\mathcal{C}_c(\mathbb{R}, \mathbb{C})$.

To see the above, let φ be an arbitrary simple function such that $\varphi|_{[-n,n]^c} = 0$ for some $n \in \mathbb{N}$ and let $\epsilon > 0$ be arbitrary. By Lusin's Theorem (Theorem 2.5.1) there exists a continuous function $f : [-n, n] \to \mathbb{C}$ such that

$$\lambda(\{x \in [-n, n] \mid f(x) \neq \varphi(x)\}) < \epsilon$$

and

$$\sup\{|f(x)| \mid x \in [-n, n]\} \leq \|\varphi\|_\infty < \infty.$$

Extend f to a continuous function $g : \mathbb{R} \to \mathbb{C}$ by defining

$$g(x) = \begin{cases} f(x) & \text{if } x \in [-n, n] \\ -\frac{f(x)}{\epsilon}(x - n) + f(x) & \text{if } x \in [n, n + \delta) \\ \frac{f(x)}{\epsilon}(x + n) + f(x) & \text{if } x \in (-n - \delta, -n] \\ 0 & \text{otherwise} \end{cases}.$$

Clearly $g \in \mathcal{C}_c(\mathbb{R}, \mathbb{C})$ and it is easy to see that $\|g\|_\infty \leq \|\varphi\|_\infty$ since we extended f to g using linear functions connecting $f(\pm n)$ to 0. Therefore, since

$$\begin{aligned}
&\int_{\mathbb{R}} |g - \varphi|^p \, d\lambda \\
&= \int_{[-n,n]} |f - \varphi|^p \, d\lambda + \int_{[n,n+\epsilon)\cup(-n-\epsilon,-n]} |g|^p \, d\lambda \\
&= \int_{\{x\in[-n,n] \mid f(x)\neq\varphi(x)\}} |f - \varphi|^p \, d\lambda + \int_{[n,n+\epsilon)\cup(-n-\epsilon,-n]} |g|^p \, d\lambda \\
&\leq \int_{\{x\in[-n,n] \mid f(x)\neq\varphi(x)\}} (2\|\varphi\|_\infty)^p \, d\lambda + \int_{[n,n+\epsilon)\cup(-n-\epsilon,-n]} \|\varphi\|_\infty^p \, d\lambda \\
&\leq (2\|\varphi\|_\infty)^p \epsilon + 2\epsilon \|\varphi\|_\infty^p \\
&= (2^p + 2) \|\varphi\|_\infty^p \, \epsilon
\end{aligned}$$

the proof is complete as $\|\varphi\|_\infty$ is fixed and $\epsilon > 0$ was arbitrary. ∎

Chapter 4

Differentiation and Integration

With our construction of integrals with respect to measures complete, we can turn our attention to studying the relation between our integral to other objects. 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 in the measure theoretic realm. Thus this section is devoted to understanding the relationship between the Lebesgue integral and differentiation of measurable functions.

After a technical lemma pertaining to covering subsets of $\mathbb{R}$ with intervals of small size, we will demonstrate that every monotone function is differentiable λ-almost everywhere and obtain a bound for the integral of the derivative. We then turn our attention to seeing if there is a version of the Fundamental Theorems of Calculus for the Lebesgue integral. In particular, if $f \in L_1([a, b], \lambda)$, what can we say about the function $F : [a, b] \to \mathbb{R}$ defined by

$$F(x) = \int_{[a,x]} f \, d\lambda?$$

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 $\lambda(J) < \operatorname{dist}(\{x\}, \bigcup_{k=1}^n J_k)$. 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 λ-almost everywhere, f' is Lebesgue measurable, $f' \geq 0$ λ-almost everywhere, and*

$$\int_{[a,b]} f' \, d\lambda \leq f(b) - f(a).$$

Proof. 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

$$\{x \in [a, b] \mid D^s f(x) \neq D^t f(x)\}$$
$$\{x \in [a, b] \mid D^s f(x) \neq D_t f(x)\}$$
$$\{x \in [a, b] \mid D_s f(x) \neq D_t f(x)\}$$

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 3.3.13, $(g_n)_{n \geq 1}$ 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 3.5.1) and Proposition 3.3.13 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 $n \in \mathbb{N}$,

$$n\int_{\left[a,a+\frac{1}{n}\right]} f\,d\lambda \geq n\int_{\left[a,a+\frac{1}{n}\right]} f(a)\,d\lambda = f(a).$$

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 $f : [a,b] \to \mathbb{R}$ is Lebesgue measurable and differentiable λ-almost everywhere, then $f' : [a,b] \to \mathbb{R}$ is Lebesgue measurable.*

Proof. For each $n \in \mathbb{N}$, let $g_n : [a,b] \to \mathbb{R}$ be defined by

$$g_n(x) = n\left(f\left(x+\frac{1}{n}\right) - f(x)\right)$$

for all $x \in [a,b]$ (where $f(y) = f(b)$ for all $y > b$). By Proposition 3.3.13 $(g_n)_{n\geq 1}$ 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 undergraduate calculus 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 $f : [a,b] \to \mathbb{C}$ is said to be of *bounded variation* if there exists an $M \in \mathbb{R}$ such that whenever $\{x_k\}_{k=0}^n$ is a partition of $[a,b]$, then

$$\sum_{k=1}^n |f(x_k) - f(x_{k-1})| \leq M.$$

Remark 4.3.2. If $f : [a,b] \to \mathbb{C}$ it is clear that f is of bounded variation if and only if $\text{Re}(f)$ and $\text{Im}(f)$ are of bounded variation. Thus we will focus on real-valued functions of bounded variation.

Example 4.3.3. 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 4.3.4. 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 4.3.5. 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 4.3.6. Let $f : [a, b] \to \mathbb{R}$ be of bounded variation. The *total variation* of f, denoted $V_f(a, b)$, is

$$V_f(a,b) = \sup\left\{ \sum_{k=1}^{n} |f(x_k) - f(x_{k-1})| \;\middle|\; \substack{n\in\mathbb{N},\\ \{x_k\}_{k=1}^n \text{ a partition of } [a,b]} \right\}.$$

If $f : [a, b] \to \mathbb{R}$ is of bounded variation, then for all $x, y \in (a, b)$ such that $x < y$ the restriction of f to $[x, y]$ is of bounded variation so $V_f(x, y)$ makes sense. Using this, we are able to prove the following.

Theorem 4.3.7 (Jordan Decomposition Theorem). *Let $f : [a, b] \to \mathbb{R}$ be of bounded variation. Define $V, D : [a, b] \to \mathbb{R}$ by $V(x) = V_f(a, x)$ (with $V(a) = 0$) and $D(x) = V(x) - f(x)$ for all $x \in [a, b]$. Then V and D are non-decreasing functions such that $f = V - D$.*

In particular, by Remark 4.3.5, 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 4.3.7), we immediately obtain information about derivatives and integrals of functions of bounded variation.

Corollary 4.3.8. *If $f : [a,b] \to \mathbb{R}$ is of bounded variation, then f is differentiable λ-almost everywhere, f' is Lebesgue measurable, and $f' \in L_1([a,b],\lambda)$.*

Proof. Since f is of bounded variation, by the Jordan Decomposition Theorem (Theorem 4.3.7) 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 $f \in L_1([a,b],\lambda)$. ∎

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 $f : [a, b] \to \mathbb{C}$ is said to be *absolutely continuous*if for all $\epsilon > 0$ there exists a $\delta > 0$ such that whenever $\{a_k\}_{k=1}^n, \{b_k\}_{k=1}^n \subseteq [a, b]$ are such that

$$a \leq a_1 < b_1 \leq a_2 < b_2 \leq \cdots \leq a_n < b_n \leq b \qquad \text{and} \qquad \sum_{k=1}^n |b_k - a_k| < \delta$$

then

$$\sum_{k=1}^n |f(b_k) - f(a_k)| < \epsilon.$$

Remark 4.4.2. Again, it is not difficult to see using the triangle inequality that a function $f : [a, b] \to \mathbb{C}$ is absolutely continuous if and only if $\mathrm{Re}(f)$ and $\mathrm{Im}(f)$ are absolutely continuous. Thus we will mainly focus on real-valued functions.

Example 4.4.3. 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 4.4.4. 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 1.4.4 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 4.4.5. *Every real-valued absolutely continuous function is continuous and of bounded variation.*

Proof. Let $f : [a, b] \to \mathbb{R}$ be absolutely continuous. It easily follows from definition that f is continuous (i.e. take $n = 1$ in Definition 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 $|z_{p_{j+1}} - z_{p_j}| < \delta$, 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 4.4.6. *If $f : [a, b] \to \mathbb{R}$ is absolutely continuous, then f is differentiable λ-almost everywhere, f' is Lebesgue measurable, and $f' \in L_1([a, b], \lambda)$.*

Proof. Since every absolutely continuous function is of bounded variation by Proposition 4.4.5, the result follows from Corollary 4.3.8. ∎

Of course, it is natural to ask whether the converse of Proposition 4.4.5 holds. To construct an example to show this is not the case, we require the following.

Proposition 4.4.7. *If $f : [a, b] \to \mathbb{R}$ is absolutely continuous and $f' = 0$ λ-almost everywhere, then f is constant.*

Proof. To see that f is constant on $[a, b]$, let $c \in (a, b]$ be arbitrary. We claim that $f(c) = f(a)$.

To see this, let $\epsilon > 0$ and recall that since $f' = 0$ λ-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 4.4.8. 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 2.1.8 and thus uniformly continuous $[0, 1]$ and of bounded variation. The fact that f is not absolutely continuous follows from Proposition 4.4.7 along with the fact that f is non-constant yet $f' = 0$ almost everywhere.

To conclude this section, we desire to show that functions defined by integrating against an L_1-function are absolutely continuous and thus the collection of absolutely continuous functions include those defined in a 'Fundamental Theorem of Calculus'-like manner. This is achieved via the following lemma.

Lemma 4.4.9. *Let $(X, \mathcal{A}, \mu)$ be a measure space and let $f \in L_1(X, \mu)$. Then for all $\epsilon > 0$ there exists a $\delta > 0$ such that if $A \in \mathcal{A}$ and $\mu(A) < \delta$ then*

$$\int_A |f| \, d\mu < \epsilon.$$

Proof. 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 4.4.10. *Let $f \in L_1([a,b], \lambda)$. If $F : [a,b] \to \mathbb{C}$ is defined by*

$$F(x) = \int_{[a,x]} f \, d\lambda$$

for all $x \in [a,b]$, then F is absolutely continuous.

Proof. First notice that F is well-defined as $f \in L_1([a,b], \lambda)$.

To see that F is absolutely continuous, let $\epsilon > 0$. Since f is Lebesgue integrable, by Lemma 4.4.9 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 4.4.10 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 4.4.10 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 $f \in L_1([a,b],\lambda)$ be real-valued and define $F : [a,b] \to \mathbb{R}$ by*

$$F(x) = \int_{[a,x]} f \, d\lambda$$

for all $x \in [a,b]$. If F is non-decreasing, then $f(x) \geq 0$ for almost every x.

Proof. Let

$$X = \{x \in [a,b] \mid f(x) < 0\},$$

which is a 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 1.4.12, 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 3.6.1) that

$$\begin{aligned} F(b) - F(a) &< \int_V f \, d\lambda \\ &= \lim_{n\to\infty} \int_{\mathbb{R}} \sum_{k=1}^{n} f_k \, d\lambda \\ &= \lim_{n\to\infty} \sum_{k=1}^{n} F(b_k) - F(a_k) \\ &\leq F(b) - F(a) \end{aligned}$$

since F is non-decreasing. As this clearly is a contradiction, we obtain that $\lambda(X) = 0$ as desired. ∎

Corollary 4.5.2. *Let $f \in L_1([a, b], \lambda)$ be real-valued and define $F : [a, b] \to \mathbb{R}$ by*

$$F(x) = \int_{[a,x]} f \, d\lambda$$

for all $x \in [a, b]$. If $F(x) = 0$ for all $x \in [a, b]$, then $f = 0$ λ-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 $f = 0$ λ-almost everywhere. ∎

Using all of the above, we arrive at our Fundamental Theorems of Calculus which complete characterize absolutely continuous functions.

Theorem 4.5.3 (Fundamental Theorem of Calculus, I). *Let $f \in L_1([a, b], \lambda)$ be real-valued. If $F : [a, b] \to \mathbb{R}$ is defined by*

$$F(x) = \int_{[a,x]} f \, d\lambda$$

for all $x \in [a, b]$, then F' exists almost everywhere and $F'(x) = f(x)$ for almost every x.

Proof. To begin, note F is absolutely continuous (and thus Lebesgue measurable) by Proposition 4.4.10. Hence F' exists λ-almost everywhere and is Lebesgue integrable by Corollary 4.4.6. To demonstrate that $F' = f$ λ-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

$$F_n(x) = n\left(F\left(x + \frac{1}{n}\right) - F(x)\right) = n \int_{\left[x, x+\frac{1}{n}\right]} f \, d\lambda$$

for all $x \in [a, b]$. Clearly each F_n is a Lebesgue measurable function by Proposition 1.4.8 since F is Lebesgue measurable. Furthermore, notice for each $n \in \mathbb{N}$ and $x \in [a, b]$ that

$$|F_n(x)| \leq n \int_{\left[x, x+\frac{1}{n}\right]} |f| \, d\lambda \leq n\left(\frac{1}{n}M\right) = M.$$

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 3.6.1) that

$$\int_{[a,c]} F' \, d\lambda = \lim_{n\to\infty} \int_{[a,c]} F_n \, d\lambda$$

for all $c \in [a, b]$. Hence

$$\begin{aligned}\int_{[a,c]} F' \, d\lambda &= \lim_{n\to\infty} n \int_{[a,c]} F\left(x + \frac{1}{n}\right) - F(x) \, d\lambda(x) \\ &= \lim_{n\to\infty} n\left(\int_{\left[c, c+\frac{1}{n}\right]} F \, d\lambda - \int_{\left[a, a+\frac{1}{n}\right]} F \, d\lambda\right)\end{aligned}$$

for all $c \in [a, b]$.

We claim that

$$\lim_{n\to\infty} n \int_{\left[c, c+\frac{1}{n}\right]} F \, d\lambda = F(c)$$

for all $c \in [a, b]$. 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 $F' - f = 0$ λ-almost everywhere. Hence $F' = f$ λ-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 $F' = f$ λ-almost everywhere by Theorem 3.1.11.

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

$$F' = F_1' - F_2' = f_1 - f_2 = f$$

λ-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 $F' \in L_1([a, b], \lambda)$ and*

$$F(x) = F(a) + \int_{[a,x]} F' \, d\lambda$$

for all $x \in [a, b]$.

Proof. To begin, recall that if $F : [a, b] \to \mathbb{R}$ is absolutely continuous, then F is differentiable almost everywhere with $F' \in L_1([a, b], \lambda)$ by Corollary 4.4.6. Define $G : [a, b] \to \mathbb{R}$ by

$$G(x) = \int_{[a,x]} F' \, d\lambda$$

for all $x \in [a, b]$. Then G is absolutely continuous by Proposition 4.4.10 and $G' = F'$ λ-almost everywhere by the First Fundamental Theorem of Calculus (Theorem 4.5.3). Thus $F - G$ is absolutely continuous and

$$(F - G)' = F' - G' = 0$$

almost everywhere. Hence Proposition 4.4.7 implies that $F - G$ is constant. Therefore, as $(F - G)(a) = F(a)$, we obtain that $F(x) - G(x) = F(a)$ for all $x \in [a, b]$ so

$$F(x) = F(a) + \int_{[a,x]} F' \, d\lambda$$

for all $x \in [a, b]$ as desired. ∎

Chapter 5

Signed Measures

We have seen that integrating L_1-functions produces exactly the class of absolutely continuous functions and integrating a positive measurable function against a measure μ produces a new measure ν with specific properties (see Corollary 3.2.6). Thus it is natural to ask, "What objects do we get by integrating L_1-functions against a measure μ?" Clearly such an object is a function on a σ-algebra that need not take only positive values and thus is not a measure.

To resolve this situation, we will extend our notion of a measure in this section. In particular, this section will focus on "real-valued measures" and the class of "complex-valued measures" easily follows and will be left as homework. After developing the theory of "real-valued measures", we will be able to completely describe the collection of "real-valued measures" that can be obtained by integration against a real-valued L_1-function. Moreover, given any two σ-finite measures μ and ν, we will demonstrate we can always write ν as a sum of a measure obtained by integrating a positive measurable function against μ and a measure that is "orthogonal" to μ.

5.1 Signed Measures

To begin, we extend our notion of a measure to allow for negative values. Since we allow measures to obtain the value ∞, we must allow our new notion of measures to obtain the value $-\infty$. However, since we desire a notion of countably additivity, we will not permit both $\pm\infty$ to be obtained.

Definition 5.1.1. Let $(X, \mathcal{A})$ be a measurable space. A function $\nu : \mathcal{A} \to [-\infty, \infty]$ is said to be a *signed measure* on $(X, \mathcal{A})$ if

1. $\nu(\emptyset) = 0$,

2. the range of ν is contained in either $[-\infty, \infty)$ or $(-\infty, \infty]$, and

3. if $\{A_n\}_{n=1}^\infty \subseteq \mathcal{A}$ are pairwise disjoint, then $\nu\left(\bigcup_{n=1}^\infty A_n\right) = \sum_{n=1}^\infty \nu(A_n)$ in the following sense:
   - If $\nu\left(\bigcup_{n=1}^\infty A_n\right) = \pm\infty$, then $\sum_{n=1}^\infty \nu(A_n)$ diverges to $\pm\infty$.
   - If $\left|\nu\left(\bigcup_{n=1}^\infty A_n\right)\right| < \infty$, then $\sum_{n=1}^\infty \nu(A_n)$ converges absolutely to $\nu\left(\bigcup_{n=1}^\infty A_n\right)$.

Remark 5.1.2. The reason we require "$\sum_{n=1}^\infty \nu(A_n)$ converges absolutely to $\nu\left(\bigcup_{n=1}^\infty A_n\right)$" in the case that $\left|\nu\left(\bigcup_{n=1}^\infty A_n\right)\right| < \infty$ is that series that do not converge absolutely (i.e. converge conditionally) can be rearranged to converge to any real number and can be rearranged to diverge to $\pm\infty$. As such, we need $\sum_{n=1}^\infty \nu(A_n)$ to converges absolutely in order to make sense of $\nu\left(\bigcup_{n=1}^\infty A_n\right)$.

Example 5.1.3. Clearly any positive measure is a signed measure. Similarly, if $\mu_1, \ldots, \mu_n$ are finite positive measures on a measurable space $(X, \mathcal{A})$ and $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$, then $\sum_{k=1}^n \alpha_k \mu_k$ is a signed measure on $(X, \mathcal{A})$.

Recall our motivation for signed measures is the following example.

Example 5.1.4. Let $(X, \mathcal{A}, \mu)$ be a measure space and let $f \in L_1(X, \mu)$ be real-valued. Define $\nu : \mathcal{A} \to [-\infty, \infty]$ by

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

for all $A \in \mathcal{A}$. We claim that ν is a signed measure. Indeed clearly $\nu(\emptyset) = 0$ and $\nu : \mathcal{A} \to (-\infty, \infty)$ since $f \in L_1(X, \mu)$. Moreover, since the final property in Definition 5.1.1 is precisely demonstrated in Corollary 3.6.4, the claim follows.

Remark 5.1.5. Notice all of the above examples of signed measures are obtained via linear combinations of positive measures. Indeed if ν is the measure from Example 5.1.4, then

$$\nu(A) = \int_A f \, d\mu = \int_A f_+ \, d\mu - \int_A f_- \, d\mu$$

for all $A \in \mathcal{A}$. Since $\mu_+, \mu_- : \mathcal{A} \to [0, \infty)$ defined by $\mu_\pm(A) = \int_A f_\pm \, d\mu$ are positive measures by Corollary 3.2.6, ν is the difference of two positive measures. This is not a coincidence as will be demonstrated in subsequent sections.

5.2 The Hahn Decomposition Theorem

To begin our analysis of signed measure and decomposing them as a linear combination of positive measures, notice in Remark 5.1.5 that ν was described as the difference of two positive measures. However, if $P = \{x \mid f(x) \geq 0\}$ and $N = \{x \mid f(x) \leq 0\}$, then for all $A \subseteq P$ and $B \subseteq N$ we see that $\nu(A) \geq 0$ and $\nu(B) \leq 0$. Such sets are essential to understanding signed measures and are described as follows.

Definition 5.2.1. Let $(X, \mathcal{A})$ be a measurable space and let $\nu : \mathcal{A} \to [-\infty, \infty]$ be a signed measure on X. A set $P \in \mathcal{A}$ is said to be *positive* for ν if $\nu(B) \geq 0$ whenever $B \in \mathcal{A}$ and $B \subseteq P$.

Similarly, a set $N \in \mathcal{A}$ is said to be *negative* for ν if $\nu(B) \leq 0$ whenever $B \in \mathcal{A}$ and $B \subseteq N$.

Finally, a set $A \in \mathcal{A}$ is said to be *null* for ν if $\nu(B) = 0$ whenever $B \in \mathcal{A}$ and $B \subseteq A$ (that is, a set is null if and only if it is both positive and negative).

Example 5.2.2. Clearly the empty set is a null set for every signed measure.

Example 5.2.3. Let $X = [-\pi, \pi]$ and define $\nu : \mathcal{M}(\mathbb{R}) \to (-\infty, \infty)$ by

$$\nu(A) = \int_A \sin(x)\, d\lambda(x)$$

for all $A \in \mathcal{M}(\mathbb{R})$. Then $[0, \pi]$ is a positive set. Indeed notice $\sin(x) \geq 0$ for all $x \in [0, \pi]$. Therefore if $B \in \mathcal{M}(\mathbb{R})$ and $B \subseteq [0, \pi]$ then

$$\nu(B) = \int_B \sin(x)\, d\lambda(x) \geq 0.$$

Similarly $[-\pi, 0]$ is a negative set. However, $[-\pi, \pi]$ is not positive, negative, nor null even though $\nu([-\pi, \pi]) = 0$ as $\nu([0, \pi]) > 0$ yet $\nu([-\pi, 0]) < 0$.

Example 5.2.4. More generally, let $(X, \mathcal{A}, \mu)$ be a measure space and let $f \in L_1(X, \mu)$ be real-valued. By Example 5.1.4 if we define $\nu : \mathcal{A} \to [-\infty, \infty]$ by

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

for all $A \in \mathcal{A}$, then ν is a signed measure on $(X, \mathcal{A})$. It is not difficult to see that $P = \{x \in X \mid f(x) \geq 0\}$ is a positive set for ν, $N = \{x \in X \mid f(x) < 0\}$ is a negative set for ν, and $\{x \in X \mid f(x) = 0\}$ is a null set for μ.

Our first goal with respect to signed measures is to demonstrate that there are 'large' positive and negative sets. In particular, notice if P and N are as in Example 5.2.4, then $P \cup N = X$ whereas $P \cap N = \emptyset$. The Hahn Decomposition Theorem (Theorem 5.2.7) will extend this idea to any signed measure. However, first we need two lemmas; the first showing we can combine positive sets to get a positive set, and the second showing we can extract a positive set from a set of positive measure.

Lemma 5.2.5. *Let $(X, \mathcal{A})$ be a measurable space and let ν be a signed measure on $(X, \mathcal{A})$. If $\{A_n\}_{n=1}^\infty \subseteq \mathcal{A}$ are positive sets for ν, then $\bigcup_{n=1}^\infty A_n$ is a positive set for ν. Similarly, if $\{A_n\}_{n=1}^\infty \subseteq \mathcal{A}$ are negative sets for ν, then $\bigcup_{n=1}^\infty A_n$ is a negative set for ν.*

Proof. Assume $\{A_n\}_{n=1}^\infty \subseteq \mathcal{A}$ are positive sets for ν. To see that $A = \bigcup_{n=1}^\infty A_n$ is positive, let $B \in \mathcal{A}$ be such that $B \subseteq A$. Let $B_1 = B \cap A_1$ and for each $n \geq 2$ define

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

Clearly $\{B_k\}_{k=1}^n$ are pairwise disjoint elements of $\mathcal{A}$ such that $B_n \subseteq A_n$ for all $n \in \mathbb{N}$ and $B = \bigcup_{n=1}^\infty B_n$ (since $B \subseteq A$). Since each A_n is positive and $B_n \subseteq A_n$ for all $n \in \mathbb{N}$, we obtain that $\nu(B_n) \geq 0$ for all $n \in \mathbb{N}$. Therefore, as $\{B_k\}_{k=1}^n$ are pairwise disjoint, we obtain that

$$\nu(B) = \nu\left(\bigcup_{n=1}^\infty B_n \right) = \sum_{n=1}^\infty \nu(B_n) \geq 0.$$

Therefore, since B was arbitrary, we obtain that A is positive.

The proof in the case that $\{A_n\}_{n=1}^\infty \subseteq \mathcal{A}$ are negative sets for ν is obtained by reversing all inequalities. ∎

Lemma 5.2.6. *Let $(X, \mathcal{A})$ be a measure space and let ν be a signed measure on $(X, \mathcal{A})$. If $A \in \mathcal{A}$ and $\nu(A) > 0$, then there exists a positive set $P \subseteq A$ such that $\nu(P) > 0$.*

Proof. Let $A \in \mathcal{A}$ be such that $\nu(A) > 0$. If A is positive, then there is nothing to prove. Hence we may assume that there exists a $B \in \mathcal{A}$ such that $B \subseteq A$ and $\nu(B) < 0$. Clearly this implies there exists an $m \in \mathbb{N}$ and a $B \in \mathcal{A}$ such that $B \subseteq A$ and $\nu(B) < -\frac{1}{m}$. Hence there exists a least natural number $m_1 \in \mathbb{N}$ such that there exists a $B \in \mathcal{A}$ such that $B \subseteq A$ and $\nu(B) < -\frac{1}{m_1}$. Choose $B_1 \in \mathcal{A}$ such that $B_1 \subseteq A$ and $\nu(B_1) < -\frac{1}{m_1}$.

Proceeding recursively, assume we have constructed $B_1, \ldots, B_n \in \mathcal{A}$ and $m_1, \ldots, m_n \in \mathbb{N}$ such that $m_k \in \mathbb{N}$ is the least natural number such that there exists a $B \in \mathcal{A}$ such that $B \subseteq A \setminus \left(\bigcup_{j=1}^{k-1} B_j \right)$ and $\nu(B) < -\frac{1}{m_k}$, $B_k \subseteq A \setminus \left(\bigcup_{j=1}^{k-1} B_j \right)$, and $\nu(B_k) < -\frac{1}{m_k}$ for all $k \in \{1, \ldots, n\}$. Since $\{B_k\}_{k=1}^n$ are pairwise disjoint, notice that

$$\begin{aligned} 0 < \nu(A) &= \nu\left(A \setminus \left(\bigcup_{k=1}^n B_k \right) \right) + \nu\left(\bigcup_{k=1}^n B_k \right) \\ &= \nu\left(A \setminus \left(\bigcup_{j=1}^n B_j \right) \right) + \sum_{k=1}^n \nu(B_k). \end{aligned}$$

Therefore, since $\nu(B_k) < 0$ for all $k \in \{1, \ldots, n\}$, we obtain that

$$\nu\left(A \setminus \left(\bigcup_{k=1}^n B_k \right) \right) > 0.$$

Therefore, if $A \setminus (\bigcup_{k=1}^n B_k)$ is positive, the proof is complete. Otherwise there exists a least natural number $m_{n+1} \in \mathbb{N}$ such that there exists a $B \in \mathcal{A}$ such that $B \subseteq A \setminus (\bigcup_{k=1}^n B_j)$ and $\nu(B) < -\frac{1}{m_{n+1}}$. Choose $B_{n+1} \in \mathcal{A}$ such that $B_{n+1} \subseteq A \setminus (\bigcup_{k=1}^n B_j)$ and $\nu(B_{n+1}) < -\frac{1}{m_{n+1}}$.

The above recursive process thereby either completes the proof or produces a collection $\{B_n\}_{n=1}^\infty \subseteq \mathcal{A}$ and a sequence $(m_n)_{n\geq 1}$ of natural numbers such that m_n is the least natural number such that there exists a $B \in \mathcal{A}$ such that $B \subseteq A \setminus \left(\bigcup_{k=1}^{n-1} B_k\right)$ and $\nu(B) < -\frac{1}{m_n}$, $B_n \subseteq A \setminus \left(\bigcup_{j=1}^{n-1} B_j\right)$, and $\nu(B_n) < -\frac{1}{m_n}$ for all $n \in \mathbb{N}$.

Let

$$P = A \setminus \left(\bigcup_{n=1}^\infty B_n\right) \in \mathcal{A}.$$

Hence

$$0 < \nu(A) = \nu(P) + \nu\left(\bigcup_{n=1}^\infty B_n\right).$$

Therefore, it must be the case that $\nu(\bigcup_{n=1}^\infty B_n) \neq -\infty$. Thus, since $\nu(B_n) < 0$ for all $n \in \mathbb{N}$, it must be the case that $\sum_{n=1}^\infty \nu(B_k)$ converges absolutely and

$$0 < \nu(A) = \nu(P) + \nu\left(\bigcup_{n=1}^\infty B_n\right) = \nu(P) + \sum_{n=1}^\infty \nu(B_k).$$

Moreover, since $\sum_{n=1}^\infty \nu(B_k)$ converges absolutely and since $\nu(B_n) < -\frac{1}{m_n}$ for all $n \in \mathbb{N}$, we obtain that $\lim_{n\to\infty} m_n = \infty$.

We claim that P is positive. To see that P is positive, suppose for the sake of a contradiction that there exists a $B' \in \mathcal{A}$ such that $B' \subseteq P$ and $\nu(B') < 0$. Since $\lim_{n\to\infty} m_n = \infty$, there exists an $N \in \mathbb{N}$ such that $\nu(B') > -\frac{1}{m_N - 1}$. However, since $m_N - 1 < m_N$ and $B' \subseteq P \subseteq A \setminus \left(\bigcup_{n=1}^{N-1} B_n\right)$, the inequality $\nu(B') < -\frac{1}{m_N-1}$ contradicts the fact that m_N is the least natural number such that there exists a $B \in \mathcal{A}$ such that $B \subseteq A \setminus \left(\bigcup_{k=1}^{N-1} B_k\right)$ and $\nu(B) < -\frac{1}{m_N}$. Therefore P is positive. ∎

Using the above, we obtain our first vital step towards understanding signed measures.

Theorem 5.2.7 (Hahn Decomposition Theorem). *Let $(X, \mathcal{A})$ be a measurable space. If ν is a signed measure on $(X, \mathcal{A})$, then there exists a positive set P and a negative set N for ν such that $X = P \cup N$ and $P \cap N = \emptyset$.*

Proof. Recall that if ν is a signed measure on $(X, \mathcal{A})$, then $-\nu$ is also a signed measure on $(X, \mathcal{A})$. Furthermore, it is elementary to see that a set $A \in \mathcal{A}$ is positive (respectively negative) for ν if and only if A is negative (respectively positive) for $-\nu$. Therefore, by replacing ν with $-\nu$ if necessary, we may assume that $\nu : \mathcal{A} \to [-\infty, \infty)$.

Let

$$\alpha = \sup\{\nu(A) \mid A \in \mathcal{A}, A \text{ positive for } \nu\}.$$

Since $\emptyset$ is a positive set for μ, we obtain that $\alpha \geq 0$ (i.e. the supremum is not over an empty set).

Choose a sequence $(A_n)_{n\geq 1}$ of positive sets for ν such that $\lim_{n\to\infty} \nu(A_n) = \alpha$ and let

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

Then P is clearly an element of $\mathcal{A}$ that is positive for μ by Lemma 5.2.5. Hence $\nu(P) \leq \alpha$ by the definition of α. However, since for each $n \in \mathbb{N}$ we have

$$\nu(P \setminus A_n) \geq 0$$

as $P \setminus A_n \subseteq P$ and P is positive, we obtain that

$$\nu(P) = \nu(A_n) + \nu(P \setminus A_n) \geq \nu(A_n)$$

for all $n \in \mathbb{N}$. Therefore, since $\lim_{n\to\infty} \nu(A_n) = \alpha$, we obtain that $\nu(P) = \alpha$. Hence $\alpha \neq \infty$.

Let $N = X \setminus P$. To complete the proof, it suffices to show that N is negative. To see this, suppose for the sake of a contradiction that N is not negative. Hence there exists a $B \in \mathcal{A}$ such that $B \subseteq N$ and $\nu(B) > 0$. Since $\nu(B) > 0$, Lemma 5.2.6 implies there exists a $P_0 \in \mathcal{A}$ such that $P_0 \subseteq B$, $\nu(P_0) > 0$, and P_0 is positive for ν. Since $P_0 \subseteq B \subseteq N$, we see that $P \cap P_0 = \emptyset$ and thus

$$\nu(P \cup P_0) = \nu(P) + \nu(P_0) > \nu(P) = \alpha.$$

However, since $P \cup P_0$ is a positive subset for ν by Lemma 5.2.5, $\nu(P \cup P_0) > \alpha$ contradicts the definition of α. Hence it must have been the case that N is negative as desired. ∎

5.3 The Jordan Decomposition Theorem

Using the Hahn Decomposition Theorem (Theorem 5.2.7), it is now not difficult to completely characterize all signed measures using positive measures.

Theorem 5.3.1 (Jordan Decomposition Theorem). *Let $(X, \mathcal{A})$ be a measurable space. If ν is a signed measure on $(X, \mathcal{A})$, then there exists measures $\nu_+, \nu_- : \mathcal{A} \to [0, \infty]$ such that $\nu(A) = \nu_+(A) - \nu_-(A)$ for all $A \in \mathcal{A}$.*

Proof. Let ν be a signed measure on $(X, \mathcal{A})$. By the Hahn Decomposition Theorem (Theorem 5.2.7), there exists a positive set P and a negative set N for ν such that $X = P \cup N$ and $P \cap N = \emptyset$.

Define $\nu_+, \nu_- : \mathcal{A} \to [0, \infty]$ by

$$\nu_+(A) = \nu(A \cap P) \qquad \text{and} \qquad \nu_-(A) = -\nu(A \cap N)$$

for all $A \in \mathcal{A}$. Clearly $\nu_+(A), \nu_-(A) \in [0, \infty]$ for all $A \in \mathcal{A}$ since P is a positive set for ν and N is a negative set for ν. Furthermore, since $X = P \cup N$ and $P \cap N = \emptyset$, we see for all $A \in \mathcal{A}$ that

$$\nu(A) = \nu((A \cap P) \cup (A \cap N)) = \nu(A \cap P) + \nu(A \cap N) = \nu_+(A) - \nu_-(A).$$

Finally the fact that ν_+ and ν_- are measures follows from the same arguments as used in Example 1.1.15. ∎

One question that arises from the Jordan Decomposition Theorem (Theorem 5.3.1 is whether or not the decomposition obtained is unique. In general, these measures need not unique (especially if the measure ν has some nice isolated set so we can add a value to $\nu_\pm$ on this set). However, based on their construction, the measures $\nu_\pm$ have an additional property.

Definition 5.3.2. Let $(X, \mathcal{A})$ be a measure space. Two signed measures ν_1 and ν_2 on $(X, \mathcal{A})$ are said to be *mutually singular*, denoted $\nu_1 \perp \nu_2$, if there exists sets $A_1, A_2 \in \mathcal{A}$ such that $X = A_1 \cup A_2$, $A_1 \cap A_2 = \emptyset$, A_1 is null for ν_1, and A_2 is null for ν_2.

Remark 5.3.3. Since null sets for measures are just sets on which the measure vanishes, we see that two measures ν_1 and ν_2 on $(X, \mathcal{A})$ are mutually singular if and only if there exists $A_1, A_2 \in \mathcal{A}$ such that $X = A_1 \cup A_2$, $A_1 \cap A_2 = \emptyset$, and $\nu_1(A_1) = \nu_2(A_2) = 0$.

Example 5.3.4. Recall if ν is a signed measure on $(X, \mathcal{A})$, then the proof of the Jordan Decomposition Theorem (Theorem 5.3.1) demonstrates that there exists a positive set P and a negative set N for ν such that $X = P \cup N$, $P \cap N = \emptyset$, and if $\nu_+, \nu_- : \mathcal{A} \to [0, \infty]$ are defined by

$$\nu_+(A) = \nu(A \cap P) \qquad \text{and} \qquad \nu_-(A) = -\nu(A \cap N)$$

for all $A \in \mathcal{A}$, then ν_+ and ν_- are measures. Since clearly $\nu_+(N) = 0 = \nu_-(P)$, ν_+ and ν_- are mutually singular measures.

Example 5.3.5. Let $x \in \mathbb{R}$ be arbitrary. Then δ_x, the point mass measure at x from Example 1.1.12, and the Lebesgue measure are mutually singular measure on the Lebesgue measurable sets since $\{x\}, \mathbb{R} \setminus \{x\}$ are Lebesgue measurable sets that are disjoint with union $\mathbb{R}$ such that $\delta_x(\mathbb{R} \setminus \{x\}) = 0 = \lambda(\{x\})$.

By adding the condition that the resulting measures in the Jordan Decomposition Theorem (Theorem 5.3.1) must be singular, we obtain the decomposition produced is unique.

Proposition 5.3.6. *Let $(X, \mathcal{A})$ be a measurable space. If ν is a signed measure on $(X, \mathcal{A})$, then there exists a unique pair (ν_+, ν_-) of mutually singular measures on $(X, \mathcal{A})$ such that $\nu(A) = \nu_+(A) - \nu_-(A)$ for all $A \in \mathcal{A}$.*

Proof. Let ν be a signed measure on $(X, \mathcal{A})$. By the Jordan Decomposition Theorem (Theorem 5.3.1) and Example 5.3.4 there exists a pair (ν_+, ν_-) of mutually singular measures on $(X, \mathcal{A})$ such that $\nu(A) = \nu_+(A) - \nu_-(A)$ for all $A \in \mathcal{A}$.

To see uniqueness, assume there exists a pair (ν_1, ν_2) of mutually singular measures on $(X, \mathcal{A})$ such that $\nu(A) = \nu_1(A) - \nu_2(A)$ for all $A \in \mathcal{A}$. By assumptions, there exists sets $P, N, B, C \in \mathcal{A}$ such that $X = P \cup N = B \cup C$, $P \cap N = B \cap C = \emptyset$, and

$$\nu_+(N) = \nu_-(P) = \nu_1(C) = \nu_2(B) = 0.$$

We desire to show that $\nu_+ = \nu_1$ and $\nu_- = \nu_2$.

To see that $\nu_+(A) = \nu_1(A)$ and $\nu_-(A) = \nu_2(A)$ for all $A \in \mathcal{A}$, first assume $A \in \mathcal{A}$ is such that $A \subseteq P \cap B$. Then $\nu_-(A) = 0$ since $A \subseteq P$ and $\nu_2(A) = 0$ since $A \subseteq B$. Hence $\nu_-(A) = \nu_2(A)$ and

$$\nu_+(A) = \nu_+(A) - \nu_-(A) = \nu(A) = \nu_1(A) - \nu_2(A) = \nu_1(A)$$

as desired.

Next assume $A \in \mathcal{A}$ is such that $A \subseteq P \cap C$. Then $\nu_-(A) = 0$ since $A \subseteq P$ and $\nu_1(A) = 0$ since $A \subseteq C$. Hence

$$\nu_+(A) = \nu_+(A) - \nu_-(A) = \nu(A) = \nu_1(A) - \nu_2(A) = -\nu_2(A).$$

However, since $\nu_+(A) \geq 0$ and $-\nu_2(A) \leq 0$, it must be the case that $\nu_+(A) = \nu_2(A) = 0$. Hence $\nu_1(A) = \nu_+(A) = 0 = \nu_2(A) = \nu_-(A)$ in this case.

Next assume $A \in \mathcal{A}$ is such that $A \subseteq N \cap B$. Then $\nu_+(A) = 0$ since $A \subseteq N$ and $\nu_2(A) = 0$ since $A \subseteq B$. Hence

$$-\nu_-(A) = \nu_+(A) - \nu_-(A) = \nu(A) = \nu_1(A) - \nu_2(A) = \nu_1(A).$$

However, since $\nu_1(A) \geq 0$ and $-\nu_-(A) \leq 0$, it must be the case that $\nu_-(A) = \nu_1(A) = 0$. Hence $\nu_1(A) = \nu_+(A) = 0 = \nu_2(A) = \nu_-(A)$ in this case.

Next assume $A \in \mathcal{A}$ is such that $A \subseteq N \cap C$. Then $\nu_+(A) = 0$ since $A \subseteq N$ and $\nu_1(A) = 0$ since $A \subseteq C$. Hence $\nu_+(A) = \nu_1(A)$ and

$$-\nu_-(A) = \nu_+(A) - \nu_-(A) = \nu(A) = \nu_1(A) - \nu_2(A) = -\nu_2(A)$$

as desired.

Finally, let $A \in \mathcal{A}$ be arbitrary. Then

$$\{A \cap P \cap B, A \cap P \cap C, A \cap N \cap B, A \cap N \cap C\}$$

are pairwise disjoint elements of $\mathcal{A}$ such that

$$A = (A \cap P \cap B) \cup (A \cap P \cap C) \cup (A \cap N \cap B) \cup (A \cap N \cap C).$$

Hence, by using the above four cases to each of these sets, we obtain that

$$\begin{aligned}
&\nu_+(A) \\
&= \nu_+(A \cap P \cap B) + \nu_+(A \cap P \cap C) + \nu_+(A \cap N \cap B) + \nu_+(A \cap N \cap C) \\
&= \nu_1(A \cap P \cap B) + \nu_1(A \cap P \cap C) + \nu_1(A \cap N \cap B) + \nu_1(A \cap N \cap C) \\
&= \nu_1(A)
\end{aligned}$$

and

$$\begin{aligned}
&\nu_-(A) \\
&= \nu_-(A \cap P \cap B) + \nu_-(A \cap P \cap C) + \nu_-(A \cap N \cap B) + \nu_-(A \cap N \cap C) \\
&= \nu_2(A \cap P \cap B) + \nu_2(A \cap P \cap C) + \nu_2(A \cap N \cap B) + \nu_2(A \cap N \cap C) \\
&= \nu_2(A).
\end{aligned}$$

Therefore, since $A \in \mathcal{A}$ was arbitrary, $\nu_+ = \nu_1$ and $\nu_- = \nu_2$ as desired. ∎

Due to Proposition 5.3.6, we make the following definition.

Definition 5.3.7. Let $(X, \mathcal{A})$ be a measurable space and let ν be a signed measure on $(X, \mathcal{A})$. The *positive* and *negative parts of* ν, denoted ν_+ and ν_- respectively, is the unique pair of mutually singular measures on $(X, \mathcal{A})$ such that $\nu(A) = \nu_+(A) - \nu_-(A)$ for all $A \in \mathcal{A}$.

Example 5.3.8. Let $(X, \mathcal{A}, \mu)$ be a measure space and let $f \in L_1(X, \mu)$ be real-valued. Recall from Example 5.1.4 that if we define $\nu : \mathcal{A} \to (-\infty, \infty)$ by

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

for all $A \in \mathcal{A}$, then ν is a signed measure on $(X, \mathcal{A})$. We claim that if we define $\nu_+, \nu_- : \mathcal{A} \to [0, \infty)$ by

$$\nu_+(A) = \int_A f_+ \, d\mu \qquad \text{and} \qquad \nu_- = \int_A f_- \, d\mu$$

for all $A \in \mathcal{A}$, then ν_+ and ν_- are the positive and negative parts of ν respectively. To see this, first note that ν_+ and ν_- are finite measures by Corollary 3.2.6 and the fact that $f_+, f_- \in L_1(X, \mu)$. Clearly $\nu = \nu_+ - \nu_-$. To see that ν_+ and ν_- are mutually singular, let

$$P = \{x \in X \mid f(x) > 0\} = \{x \in X \mid f_+(x) > 0, f_-(x) = 0\}.$$

Clearly $P \in \mathcal{A}$ since f is measurable. Since

$$\nu_-(P) = \int_P f_- \, d\mu = 0 \qquad \text{and} \qquad \nu_+(X \setminus P) = \int_{X \setminus P} f_+ \, d\mu = 0$$

we obtain that ν_+ and ν_- are mutually singular as desired.

5.4 Finite Signed Measures

Based on the uniqueness of the Jordan Decomposition, given any signed measure there is a very natural associated positive measure that reveals substantial information about the signed measure.

Definition 5.4.1. Let $(X, \mathcal{A})$ be a measurable space and let ν be a signed measure on $(X, \mathcal{A})$. The *total variation* (or *absolute value*) of ν, denoted $|\nu|$, is the positive measure on $(X, \mathcal{A})$ defined by

$$|\nu|(A) = \nu_+(A) + \nu_-(A)$$

for all $A \in \mathcal{A}$, where (ν_+, ν_-) are the positive and negative parts of ν respectively.

Example 5.4.2. Let $(X, \mathcal{A}, \mu)$ be a measure space and let $f \in L_1(X, \mu)$ be real-valued. Recall from Example 5.3.8 that if we define $\nu : \mathcal{A} \to (-\infty, \infty)$ by

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

for all $A \in \mathcal{A}$ and we define $\nu_+, \nu_- : \mathcal{A} \to [0, \infty)$ by

$$\nu_+(A) = \int_A f_+ \, d\mu \qquad \text{and} \qquad \nu_- = \int_A f_- \, d\mu$$

for all $A \in \mathcal{A}$, then ν is a signed measure with positive and negative parts ν_+ and ν_- respectively. It is clear that

$$|\nu|(A) = \int_A f_+ \, d\mu + \int_A f_- \, d\mu = \int_A |f| \, d\mu$$

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

The total variation of a signed measure has another description that can be useful (especially with complex-valued measures).

Proposition 5.4.3. *Let $(X, \mathcal{A})$ be a measurable space and let ν be a signed measure on $(X, \mathcal{A})$. Let $\mathcal{P}$ denote all countable collections $\{A_n\}_{n=1}^\infty$ of pairwise disjoint measurable sets such that $X = \bigcup_{n=1}^\infty A_n$. Then for all $A \in \mathcal{A}$,*

$$|\nu|(A) = \sup_{\{A_n\}_{n=1}^\infty \in \mathcal{P}} \sum_{n=1}^\infty |\nu(A \cap A_n)|.$$

Proof. Since ν_+ and ν_- are mutually singular, there exists $P, N \in \mathcal{A}$ such that $X = P \cup N$, $P \cap N = \emptyset$, and $\nu_+(N) = \nu_-(P) = 0$. Thus $\{P, N\} \in \mathcal{P}$ and

$$|\nu(A \cap P)| + |\nu(A \cap N)| = \nu_+(A) + \nu_-(A) = |\nu|(A).$$

Hence

$$|\nu|(A) \geq \sup_{\{A_n\}_{n=1}^\infty \in \mathcal{P}} \sum_{n=1}^\infty |\nu(A \cap A_n)|.$$

Conversely, for any $\{A_n\}_{n=1}^\infty \in \mathcal{P}$ we have

$$\begin{aligned}
\sum_{n=1}^\infty |\nu(A \cap A_n)| &= \sum_{n=1}^\infty |\nu_+(A \cap A_n) - \nu_-(A \cap A_n)| \\
&\leq \sum_{n=1}^\infty \nu_+(A \cap A_n) + \nu_-(A \cap A_n) \\
&= \nu_+\left(A \cap \left(\bigcup_{n=1}^\infty A_n\right)\right) + \nu_-\left(A \cap \left(\bigcup_{n=1}^\infty A_n\right)\right) \\
&= \nu_+(A) + \nu_-(A) = |\nu|(A).
\end{aligned}$$

Hence the inequality follows. ∎

The total variation of a signed measure immediately produces a bounded for the value of the signed measure on a set.

Lemma 5.4.4. *Let $(X, \mathcal{A})$ be a measurable space. If ν is a signed measure on $(X, \mathcal{A})$, then $|\nu(A)| \leq |\nu|(A)$ for all $A \in \mathcal{A}$.*

Proof. Let ν be a signed measure on $(X, \mathcal{A})$ with positive and negative parts ν_+ and ν_- respectively. Clearly

$$|\nu(A)| = |\nu_+(A) - \nu_-(A)| = \nu_+(A) + \nu_-(A) = |\nu|(A)$$

for all $A \in \mathcal{A}$. ∎

We have seen some instances where σ-finite measures are preferable over arbitrary measures (with more instances to occur soon). Furthermore clearly finite measures are even nicer. Thus we introduce the following.

Definition 5.4.5. Let $(X, \mathcal{A})$ be a measurable space. A signed measure ν on $(X, \mathcal{A})$ is said to be *finite* if $|\nu(A)| < \infty$ for all $A \in \mathcal{A}$.

Lemma 5.4.6. *Let $(X, \mathcal{A})$ be a measurable space and let ν be a signed measure on $(X, \mathcal{A})$. The following are equivalent:*

(1) ν is finite.

(2) ν_+ and ν_- are finite.

(3) $|\nu|$ is finite.

(4) $\nu(X) \neq \pm\infty$.

Proof. First we claim that (2) and (3) are equivalent. Indeed recall that

$$|\nu|(X) = \nu_+(X) + \nu_-(X).$$

Therefore, since $|\nu|(X), \nu_+(X), \nu_-(X) \in [0,\infty]$, we see that $|\nu|(X) < \infty$ if and only if $\nu_+(X), \nu_-(X) < \infty$. Hence $|\nu|$ is finite if and only if ν_+ and ν_- are finite. Thus (2) and (3) are equivalent.

Next assume that (3) holds. To see that (1) holds, notice by Lemma 5.4.4 that

$$|\nu(A)| \leq |\nu|(A) \leq |\nu|(X)$$

for all $A \in \mathcal{A}$. Hence ν is finite by definition. Thus (3) implies (1).

Next assume that (1) holds. To see that (4) holds, recall that ν_+ and ν_- are mutually singular so there exists $P, N \in \mathcal{A}$ such that $X = P \cup N$, $P \cap N = \emptyset$, and $\nu_+(N) = \nu_-(P) = 0$. Since

$$\nu_+(X) = \nu_+(X \cap P) = \nu_+(X \cap P) - \nu_-(X \cap P) = \nu(P) \leq |\nu(P)| < \infty,$$

and since

$$\nu_-(X) = \nu_-(X \cap N) = \nu_-(X \cap N) - \nu_+(X \cap N) = -\nu(N) \leq |\nu(N)| < \infty,$$

we see that $\nu(X) = \nu_+(X) - \nu_-(X) \neq \pm\infty$. Hence (4) holds.

Finally, to see that (4) implies (2), assume that (2) fails. Hence either $\nu_+(X) = \infty$ or $\nu_-(X) = \infty$. If $\nu_+(X) = \nu_-(X) = \infty$, then if $P, N \in \mathcal{A}$ are such that $X = P \cup N$, $P \cap N = \emptyset$, and $\nu_+(N) = \nu_-(P) = 0$, we have that $\nu_+(P) = \infty$ and $\nu_-(N) = \infty$. Hence

$$\nu(P) = \nu_+(P) - \nu_-(P) = \nu_+(P) = \infty$$

whereas

$$\nu(N) = \nu_+(N) - \nu_-(N) = -\nu_+(N) = -\infty$$

which contradicts the fact that a signed measure can only take one value from $\{\pm\infty\}$. Otherwise, if $\nu_+(X) = \infty$ but $\nu_-(X) \neq \infty$, then $\nu(X) = \infty$ and thus (4) fails. Similarly if $\nu_-(X) = \infty$ but $\nu_+(X) \neq \infty$, then $\nu(X) = -\infty$ and thus (4) fails. Hence (4) implies (2) thereby completing the proof. ∎

In fact, the collection of finite signed measures has a nice normed linear space structure.

Proposition 5.4.7. *Let $(X, \mathcal{A})$ be a measurable space and let*

$$Meas(X, \mathcal{A}) = \{\nu \mid \nu \text{ a finite signed measure on } (X, \mathcal{A})\}.$$

Then $Meas(X, \mathcal{A})$ is a vector space over $\mathbb{R}$ with the operations of pointwise addition and scalar multiplication. Furthermore, if $\|\cdot\|_{Meas} : Meas(X, \mathcal{A}) \to [0,\infty)$ is defined by

$$\|\nu\|_{Meas} = |\nu|(X)$$

for all $\nu \in Meas(X, \mathcal{A})$, then $\|\cdot\|_{Meas}$ is a norm on $Meas(X, \mathcal{A})$.

Proof. It is elementary to see that if $\nu_1, \nu_2 \in \mathrm{Meas}(X, \mathcal{A})$ and $\lambda_1, \lambda_2 \in \mathbb{R}$ then the signed measure $\lambda_1\nu_1 + \lambda_2\nu_2$ defined by

$$(\lambda_1\nu_1 + \lambda_2\nu_2)(A) = \lambda_1\nu_1(A) + \lambda_2\nu_2(A)$$

is a finite signed measure. Hence $\mathrm{Meas}(X, \mathcal{A})$ is a vector space over $\mathbb{R}$ (seeing as it is a subspace of the vector space of all real-valued functions on $\mathcal{A}$).

To see that $\|\cdot\|_{\mathrm{Meas}}$ is a norm on $\mathrm{Meas}(X, \mathcal{A})$, recall that the zero element of $\mathrm{Meas}(X, \mathcal{A})$ is the signed measure ν_0 where $\nu_0(A) = 0$ for all $A \in \mathcal{A}$. Clearly $|\nu_0|(X) = 0$ so $\|\nu_0\|_{\mathrm{Meas}} = 0$. Furthermore, if $\nu \in \mathrm{Meas}(X, \mathcal{A})$ is such that $\|\nu\|_{\mathrm{Meas}} = 0$, then $|\nu|(X) = 0$. Hence, by Lemma 5.4.4, we see that

$$|\nu(A)| \leq |\nu|(A) \leq |\nu|(X) = 0$$

for all $A \in \mathcal{A}$. Therefore $\nu = \nu_0$ as required.

Next let $\nu \in \mathrm{Meas}(X, \mathcal{A})$ and $\alpha \in \mathbb{R}$ be arbitrary. If $\alpha \geq 0$, clearly $(\alpha\nu)_+ = \alpha\nu_+$ and $(\alpha\nu)_- = \alpha\nu_-$ so

$$\begin{aligned} \|\alpha\nu\|_{\mathrm{Meas}} &= (\alpha\nu)_+(X) + (\alpha\nu)_-(X) \\ &= \alpha\nu_+(X) + \alpha\nu_-(X) \\ &= \alpha|\nu|(X) \\ &= |\alpha|\,\|\nu\|_{\mathrm{Meas}}. \end{aligned}$$

Alternatively, if $\alpha < 0$ then clearly $(\alpha\nu)_+ = -\alpha\nu_-$ and $(\alpha\nu)_- = -\alpha\nu_+$ so

$$\begin{aligned} \|\alpha\nu\|_{\mathrm{Meas}} &= (\alpha\nu)_+(X) + (\alpha\nu)_-(X) \\ &= -\alpha\nu_-(X) - \alpha\nu_+(X) \\ &= -\alpha|\nu|(X) \\ &= |\alpha|\,\|\nu\|_{\mathrm{Meas}}. \end{aligned}$$

Hence $\|\alpha\nu\|_{\mathrm{Meas}} = |\alpha|\,\|\nu\|_{\mathrm{Meas}}$ in all cases as required.

Finally, to see that $\|\cdot\|_{\mathrm{Meas}}$ satisfies the triangle inequality, let $\nu, \gamma \in \mathrm{Meas}(X, \mathcal{A})$ be arbitrary. Then

$$\nu + \gamma = (\nu_+ + \gamma_+) - (\nu_- + \gamma_-).$$

Consider $(\nu + \gamma)_+$ and $(\nu + \gamma)_-$. Since $(\nu + \gamma)_+$ and $(\nu + \gamma)_-$ are mutually singular, there exists $P, N \in \mathcal{A}$ such that $X = P \cup N$, $P \cap N = \emptyset$, and $(\nu + \gamma)_+(N) = (\nu + \gamma)_-(P) = 0$. Then

$$\begin{aligned} (\nu + \gamma)_+(X) &= (\nu + \gamma)_+(P) + (\nu + \gamma)_+(N) \\ &= (\nu + \gamma)_+(P) \\ &= (\nu + \gamma)_+(P) - (\nu + \gamma)_-(P) \\ &= (\nu + \gamma)(P) \\ &= (\nu_+(P) + \gamma_+(P)) - (\nu_-(P) + \gamma_-(P)) \\ &\leq \nu_+(P) + \gamma_+(P) \\ &\leq \nu_+(X) + \gamma_+(X). \end{aligned}$$

Similarly

$$\begin{aligned}(\nu+\gamma)_-(X) &= (\nu+\gamma)_-(P) + (\nu+\gamma)_-(N) \\ &= (\nu+\gamma)_-(N) \\ &= (\nu+\gamma)_-(N) - (\nu+\gamma)_+(N) \\ &= -(\nu+\gamma)(N) \\ &= (\nu_-(N)+\gamma_-(N)) - (\nu_+(N)+\gamma_+(N)) \\ &\leq \nu_-(N)+\gamma_-(N) \\ &\leq \nu_-(X)+\gamma_-(X).\end{aligned}$$

Therefore

$$\begin{aligned}\|\nu+\gamma\|_{\text{Meas}} &= (\nu+\gamma)_+(X) + (\nu+\gamma)_-(X) \\ &\leq (\nu_+(X)+\gamma_+(X)) + (\nu_-(X)+\gamma_-(X)) \\ &= \|\nu\|_{\text{Meas}} + \|\gamma\|_{\text{Meas}}\end{aligned}$$

as desired. ∎

Moreover, we have the following.

Theorem 5.4.8. *If $(X,\mathcal{A})$ is a measurable space, then $(Meas(X,\mathcal{A}), \|\cdot\|_{Meas})$ is a Banach space.*

Proof. To see that $(\text{Meas}(X,\mathcal{A}), \|\cdot\|_{\text{Meas}})$ is a Banach space, let $(\nu_n)_{n\geq 1}$ be an arbitrary Cauchy sequence in $(\text{Meas}(X,\mathcal{A}), \|\cdot\|_{\text{Meas}})$. Notice by Lemma 5.4.4 that

$$\begin{aligned}|\nu_n(A)-\nu_m(A)| &= |(\nu_n-\nu_m)(A)| \\ &\leq |\nu_n-\nu_m|(A) \\ &\leq |\nu_n-\nu_m|(X) \\ &= \|\nu_n-\nu_m\|_{\text{Meas}}\end{aligned}$$

for all $A\in\mathcal{A}$ and $n,m\in\mathbb{N}$. Hence $(\nu_n(A))_{n\geq 1}$ is Cauchy in $\mathbb{R}$ for all $A\in\mathcal{A}$. Therefore, since $\mathbb{R}$ is complete, $\lim_{n\to\infty}\nu_n(A)$ exists for all $A\in\mathcal{A}$.

Define $\nu:\mathcal{A}\to\mathbb{R}$ by $\nu(A)=\lim_{n\to\infty}\nu_n(A)$ for all $A\in\mathcal{A}$. We claim that ν is a finite signed measure on $(X,\mathcal{A})$ and that $(\nu_n)_{n\geq 1}$ converges to ν with respect to $\|\cdot\|_{\text{Meas}}$. To see these claims, first notice that

$$\nu(\emptyset) = \lim_{n\to\infty}\nu_n(\emptyset) = \lim_{n\to\infty} 0 = 0.$$

Furthermore, we claim that ν is finitely additive. Indeed if $\{A_k\}_{k=1}^N$ is a finite pairwise disjoint collection of elements of $\mathcal{A}$, then clearly

$$\nu\left(\bigcup_{k=1}^N A_k\right) = \lim_{n\to\infty}\nu_n\left(\bigcup_{k=1}^N A_k\right) = \lim_{n\to\infty}\sum_{k=1}^N \nu_n(A_k) = \sum_{k=1}^N \nu_n(A_k).$$

Next we desire a bound on ν. Since $(\nu_n)_{n\geq 1}$ is Cauchy, $(\nu_n)_{n\geq 1}$ is bounded with respect to $\|\cdot\|_{\mathrm{Meas}}$. Hence there exists an $M \in \mathbb{N}$ such that $\|\nu_n\|_{\mathrm{Meas}} \leq M$ for all $n \in \mathbb{N}$. Hence for all $A \in \mathcal{A}$

$$|\nu_n(A)| \leq |\nu_n|(A) \leq |\nu_n|(X) \leq M.$$

Therefore, since $\nu(A) = \lim_{n\to\infty} \nu_n(A)$ for all $A \in \mathcal{A}$, we obtain that $|\nu(A)| \leq M$ for all $A \in \mathcal{A}$. Note this demonstrates that ν will be a finite signed measure provided we can demonstrate that ν is a signed measure.

Before proceeding to show that ν is a signed measure, we claim that

$$\lim_{n\to\infty} \sup\{|\nu(A) - \nu_n(A)| \mid A \in \mathcal{A}\} = 0.$$

To see this, recall from the above computation that notice for all $m \geq n$ and $A \in \mathcal{A}$ that

$$|\nu_m(A) - \nu_n(A)| \leq \|\nu_n - \nu_m\|_{\mathrm{Meas}}.$$

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

$$|\nu(A) - \nu_n(A)| = \limsup_{m\to\infty} |\nu_m(A) - \nu_n(A)| \leq \limsup_{m\to\infty} \|\nu_n - \nu_m\|_{\mathrm{Meas}}.$$

Hence

$$\sup\{|\nu(A) - \nu_n(A)| \mid A \in \mathcal{A}\} \leq \limsup_{m\to\infty} \|\nu_n - \nu_m\|_{\mathrm{Meas}}.$$

However, since $(\nu_n)_{n\geq 1}$ is Cauchy, we see by the definition of a Cauchy sequence that

$$\limsup_{n\to\infty} \limsup_{m\to\infty} \|\nu_n - \nu_m\|_{\mathrm{Meas}} = 0.$$

Hence the claim is complete.

To see that ν is countably additive, let $\{A_n\}_{n=1}^\infty$ be an arbitrary collection of pairwise disjoint subsets of $\mathcal{A}$. First we must demonstrate that $\sum_{k=1}^\infty \nu(A_k)$ converges absolutely. To see this, note for all $N \in \mathbb{N}$ and $n \in \mathbb{N}$ that

$$\sum_{k=1}^N |\nu_n(A_k)| \leq \sum_{k=1}^N |\nu_n|(A_k) = |\nu_n|\left(\bigcup_{k=1}^N A_k\right) \leq M.$$

Hence

$$\sum_{k=1}^N |\nu(A_k)| = \lim_{n\to\infty} \sum_{k=1}^N |\nu_n(A_k)| \leq M$$

for all $N \in \mathbb{N}$ so $\sum_{k=1}^\infty \nu(A_k)$ converges absolutely.

To see that $\nu\left(\bigcup_{k=1}^\infty A_k\right) = \sum_{k=1}^\infty \nu(A_k)$, let $\epsilon > 0$ be arbitrary. Since $(\nu_n)_{n\geq 1}$ is Cauchy there exists an $N_1 \in \mathbb{N}$ such that $\|\nu_n - \nu_m\|_{\mathrm{Meas}} < \frac{\epsilon}{4}$ for all $n, m \geq N_1$. Furthermore, since

$$\lim_{n\to\infty} \sup\{|\nu(A) - \nu_n(A)| \mid A \in \mathcal{A}\} = 0,$$

there exists an $N_2 \in \mathbb{N}$ such that $|\nu(A) - \nu_n(A)| < \frac{\epsilon}{4}$ for all $A \in \mathcal{A}$ and $n \geq N_2$. Let $N = \max\{N_1, N_2\}$. Since $\nu_N\left(\bigcup_{k=1}^\infty A_k\right) = \sum_{k=1}^\infty \nu_N(A_k)$ and since $\sum_{k=1}^\infty \nu(A_k)$ converges absolutely, there exists an $M \in \mathbb{N}$ such that

$$\left|\nu_N\left(\bigcup_{k=M}^\infty A_k\right)\right| = \left|\sum_{k=M}^\infty \nu_N(A_k)\right| < \frac{\epsilon}{4}$$

and

$$\left|\sum_{k=M}^\infty \nu(A_k)\right| < \frac{\epsilon}{4}.$$

Hence, since ν has been verified to be finitely additive,

$$\begin{aligned}
&\left|\nu\left(\bigcup_{k=1}^\infty A_k\right) - \sum_{k=1}^\infty \nu(A_k)\right| \\
&\leq \left|\nu\left(\bigcup_{k=1}^\infty A_k\right) - \nu_N\left(\bigcup_{k=1}^\infty A_k\right)\right| + \left|\nu_N\left(\bigcup_{k=1}^\infty A_k\right) - \nu_N\left(\bigcup_{k=1}^{M-1} A_k\right)\right| \\
&\quad + \left|\nu_N\left(\bigcup_{k=1}^{M-1} A_k\right) - \sum_{k=1}^{M-1} \nu(A_k)\right| + \left|\sum_{k=1}^{M-1} \nu(A_k) - \sum_{k=1}^\infty \nu(A_k)\right| \\
&\leq \frac{\epsilon}{4} + \left|\sum_{k=M}^\infty \nu_N(A_k)\right| + \left|\nu_N\left(\bigcup_{k=1}^{M-1} A_k\right) - \nu\left(\bigcup_{k=1}^{M-1} A_k\right)\right| + \frac{\epsilon}{4} \\
&\leq \frac{\epsilon}{4} + \frac{\epsilon}{4} + \frac{\epsilon}{4} + \frac{\epsilon}{4} = \epsilon.
\end{aligned}$$

Therefore, since $\epsilon > 0$ was arbitrary, we obtain that $\nu\left(\bigcup_{k=1}^\infty A_k\right) = \sum_{k=1}^\infty \nu(A_k)$ as desired. Hence $\nu \in \mathrm{Meas}(X, \mathcal{A})$.

Finally, to see that $(\nu_n)_{n\geq 1}$ converges to ν with respect to $\|\cdot\|_{\mathrm{Meas}}$, fix $m \in \mathbb{N}$. Since $\nu - \nu_m$ is a signed measure, there exists $P, N \in \mathcal{A}$ such that $X = P \cup N$, $P \cap N = \emptyset$, and $(\nu - \nu_m)_+(N) = (\nu - \nu_m)_-(P) = 0$. Therefore

$$\begin{aligned}
\|\nu - \nu_m\|_{\mathrm{Meas}} &= |\nu - \nu_m|(X) \\
&= (\nu - \nu_m)_+(X) + (\nu - \nu_m)_-(X) \\
&= (\nu - \nu_m)_+(P) + (\nu - \nu_m)_-(N) \\
&= (\nu - \nu_m)(P) - (\nu - \nu_m)(N) \\
&= \lim_{n\to\infty} |(\nu_n - \nu_m)(P) - (\nu_n - \nu_m)(N)| \\
&\leq \limsup_{n\to\infty} |(\nu_n - \nu_m)(P)| + |(\nu_n - \nu_m)(N)| \\
&\leq \limsup_{n\to\infty} |\nu_n - \nu_m|(P) + |\nu_n - \nu_m|(N) \\
&\leq \limsup_{n\to\infty} 2|\nu_n - \nu_m|(X) \\
&\leq \limsup_{n\to\infty} 2\,\|\nu_n - \nu_m\|_{\mathrm{Meas}}.
\end{aligned}$$

Therefore, since m was arbitrary and since $\limsup_{n\to\infty} 2\,\|\nu_n - \nu_m\|_{\mathrm{Meas}}$ tends to zero as m tends to infinity as $(\nu_n)_{n\geq 1}$ is Cauchy, the result follows. ∎

5.5 The Radon-Nikodym Theorem

In this section, we will prove one of the most important theorems for finite signed measures, the Radon-Nikodym Theorem (Theorem 5.5.5). The Radon-Nikodym Theorem will completely characterize which measures can be obtained by integrating L_1-functions against another fixed measure via the following condition (which was precisely the additional property we observed in Corollary 3.2.6).

Definition 5.5.1. Let $(X, \mathcal{A}, \mu)$ be a measure space and let ν be a signed measure on $(X, \mathcal{A})$. It is said that ν is *absolutely continuous* with respect to μ, denoted $\nu \ll \mu$, if $A \in \mathcal{A}$ and $\mu(A) = 0$ implies $\nu(A) = 0$.

Remark 5.5.2. In fact, if $(X, \mathcal{A}, \mu)$ is a measure space and ν is a signed measure that is absolutely continuous with respect to μ, then if $A \in \mathcal{A}$ is such that $\mu(A) = 0$ then A is a null set for ν. To see this, assume $A \in \mathcal{A}$ is such that $\mu(A) = 0$. To see that A is null for ν we simply observe since $\nu \ll \mu$ that if $B \in \mathcal{A}$ and $B \subseteq A$ then $\mu(B) = 0$ so $\nu(B) = 0$.

Example 5.5.3. Let $(X, \mathcal{A}, \mu)$ be a measure space and let $f \in L_1(X, \mu)$ be real valued. Recall from Example 5.1.4 that if we define $\nu : \mathcal{A} \to [-\infty, \infty]$ by

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

for all $A \in \mathcal{A}$, then ν is a signed measure. Since $A \in \mathcal{A}$ and $\mu(A) = 0$ implies

$$\nu(A) = \int_A f \, d\mu = 0,$$

we see that $\nu \ll \mu$.

Perhaps unsurprisingly, the notion of absolute continuity of measure plays well with respect to the positive and negative parts of signed measures.

Lemma 5.5.4. *Let $(X, \mathcal{A}, \mu)$ be a measure space and let ν be a signed measure on $(X, \mathcal{A})$. Then $\nu \ll \mu$ if and only if $\nu_+ \ll \mu$ and $\nu_- \ll \mu$.*

Proof. Clearly if $\nu_+ \ll \mu$ and $\nu_- \ll \mu$ then $\nu \ll \mu$. Conversely, assume that $\nu \ll \mu$. Since ν_+ and ν_- are mutually singular, there exists $P, N \in \mathcal{A}$ such that $X = P \cup N$, $P \cap N = \emptyset$, and $\nu_+(N) = \nu_-(P) = 0$. To see that $\nu_+ \ll \mu$ and $\nu_- \ll \mu$, let $A \in \mathcal{A}$ be such that $\mu(A) = 0$. Then $\mu(A \cap P) = 0$ so, since $\nu \ll \mu$,

$$0 = \nu(A \cap P) = \nu_+(A \cap P) - \nu_-(A \cap P) = \nu_+(A).$$

Similarly, since $\mu(A \cap N) = 0$ we obtain that

$$0 = \nu(A \cap N) = \nu_+(A \cap N) - \nu_-(A \cap N) = -\nu_-(A).$$

Hence, since $A \in \mathcal{A}$ with $\mu(A) = 0$ was arbitrary, we obtain that $\nu_+ \ll \mu$ and $\nu_- \ll \mu$. ∎

We arrive at the centrepiece to understanding the relations between absolutely continuous measures.

Theorem 5.5.5 (The Radon-Nikodym Theorem). *Let $(X, \mathcal{A})$ be a measurable space. If μ and ν are measures on $(X, \mathcal{A})$ such that μ is σ-finite and $\nu \ll \mu$, then there exists a measurable function $f : X \to [0, \infty]$ such that*

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

for all $A \in \mathcal{A}$. Furthermore, if $g : X \to [0, \infty]$ is a measurable function such that

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

for all $A \in \mathcal{A}$, then $f = g$ almost everywhere.

The proof of the Radon-Nikodym Theorem (Theorem 5.5.5) is not a simple one. The idea is to build up the measurable function f from knowledge of ν and the fact that $\nu \ll \mu$. Since measure theory only works well for countable collections of sets, we must build up f using countable collections. The following lemma is our first step.

Lemma 5.5.6. *Let $(X, \mathcal{A})$ be a measurable space and let $Q \subseteq \mathbb{R}$ be a countable set. Assume $\{A_q\}_{q \in Q} \subseteq \mathcal{A}$ are such that if $q_1, q_2 \in Q$ and $q_1 \leq q_2$ then $A_{q_1} \subseteq A_{q_2}$. There exists a measurable function $f : X \to [-\infty, \infty]$ such that $f(x) \geq q$ for all $x \in A_q^c$ and $q \in Q$ and $f(x) \leq q$ for all $x \in A_q$ and $q \in Q$.*

Proof. Define $f : X \to [-\infty, \infty]$ by

$$f(x) = \inf\{q \in Q \mid x \in A_q\}.$$

Clearly f is well-defined. Moreover, if $x \in X$ and $x \in A_q$ for some $q \in Q$, then clearly $f(x) \leq q$ by definition. Furthermore, if $x \in X$ and $x \in A_q^c$ for some $q \in Q$, then $x \notin A_{q'}$ for all $q' \leq q$ so $f(x) \geq q$. Thus, to complete the proof, it suffices to show that f is measurable.

To see that f is measurable, it suffices to show that

$$\{x \in X \mid f(x) < t\} \in \mathcal{A}$$

for all $t \in \mathbb{R}$. However, since clearly

$$\{x \in X \mid f(x) < t\} = \bigcup_{\substack{q \in Q \\ q < t}} A_q,$$

we see the desired sets are measurable as $\{A_q\}_{q \in Q} \subseteq \mathcal{A}$ and Q is countable. ■

Next we extend Lemma 5.5.6 thereby weakening the conditions required to construct our function at the cost of only having a bound almost everywhere (which we do not care about if we are going to integrate the function).

Lemma 5.5.7. *Let $(X, \mathcal{A}, \mu)$ be a measure space and let $Q \subseteq \mathbb{R}$ be a countable set. Assume $\{A_q\}_{q\in Q} \subseteq \mathcal{A}$ are such that if $q_1, q_2 \in Q$ and $q_1 \leq q_2$ then*

$$\mu(A_{q_1} \setminus A_{q_2}) = 0.$$

There exists a measurable function $f : X \to [-\infty, \infty]$ such that, for all $q \in Q$, $f(x) \geq q$ for μ-almost every $x \in A_q^c$ and for all $q \in Q$ and $f(x) \leq q$ for μ-almost every $x \in A_q$.

Proof. Let

$$Z = \bigcup_{\substack{q_1,q_2\in Q\\ q_1<q_2}} A_{q_1} \setminus A_{q_2}.$$

Then $Z \in \mathcal{A}$ since $\{A_q\}_{q\in Q} \subseteq \mathcal{A}$ and $Q \times Q$ is countable. Furthermore, by the assumptions on $\{A_q\}_{q\in Q} \subseteq \mathcal{A}$, we see that

$$0 \leq \mu(Z) \leq \sum_{\substack{q_1,q_2\in Q\\ q_1<q_2}} \mu(A_{q_1} \setminus A_{q_2}) = 0.$$

For each $q \in Q$, let $B_q = A_q \cup Z$. Notice if $q_1, q_2 \in Q$ and $q_1 \leq q_2$ then

$$B_{q_1} = A_{q_1} \cup Z = (A_{q_1} \cap A_{q_2}) \cup (A_{q_1} \setminus A_{q_2}) \cup Z = (A_{q_1} \cap A_{q_2}) \cup Z \subseteq A_{q_2} \cup Z = B_{q_2}.$$

Therefore, by Lemma 5.5.6 there exists a measurable function $f : X \to [-\infty, \infty]$ such that $f(x) \geq q$ for all $x \in B_q^c$ and $q \in Q$ and $f(x) \leq q$ for all $x \in B_q$ and $q \in Q$.

We claim that f satisfies the desired properties. Indeed if $q \in Q$ and $x \in A_q$ then $x \in B_q$ so $f(x) \leq q$. Furthermore, if $q \in Q$ and $x \in A_q^c \setminus Z$, then, since

$$A_q^c \setminus Z = A_q^c \cap Z^c = (A_q \cup Z)^c = B_q^c,$$

we see that $f(x) \geq q$. Therefore, since $\mu(Z) = 0$, we obtain that if $q \in Q$ then $f(x) \geq q$ for almost every $x \in A_q^c$ as desired. ∎

Proof of the Radon-Nikodym Theorem (Theorem 5.5.5). The proof will proceed by first assuming that μ is finite. We will then use the finite case to prove the σ-finite case. Finally, we recall the uniqueness claim is precisely Proposition 3.2.4.

Case 1: μ is finite. We desire to use Lemma 5.5.7 to construct the desired function. To do so, we will need to construct the appropriate sets.

For each $q \in \mathbb{Q}$, notice that $\nu - q\mu$ is a signed measure since ν is positive and μ is finite (i.e. the value $-\infty$ cannot be obtained). Hence, by the Hahn Decomposition Theorem (Theorem 5.2.7) there exists $P_q, N_q \in \mathcal{A}$ such that

P_q is a positive set for $\nu - q\mu$, N_q is a negative set for $\nu - q\mu$, $X = P_q \cup N_q$ and $P_q \cap N_q = \emptyset$. For the case $q = 0$, we take $P_0 = X$ and $N_0 = \emptyset$. Note we are interested in P_q and N_q as $\nu \geq q\mu$ on P_q and $\nu \leq q\mu$ on N_q.

Assume $q, r \in \mathbb{Q}$ are such that $q < r$. Then, since N_q is a negative set for $\nu - q\mu$, we see that

$$\nu(N_q \setminus N_r) - q\mu(N_q \setminus N_r) = (\nu - q\mu)(N_q \setminus N_r) \leq 0.$$

Therefore

$$\nu(N_q \setminus N_r) \leq q\mu(N_q \setminus N_r) \leq q\mu(X) < \infty.$$

Furthermore, since $N_q \setminus N_r \subseteq N_r^c = P_r$ and P_r is a positive set for $\nu - r\mu$, we see that

$$\nu(N_q \setminus N_r) - r\mu(N_q \setminus N_r) = (\nu - r\mu)(N_q \setminus N_r) \geq 0.$$

Hence by combining the two inequalities above and by using the fact that $\nu(N_q \setminus N_r) < \infty$, we see that

$$\begin{aligned} &\nu(N_q \setminus N_r) - q\mu(N_q \setminus N_r) \leq 0 \leq \nu(N_q \setminus N_r) - r\mu(N_q \setminus N_r) \\ &\quad\Rightarrow (r - q)\mu(N_q \setminus N_r) \leq 0 \end{aligned}$$

However, since $q < r$ and as μ is positive, this implies that $\mu(N_q \setminus N_r) = 0$.

Since $\mu(N_q \setminus N_r) = 0$ for all $q, r \in \mathbb{Q}$ with $q \leq r$, Lemma 5.5.7 implies there exists a measurable function $f : X \to [-\infty, \infty]$ such that for all $q \in Q$ we have $f(x) \geq q$ for almost every $x \in P_q$ whereas $f(x) \leq q$ for almost every $x \in N_q$. We claim this is the function we are search for.

To begin, recall with $q = 0$ that $N_0 = \emptyset$ and $P_0 = X$. Hence $f(x) \geq 0$ for almost every $x \in X$. Thus we may assume that $f : X \to [0, \infty]$.

Before we proceed with the proof, we require an observation that is the crux of the proof. Assume $q, r \in \mathbb{Q}$ are such that $q < r$. If $A \in \mathcal{A}$ and $A \subseteq P_q \cap N_r$ then $A \subseteq P_q$ so, since P_q is a positive set for $\nu - q\mu$,

$$0 \leq (\nu - q\mu)(A) = \nu(A) - q\mu(A),$$

and $A \subseteq N_r$ so, since N_r is a negative set for $\nu - r\mu$,

$$0 \geq (\nu - r\mu)(A) = \nu(A) - r\mu(A).$$

Hence

$$q\mu(A) \leq \nu(A) \leq r\mu(A)$$

whenever $A \subseteq P_q \cap N_r$. Similarly, if $A \in \mathcal{A}$ and $A \subseteq P_q \cap N_r$ then $f(x) \geq q$ for almost every $x \in A$ as $A \subseteq P_q$, and $f(x) \leq r$ for almost every $x \in A$ as $A \subseteq N_r$. Therefore

$$q\mu(A) \leq \int_A f \, d\mu \leq r\mu(A).$$

The remainder of the proof is some simple analysis to show that we can squeeze these quantities to show that $\nu(A) = \int_A f \, d\mu$ for every $A \in \mathcal{A}$.

To complete this case of the proof, let $A \in \mathcal{A}$ be arbitrary. For each fixed $m \in \mathbb{N}$, let

$$A_{m,n} = A \cap \left(N_{\frac{n+1}{m}} \setminus \left(\bigcup_{k=0}^{n} N_{\frac{k}{m}} \right) \right) \in \mathcal{A}$$

for all $n \in \mathbb{N} \cup \{0\}$. Furthermore, let

$$A_{m,\infty} = A \setminus \bigcup_{n=0}^{\infty} A_{m,n} = A \setminus \bigcup_{n=0}^{\infty} N_{\frac{n}{m}} \in \mathcal{A}.$$

Hence $\{A_{m,\infty}\} \cup \{A_{m,n}\}_{n=0}^{\infty}$ is a pairwise disjoint collection of elements of $\mathcal{A}$ such that

$$A = A_{m,\infty} \cup \left(\bigcup_{n=0}^{\infty} A_{m,n} \right).$$

Consider $A_{m,\infty}$ and the case that $\mu(A_{m,\infty}) > 0$. In this case, notice if $x \in A_{m,\infty}$ then $x \in N_{\frac{n}{m}}^c = P_{\frac{n}{m}}$ for all $n \in \mathbb{N}$ so $f(x) \geq \frac{n}{m}$ for all $n \in \mathbb{N}$ and thus $f(x) = \infty$. Therefore, since $\mu(A_{m,\infty}) > 0$ and f is non-negative, we obtain that

$$\infty = \int_{A_{m,\infty}} f \, d\mu \leq \int_A f \, d\mu \leq \infty.$$

On the other hand, since $A_{m,\infty} \subseteq N_{\frac{n}{m}}^c = P_{\frac{n}{m}}$ for all $n \in \mathbb{N}$, the above computations show that $\nu(A_{m,\infty}) \geq \frac{n}{m}\mu(A_{m,\infty})$ for all $n \in \mathbb{N}$. Hence, since $\mu(A_{m,\infty}) > 0$, $\nu(A_{m,\infty}) = \infty$, which clearly implies $\nu(A) = \infty$. Therefore, if $\mu(A_\infty) > 0$ then

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

and the proof would be complete.

Thus to complete the proof, we may assume that $\mu(A_{m,\infty}) = 0$. Since $\nu \ll \mu$, this implies that $\nu(A_{m,\infty}) = 0$. Therefore, since $\mu(A_{m,\infty}) = 0$, we see that

$$\int_{A_{m,\infty}} f \, d\mu = 0 = \nu(A_{m,\infty}).$$

Furthermore, for each $n \in \mathbb{N} \cup \{0\}$ we see that $A_{m,n} \subseteq P_{\frac{n}{m}} \cap N_{\frac{n+1}{m}}$ so by the previous computations

$$\frac{n}{m}\mu(A_{m,n}) \leq \nu(A_{m,n}) \leq \frac{n+1}{m}\mu(A_{m,n})$$

and

$$\frac{n}{m}\mu(A_{m,n}) \leq \int_{A_{m,n}} f \, d\mu \leq \frac{n+1}{m}\mu(A_{m,n}).$$

Hence

$$\left| \nu(A_{m,n}) - \int_{A_{m,n}} f \, d\mu \right| \leq \frac{1}{m}\mu(A_{m,n}).$$

Therefore, either both $\nu(A)$ and $\int_A f\,d\mu$ are infinite, or otherwise all terms in the following computation are finite:

$$\begin{aligned}
\left|\nu(A) - \int_A f\,d\mu\right| &= \left|\nu(A_{m,\infty}) + \sum_{n=0}^{\infty} \nu(A_{m,n}) - \int_{A_{m,\infty}} f\,d\mu - \sum_{n=0}^{\infty}\int_{A_{m,n}} f\,d\mu\right| \\
&= \left|\sum_{n=0}^{\infty} \nu(A_{m,n}) - \sum_{n=0}^{\infty}\int_{A_{m,n}} f\,d\mu\right| \\
&\leq \sum_{n=0}^{\infty}\left|\nu(A_{m,n}) - \int_{A_{m,n}} f\,d\mu\right| \\
&\leq \sum_{n=0}^{\infty} \frac{1}{m}\mu(A_{m,n}) \\
&= \frac{1}{m}\mu\left(\bigcup_{n=0}^{\infty} A_{m,n}\right) \\
&\leq \frac{1}{m}\mu(X).
\end{aligned}$$

Therefore, since the above holds for all $m \in \mathbb{N}$ and as $\mu(X) < \infty$, we obtain that

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

Therefore, since $A \in \mathcal{A}$ was arbitrary, the proof is complete in this case.

Case 2: μ is σ-finite. Since μ is σ-finite, Remark 1.1.21 implies there exists a pairwise disjoint collection $\{X_n\}_{n=1}^{\infty} \subseteq \mathcal{A}$ such that $X = \bigcup_{n=1}^{\infty} X_n$, and $\mu(X_n) < \infty$ for all $n \in \mathbb{N}$.

For each $n \in \mathbb{N}$, define $\mu_n, \nu_n : \mathcal{A} \to [0,\infty]$ by

$$\mu_n(A) = \mu(A \cap X_n) \qquad \text{and} \qquad \nu_n(A) = \nu(A \cap X_n)$$

for all $A \in \mathcal{A}$. Clearly μ_n and ν_n are well-defined measures on $(X, \mathcal{A})$. Furthermore, since

$$\mu_n(X) = \mu(X_n) < \infty,$$

μ_n is a finite measure. Moreover, if $A \in \mathcal{A}$ and $\mu_n(A) = 0$, then $\mu(A \cap X_n) = 0$ so $\nu_n(A) = \nu(A \cap X_n) = 0$ as $\nu \ll \mu$. Hence $\nu_n \ll \mu_n$.

By the previous case, for each $n \in \mathbb{N}$ there exists a measurable function $f_n : X \to [0,\infty]$ such that

$$\nu_n(A) = \int_A f_n\,d\mu_n$$

for all $A \in \mathcal{A}$. Notice by the definition of μ_n and the definition of the integral that

$$\int_A f_n\,d\mu_n = \int_{A\cap X_n} f_n\,d\mu$$

for all $A \in \mathcal{A}$ (i.e. clearly holds for all characteristic functions, thus simple functions, and thus non-negative measurable functions).

Let $f : X \to [0, \infty]$ be defined by

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

for all $x \in X$. Then f is a measurable function by Proposition 2.1.22. Furthermore, we see for all $A \in \mathcal{A}$ that

$$\begin{aligned}\nu(A) &= \nu\left(\bigcup_{n=1}^{\infty} A \cap X_n\right) \\ &= \sum_{n=1}^{\infty} \nu(A \cap X_n) \\ &= \sum_{n=1}^{\infty} \nu_n(A) \\ &= \sum_{n=1}^{\infty} \int_A f_n \, d\mu_n \\ &= \sum_{n=1}^{\infty} \int_{A \cap X_n} f_n \, d\mu \\ &= \sum_{n=1}^{\infty} \int_A f_n \chi_{X_n} \, d\mu \\ &= \int_A \sum_{n=1}^{\infty} f_n \chi_{X_n} \, d\mu \\ &= \int_A f \, d\mu\end{aligned}$$

(where the second last inequality follows from Corollary 3.2.5). Hence the proof of existence of an f is complete. ∎

Using our knowledge of signed measures, we can extend the Radon-Nikodym Theorem (Theorem 5.5.5) to signed measures.

Corollary 5.5.8. *Let $(X, \mathcal{A}, \mu)$ be a σ-finite measure space. If ν is a finite signed measure on $(X, \mathcal{A})$ such that $\nu \ll \mu$, then there exists a unique real-valued function $f \in L_1(X, \mu)$ such that*

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

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

Proof. Since ν is a finite signed measure on $(X, \mathcal{A})$ such that $\nu \ll \mu$, we see that ν_+ and ν_- are finite positive measure on $(X, \mathcal{A})$ that are absolutely continuous with respect to μ by Lemma 5.4.6 and Lemma 5.5.4. Hence by the Radon-Nikodym Theorem (Theorem 5.5.5) there exists positive measurable functions $f_1, f_2 : X \to [0, \infty]$ such that

$$\nu_+(A) = \int_A f_1 \, d\mu \qquad \text{and} \qquad \nu_-(A) = \int_A f_2 \, d\mu$$

for all $A \in \mathcal{A}$. Since ν_+ and ν_- are finite, we see that

$$\int_X |f_1| \, d\mu = \int_X f_1 \, d\mu = \nu_+(X) < \infty$$

and

$$\int_X |f_2| \, d\mu = \int_X f_2 \, d\mu = \nu_-(X) < \infty$$

so $f_1, f_2 \in L_1(X, \mu)$. Hence $f_1 - f_2 \in L_1(X, \mu)$. Therefore, since

$$\int_A f_1 - f_2 \, d\mu = \int_A f_1 - f_2 \, d\mu = \nu(A)$$

for all $A \in \mathcal{A}$, the existence proof is complete.

For uniqueness, notice if $g \in L_1(X, \mu)$ is such that

$$\int_A f \, d\mu = \int_A g \, d\mu$$

for all $A \in \mathcal{A}$, then as $f_+, f_-, g_+, g_- \in L_1(X, \mu)$ we obtain that

$$\int_A f_+ + g_- \, d\mu = \int_A g_+ + f_- \, d\mu$$

for all $A \in \mathcal{A}$. Hence Proposition 3.2.4 implies that $f_+ + g_- = g_+ + f_-$ almost everywhere. Therefore, since $f_+, f_-, g_+, g_- \in L_1(X, \mu)$ implies f_+, f_-, g_+, g_- are finite almost everywhere, we obtain that $f = f_+ - f_- = g_+ - g_- = g$ almost everywhere. ∎

Example 5.5.9. To see why the assumption that μ is σ-finite is required in the Radon-Nikodym Theorem, let μ be the counting measure on $\mathbb{R}$ restricted to the Lebesgue measurable sets and let $\nu = \lambda$. We claim that $\lambda \ll \mu$. Indeed if $A \in \mathcal{M}(\mathbb{R})$ is such that $\mu(A) = 0$, then $A = \emptyset$ so $\lambda(A) = 0$.

To see that the Radon-Nikodym Theorem fails in this setting, suppose for the sake of a contradiction that there exists a Lebesgue measurable function $f : \mathbb{R} \to [0, \infty]$ such that

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

for all $A \in \mathcal{M}(\mathbb{R})$. Then

$$0 = \lambda(\{x\}) = \int_{\{x\}} f \, d\mu = f(x)\mu(\{x\}) = f(x)$$

for all $x \in X$. Hence $f = 0$ so

$$\lambda(A) = \int_A f \, d\mu = 0$$

for all $A \in \mathcal{M}(\mathbb{R})$, which is absurd. Hence the Radon-Nikodym Theorem potentially fails when μ is not σ-finite.

Due to the uniqueness portion of the Radon-Nikodym Theorem (Theorem 5.5.5), we make the following definition.

Definition 5.5.10. Let $(X, \mathcal{A}, \mu)$ be a σ-finite measure space and let ν be either a measure or a finite signed measure on $(X, \mathcal{A})$ such that $\nu \ll \mu$. The *Radon-Nikodym derivative of ν with respect to μ*, denoted $\frac{d\nu}{d\mu}$, is the unique measurable function (positive if ν is a measure and in $L_1(X, \mu)$ if ν is finite) such that

$$\nu(A) = \int_A \frac{d\nu}{d\mu} \, d\mu$$

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

Unsurprisingly, there is a connection between measures that are absolutely continuous with respect to the Lebesgue measure and absolutely continuous functions.

Proposition 5.5.11. *Let $F : \mathbb{R} \to \mathbb{R}$ be non-decreasing function with the limit conditions required by Example 1.3.10 and let λ_F denote the Borel-Stieltjes measure associated to F (the restriction of the Lebesgue-Stieltjes measure to $\mathfrak{B}(\mathbb{R})$). Then F is absolutely continuous on every closed interval in $\mathbb{R}$ if and only if $\lambda_F \ll \lambda$. Furthermore, if F is absolutely continuous, then $\frac{d\lambda_F}{d\lambda} = F'$.*

Proof. To begin, assume F is absolutely continuous on every closed interval in $\mathbb{R}$. Therefore, since F is non-decreasing, F' exists on $\mathbb{R}$ and $F' \geq 0$ by the Lebesgue Differentiation Theorem (Theorem 4.2.2).

Define $\nu : \mathfrak{B}(\mathbb{R}) \to [0, \infty]$ by

$$\nu(A) = \int_A F' \, d\lambda$$

for all $A \in \mathfrak{B}(\mathbb{R})$. Clearly ν is a well-defined measure by Corollary 3.2.6. To see that $\lambda_F \ll \lambda$ and that $\frac{d\lambda_F}{d\lambda} = F'$, it suffices by the uniqueness portion of the Radon-Nikodym Theorem (Theorem 5.5.5) to show that $\nu = \lambda_F$.

To begin, notice for all $(a, b] \in \mathbb{R}$ that

$$\lambda_F((a, b]) = F(b) - F(a) = \int_{(a,b]} F' \, d\lambda = \nu((a, b])$$

by the second Fundamental Theorem of Calculus (Theorem 4.5.4). Hence, by taking the limit as $a \to -\infty$ or as $b \to \infty$, we obtain that $\lambda_F(A) = \nu(A)$

for all $A \in \mathcal{F}$ where $\mathcal{F}$ is defined as in Example 1.3.10. Therefore, since ν is a measure, we obtain that $\lambda_F = \nu$ by the uniqueness portion of Example 1.3.10.

Conversely, assume that $\lambda_F \ll \lambda$. Therefore, by the Radon-Nikodym Theorem there exists a Borel function $g : \mathbb{R} \to [0, \infty]$ such that

$$\lambda_F(A) = \int_A g \, d\lambda$$

for all $A \in \mathfrak{B}(\mathbb{R})$. Since $F(x) \in \mathbb{R}$ for all $x \in \mathbb{R}$, we see that

$$\int_{[a,b]} |g| \, d\lambda = \int_{[a,b]} g \, d\lambda = \lambda_F((a, b]) = F(b) - F(a) < \infty$$

so $g \in L_1([a, b], \lambda)$ for all $a, b \in \mathbb{R}$. Therefore, since for all $x, a, b \in \mathbb{R}$ with $a < x < b$ we have that

$$F(x) - F(a) = \lambda_F((a, x]) = \int_{[a,x]} g d\lambda,$$

we see that F is absolutely continuous $[a, b]$ by Proposition 4.4.10. ∎

In fact, the Radon-Nikodym derivative allows an extension of 'integration by change of variables'.

Proposition 5.5.12. *Let $(X, \mathcal{A}, \mu)$ be a σ-finite measure space and let ν be a measure on $(X, \mathcal{A})$ such that $\nu \ll \mu$. If $f : X \to [0, \infty]$ is measurable, then*

$$\int_X f \, d\nu = \int_X f \frac{d\nu}{d\mu} \, d\mu.$$

Proof. First, since $\nu \ll \mu$, we know that $\frac{d\nu}{d\mu}$ is a well-defined, non-negative measurable function on $(X, \mathcal{A})$ by the Radon-Nikodym Theorem.

To begin, assume $f = \chi_A$ for some $A \in \mathcal{A}$. Then clearly

$$\int_X \chi_A \, d\nu = \nu(A) = \int_A \frac{d\nu}{d\mu} \, d\mu = \int_X \chi_A \frac{d\nu}{d\mu} \, d\mu.$$

Hence, by linearity, we obtain that

$$\int_X f \, d\nu = \int_X f \frac{d\nu}{d\mu} \, d\mu$$

for all simple functions f. Therefore, since every measurable function is the increasing limit of simple functions and since $\frac{d\nu}{d\mu}$ is non-negative, we obtain by two applications of the Monotone Convergence Theorem that

$$\int_X f \, d\nu = \int_X f \frac{d\nu}{d\mu} \, d\mu$$

for all measurable functions $f : X \to [0, \infty]$.

Now assume $f : X \to \mathbb{C}$ is measurable. Since the above demonstrates that

$$\int_X \left| f \frac{d\nu}{d\mu} \right| \, d\mu = \int_X |f| \frac{d\nu}{d\mu} \, d\mu = \int_X |f| \, d\nu,$$

we clearly see that f is integrable with respect to ν if and only if $f\frac{d\nu}{d\mu}$ is integrable with respect to μ.

Finally, assume that $f : X \to \mathbb{C}$ is measurable and integrable with respect to ν. Thus we can write $f = \sum_{k=1}^{4} i^k f_k$ where $f_k : X \to [0, \infty)$ are measurable functions that are integrable with respect to ν. Hence $f_k \frac{d\nu}{d\mu}$ are all integrable with respect to μ so the above implies

$$\int_X f \, d\nu = \sum_{k=1}^{4} i^k \int_X f_k \, d\nu = \sum_{k=1}^{4} i^k \int_X f_k \frac{d\nu}{d\mu} \, d\mu = \int_X f \frac{d\nu}{d\mu} \, d\mu$$

as desired. ∎

Combining the above two results, we have the following.

Corollary 5.5.13. *Let $F : \mathbb{R} \to \mathbb{R}$ be non-decreasing with the limit conditions required by Example 1.3.10 that is absolutely continuous on every closed interval and let λ_F denote the Borel-Stieltjes measure associated to F. If $f : X \to [0, \infty]$ is Lebesgue measurable, then*

$$\int_{\mathbb{R}} f \, d\lambda_F = \int_{\mathbb{R}} f F' \, d\lambda.$$

5.6 The Lebesgue Decomposition Theorem

To conclude this chapter, we demonstrate that given two σ-finite measures μ and ν on a measurable space, we can always decompose ν into a portion that is absolutely continuous with respect to μ and a portion that is orthogonal to μ.

Theorem 5.6.1 (The Lebesgue Decomposition Theorem). *Let $(X, \mathcal{A})$ be a measurable space. If μ and ν are σ-finite measures on $(X, \mathcal{A})$, then there exists a unique pair of measures (ν_a, ν_s) on $(X, \mathcal{A})$ such that $\nu_a \ll \mu$, $\nu_s \perp \mu$, and $\nu = \nu_a + \nu_s$.*

Proof. Let μ and ν are σ-finite measures on $(X, \mathcal{A})$. We claim that $\mu + \nu$ is σ-finite. Indeed since μ and ν are σ-finite, there exist $\{X_n\}_{n=1}^\infty \subseteq \mathcal{A}$ and $\{Y_m\}_{n=1}^\infty \subseteq \mathcal{A}$ such that $X = \bigcup_{n=1}^\infty X_n = \bigcup_{m=1}^\infty Y_n$ and $\mu(X_n), \nu(Y_m) < \infty$ for all $n, m \in \mathbb{N}$. For each $n, m \in \mathbb{N}$ let $Z_{n,m} = X_n \cap Y_m$. Then

$$\{Z_{n,m} \mid n, m \in \mathbb{N}\}$$

is a countable collection of elements of $\mathcal{A}$ such that $X = \bigcup_{n,m=1}^{\infty} Z_{n,m}$ and

$$(\mu + \nu)(Z_{n,m}) = \mu(Z_{n,m}) + \nu(Z_{n,m}) \leq \mu(X_n) + \nu(Y_m) < \infty$$

for all $n, m \in \mathbb{N}$. Hence $\mu + \nu$ is σ-finite.

Notice if $A \in \mathcal{A}$ and $(\mu + \nu)(A) = 0$ then clearly $\mu(A) = 0$. Hence $\mu \ll \mu + \nu$. Therefore, since $\mu + \nu$ is σ-finite, the Radon-Nikodym Theorem (Theorem 5.5.5) implies there exists a measurable function $f : X \to [0, \infty]$ such that

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

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

Let

$$P = \{x \in X \mid f(x) > 0\} \qquad \text{and} \qquad N = \{x \mid f(x) = 0\}.$$

Clearly $P, N \in \mathcal{A}$ are such that $P \cap N = \emptyset$ and $P \cup N = X$. Furthermore, P is a positive set for μ and N is a null set for μ by Example 5.2.4.

Define $\nu_a, \nu_s : \mathcal{A} \to [0, \infty]$ by

$$\nu_a(A) = \nu(A \cap P) \qquad \text{and} \qquad \nu_s(A) = \nu(A \cap N)$$

for all $A \in \mathcal{A}$. We claim that (ν_a, ν_s) are the pair of measures we are looking for. Clearly ν_a and ν_s are measures on $(X, \mathcal{A})$ by Example 1.1.15. Furthermore, as $P \cap N = \emptyset$ and $P \cup N = X$, we see for all $A \in \mathcal{A}$ that

$$\nu(A) = \nu(A \cap P) + \nu(A \cap N) = \nu_a(A) + \nu_s(A)$$

so $\nu = \nu_a + \nu_s$ as desired.

To see that $\nu_a \ll \mu$, let $A \in \mathcal{A}$ such that $\mu(A) = 0$ be arbitrary. Then

$$0 = \mu(A \cap P) = \int_{A \cap P} f \, d(\mu + \nu).$$

However, since $f(x) > 0$ for all $x \in P$, the above equation implies $(\mu + \nu)(A \cap P) = 0$. Therefore $\nu_a(A) = \nu(A \cap P) = 0$. Hence, since $A \in \mathcal{A}$ such that $\mu(A) = 0$ was arbitrary, $\nu_a \ll \mu$.

To see that $\nu_s \perp \mu$, notice that

$$\mu(N) = \int_N f \, d(\mu + \nu) = 0$$

as $f(x) = 0$ for all $x \in N$ and

$$\nu_s(P) = \nu(A \cap N \cap P) = \nu(\emptyset) = 0.$$

Therefore since $P, N \in \mathcal{A}$ are such that $P \cap N = \emptyset$ and $P \cup N = X$, the claim existence claim follows.

To see uniqueness of the pair (ν_a, ν_s), assume (ν_a', ν_s') is a pair of measures on $(X, \mathcal{A})$ such that $\nu_a' \ll \mu$, $\nu_s' \perp \mu$, and $\nu = \nu_a' + \nu_s'$. Since ν is σ-finite, Remark 1.1.21 implies there exists a pairwise disjoint collection $\{Y_n\}_{n=1}^\infty \subseteq \mathcal{A}$ such that $X = \bigcup_{n=1}^\infty Y_n$ and $\nu(Y_n) < \infty$ for all $n \in \mathbb{N}$. Therefore, since $\nu = \nu_a + \nu_s = \nu_a' + \nu_s'$, we obtain for all $n \in \mathbb{N}$ and $A \in \mathcal{A}$ that

$$\nu_a'(A \cap Y_n) - \nu_a(A \cap Y_n) = \nu_s'(A \cap Y_n) - \nu_s(A \cap Y_n)$$

as all terms are finite.

Since $\nu_s \perp \mu$ and $\nu_s' \perp \mu$, there exists $N, B \in \mathcal{A}$ such that

$$\nu_s(N^c) = \mu(N) = 0 = \mu(B) = \nu_s'(B^c).$$

Hence, if $Y = B \cup N \in \mathcal{A}$, then $\nu_s(Y^c) = \nu_s'(Y^c) = \mu(Y) = 0$. Hence, for all $A \in \mathcal{A}$ and $n \in \mathbb{N}$

$$\begin{aligned}\nu_s'(A \cap Y_n) - \nu_s(A \cap Y_n) &= \nu_s'(A \cap Y_n \cap Y) - \nu_s(A \cap Y_n \cap Y) \\ &= \nu_s'((A \cap Y) \cap Y_n) - \nu_s((A \cap Y) \cap Y_n) \\ &= \nu_a((A \cap Y) \cap Y_n) - \nu_a'((A \cap Y) \cap Y_n) = 0.\end{aligned}$$

Hence $\nu_s'(A \cap Y_n) = \nu_s(A \cap Y_n)$ and thus $\nu_a'(A \cap Y_n) = \nu_a(A \cap Y_n)$ for all $A \in \mathcal{A}$ and $n \in \mathbb{N}$. Therefore, for all $A \in \mathcal{A}$ we have that

$$\begin{aligned}\nu_a(A) &= \nu_a\left(\bigcup_{n=1}^\infty (Y_n \cap A)\right) \\ &= \sum_{n=1}^\infty \nu_a(Y_n \cap A) \\ &= \sum_{n=1}^\infty \nu_a'(Y_n \cap A) \\ &= \nu_a'\left(\bigcup_{n=1}^\infty (Y_n \cap A)\right) = \nu_a'(A)\end{aligned}$$

and similarly $\nu_s(A) = \nu_s'(A)$. Hence $\nu_a' = \nu_a$ and $\nu_s' = \nu_s$ as desired. ∎

There are several quick corollaries to the Lebesgue Decomposition Theorem (Theorem 5.6.1).

Corollary 5.6.2. *Let $F : [0, 1] \to [0, 1]$ denote the Cantor ternary function (with $F(x) = 0$ if $x < 0$ and $F(x) = 1$ if $x > 1$) and let λ_F denote the Borel-Stieltjes measure associated to F. Then $\lambda_F \perp \lambda$.*

Proof. To see this, note that λ_F and λ are σ-finite measures on $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$. Thus the Lebesgue Decomposition Theorem (Theorem 5.6.1) implies that there exists a unique pair of measure (ν_a, ν_s) on $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$ such that $\nu_a \ll \lambda$, $\nu_s \perp \lambda$, and $\lambda_F = \nu_a + \nu_s$.

By the Radon-Nikodym Theorem (Theorem 5.5.5) there exists a measurable function $f : \mathbb{R} \to [0, \infty]$ such that

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

for all $A \in \mathfrak{B}(\mathbb{R})$. However, if $A \in \mathfrak{B}(\mathbb{R})$ and A is a subset an interval in the complement of the Cantor set, we see that

$$0 = \lambda_F(A) = \nu_s(A) + \int_A f \, d\lambda.$$

Hence $\int_A f \, d\lambda = 0$ for all such A. Therefore we obtain that $f|_{\mathcal{C}^c} = 0$ λ-almost everywhere. Therefore, since $\lambda(\mathcal{C}) = 0$, we obtain that $f = 0$ λ-almost everywhere. Hence $\nu_a = 0$ so $\lambda_F = \nu_s$ is mutually singular to λ as desired. ∎

Corollary 5.6.3. *Let $(X, \mathcal{A}, \mu)$ be a σ-finite measure space. If ν is a finite signed measure on $(X, \mathcal{A})$, then then there exists a unique pair of finite signed measures (ν_a, ν_s) on $(X, \mathcal{A})$ such that $\nu_a \ll \mu$, $\nu_s \perp \mu$, and $\nu = \nu_a + \nu_s$.*

Proof. Since ν is a finite signed measure, ν_+ and ν_- are finite measures by Lemma 5.4.6. Therefore, the Lebesgue Decomposition Theorem (Theorem 5.6.1) implies there exists measures $\nu_{+,a}, \nu_{+,s}, \nu_{-,a}$ and $\nu_{-,s}$ on $(X, \mathcal{A})$ such that $\nu_{+,a}, \nu_{-,a} \ll \mu$, $\nu_{+,s} \perp \mu$, $\nu_{-,s} \perp \mu$, $\nu_+ = \nu_{+,a} + \nu_{+,s}$, and $\nu_- = \nu_{-,a} + \nu_{-,s}$. Furthermore, since ν_+ and ν_- are finite measures, we see that $\nu_{+,a}, \nu_{+,s}, \nu_{-,a}$ and $\nu_{-,s}$ are all finite measures.

Let $\nu_a = \nu_{+,a} - \nu_{-,a}$ and $\nu_s = \nu_{+,s} - \nu_{-,s}$. Clearly ν_a and ν_s are finite signed measures on $(X, \mathcal{A})$ such that $\nu = \nu_a + \nu_s$. To see that $\nu_a \ll \mu$, let $A \in \mathcal{A}$ be an arbitrary element such that $\mu(A) = 0$. Then, as $\nu_{+,a}, \nu_{-,a} \ll \mu$, we clearly have $\nu_a(A) = 0$. Hence, since $A \in \mathcal{A}$ was arbitrary, $\nu_a \ll \mu$. Similarly, since $\nu_{+,s} \perp \mu$ and $\nu_{-,s} \perp \mu$ there exists $N, B \in \mathcal{A}$ such that $\nu_{+,s}(N^c) = \mu(N) = 0 = \mu(B) = \nu_{-,s}(B^c)$. Hence, if $Y = B \cup N \in \mathcal{A}$, then $\nu_{+,s}(Y^c) = \nu_{-,s}(Y^c) = \mu(Y) = 0$. Hence $\nu_s(Y^c) = 0 = \mu(Y)$ so $\nu_s \perp \mu$ as desired.

Finally, the proof of uniqueness may be repeated verbatim from the proof of uniqueness used in the Lebesgue Decomposition Theorem (Theorem 5.6.1). ∎

Chapter 6

Product Measures and Fubini's Theorem

One idea from undergraduate studies that has yet to be discussed in the context of measure theory is integrating functions of multiple variables. Of course, we have discussed arbitrary measure spaces so we have a notion of integrating functions on $\mathbb{R}^n$ with respect to λ_n. However, we do not have any nice methods for computing integrals of functions on $\mathbb{R}^n$ with respect to λ_n other than definitions. It seems like λ_n should be related to λ and we should be able to integrate functions on $\mathbb{R}^n$ with respect to λ_n by performing some integrals using λ; that is, we should have a Fubini's Theorem to compute the integrals by integrating over the variables one at a time.

Thus the focus of this chapter is to develop and prove a version of Fubini's Theorem in the measure-theoretic context. Before we can do that, we need to discuss how to construct measures on a product of measure spaces. Subsequently, we will be able to prove Fubini's Theorem and a theorem by Tonelli, which is often first necessary to use in order to invoke Fubini's Theorem.

6.1 Product Measures

To begin, given two measure spaces $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$, we desire a notion of the product of μ and ν so that the measure of a Cartesian product an element $A \in \mathcal{A}$ and $B \in \mathcal{B}$ is $\mu(A)\nu(B)$ (i.e. the measure of a rectangle is the products of the measures of the length and width). This is also exactly what we need to study independence in probability theory; that is, the joint probability measure of two independent random variables will be this product of the two individual probability measures. Moreover, the hope is then that the product of n-copies of λ is λ_n. In order to define such a measure, we first need to construct a σ-algebra. Since we want $A \times B$ to be measurable for all $A \in \mathcal{A}$ and $B \in \mathcal{B}$, a natural place to start is the following.

Definition 6.1.1. Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be measurable spaces. The *set of measurable rectangles of* $(X, \mathcal{A})$ *and* $(Y, \mathcal{B})$, denote $\mathcal{R}(\mathcal{A} \times \mathcal{B})$, is the set of all subsets of $X \times Y$ of the form $A \times B$ where $A \in \mathcal{A}$ and $B \in \mathcal{B}$. Elements of $\mathcal{R}(\mathcal{A} \times \mathcal{B})$ are called *measurable rectangles*.

There is a natural algebra (not σ-algebra) once we have the measurable rectangles.

Lemma 6.1.2. *Let* $(X, \mathcal{A})$ *and* $(Y, \mathcal{B})$ *be measurable spaces and let*

$$\mathfrak{A}(\mathcal{A} \times \mathcal{B}) = \{E \subseteq X \times Y \mid E \textit{ is a finite union of elements of } \mathcal{R}(\mathcal{A} \times \mathcal{B})\}.$$

Then $\mathfrak{A}(\mathcal{A} \times \mathcal{B})$ *is an algebra on* $X \times Y$. *Furthermore, if* $Z \in \mathfrak{A}(\mathcal{A} \times \mathcal{B})$, *then there exists a pairwise disjoint collection* $\{R_k\}_{k=1}^n \subseteq \mathcal{R}(\mathcal{A} \times \mathcal{B})$ *such that* $Z = \bigcup_{k=1}^n R_k$.

Proof. To see that $\mathfrak{A}(\mathcal{A} \times \mathcal{B})$ is an algebra on $X \times Y$, first notice that $\emptyset = \emptyset \times \emptyset \in \mathcal{R}(\mathcal{A} \times \mathcal{B}) \subseteq \mathfrak{A}(\mathcal{A} \times \mathcal{B})$ and $X \times Y \in \mathcal{R}(\mathcal{A} \times \mathcal{B}) \subseteq \mathfrak{A}(\mathcal{A} \times \mathcal{B})$ by definition. Furthermore, clearly $\mathfrak{A}(\mathcal{A} \times \mathcal{B})$ is closed under finite unions by definition.

To see that $\mathfrak{A}(\mathcal{A} \times \mathcal{B})$ is closed under complements and thus an algebra on $X \times Y$, we first claim that $\mathfrak{A}(\mathcal{A} \times \mathcal{B})$ is closed under finite intersections. Indeed $\mathcal{R}(\mathcal{A} \times \mathcal{B})$ is clearly closed under intersections as

$$(A \times B) \cap (A' \times B') = (A \cap A') \times (B \cap B').$$

Therefore de Morgan's laws implies that $\mathfrak{A}(\mathcal{A} \times \mathcal{B})$ is closed under finite intersections.

Next notice if $A \in \mathcal{A}$ and $B \in \mathcal{B}$ then

$$(A \times B)^c = (A^c \times B) \cup (A \times B^c) \cup (A^c \times B^c) \in \mathfrak{A}(\mathcal{A} \times \mathcal{B}).$$

To complete the claim that $\mathfrak{A}(\mathcal{A} \times \mathcal{B})$ is closed under complements, let $Z \in \mathfrak{A}(\mathcal{A} \times \mathcal{B})$ be arbitrary. Then we may write $Z = \bigcup_{k=1}^n A_k \times B_k$ where $\{A_k\}_{k=1}^n \subseteq \mathcal{A}$ and $\{B_k\}_{k=1}^n \subseteq \mathcal{B}$ by the definition of $\mathfrak{A}(\mathcal{A} \times \mathcal{B})$. Therefore

$$Z^c = \bigcap_{k=1}^n \left((A_k^c \times B_k) \cup (A_k \times B_k^c) \cup (A_k^c \times B_k^c)\right).$$

Hence, since $\mathcal{A}$ and $\mathcal{B}$ are closed under complements and since $\mathfrak{A}(\mathcal{A} \times \mathcal{B})$ is closed under finite unions and intersections, we obtain that $Z^c \in \mathfrak{A}(\mathcal{A} \times \mathcal{B})$ as desired. Hence $\mathfrak{A}(\mathcal{A} \times \mathcal{B})$ is an algebra.

For the other claim, let $Z \in \mathfrak{A}(\mathcal{A} \times \mathcal{B})$ be arbitrary. Hence we may write $Z = \bigcup_{k=1}^n A_k \times B_k$ where $\{A_k\}_{k=1}^n \subseteq \mathcal{A}$ and $\{B_k\}_{k=1}^n \subseteq \mathcal{B}$ by the definition of $\mathfrak{A}(\mathcal{A} \times \mathcal{B})$. We will proceed by recursion on m to show that $\bigcup_{k=1}^m A_k \times B_k$ can be written as a disjoint union of elements of $\mathcal{R}(\mathcal{A} \times \mathcal{B})$. Clearly the case $m = 1$ is trivial.

Assume it has been demonstrated for some $m \geq 1$ that $\bigcup_{k=1}^{m} A_k \times B_k = \bigcup_{k=1}^{M} A'_k \times B'_k$ where $\{A'_k \times B'_k\}_{k=1}^{M}$ are pairwise disjoint elements of $\mathcal{R}(\mathcal{A} \times \mathcal{B})$. To see that $\bigcup_{k=1}^{m+1} A_k \times B_k$ can be written as a disjoint union of elements of $\mathcal{R}(\mathcal{A} \times \mathcal{B})$, consider

$$X_1 = (A_{m+1} \times B_{m+1}) \setminus (A'_1 \times B'_1).$$

By the above computations, the set X_1 can be written as the disjoint union of at most three measurable rectangles, namely

$$\begin{aligned} R_1 &= (A_{m+1} \setminus A'_1) \times (B_{m+1} \cap B'_1) \\ R_2 &= (A_{m+1} \cap A'_1) \times (B_{m+1} \setminus B'_1) \\ R_3 &= (A_{m+1} \setminus A'_1) \times (B_{m+1} \setminus B'_1). \end{aligned}$$

Then, for each $k \in \{1, 2, 3\}$, consider

$$X_{2,k} = R_k \setminus (A'_2 \times B'_2).$$

Hence, for each k, $X_{2,k}$ can be written as the disjoint union of at most three measurable rectangles, and the collection of all measurable rectangles obtained over all $k \in \{1, 2, 3\}$ is pairwise disjoint as R_1, R_2, and R_3 are pairwise disjoint. Therefore, by repeating this process ad nauseum, we obtain that $\bigcup_{k=1}^{m+1} A_k \times B_k$ can be written as a disjoint union of elements of $\mathcal{R}(\mathcal{A} \times \mathcal{B})$. ∎

In order to prove our definition of a measure on the algebra $\mathfrak{A}(\mathcal{A} \times \mathcal{B})$ is indeed a measure, we will need the following.

Lemma 6.1.3. *Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be measure spaces and let $A \times B \in \mathcal{R}(\mathcal{A} \times \mathcal{B})$. If $\{A_k \times B_k\}_{k=1}^{\infty}$ are pairwise disjoint elements of $\mathcal{R}(\mathcal{A} \times \mathcal{B})$ such that*

$$A \times B = \bigcup_{k=1}^{\infty} A_k \times B_k,$$

then

$$\mu(A)\nu(B) = \sum_{k=1}^{\infty} \mu(A_k)\nu(B_k).$$

Proof. To begin, notice for each $(x, y) \in A \times B$ that (x, y) is contained in exactly one element of $\{A_k \times B_k\}_{k=1}^{\infty}$. Therefore, for each $x \in A$

$$B = \bigcup_{k \text{ such that } x \in A_k} B_k.$$

Therefore, since for each $x \in A$ the collection

$$\{B_k \mid k \in \{1, \dots, n\} \text{ is such that } x \in A_k\}$$

is pairwise disjoint and since ν is a measure, we obtain that

$$\nu(B) = \sum_{k \text{ such that } x \in A_k} \nu(B_k)$$

Hence

$$\nu(B)\chi_A(x) = \sum_{k=1}^{\infty} \nu(B_k)\chi_{A_k}(x)$$

for each $x \in A$. However, since $A_k \subseteq A$ for all $k \in \mathbb{N}$, we obtain that

$$\nu(B)\chi_A = \sum_{k=1}^{\infty} \nu(B_k)\chi_{A_k}$$

as functions on X. Therefore, since all terms are positive, we obtain by Corollary 3.2.5 that

$$\begin{aligned}
\mu(A)\nu(B) &= \int_X \nu(B)\chi_A \, d\mu \\
&= \int_X \sum_{k=1}^{\infty} \nu(B_k)\chi_{A_k} \, d\mu \\
&= \sum_{k=1}^{\infty} \int_X \nu(B_k)\chi_{A_k} \, d\mu \\
&= \sum_{k=1}^{\infty} \mu(A_k)\nu(B_k)
\end{aligned}$$

as desired. ∎

We now can construct a pre-measure.

Lemma 6.1.4. *Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be measure spaces. Then there exists a unique pre-measure $\mu \cdot \nu$ on the algebra $\mathfrak{A}(\mathcal{A} \times \mathcal{B})$ from Lemma 6.1.2 such that*

$$(\mu \cdot \nu)(A \times B) = \mu(A)\nu(B)$$

for all $A \in \mathcal{A}$ and $B \in \mathcal{B}$.

Proof. We desire to define $\mu \cdot \nu : \mathfrak{A}(\mathcal{A} \times \mathcal{B}) \to [0, \infty]$ such that $(\mu \cdot \nu)(\emptyset) = 0$ and

$$(\mu \cdot \nu)(Z) = \sum_{k=1}^{n} \gamma(A_k \times B_k) = \sum_{k=1}^{n} \mu(A_k)\nu(B_k)$$

whenever $Z \in \mathfrak{A}(\mathcal{A} \times \mathcal{B})$ and $\{A_k \times B_k\}_{k=1}^{n} \subseteq \mathcal{R}(\mathcal{A} \times \mathcal{B})$ are pairwise disjoint such that $Z = \bigcup_{k=1}^{n} A_k \times B_k$. However, since there are multiple ways to write an element $Z \in \mathfrak{A}(\mathcal{A} \times \mathcal{B})$ as a disjoint union of elements of $\mathcal{R}(\mathcal{A} \times \mathcal{B})$, it is necessary to show that the above definition of $(\mu \cdot \nu)$ is well-defined. However, since $\mu \cdot \nu$ is clearly well-defined on elements of $\mathcal{R}(\mathcal{A} \times \mathcal{B})$ by Lemma 6.1.3, we see that $\mu \cdot \nu$ is well-defined on $\mathfrak{A}(\mathcal{A} \times B)$.

To see that $\mu \cdot \nu$ is a pre-measure on $\mathfrak{A}(\mathcal{A} \times \mathcal{B})$, we first see that $(\mu \cdot \nu)(\emptyset) = 0$. Furthermore, if $\{Z_n\}_{n=1}^\infty \subseteq \mathfrak{A}(\mathcal{A} \times \mathcal{B})$ are pairwise disjoint such that $Z = \bigcup_{k=1}^\infty Z_n \in \mathfrak{A}(\mathcal{A} \times \mathcal{B})$, we easily see via the definition of $\mu \cdot \nu$ and Lemma 6.1.3 that

$$(\mu \cdot \nu)(Z) = \sum_{n=1}^\infty (\mu \cdot \nu)(Z_n).$$

Hence $\mu \cdot \nu$ is a pre-measure on $\mathfrak{A}(\mathcal{A} \times \mathcal{B})$.

To see uniqueness, assume $\gamma : \mathfrak{A}(\mathcal{A} \times \mathcal{B}) \to [0, \infty]$ is a pre-measure on $\mathfrak{A}(\mathcal{A} \times \mathcal{B})$ such that $\gamma(A \times B) = \mu(A)\nu(B)$ for all $A \times B \in \mathcal{R}(\mathcal{A} \times \mathcal{B})$. If $Z \in \mathfrak{A}(\mathcal{A} \times \mathcal{B})$ is arbitrary then Lemma 6.1.2 implies pairwise disjoint collection $\{A_k \times B_k\}_{k=1}^n \subseteq \mathcal{R}(\mathcal{A} \times \mathcal{B})$ such that $Z = \bigcup_{k=1}^n A_k \times B_k$. Hence, since γ is a pre-measure,

$$\gamma(Z) = \sum_{k=1}^n \gamma(A_k \times B_k) = \sum_{k=1}^n \mu(A_k)\nu(B_k) = (\mu \cdot \nu)(Z).$$

Therefore, since $Z \in \mathfrak{A}(\mathcal{A} \times \mathcal{B})$ was arbitrary, $\gamma = \mu \cdot \nu$ as desired. ■

Using Chapter 1, we obtain the measure we are looking for.

Definition 6.1.5. Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be measure spaces. The *product measure of* μ *and* ν, denoted $\mu \times \nu$, is the Carathéodory extension of the pre-measure $\mu \cdot \nu$ on the algebra $\mathfrak{A}(\mathcal{A} \times \mathcal{B})$ from Lemma 6.1.4. The collection of $\mu \times \nu$ measurable sets is denoted $\mathcal{M}(\mu \times \nu)$.

Remark 6.1.6. Recall from the Carathéodory-Hahn Extension Theorem (Theorem 1.3.7) that

$$\mathcal{R}(\mathcal{A} \times \mathcal{B}) \subseteq \mathfrak{A}(A \times \mathcal{B}) \subseteq \mathcal{M}(\mu \times \nu)$$

and that

$$(\mu \times \nu)(A \times B) = (\mu \cdot \nu)(A \times B) = \mu(A)\nu(B)$$

for all $A \in \mathcal{A}$ and $B \in \mathcal{B}$. Furthermore, since $\mu \times \nu$ is a Carathéodory Extension, $\mu \times \nu$ is automatically complete by Proposition 1.2.14.

Example 6.1.7. We claim that $\lambda \times \lambda = \lambda_2$; that is, the product measure of two one-dimensional Lebesgue measures is the two-dimensional Lebesgue measure. Indeed recall by construction that for all $Z \in \mathbb{R} \times \mathbb{R} = \mathbb{R}^2$ we have

$$\begin{aligned}
&(\lambda \cdot \lambda)^*(Z) \\
&= \inf\left\{ \sum_{n=1}^{\infty} (\lambda \cdot \lambda)(Z_n) \,\middle|\, \{Z_n\}_{n=1}^{\infty} \subseteq \mathfrak{A}(\mathcal{M}(\mathbb{R}) \times \mathcal{M}(\mathbb{R})), Z \subseteq \bigcup_{n=1}^{\infty} Z_n \right\} \\
&= \inf\left\{ \sum_{n=1}^{\infty} (\lambda \cdot \lambda)(Z_n) \,\middle|\, \{Z_n\}_{n=1}^{\infty} \subseteq \mathcal{R}(\mathcal{M}(\mathbb{R}) \times \mathcal{M}(\mathbb{R})), Z \subseteq \bigcup_{n=1}^{\infty} Z_n \right\} \\
&= \inf\left\{ \sum_{n=1}^{\infty} \lambda(A_n)\lambda(B_n) \,\middle|\, \{A_n, B_n\}_{n=1}^{\infty} \subseteq \mathcal{M}(\mathbb{R}), Z \subseteq \bigcup_{n=1}^{\infty} A_n \times B_n \right\}.
\end{aligned}$$

However, since

$$\lambda(A) = \inf\left\{ \sum_{n=1}^{\infty} \lambda(I_n) \,\middle|\, \{I_n\}_{n=1}^{\infty} \text{ open intervals}, A \subseteq \bigcup_{n=1}^{\infty} I_n \right\},$$

we see that

$$\begin{aligned}
&(\lambda \cdot \lambda)^*(Z) \\
&= \inf\left\{ \sum_{n=1}^{\infty} \lambda(I_n)\lambda(J_n) \,\middle|\, \{I_n, J_n\}_{n=1}^{\infty} \text{ open intervals}, Z \subseteq \bigcup_{n=1}^{\infty} I_n \times J_n \right\} \\
&= \lambda_2^*(Z)
\end{aligned}$$

by the definition of the two-dimensional Lebesgue outer measure from Definition 1.2.6. Hence both $(\lambda \cdot \lambda)^*$ and λ_2^* are the outer measure associated to the pre-measure $\lambda \cdot \lambda$ on the algebra $\mathfrak{A}(\mathcal{M}(\mathbb{R}) \times \mathcal{M}(\mathbb{R}))$. Therefore, since $\lambda \cdot \lambda$ is clearly σ-finite in the sense of the Carathéodory-Hahn Extension Theorem (Theorem 1.3.7), the Carathéodory-Hahn Extension Theorem implies that $\mathcal{M}(\lambda \times \lambda)$ is the two-dimensional Lebesgue measurable sets and $\lambda \times \lambda = \lambda_2$ as desired.

Remark 6.1.8. It is not difficult to see that the operation of taking the product of measures is associative in the sense that if $\{(X_k, \mathcal{A}_k, \mu_k)\}_{k=1}^n$ are measure spaces, then $\mathcal{M}((\mu_1 \times \mu_2) \times \mu_3) = \mathcal{M}(\mu_1 \times (\mu_2 \times \mu_3))$ and $(\mu_1 \times \mu_2) \times \mu_3 = \mu_1 \times (\mu_2 \times \mu_3)$. Indeed if

$$\mathfrak{A} = \left\{ Z \subseteq X_1 \times X_2 \times X_3 \,\middle|\, Z = \bigcup_{k=1}^{n} Z_k \text{ where } Z_k \in \mathcal{A}_1 \times \mathcal{A}_2 \times \mathcal{A}_3 \right\}$$

then $\mathfrak{A}$ is an algebra by similar arguments to those used in Lemma 6.1.2. By using our knowledge of product measures, it is possible to extend Lemma 6.1.3 to triples (i.e. take one measure in the assumptions of Lemma 6.1.3 to be a product measure) and then extend Lemma 6.1.4 to construct a unique pre-measure $\mu_1 \cdot \mu_2 \cdot \mu_3$ on $\mathfrak{A}$ such that

$$(\mu_1 \cdot \mu_2 \cdot \mu_3)(A_1 \times A_2 \times A_3) = \mu_1(A_1)\mu_2(A_2)\mu_3(A_3)$$

for all $A_k \in \mathcal{A}_k$. Then, by similar arguments to those used in Example 6.1.7, it is possible to see that both $(\mu_1 \times \mu_2) \times \mu_3$ and $\mu_1 \times (\mu_2 \times \mu_3)$ are the Carathéodory Extension of $\mu_1 \cdot \mu_2 \cdot \mu_3$. In particular, by extending the proof of Example 6.1.7, one can see that

$$\underbrace{\lambda \times \lambda \times \cdots \times \lambda}_{n \text{ times}} = \lambda_n.$$

Not only are product measures automatically complete, product measures inherit properties from their underlying measures.

Proposition 6.1.9. *Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be measurable spaces. If both μ and ν are finite, then $\mu \times \nu$ is finite. Similarly, if both μ and ν are σ-finite, then $\mu \times \nu$ is σ-finite.*

Proof. Notice if $A \in \mathcal{A}$ and $B \in \mathcal{B}$ are such that $\mu(A), \nu(B) < \infty$, then $(\mu \times \nu)(A \times B) < \infty$. Therefore, clearly $\mu \times \nu$ is finite if μ and ν are finite.

Assume μ and ν are σ-finite. Then there exists collections $\{A_n\}_{n=1}^\infty \subseteq \mathcal{A}$ and $\{B_n\}_{n=1}^\infty \subseteq \mathcal{B}$ such that

$$X = \bigcup_{n=1}^\infty A_n, \quad Y = \bigcup_{n=1}^\infty B_n, \quad \text{and} \quad \mu(A_n), \nu(B_n) < \infty$$

for all $n \in \mathbb{N}$. For each $n, m \in \mathbb{N}$, let

$$Z_{n,m} = A_n \times B_m.$$

Clearly $\{Z_{n,m}\}_{n,m=1}^\infty$ is a countable collection of elements of $\mathcal{M}(\mu \times \nu)$ such that $(\mu \times \nu)(Z_{n,m}) < \infty$ for all $n, m \in \mathbb{N}$ and $X \times Y = \bigcup_{n,m=1}^\infty Z_{n,m}$. Hence $\mu \times \nu$ is σ-finite as desired. ∎

6.2 Tonelli's and Fubini's Theorem

In this section, we will state our two main theorems. The proofs will be postponed until the next section in order to provide examples to see why all of the conditions are necessary.

Theorem 6.2.1 (Fubini's Theorem). *Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be complete measure spaces. If $f \in L_1(X \times Y, \mu \times \nu)$, then:*

1. *for μ-almost every $x \in X$ the function $f_x : Y \to \mathbb{C}$ defined by $f_x(y) = f(x, y)$ for all $y \in Y$ is a well-defined element of $L_1(Y, \nu)$ and for ν-almost every $y \in Y$ the function $f_y : X \to \mathbb{C}$ defined by $f_y(x) = f(x, y)$ for all $x \in X$ is a well-defined element of $L_1(X, \mu)$,*

2. *the function $\Phi : X \to \mathbb{C}$ defined by $\Phi(x) = \int_Y f_x \, d\nu$ is a well-defined element of $L_1(X, \mu)$ and the function $\Psi : Y \to \mathbb{C}$ defined by $\Psi(y) = \int_X f_y \, d\mu$ is a well-defined element of $L_1(Y, \nu)$, and*

3. $\int_{X\times Y} f\,d(\mu\times\nu) = \int_X \Phi\,d\mu = \int_Y \Psi\,d\nu$; *that is*

$$\int_{X\times Y} f\,d(\mu\times\nu) = \int_X \left(\int_Y f(x,y)\,d\nu(y)\right) d\mu(x)$$
$$= \int_Y \left(\int_X f(x,y)\,d\mu(x)\right) d\nu(y).$$

Theorem 6.2.2 (Tonelli's Theorem). *Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be complete, σ-finite measure spaces. If $f : X\times Y \to [0,\infty]$ is $\mu\times\nu$-measurable, then:*

1. *for μ-almost every $x \in X$ the function $f_x : Y \to [0,\infty]$ defined by $f_x(y) = f(x,y)$ for all $y \in Y$ is a well-defined ν-measurable function and for ν-almost every $y \in Y$ the function $f_y : X \to [0,\infty]$ defined by $f_y(x) = f(x,y)$ for all $x \in X$ is a well-defined μ-measurable function,*

2. *the function $\Phi : X \to [0,\infty]$ defined by $\Phi(x) = \int_Y f_x\,d\nu$ is a well-defined μ-measurable function and the function $\Psi : Y \to [0,\infty]$ defined by $\Psi(y) = \int_X f_y\,d\mu$ is a well-defined ν-measurable function, and*

3. $\int_{X\times Y} f\,d(\mu\times\nu) = \int_X \Phi\,d\mu = \int_Y \Psi\,d\nu$; *that is*

$$\int_{X\times Y} f\,d(\mu\times\nu) = \int_X \left(\int_Y f(x,y)\,d\nu(y)\right) d\mu(x)$$
$$= \int_Y \left(\int_X f(x,y)\,d\mu(x)\right) d\nu(y).$$

Remark 6.2.3. It is clear that Fubini's Theorem (Theorem 6.2.1) is more general that Tonelli's Theorem (Theorem 6.2.2) 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 $f \in L_1(\mu\times\nu)$, one generally verifies that

$$\int_{X\times Y} |f|\,d(\mu\times\nu) < \infty$$

using Tonelli's Theorem!

To see why all of the conditions are necessary in Fubini's Theorem (Theorem 6.2.1) and Tonelli's Theorem (Theorem 6.2.2), we exhibit the following three examples.

Example 6.2.4. Let $X = Y = \mathbb{N}$ and let μ and ν be the counting measure on $\mathbb{N}$. Consider the function $f : \mathbb{N}\times\mathbb{N} \to \mathbb{R}$ defined by

$$f(n,m) = \begin{cases} 1 & \text{if } n = m \\ -1 & \text{if } m = n+1 \\ 0 & \text{otherwise} \end{cases}.$$

Then

$$\int_X \left(\int_Y f(n,m)\, d\nu(m) \right) d\mu(n) = \int_X 0\, d\mu(n) = 0$$

(since for any n, m obtains the values n and $n+1$ as m varies over $\mathbb{N}$) whereas

$$\int_Y \left(\int_X f(n,m)\, d\mu(n) \right) d\nu(m) = \int_Y \chi_{m=1}\, d\nu(m) = 1$$

(i.e. if $m = 1$ then n takes the value m once and does not take the value $m-1$ as n varies over $\mathbb{N}$, whereas if $m \geq 2$ then n takes the values m and $m-1$ exactly once each as n varies over $\mathbb{N}$). Hence the final conclusions of Fubini's Theorem (Theorem 6.2.1) and Tonelli's Theorem (Theorem 6.2.2) do not hold in this setting.

To see why this does not contradict the theorems, notice Tonelli's Theorem does not apply as f is not positive. To see why Fubini's Theorem does not apply, first notice by the properties of the product measure that $\mu \times \nu$ is the counting measure on $\mathbb{N} \times \mathbb{N}$. Therefore

$$\int_{X \times Y} |f|\, d(\mu \times \nu) = \infty$$

as $|f(n,m)| = 1$ for infinitely many $(n,m) \in \mathbb{N}^2$. Hence $f \notin L_1(X \times Y, \mu \times \nu)$ so Fubini's Theorem does not apply.

Example 6.2.5. Let $X = Y = [0,1]$, let $\mu = \lambda$, and let ν be the counting measure on Y. Consider the function $f : X \times Y \to \mathbb{R}$ defined by

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

Then

$$\int_X \left(\int_Y f(x,y)\, d\nu(y) \right) d\mu(x) = \int_X 1\, d\mu(x) = 1$$

(since ν is the counting measure and since for each $x \in X$ there exists a unique $y \in Y$ such that $f(x,y) \neq 0$ in which case $f(x,y) = 1$) whereas

$$\int_Y \left(\int_X f(x,y)\, d\mu(x) \right) d\nu(y) = \int_Y 0\, d\nu(y) = 0$$

(since $f(x,y) = 0$ for λ-almost every x for each $y \in Y$). Hence the final conclusions of Fubini's Theorem (Theorem 6.2.1) and Tonelli's Theorem (Theorem 6.2.2) do not hold in this setting.

To see why this does not contradict the theorems, notice Tonelli's Theorem does not apply since ν is not σ-finite on Y. To see why Fubini's Theorem does not apply, first notice that $f = \chi_\Delta$ where

$$\Delta = \{(x,x) \mid x \in [0,1]\}.$$

We claim that

$$(\mu \times \nu)(\Delta) = \infty$$

which implies $f \notin L_1(X \times Y, \mu \times \nu)$ so Fubini's Theorem does not apply. To see the claim recall that

$$(\mu \times \nu)(\Delta) = \inf \left\{ \sum_{n=1}^{\infty} \mu(A_n)\nu(B_n) \,\middle|\, \substack{\Delta \subseteq \bigcup_{n=1}^{\infty} A_n \times B_n, \\ \{A_n\}_{n=1}^{\infty} \subseteq \mathcal{M}(\mathbb{R}), \{B_n\}_{n=1}^{\infty} \subseteq \mathcal{P}([0,1])} \right\}.$$

Assume $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{M}(\mathbb{R})$ and $\{B_n\}_{n=1}^{\infty} \subseteq \mathcal{P}([0,1])$ are arbitrary collections such that $\Delta \subseteq \bigcup_{n=1}^{\infty} A_n \times B_n$. Then for each $x \in [0,1]$ we have

$$(x, x) \in \Delta \subseteq \bigcup_{n=1}^{\infty} A_n \times B_n$$

so there exists an $n \in \mathbb{N}$ such that $x \in A_n \cap B_n$. Therefore $[0,1] \subseteq \bigcup_{n=1}^{\infty}(A_n \cap B_n)$. Therefore, since

$$1 = \lambda([0,1]) \leq \sum_{n=1}^{\infty} \lambda^*(A_n \cap B_n),$$

it must be the case that there exists an $n_0 \in \mathbb{N}$ such that $\lambda^*(A_{n_0} \cap B_{n_0}) > 0$. This implies $\lambda(A_{n_0}) > 0$ and B_{n_0} is infinite. Hence $\lambda(A_{n_0})\mu(B_{n_0}) = \infty$. Therefore, since $\{A_n\}_{n=1}^{\infty}$ and $\{B_n\}_{n=1}^{\infty}$ were arbitrary, the claim is complete.

Example 6.2.6. Let $X = Y = [0,1]$ and let $\mu = \nu = \lambda$ on $[0,1]$. By the Well-Ordering Principle and by assuming the Continuum Hypothesis (i.e. so the cardinality of $[0,1]$ is the first uncountable cardinal $\aleph_1$), there exists a well-ordering $\preceq$ on $[0,1]$ such that $\{x \in [0,1] \mid x \preceq y\}$ is countable for all $y \in [0,1]$ except for the one corresponding to $\aleph_1$.

Let $Z = \{(x,y) \in [0,1]^2 \mid x \preceq y\}$ and let $f = \chi_Z$. Then

$$\begin{aligned} \int_X \left(\int_Y f(x,y)\, d\nu(y) \right) d\mu(x) &= \int_X \lambda(\{y \in [0,1] \mid x \preceq y\})\, d\mu(x) \\ &= \int_X 1 \, d\mu(x) = 1 \end{aligned}$$

(since $\{y \in [0,1] \mid x \preceq y\}$ contains all but a countable number of elements of $[0,1]$ for λ-almost every x) whereas

$$\begin{aligned} \int_Y \left(\int_X f(x,y)\, d\mu(x) \right) d\nu(y) &= \int_Y \lambda(\{x \in [0,1] \mid x \preceq y\})\, d\nu(y) \\ &= \int_Y 0 \, d\nu(y) = 0 \end{aligned}$$

(since $\{x \in [0,1] \mid x \preceq y\}$ is countable for all but one $y \in [0,1]$). Hence the final conclusions of Fubini's Theorem (Theorem 6.2.1) and Tonelli's Theorem (Theorem 6.2.2) do not hold in this setting.

It turns out that Z is not a measurable set for $\lambda \times \lambda$ (and thus not two-dimensional Lebesgue measurable by Example 6.1.7). Indeed suppose for the sake of a contradiction that Z is measurable. Since $\lambda \times \lambda$ is a finite measure on $[0, 1]$ by Proposition 6.1.9, $(\lambda \times \lambda)(Z) < \infty$ and thus $f = \chi_Z \in L_1(\lambda \times \lambda)$. Since λ is complete, all the assumptions of Fubini's Theorem then holds thereby contradicting the above computations. Hence Z is not a measurable set for $\lambda \times \lambda$.

Thus this really isn't an example of why the hypotheses of Fubini and Tonelli's Theorem must be satisfied, but another method of constructing a set that is not measurable.

6.3 Proof of Tonelli's and Fubini's Theorem

In this section we will prove Fubini's Theorem (Theorem 6.2.1) and Tonelli's Theorem (Theorem 6.2.2). Since the proofs are long and complicated, we will divided the proofs into several lemmata. The idea of the proof is to first prove the results for the characteristic functions of specific sets, then the characteristic functions of all sets. This gives that the results hold for all simple function from which we will obtain the full results.

To reduce assumptions, we will assume throughout that we are working on fixed complete measure spaces $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ and we establish the following notation.

Notation 6.3.1. Given $Z \subseteq X \times Y$, $x \in X$, and $y \in Y$, denote

$$Z_x = \{w \in Y \mid (x, w) \in Z\} \qquad \text{and} \qquad Z_y = \{z \in X \mid (z, y) \in Z\}.$$

Similarly, given a function $f : X \times Y \to \mathbb{C}$, $x \in X$, and $y \in Y$, let $f_x : Y \to \mathbb{C}$ and $f_y : X \to \mathbb{C}$ denote the functions such that

$$f_x(w) = f(x, w) \qquad \text{and} \qquad f_y(z) = f(z, y)$$

for all $z \in X$ and $w \in Y$. Finally, let

$$\mathcal{R}_\sigma(\mathcal{A} \times \mathcal{B}) = \left\{ Z \subseteq X \times Y \;\middle|\; \substack{Z \text{ is a countable union} \\ \text{of elements of } \mathcal{R}(\mathcal{A} \times \mathcal{B})} \right\}$$
$$\mathcal{R}_{\sigma\delta}(\mathcal{A} \times \mathcal{B}) = \left\{ Z \subseteq X \times Y \;\middle|\; \substack{Z \text{ is a countable intersection} \\ \text{of elements of } \mathcal{R}_\sigma(\mathcal{A} \times \mathcal{B})} \right\}$$

Note since $\mathcal{R}(\mathcal{A} \times \mathcal{B}) \subseteq \mathcal{M}(\mu \times \nu)$ and since $\mathcal{M}(\mu \times \nu)$ is a σ-algebra that

$$\mathcal{R}(\mathcal{A} \times \mathcal{B}) \subseteq \mathfrak{A}(\mathcal{A} \times \mathcal{B}) \subseteq R_\sigma(\mathcal{A} \times B) \subseteq \mathcal{R}_{\sigma\delta}(\mathcal{A} \times B) \subseteq \mathcal{M}(\mu \times \nu).$$

Lemma 6.3.2. *If $Z \in \mathcal{R}_{\sigma\delta}(\mathcal{A} \times \mathcal{B})$, then Z_x is ν-measurable for every $x \in X$.*

Proof. The proof will be divided into three cases of increasing generality.

Case 1: $Z \in \mathcal{R}(\mathcal{A} \times \mathcal{B})$. In this case, we may write $Z = A \times B$ for some $A \in \mathcal{A}$ and $B \in \mathcal{B}$. Therefore, for each $x \in X$,

$$Z_x = \begin{cases} \emptyset & \text{if } x \notin A \\ B & \text{if } x \in A \end{cases}.$$

Since $\emptyset, B$ are elements of $\mathcal{B}$ and thus ν-measurable, Z_x is ν-measurable in this case.

Case 2: $Z \in \mathcal{R}_\sigma(\mathcal{A} \times \mathcal{B})$. In this case, we may write $Z = \bigcup_{n=1}^\infty R_n$ for some collection $\{R_n\}_{n=1}^\infty \subseteq \mathcal{R}(\mathcal{A} \times \mathcal{B})$. Since

$$Z_x = \bigcup_{n=1}^\infty (R_n)_x$$

for each $x \in X$, and since $(R_n)_x \in \mathcal{B}$ for all $n \in \mathbb{N}$ and $x \in X$ by Case 1, we see that $Z_x \in \mathcal{B}$ for all $x \in X$ and thus Z_x is ν-measurable in this case.

Case 3: $Z \in \mathcal{R}_{\sigma\delta}(\mathcal{A} \times \mathcal{B})$. In this case, we may write $Z = \bigcap_{n=1}^\infty Z_n$ for some collection $\{Z_n\}_{n=1}^\infty \subseteq \mathcal{R}_\sigma(\mathcal{A} \times \mathcal{B})$. Since

$$Z_x = \bigcap_{n=1}^\infty (Z_n)_x$$

for each $x \in X$, and since $(Z_n)_x \in \mathcal{B}$ for all $n \in \mathbb{N}$ and $x \in X$ by Case 2, we see that $Z_x \in \mathcal{B}$ for all $x \in X$ and thus Z_x is ν-measurable. Since this is the most general case, the proof is complete. ■

Lemma 6.3.3. *Let $Z \in \mathcal{R}_{\sigma\delta}$ be such that $(\mu \times \nu)(Z) < \infty$ and define $g : X \to [0, \infty]$ by $g(x) = \nu(Z_x)$ for all $x \in X$. Then g is μ-measurable and*

$$\int_X g \, d\mu = (\mu \times \nu)(Z) = \int_{X \times Y} \chi_Z \, d(\mu \times \nu).$$

In particular $\nu(Z_x) < \infty$ for μ-almost every $x \in X$.

Proof. First notice that g is well-defined since Z_x is ν-measurable for all $x \in X$ by Lemma 6.3.2. Furthermore, the equality

$$(\mu \times \nu)(Z) = \int_{X \times Y} \chi_Z \, d(\mu \times \nu)$$

is trivial.

The remainder of the proof will be divided into three cases of increasing generality.

Case 1: $Z \in \mathcal{R}(\mathcal{A} \times \mathcal{B})$. In this case, we may write $Z = A \times B$ for some $A \in \mathcal{A}$ and $B \in \mathcal{B}$. Since

$$Z_x = \begin{cases} \emptyset & \text{if } x \notin A \\ B & \text{if } x \in A \end{cases}$$

for all $x \in X$, we see that

$$g(x) = \nu(Z_x) = \nu(B)\chi_A(x)$$

for all $x \in X$. Hence g is clearly μ-measurable and

$$\int_X g\,d\mu = \nu(B)\int_X \chi_A\,d\mu = \mu(A)\nu(B) = (\mu\times\nu)(Z)$$

as desired.

Case 2: $Z \in \mathcal{R}_\sigma(\mathcal{A}\times\mathcal{B})$. In this case, we may write $Z = \bigcup_{n=1}^\infty R_n$ for some collection $\{R_n\}_{n=1}^\infty \subseteq \mathcal{R}(\mathcal{A}\times\mathcal{B})$. Since $\mathfrak{A}(\mathcal{A}\times\mathcal{B})$ is an algebra constructed by taking all finite unions of elements of $\mathcal{R}(\mathcal{A}\times\mathcal{B})$, we notice the proof of Lemma 6.1.2 implies that we may assume that the collection $\{R_n\}_{n=1}^\infty \subseteq \mathcal{R}(\mathcal{A}\times\mathcal{B})$ is pairwise disjoint. Hence $\{(R_n)_x\}_{n=1}^\infty \subseteq \mathcal{B}$ is pairwise disjoint for all $x \in X$. Therefore

$$g(x) = \nu(Z_x) = \nu\left(\bigcup_{n=1}^\infty (R_n)_x\right) = \sum_{n=1}^\infty \nu((R_n)_x)$$

for all $x \in X$. Therefore, by Case 1, g is a countable sum of non-negative μ-measurable functions and hence is μ-measurable by Proposition 2.1.22. Moreover, by Corollary 3.2.5,

$$\begin{aligned}\int_X g\,d\mu &= \int_X \sum_{n=1}^\infty \nu((R_n)_x)\,d\mu(x)\\ &= \sum_{n=1}^\infty \int_X \nu((R_n)_x)\,d\mu(x)\\ &= \sum_{n=1}^\infty (\mu\times\nu)(R_n)\\ &= (\mu\times\nu)\left(\bigcup_{n=1}^\infty R_n\right)\\ &= (\mu\times\nu)(Z)\end{aligned}$$

as desired.

Case 3: $Z \in \mathcal{R}_{\sigma\delta}(\mathcal{A}\times\mathcal{B})$. In this case, we may write $Z = \bigcap_{n=1}^\infty Z_n$ for some collection $\{Z_n\}_{n=1}^\infty \subseteq \mathcal{R}_\sigma(\mathcal{A}\times\mathcal{B})$. Since $(\mu\times\nu)(Z) < \infty$, the definition of $\mu\times\nu$ implies there exists a $\{C_n\}_{n=1}^\infty \in \mathfrak{A}(\mathcal{A}\times\mathcal{B})$ such that $Z \subseteq \bigcup_{n=1}^\infty C_n$ and

$$(\mu\times\nu)(Z) \leq \sum_{n=1}^\infty (\mu\times\nu)(C_n) < \infty.$$

Therefore, if $Z_0' = \bigcup_{n=1}^\infty C_n$, then $Z_0' \in \mathcal{R}_\sigma(\mathcal{A}\times\mathcal{B})$ and $(\mu\times\nu)(Z_0) < \infty$. Moreover, since the intersection of any two elements of $\mathcal{R}(\mathcal{A}\times\mathcal{B})$ is an element of $\mathcal{R}(\mathcal{A}\times\mathcal{B})$ (i.e. $(A_1\times B_1)\cap(A_2\times B_2) = (A_1\cap A_2)\times(B_1\cap B_2)$), since each element of $\mathcal{R}_\sigma(\mathcal{A}\times\mathcal{B})$ is a countable union of elements of $\mathcal{R}(\mathcal{A}\times\mathcal{B})$, 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 $\{Z'_n\}_{n=0}^\infty \subseteq \mathcal{R}_\sigma(\mathcal{A}\times\mathcal{B})$, $Z = \bigcap_{n=0}^\infty Z'_n$, $(\mu\times\nu)(Z'_0) < \infty$, and $Z'_n \subseteq Z'_{n-1}$ for all $n \in \mathbb{N}$.

For each $n \in \mathbb{N}\cup\{0\}$, define $g_n : X \to [0,\infty]$ by $g_n(x) = \nu((Z'_n)_x)$ for all $x \in X$. Clearly each g_n is μ-measurable by Case 2. Moreover, Case 2 implies that

$$0 \leq \int_X \nu((Z'_0)_x)\, d\mu(x) = \int_X g_0\, d\mu = (\mu\times\nu)(Z'_0) < \infty$$

and thus $\nu((Z'_0)_x) < \infty$ for μ-almost every x. In addition, notice that $Z_x = \bigcap_{n=1}^\infty (Z'_n)_x$ for all $x \in X$ 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 $x \in X$. Therefore, we obtain by the Monotone Convergence Theorem for Measures (Theorem 1.1.23) that

$$\lim_{n\to\infty} g_n(x) = \lim_{n\to\infty} \nu((Z'_n)_x) = \nu(Z_x) = g(x)$$

for μ-almost every x. Therefore, since μ is complete, Corollary 2.1.28 implies that g is μ-measurable.

Next, since $(Z'_n)_x \subseteq (Z'_{n-1})_x$ for all $n \in \mathbb{N}$ and $x \in X$, we see that $g_n(x) \leq g_0(x)$ for all $x \in X$. However, since g_0 is non-negative and μ-measurable, we see by Case 2 that

$$\int_X g_0\, d\mu = (\mu\times\nu)(Z'_0) < \infty$$

and thus $g_0 \in L_1(X,\mu)$. Therefore, by the Dominated Convergence Theorem (Theorem 3.6.1) and Case 2, we obtain that

$$\begin{aligned}\int_X g\, d\mu &= \lim_{n\to\infty}\int_X g_n\, d\mu \\ &= \lim_{n\to\infty}\int_X \nu((Z'_n)_x)\, d\mu(x) \\ &= \lim_{n\to\infty}(\mu\times\nu)(Z'_n).\end{aligned}$$

However, since $(\mu\times\nu)(Z'_0) < \infty$, $Z'_n \subseteq Z'_{n-1}$ for all $n \in \mathbb{N}$, and $Z = \bigcap_{n=1}^\infty Z'_n$, we obtain by the Monotone Convergence Theorem for Measures (Theorem 1.1.23) that

$$\int_X g\, d\mu = \lim_{n\to\infty}(\mu\times\nu)(Z'_n) = (\mu\times\nu)(Z)$$

as desired. Since this is the most general case, the proof is complete. ∎

Note Lemma 6.3.3 proves the desired result (i.e. Fubini's Theorem (Theorem 6.2.1)) 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.

Lemma 6.3.4. *If $Z \in \mathcal{M}(\mu \times \nu)$ is such that $(\mu \times \nu)(Z) < \infty$, then there exists an $G \in \mathcal{R}_{\sigma\delta}(\mathcal{A} \times \mathcal{B})$ such that $Z \subseteq G$ and $(\mu \times \nu)(G \setminus Z) = 0$.*

Proof. Fix $\epsilon > 0$. By the definition of $\mu \times \nu$ there exists a collection $\{C_n\}_{n=1}^\infty \subseteq \mathfrak{A}(\mathcal{A} \times \mathcal{B})$ such that $Z \subseteq \bigcup_{n=1}^\infty C_n$ and

$$\sum_{n=1}^\infty (\mu \times \nu)(C_n) \leq (\mu \times \nu)(Z) + \epsilon.$$

Since each element of $\mathfrak{A}(\mathcal{A} \times \mathcal{B})$ is a finite union of elements of $\mathcal{R}(\mathcal{A} \times \mathcal{B})$, since the countable union of countable sets is countable, and since $\mu \times \nu$ is subadditive, there exists a collection $\{R_n\}_{n=1}^\infty \subseteq \mathcal{R}(\mathcal{A} \times \mathcal{B})$ such that $Z \subseteq \bigcup_{n=1}^\infty R_n$ and

$$\sum_{n=1}^\infty (\mu \times \nu)(R_n) \leq (\mu \times \nu)(Z) + \epsilon.$$

Let $G_\epsilon = \bigcup_{n=1}^\infty R_n \in R_\sigma(\mathcal{A} \times \mathcal{B})$. Then clearly $Z \subseteq G_\epsilon$ and

$$(\mu \times \nu)(Z) \leq (\mu \times \nu)(G_\epsilon) \leq \sum_{n=1}^\infty (\mu \times \nu)(R_n) \leq (\mu \times \nu)(Z) + \epsilon.$$

Let $G = \bigcap_{n=1}^\infty G_{\frac{1}{n}} \in R_{\sigma\delta}(\mathcal{A} \times \mathcal{B})$. Clearly $Z \subseteq G$ as $Z \subseteq G_{\frac{1}{n}}$ for all $n \in \mathbb{N}$. Moreover, clearly

$$(\mu \times \nu)(Z) \leq (\mu \times \nu)(G) \leq (\mu \times \nu)\left(G_{\frac{1}{n}}\right) \leq (\mu \times \nu)(Z) + \frac{1}{n}$$

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

$$(\mu \times \nu)(Z) = (\mu \times \nu)(G).$$

Therefore, since $(\mu \times \nu)(Z) < \infty$ and $Z \subseteq G$, we obtain that

$$(\mu \times \nu)(G \setminus Z) = 0$$

as desired. ∎

Lemma 6.3.5. *If $Z \in \mathcal{M}(\mu \times \nu)$ is such that $(\mu \times \nu)(Z) = 0$, then Z_x is ν-measurable with $\nu(Z_x) = 0$ for μ-almost every $x \in X$.*

Proof. By Lemma 6.3.4 there exists an $G \in R_{\sigma\delta}(\mathcal{A} \times \mathcal{B})$ such that $Z \subseteq G$ and $(\mu \times \nu)(G \setminus Z) = 0$. Hence $(\mu \times \nu)(G) = 0$.

Recall from Lemma 6.3.3 that if we define $g : X \to [0, \infty]$ by $g(x) = \nu(G_x)$ for all $x \in X$, then g is μ-measurable and

$$\int_X g \, d\mu = (\mu \times \nu)(G) = 0.$$

Therefore $0 = g(x) = \nu(G_x)$ for μ-almost every $x \in X$. Since $Z \subseteq G$ so $Z_x \subseteq G_x$ for all $x \in X$ and since ν is complete, we obtain that Z_x is ν-measurable with $\nu(Z_x) = 0$ for μ-almost every $x \in X$. ∎

Lemma 6.3.6. *If $Z \in \mathcal{M}(\mu \times \nu)$ is such that $(\mu \times \nu)(Z) < \infty$, then Fubini's Theorem (Theorem 6.2.1) holds for the function $f = \chi_Z$.*

Proof. Fix $Z \in \mathcal{M}(\mu \times \nu)$ such that $(\mu \times \nu)(Z) < \infty$. By Lemma 6.3.4 there exists an $G \in \mathcal{R}_{\sigma\delta}(\mathcal{A} \times \mathcal{B})$ such that $Z \subseteq G$ and $(\mu \times \nu)(G \setminus Z) = 0$.

Notice for all $x \in X$ that

$$Z_x = (G_x) \setminus (G \setminus Z)_x.$$

Since $(\mu \times \nu)(G) = (\mu \times \nu)(Z) < \infty$ we know that G_x is ν-measurable for all $x \in X$ by Lemma 6.3.3. Moreover, since $(\mu \times \nu)(G \setminus Z) = 0$, we know that $(G \setminus Z)_x$ is ν-measurable for μ-almost every $x \in X$ by Lemma 6.3.5. Hence we obtain that Z_x is ν-measurable for μ-almost every $x \in X$. Moreover, by Lemma 6.3.5,

$$\nu(Z_x) = \nu(G_x) - \nu((G \setminus Z)_x) = \nu(G_x).$$

for μ-almost every $x \in X$.

Let $f = \chi_Z$ and notice that $f_x : Y \to [0, \infty]$ is defined by

$$f_x(y) = f(x, y) = \chi_Z(x, y) = \chi_{Z_x}(y).$$

Therefore, since Z_x is ν-measurable for μ-almost every $x \in X$, f_x is ν-measurable for μ-almost every $x \in X$. Moreover

$$\int_Y f_x \, d\nu = \int_Y \chi_{Z_x} \, d\nu = \nu(Z_x) = \nu(G_x)$$

for μ-almost every $x \in X$. However, by Lemma 6.3.3,

$$\int_X \nu(G_x) \, d\mu(x) = (\mu \times \nu)(G) < \infty$$

so that $\nu(G_x) < \infty$ for μ-almost every $x \in X$. Hence $\int_Y f_x \, d\nu < \infty$ for μ-almost every $x \in X$ so $f_x \in L_1(Y, \nu)$ for μ-almost every $x \in X$ as desired.

Next recall that $\Phi : X \to [0, \infty]$ is defined by

$$\Phi(x) = \int_Y f_x \, d\nu = \nu(Z_x) = \nu(G_x)$$

for all $x \in X$. Therefore, by Lemma 6.3.3, Φ is μ-measurable and

$$\int_X \Phi \, d\mu = \int_X \nu(G_x) \, d\mu(x) = (\mu \times \nu)(G) < \infty.$$

Hence $\Phi \in L_1(X, \mu)$ as desired.

Finally, by Lemma 6.3.3,

$$\int_X \Phi \, d\mu = (\mu \times \nu)(G) = (\mu \times \nu)(Z) = \int_{X \times Y} \chi_Z \, d(\mu \times \nu)$$

as desired. The remainder of the proof of Fubini's Theorem (Theorem 6.2.1) in this case holds by symmetry (i.e. repeat Lemmata 6.3.2, 6.3.3, and 6.3.5 with y in place of x, and μ and ν interchanged). ■

Finally, we can complete the proof of Fubini's Theorem (Theorem 6.2.1).

Proof of Fubini's Theorem (Theorem 6.2.1). To begin, note Lemma 6.3.6 implies Fubini's Theorem holds for characteristic functions of finite $(\mu \times \nu)$-measure. Therefore, since simple functions are linear combinations of simple functions, it is elementary to see that Fubini's Theorem holds for simple functions which vanish off a set of finite $(\mu \times \nu)$-measure.

Next recall that every element of $L_1(\mu \times \nu)$ is a linear combination of four non-negative elements of $L_1(\mu \times \nu)$. 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 2.2.4 implies there exists a sequence $(\varphi_n)_{n\geq 1}$ of simple functions on $(X \times Y, \mu \times \nu)$ 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 3.2.1) implies that

$$\lim_{n\to\infty} \int_{X\times Y} \varphi_n \, d(\mu \times \nu) = \int_{X\times Y} f \, d(\mu \times \nu) < \infty.$$

Moreover, since $0 \leq \varphi_n \leq f$, we see that $\varphi_n \in L_1(X \times Y, \mu \times \nu)$. Therefore, the proof of Theorem 3.7.24 implies that each φ_n vanishes off a set of finite $(\mu \times \nu)$-measure and thus Fubini's Theorem holds for each φ_n.

To see that f_x is ν-measurable for μ-almost every $x \in X$, 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 ν is complete and since $y \mapsto (\varphi_n)_x(y)$ is ν-measurable for μ-almost every $x \in X$, we obtain by Proposition 2.1.22 that f_x is ν-measurable for μ-almost every $x \in X$. 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 3.2.1) implies that

$$\Phi(x) = \int_Y f_x \, d\nu = \lim_{n\to\infty} \int_Y (\varphi_n)_x \, d\nu$$

for μ-almost every $x \in X$. Hence, since $\theta_n : X \to [0, \infty]$ defined by

$$\theta_n(x) = \int_Y (\varphi_n)_x \, d\mu$$

is μ-measurable for every n, Proposition 2.1.22 implies that Φ is μ-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 μ-almost every $x \in X$, we again obtain that

$$\begin{aligned}\int_X \Phi \, d\mu &= \lim_{n\to\infty} \int_X \theta_n \, d\mu \quad \text{by the Monotone Convergence Theorem} \\ &= \lim_{n\to\infty} \int_X \left(\int_Y (\varphi_n)_x(y) \, d\nu(y) \right) d\mu(x) \\ &= \lim_{n\to\infty} \int_{X\times Y} \varphi_n \, d(\mu \times \nu) \quad \text{as Fubini's Theorem holds for } \varphi_n \\ &= \int_{X\times Y} f \, d(\mu\times\nu) \quad \text{by the Monotone Convergence Theorem.}\end{aligned}$$

Therefore, since $\int_{X\times Y} f \, d(\mu \times \nu) < \infty$, we see that $\Phi \in L_1(X, \mu)$. Since $\Phi \in L_1(X, \mu)$ implies that $\Phi(x) < \infty$ for μ-almost every $x \in X$, we see that $\int_Y f_x \, d\nu < \infty$ for μ-almost every $x \in X$. Hence $f_x \in L_1(Y, \nu)$ for μ-almost every $x \in X$ as desired.

The proof is then completed by interchanging x and y and interchanging μ and ν to obtain the results for f_y and Ψ. ■

Proof of Tonelli's Theorem (Theorem 6.2.2). To begin, note that all simple functions which vanish off a set of finite $(\mu \times \nu)$-measure are non-negative elements of $L_1(\mu \times \nu)$ so Fubini's Theorem (Theorem 6.2.1) holds for them. Hence it is elementary to see that Tonelli's Theorem holds for simple functions which vanish off a set of finite $(\mu \times \nu)$-measure.

Let f be as in Tonelli's Theorem. Since f is non-negative, Theorem 2.2.4 implies there exists a sequence $(\varphi_n)_{n\geq 1}$ of simple functions on $(X \times Y, \mu \times \nu)$ such that $\varphi_n \leq \varphi_{n+1}$ for all $n \in \mathbb{N}$ and $(\varphi_n)_{n\geq1}$ converges to f pointwise.

Notice that if each φ_n vanishes off a set of finite $(\mu \times \nu)$-measure, then the proof of Fubini's Theorem (Theorem 6.2.1) carries forward verbatim to complete the proof. Hence it suffices to show we can take each φ_n to vanish off a set of finite $(\mu \times \nu)$-measure.

Since μ and ν are σ-finite, $\mu \times \nu$ is σ-finite by Proposition 6.1.9. Hence Remark 1.1.21 implies there exists $\{Z_n\}_{n=1}^\infty \subseteq \mathcal{M}(\mathcal{A} \times \mathcal{B})$ such that $X \times Y = \bigcup_{n=1}^\infty Z_n$ and $Z_n \subseteq Z_{n+1}$ for all $n \in \mathbb{N}$.

For each $n \in \mathbb{N}$ let $\psi_n = \varphi_n \chi_{Z_n}$. Then $(\psi_n)_{n\geq1}$ is a sequence of simple functions on $X \times Y$ each of which vanishes off a set of finite $(\mu \times \nu)$-measure such that, by construction, $\psi_n \leq \psi_{n+1}$ for all $n \in \mathbb{N}$ and $(\psi_n)_{n\geq1}$ converges to f pointwise. Hence the proof of Fubini's Theorem (Theorem 6.2.1) carries forward verbatim to complete the proof. ■

Chapter 7

Riesz Representation Theorems

In this final chapter, analyze various the Banach space structures of the objects we have seen in this course. In particular, we desire to better understand all L_p-space even though we have mainly studied $L_1(X, \mu)$. One reason why one might be interested in understanding L_p-spaces is that the p-norms are very natural norms and occur with regularity in analysis.

In order to better understand L_p-spaces and the p-norms, we turn to some ideas from functional analysis. By understanding the continuous linear functionals on a Banach space, one can understand many structural properties and behaviours of the Banach space. Thus in this chapter we will describe the collection of bounded linear functionals on L_p-spaces. This will lead to an alternate yet highly useful description of the p-norms.

7.1 Dual Spaces

In this section, we will recall some necessary facts pertaining to continuous linear functionals on Banach spaces. For some detailed proofs, we refer the reader to Appendix C. Recall that $\mathbb{K}$ denotes either $\mathbb{R}$ or $\mathbb{C}$.

Definition 7.1.1. Let $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ be a normed linear space over $\mathbb{K}$. A *linear functional* on $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ is a linear map $T : \mathcal{X} \to \mathbb{K}$.

Remark 7.1.2. It is not difficult to see that a linear functional T on a normed linear space $(\mathcal{X}, \|\cdot\|)$ is continuous if and only if there exists an $M > 0$ such that $|T(\vec{x})| \leq M$ whenever $\vec{x} \in \mathcal{X}$ and $\|\vec{x}\|_{\mathcal{X}} \leq 1$ (see Theorem C.3.24). In particular, if T is a continuous linear functional and we define

$$\|T\| = \sup\{|T(\vec{x})| \mid \vec{x} \in \mathcal{X}, \|\vec{x}\|_{\mathcal{X}} \leq 1\}$$

then $\|T\| < \infty$ and

$$|T(\vec{x})| \leq \|T\| \, \|\vec{x}\|_{\mathcal{X}}$$

for all $\vec{x} \in \mathcal{X}$.

Definition 7.1.3. Let $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ be a normed linear space over $\mathbb{K}$. The *dual space of* $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$, denoted $\mathcal{X}^*$, is the set of all continuous linear functional on $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$.

Theorem 7.1.4. *If $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ is a normed linear space over $\mathbb{K}$, then the dual space $\mathcal{X}^*$ is a Banach space with respect to the norm $\|\cdot\| : \mathcal{X}^* \to [0,\infty)$ defined by*

$$\|T\| = \sup\{|T(\vec{x})| \mid \vec{x} \in \mathcal{X}, \|\vec{x}\|_{\mathcal{X}} \leq 1\}$$

for all $T \in \mathcal{X}^*$.

Proof. See Theorem C.5.8. ∎

Example 7.1.5. Let $(\mathcal{X}, d)$ be a compact metric space, let $\mathcal{C}(\mathcal{X}, \mathbb{R})$ denote continuous, real-valued functions on $\mathcal{X}$, and let ν be a finite signed measure on the Borel subsets of $\mathcal{X}$. Recall that $\mathcal{C}(\mathcal{X}, \mathbb{R})$ is a normed linear space with respect to the norm $\|\cdot\|_\infty : \mathcal{C}(\mathcal{X}, \mathbb{R}) \to [0,\infty)$ defined by

$$\|f\|_\infty = \sup\{|f(x)| \mid x \in \mathcal{X}\}$$

for all $f \in \mathcal{C}(\mathcal{X}, \mathbb{R})$.

Define $T : \mathcal{C}(\mathcal{X}, \mathbb{R}) \to \mathbb{R}$ by

$$T(f) = \int_{\mathcal{X}} f \, d\nu = \int_{\mathcal{X}} f \, d\nu_+ - \int_{\mathcal{X}} f \, d\nu_-$$

for all $f \in \mathcal{C}(\mathcal{X}, \mathbb{R})$. To see that T is well-defined recall that ν is finite so ν_+ and ν_- are finite Borel measures on $(\mathcal{X}, d)$ by Lemma 5.4.6. Therefore, since every continuous function on a compact metric space is bounded by the Extreme Value Theorem and since bounded functions are integrable with respect to any finite measure, T is well-defined. Furthermore, it is clear that T is a linear functional.

We claim that T is continuous. To see this, notice for all $f \in \mathcal{C}(\mathcal{X}, \mathbb{R})$ that

$$\begin{aligned} |T(f)| &= \left| \int_{\mathcal{X}} f \, d\nu_+ - \int_{\mathcal{X}} f \, d\nu_- \right| \\ &\leq \int_{\mathcal{X}} |f| \, d\nu_+ + \int_{\mathcal{X}} |f| \, d\nu_- \\ &\leq \|f\|_\infty \, \nu_+(\mathcal{X}) + \|f\|_\infty \, \nu_-(\mathcal{X}) \\ &= \|f\|_\infty \, \|\nu\| . \end{aligned}$$

Therefore T is continuous with $\|T\| \leq \|\nu\|$.

Example 7.1.6. Let $(X, \mathcal{A}, \mu)$ be a σ-finite measure space, let $p, q \in [1, \infty]$ be such that $\frac{1}{p} + \frac{1}{q} = 1$, and let $g \in L_q(X, \mu)$. Define $\Psi_g : L_p(X, \mu) \to \mathbb{C}$ by

$$\Psi_g(f) = \int_X fg \, d\mu$$

for all $f \in L_p(X, \mu)$. To see that T is well-defined, notice since $g \in L_q(X, \mu)$ and $\frac{1}{p} + \frac{1}{q} = 1$ that $fg \in L_1(X, \mu)$ for all $f \in L_p(X, \mu)$ by Hölder's Inequality (Theorems 3.7.8 and 3.7.22). Hence Ψ_g is well-defined. Furthermore, clearly Ψ_g is linear.

To see that Ψ_g is continuous, notice for all $f \in L_p(X, \mu)$ that

$$|\Psi_g(f)| = \left| \int_X fg \, d\mu \right| \leq \int_X |fg| \, d\mu \leq \|f\|_p \, \|g\|_q$$

by Hölder's Inequality. Hence $\Psi_g \in L_p(X, \mu)^*$ with $\|\Psi_g\| \leq \|g\|_q$.

We claim that $\|\Psi_g\| = \|g\|_q$. To see this, we divide the discussion into three cases.

Case 1: $q = 1$. In this case $p = \infty$. Consider $f : X \to \mathbb{C}$ defined by

$$f(x) = \mathrm{sgn}(g)(x) = \begin{cases} \frac{|g(x)|}{g(x)} & \text{if } g(x) \neq 0 \\ 1 & \text{if } g(x) = 0 \end{cases}$$

for all $x \in X$. It is not difficult to see that f is measurable with $|f(x)| = 1$ for all $x \in X$ and thus $f \in L_\infty(X, \mu)$ with $\|f\|_\infty = 1$. Therefore, since

$$\Psi_g(f) = \int_X fg \, d\mu = \int_X |g| \, d\mu = \|g\|_1,$$

we see that $\|\Psi_g\| \geq \|g\|_1$ and thus $\|\Psi_g\| = \|g\|_1$ as desired.

Case 2: $1 < q < \infty$. In this case $1 < p < \infty$. Let $f = \mathrm{sgn}(g)|g|^{\frac{q}{p}}$. Clearly f is a well-defined measurable function since $1 < p, q < \infty$. We claim that $f \in L_p(X, \mu)$. To see this, notice

$$\left(\int_X |f|^p \, d\mu \right)^{\frac{1}{p}} = \left(\int_X |g|^q \, d\mu \right)^{\frac{1}{p}} = \|g\|_q^{\frac{q}{p}} < \infty$$

since $|\mathrm{sgn}(g)| = 1$ and $g \in L_q(X, \mu)$. Hence $f \in L_p(X, \mu)$ with $\|f\|_p = \|g\|_q^{\frac{q}{p}}$. Therefore, since

$$\frac{1}{p} + \frac{1}{q} = 1 \quad \Longrightarrow \quad \frac{q}{p} + 1 = q$$

we see that

$$\begin{aligned}\Psi_g(f) &= \int_X fg\,d\mu \\ &= \int_X |g|^{\frac{q}{p}+1}\,d\mu \\ &= \int_X |g|^q\,d\mu \\ &= \|g\|_q^q \\ &= \|g\|_q \|g\|_q^{\frac{q}{p}} \\ &= \|g\|_q \|f\|_p .\end{aligned}$$

If $f = 0$ then clearly $g = 0$ and the result follows. Otherwise if $h = \frac{1}{\|f\|_p} f$ then $h \in L_p(X, \mu)$, $\|h\|_p = 1$, and the above computation implies that

$$\Psi_g(h) = \|g\|_q .$$

Therefore $\|\Psi_g\| \geq \|g\|_q$ and thus $\|\Psi_g\| = \|g\|_q$ as desired.

Case 3: $q = \infty$. In this case $p = 1$. Notice the previous cases did not require μ to be σ-finite whereas we will need to use σ-finiteness here. To begin, since μ is σ-finite there exists a collection $\{X_n\}_{n=1}^\infty \subseteq \mathcal{A}$ such that $\mu(X_n) < \infty$ for all $n \in \mathbb{N}$, $X = \bigcup_{n=1}^\infty X_n$, and $X_n \subseteq X_{n+1}$ for all $n \in \mathbb{N}$ by Remark 1.1.21.

Let $\epsilon > 0$ be arbitrary and let

$$A_\epsilon = \{x \in X \mid |g(x)| > \|g\|_\infty - \epsilon\}.$$

Since $g \in L_\infty(X, \mu)$, we know that $\mu(A_\epsilon) > 0$. For each $n \in \mathbb{N}$ let $B_n = A_\epsilon \cap X_n$. Then clearly $A_\epsilon = \bigcup_{n=1}^\infty B_n$ and $B_n \subseteq B_{n+1}$ for all $n \in \mathbb{N}$. Therefore $\mu(A_\epsilon) = \lim_{n\to\infty} \mu(B_n)$ by the Monotone Convergence Theorem for Measures (Theorem 1.1.23). Moreover, since $0 \leq \mu(B_n) \leq \mu(X_n) < \infty$ for all $n \in \mathbb{N}$, there exists an $N \in \mathbb{N}$ such that

$$0 < \mu(B_N) < \infty.$$

Let $f = \frac{1}{\mu(B_N)} \chi_{B_N} \mathrm{sgn}(g)$. Then f is clearly measurable with

$$\int_X |f|\,d\mu = \frac{1}{\mu(B_N)} \int_X \chi_{B_N}\,d\mu = 1.$$

Therefore, since

$$\begin{aligned}\Psi_g(f) &= \int_X fg\,d\mu \\ &= \frac{1}{\mu(B_N)} \int_X \chi_{B_N} |g|\,d\mu \\ &\geq \frac{1}{\mu(B_N)} \int_X \chi_{B_N} (\|g\|_\infty - \epsilon)\,d\mu \\ &= \|g\|_\infty - \epsilon\end{aligned}$$

as $B_N \subseteq A_\epsilon$, we obtain that $\|\Psi_g\| \geq \|g\|_\infty - \epsilon$. Hence, as $\epsilon > 0$ was arbitrary, the result follows.

Remark 7.1.7. Notice as a direct corollary Example 7.1.6 that if $(X, \mathcal{A}, \mu)$ is a σ-finite measure space and $p, q \in [1, \infty]$ are such that $\frac{1}{p} + \frac{1}{q} = 1$, then

$$\|g\|_q = \|\Psi_g\| = \sup\left\{ \left| \int_X fg\, d\mu \right| \,\middle|\, f \in L_p(X, \mu), \|f\|_p \leq 1 \right\}$$

for all $g \in L_q(X, \mu)$. This alternative way to compute the norm can be useful on occasion.

7.2 The L_p-Riesz Representation Theorem

In the previous section, we saw various continuous linear functionals on both $\mathcal{C}(\mathcal{X}, \mathbb{R})$ and $L_p(X, \mu)$. In this section, we will completely characterize the continuous linear functionals on $L_p(X, \mu)$ (for $p \neq \infty$) and thereby develop a method for verifying a function is in $L_q(X, \mu)$. In particular, our main goal is to prove the following which shows that $L_p(X, \mu)$ and $L_q(X, \mu)$ are 'dual' to each other and serves as demonstrating that the continuous linear functionals on $L_p(X, \mu)$ from Example 7.1.6 are all the continuous linear functionals there are.

Theorem 7.2.1 (Riesz Representation Theorem for L_p-Spaces). *Let $(X, \mathcal{A}, \mu)$ be a σ-finite measure space, let $1 \leq p < \infty$, and let $1 < q \leq \infty$ be such that $\frac{1}{p} + \frac{1}{q} = 1$. If $\Psi \in L_p(X, \mu)^*$ then there exists a unique $g \in L_q(X, \mu)$ such that*

$$\Psi(f) = \int_X fg\, d\mu$$

for all $f \in L_p(X, \mu)$. Moreover $\|\Psi\| = \|g\|_q$. In particular, $L_p(X, \mu)^* = L_q(X, \mu)$.

First note that the norm estimates in the Riesz Representation Theorem for L_p-spaces (Theorem 7.2.1) immediately follow for Example 7.1.6. Thus it suffices to prove given a continuous linear functional on $L_p(X, \mu)$ that there is one and exactly one element of $L_q(X, \mu)$ that, via Example 7.1.6, produces the continuous linear functional.

To begin, we desire to reduce to the setting that our functions are real-valued. Thus, let $L_p(X, \mu)_\mathbb{R}$ denote the real-valued p-integrable functions and consider the following.

Lemma 7.2.2. *Let $(X, \mathcal{A}, \mu)$ be a σ-finite measure space, let $1 \leq p < \infty$, and let $\Psi \in L_p(X, \mu)^*$. Then there exists continuous functions*

$$\psi_1, \psi_2 : L_p(X, \mu)_\mathbb{R} \to \mathbb{R}$$

such that ψ_1 and ψ_2 are (real-)linear and

$$\Psi(f) = \psi_1(\mathrm{Re}(f)) + i\psi_1(\mathrm{Im}(f)) + i\psi_2(\mathrm{Re}(f)) - \psi_2(\mathrm{Im}(f))$$

for all $f \in L_p(X, \mu)$.

Proof. Given a function $f \in L_p(X, \mu)$, recall the complex conjugate of f, denoted $\overline{f}$, is an element of $L_p(X, \mu)$. Define $\psi_1, \psi_2 : L_p(X, \mu)_{\mathbb{R}} \to \mathbb{R}$ by

$$\psi_1(f) = \mathrm{Re}(\Psi(f)) \qquad \text{and} \qquad \psi_2(f) = \mathrm{Im}(\Psi(f))$$

for all $f \in L_p(X, \mu)_{\mathbb{R}}$. Since Ψ is complex linear and continuous, it is elementary to see that ψ_1 and ψ_2 are real linear and continuous. Moreover, the equation

$$\Psi(f) = \psi_1(\mathrm{Re}(f)) + i\psi_1(\mathrm{Im}(f)) + i\psi_2(\mathrm{Re}(f)) - \psi_2(\mathrm{Im}(f))$$

for all $f \in L_p(X, \mu)$ is then trivial to verify. ∎

Next we require a method for verifying that a function is in $L_q(X, \mu)$ based on knowledge of its integral against elements of $L_p(X, \mu)$. This is achieved via the following two lemma (one for $p \in (1, \infty)$ and one for $p = 1$). Note this has significance outside the proof of the Riesz Representation Theorem (Theorem 7.2.1) as it enables us to deduce a function is in $L_q(X, \mu)$ and obtain a bound on its norm based on integration.

Lemma 7.2.3. *Let $(X, \mathcal{A}, \mu)$ be a finite measure space, let $1 < p < \infty$, and $1 < q < \infty$ be such that $\frac{1}{p} + \frac{1}{q} = 1$, and let $g \in L_1(X, \mu)_{\mathbb{R}}$. If there exists an $M \in \mathbb{R}$ such that*

$$\left| \int g\varphi \, d\mu \right| \leq M \left\| \varphi \right\|_p$$

for all measurable functions $\varphi : X \to \mathbb{R}$ of finite range, then $g \in L_q(X, \mu)$ with $\left\| g \right\|_q \leq M$.

Proof. Since $|g|^q$ is a measurable function, Theorem 2.2.4 implies there exists an increasing sequence $(\varphi_n)_{n \geq 1}$ of simple functions that converges to $|g|^q$ pointwise. For each $n \in \mathbb{N}$ let

$$\psi_n = \varphi_n^{\frac{1}{p}} \mathrm{sgn}(g).$$

Since g is real-valued so $\mathrm{sgn}(g)$ obtains a finite number of values, it is elementary to see that each ψ_n is a measurable function of finite range. Moreover, notice for all $n \in \mathbb{N}$ that

$$\left\| \psi_n \right\|_p = \left(\int_X |\psi_n|^p \, d\mu \right)^{\frac{1}{p}} = \left(\int_X \varphi_n \, d\mu \right)^{\frac{1}{p}}$$

and
$$0 \leq \varphi_n = \varphi_n^{\frac{1}{p}} \varphi_n^{\frac{1}{q}} \leq \varphi_n^{\frac{1}{p}} |g| = \varphi_n^{\frac{1}{p}} \operatorname{sgn}(g) g = \psi_n g.$$
Therefore, for all $n \in \mathbb{N}$
$$0 \leq \int_X \varphi_n \, d\mu \leq \int_X g\psi_n \, d\mu \leq M \left\|\psi_n\right\|_p = M \left(\int_X \varphi_n \, d\mu \right)^{\frac{1}{p}}.$$
Since μ is finite so all simple functions are integrable, we know that
$$\int_X \varphi_n \, d\mu < \infty$$
for all $n \in \mathbb{N}$. Hence the above equation implies that
$$\left(\int_X \varphi_n \, d\mu \right)^{\frac{1}{q}} = \left(\int_X \varphi_n \, d\mu \right)^{1-\frac{1}{p}} \leq M.$$
However, by the Monotone Convergence Theorem
$$\left(\int_X |g|^q \, d\mu \right)^{\frac{1}{q}} = \left(\lim_{n\to\infty} \int_X \varphi_n \, d\mu \right)^{\frac{1}{q}} \leq M$$
and thus $\left\|g\right\|_q \leq M$. Hence $g \in L_q(X, \mu)$ as desired. ∎

Lemma 7.2.4. *Let $(X, \mathcal{A}, \mu)$ be a finite measure space, and let $g \in L_1(X, \mu)_{\mathbb{R}}$. If there exists an $M \in \mathbb{R}$ such that*
$$\left| \int g\varphi \, d\mu \right| \leq M \left\|\varphi\right\|_1$$
for all measurable functions $\varphi : X \to \mathbb{R}$ of finite range, then $g \in L_\infty(X, \mu)$ with $\left\|g\right\|_\infty \leq M$.

Proof. Let $\epsilon > 0$ be arbitrary. Consider the set
$$A_\epsilon = \{x \in X \mid |g(x)| \geq M + \epsilon\}.$$
Clearly A_ϵ is measurable. Hence
$$\begin{aligned} (M+\epsilon)\mu(A_\epsilon) &\leq \int_{A_\epsilon} |g| \, d\mu \\ &= \int_X \operatorname{sgn}(g)\chi_{A_\epsilon} g \, d\mu \\ &\leq M \left\|\operatorname{sgn}(g)\chi_{A_\epsilon}\right\|_1 \\ &= M\mu(A_\epsilon) \end{aligned}$$
since $\operatorname{sgn}(g)\chi_{A_\epsilon}$ is a measurable function of finite range (as g is real-valued). Therefore $\epsilon\mu(A_\epsilon) \leq 0$ so $\mu(A_\epsilon) = 0$. Therefore, since $\epsilon > 0$ was arbitrary, we obtain that $g \in L_\infty(X, \mu)$ with $\left\|g\right\|_\infty \leq M$. ∎

Proof of the Riesz Representation Theorem for L_p-Spaces (Theorem 7.2.1). Recall from Example 7.1.6 that if $\Psi_g : L_p(X, \mu) \to \mathbb{C}$ is defined by

$$\Psi_g(f) = \int_X fg \, d\mu$$

for all $f \in L_p(X, \mu)$, then $\Psi_g \in L_p(X, \mu)^*$ and $\|\Psi_g\| = \|g\|_q$. Furthermore, notice if $g_1, g_2 \in L_q(X, \mu)$ are such that $\Psi_{g_1} = \Psi_{g_2}$, then

$$0 = \Psi_{g_1}(f) - \Psi_{g_2}(f) = \int_X fg_1 \, d\mu - \int_X fg_2 \, d\mu = \int_X f(g_1 - g_2) \, d\mu = \Psi_{g_1 - g_2}(f)$$

for all $f \in L_p(X, \mu)$. Therefore $0 = \|\Psi_{g_1 - g_2}\| = \|g_1 - g_2\|_q$ so $g_1 = g_2$. Hence, to complete the proof, it suffices to show that if $\Psi \in L_p(X, \mu)^*$ then there exists a $g \in L_q(X, \mu)$ such that $\Psi = \Psi_g$ (as the above produces the value of the norm and uniqueness).

Fix $\Psi \in L_p(X, \mu)^*$. Recall by Lemma 7.2.2 that there exists continuous real-linear functions $\psi_1, \psi_2 : L_p(X, \mu)_{\mathbb{R}} \to \mathbb{R}$ such that

$$\Psi(f) = \psi_1(\mathrm{Re}(f)) + i\psi_1(\mathrm{Im}(f)) + i\psi_2(\mathrm{Re}(f)) - \psi_2(\mathrm{Im}(f))$$

for all $f \in L_p(X, \mu)$. If we demonstrate that there exists $g_1, g_2 \in L_q(X, \mu)_{\mathbb{R}}$ such that

$$\psi_1(h) = \int_X hg_1 \, d\mu \qquad \text{and} \qquad \psi_2(h) = \int_X hg_2 \, d\mu$$

for all $h \in L_p(X, \mu)_{\mathbb{R}}$, then we obtain (using complex linearity) that

$$\Psi(f) = \int_X f(g_1 + ig_2) \, d\mu$$

for all $f \in L_p(X, \mu)$, which would complete the proof as $g_1 + ig_2 \in L_q(X, \mu)$. Therefore, it suffices to show that if $\psi : L_p(X, \mu)_{\mathbb{R}} \to \mathbb{R}$ is continuous and real-linear then there exists a $g \in L_q(X, \mu)_{\mathbb{R}}$ such that

$$\psi(f) = \int_X fg \, d\mu$$

for all $f \in L_p(X, \mu)$.

To see the above claim, we will divide the proof into two cases.

Case 1: μ is finite. Since μ is finite, $\chi_A \in L_p(X, \mu)$ for all $A \in \mathcal{A}$. Hence define $\nu : \mathcal{A} \to \mathbb{R}$ by

$$\nu(A) = \psi(\chi_A)$$

for all $A \in \mathcal{A}$. We claim that ν is a finite signed measure that is absolutely continuous with respect to μ. To see this, first notice that

$$\nu(\emptyset) = \psi(\chi_\emptyset) = \psi(0) = 0$$

as ψ is linear. Moreover, clearly ν does not obtain the values $\pm\infty$ by definition.

To see that ν is countably additive, let $\{A_k\}_{k=1}^\infty \subseteq \mathcal{A}$ be pairwise disjoint and let $A = \bigcup_{k=1}^\infty A_k$. Since μ is a finite measure,

$$\mu(A) = \sum_{k=1}^\infty \mu(A_k) < \infty.$$

Hence

$$\lim_{n\to\infty} \sum_{k=n}^\infty \mu(A_k) = 0.$$

Therefore

$$\lim_{n\to\infty} \left\| \chi_A - \sum_{k=1}^n \chi_{A_k} \right\|_p = \lim_{n\to\infty} \left(\sum_{k=n}^\infty \mu(A_k) \right)^{\frac{1}{p}} = 0.$$

Hence $\chi_A = \sum_{k=1}^\infty \chi_{A_k}$ as a sum of vectors in $L_p(X, \mu)$. Therefore, since ψ is a continuous linear functional, we obtain that

$$\nu(A) = \psi(\chi_A) = \sum_{k=1}^\infty \psi(\chi_{A_k}) = \sum_{k=1}^\infty \nu(A_k).$$

Thus ν is countably additive. However, to show that ν is a signed measure, it is necessary to show that the sum converges absolutely.

For each $n \in \mathbb{N}$ let $c_n = \mathrm{sgn}(\nu(A_n))$ and let $f_n = \sum_{k=1}^n c_k \chi_{A_k}$. Then for all $n, m \in \mathbb{N}$ with $n \geq m$

$$\begin{aligned} \|f_n - f_m\|_p &= \left\| \sum_{k=m+1}^n c_k \chi_{A_k} \right\|_p \\ &= \left(\sum_{k=m+1}^n \mu(A_k) \right)^{\frac{1}{p}} \\ &\leq \left(\sum_{k=m}^\infty \mu(A_k) \right)^{\frac{1}{p}}. \end{aligned}$$

Therefore, since $\lim_{m\to\infty} \left(\sum_{k=m}^\infty \mu(A_k)\right)^{\frac{1}{p}} = 0$, $(f_n)_{n\geq 1}$ is Cauchy in $L_p(X, \mu)$. Since $L_p(X, \mu)$ is complete by the Riesz-Fisher Theorem (Theorems 3.7.12 and 3.7.21), there exists an $f \in L_p(X, \mu)$ such that $f = \lim_{n\to\infty} f_n$ in $L_p(X, \mu)$. Therefore, since ψ is a continuous linear functional, we obtain that

$$\psi(f) = \lim_{n\to\infty} \psi(f_n) = \lim_{n\to\infty} \sum_{k=1}^n |\nu(A_n)|.$$

Therefore, since $\psi(f) \in \mathbb{R}$, we see that the sum converges absolutely.

To see that ν is finite, notice since μ is finite that for all $A \in \mathcal{A}$

$$|\nu(A)| = |\psi(\chi_A)| \leq \|\psi\| \, \|\chi_A\|_p = \|\psi\| \, \mu(A)^{\frac{1}{p}} < \infty.$$

Hence ν is finite. Finally, to see that ν is absolutely continuous with respect to μ, notice if $A \in \mathcal{A}$ is such that $\mu(A) = 0$, then $\chi_A = 0$ as an element of $L_p(X, \mu)$ and thus

$$\nu(A) = \psi(\chi_A) = \psi(0) = 0$$

as ψ is linear. Hence ν is a finite signed measure that is absolutely continuous with respect to μ.

By the Radon-Nikodym Theorem for signed measures (Corollary 5.5.8) there exists a real-valued function $g \in L_1(X, \mu)$ such that

$$\psi(\chi_A) = \nu(A) = \int_A g \, d\mu = \int_X g\chi_A \, d\mu$$

for all $A \in \mathcal{A}$. Using the linearity of the integral and of ψ, we obtain for any measurable function $\varphi : X \to \mathbb{R}$ with finite range that

$$\psi(\varphi) = \int_X \varphi g \, d\mu.$$

However, this implies that

$$\left| \int_X \varphi g \, d\mu \right| = |\psi(\varphi)| \leq \|\psi\| \, \|\varphi\|_p$$

for all measurable functions $\varphi : X \to \mathbb{R}$ with finite range. Hence Lemma 7.2.3 or Lemma 7.2.4 implies that $g \in L_q(X, \mu)_{\mathbb{R}}$.

Since

$$\psi(\varphi) = \int_X \varphi g \, d\mu$$

for all simple functions in $L_p(X, \mu)$, we obtain by linearity that

$$\psi(\varphi) = \int_X \varphi g \, d\mu$$

for all φ which are linear combinations of simple functions in $L_p(X, \mu)$. Therefore Theorem 3.7.24 (along with continuity) implies that

$$\psi(f) = \int_X fg \, d\mu$$

for all $f \in L_p(X, \mu)$ as desired.

Case 2: μ is σ-finite. By Remark 1.1.21 there exists $\{X_n\}_{n=1}^\infty \subseteq \mathcal{A}$ such that $X = \bigcup_{n=1}^\infty X_n$, $\mu(X_n) < \infty$ for all $n \in \mathbb{N}$, and $X_n \subseteq X_{n+1}$ for all $n \in \mathbb{N}$.

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

$$\mathcal{A}_n = \{A \cap X_n \mid A \in \mathcal{A}\}$$

and let $\mu_n = \mu|_{\mathcal{A}_n}$. It is elementary to verify that $\mathcal{A}_n$ is a σ-algebra on X_n and that μ_n is a measure on $(X_n, \mathcal{A}_n)$. Notice if $f \in L_p(X_n, \mu_n)$, we can view f as an element of $L_p(X, \mu)$ by extending f to be zero on X_n^c. Hence, for each $n \in \mathbb{N}$, we can define $\psi_n : L_p(X_n, \mu_n) \to \mathbb{R}$ by

$$\psi_n(f) = \psi(f)$$

for all $f \in L_p(X_n, \mu_n) \subseteq L_p(X, \mu)$. It is elementary to verify that ψ_n is a continuous linear functional on $L_p(X_n, \mu_n)$ with norm at most $\|\psi\|$ as the norms on $L_p(X_n, \mu_n)$ and $L_p(X, \mu)$ agree on elements of $L_p(X_n, \mu_n)$.

Since $(X_n, \mathcal{A}_n, \mu_n)$ is a finite measure space, the first case of this proof implies there exists a unique function $g_n \in L_q(X_n, \mu_n)$ such that

$$\int_{X_n} f g_n \, d\mu_n = \psi_n(f) = \psi(f)$$

for all $f \in L_p(X_n, \mu_n)$. Moreover $\|g_n\|_{L_q(X_n, \mu_n)} = \|\psi_n\| \leq \|\psi\|$.

Extend each g_n to be zero on X_n^c. Hence $g_n \in L_q(X, \mu)$ for all $n \in \mathbb{N}$, $\|g_n\|_{L_q(X_n, \mu_n)} = \|g_n\|_q$, and

$$\psi_n(f) = \int_X f g_n \, d\mu$$

for all $f \in L_p(X_n, \mu_n)$. Moreover, notice for all $n \in \mathbb{N}$ and $f \in L_p(X_n, \mu_n) \subseteq L_p(X_{n+1}, \mu_{n+1})$ that

$$\int_X f g_{n+1} \, d\mu = \psi_{n+1}(f) = \psi_n(f) = \int_X f g_n \, d\mu.$$

Therefore, due to the uniqueness of g_n, we obtain that $g_{n+1}|_{X_n} = g_n$.

Define $g : X \to \mathbb{R}$ by $g(x) = g_n(x)$ whenever $x \in X_n$. Since $g_{n+1}|_{X_n} = g_n$ and since $X = \bigcup_{n=1}^\infty X_n$, g is well-defined up to a set of measure zero and defines a measurable function (as it is the pointwise limit of $(g_n)_{n\geq 1}$). If $q = \infty$ then, since $\|g_n\|_\infty \leq \|\psi\|$ for all $n \in \mathbb{N}$, we easily see that $\|g\|_\infty \leq \|\psi\| < \infty$ and thus $g \in L_\infty(X, \mu)$. Otherwise, if $q \neq \infty$, notice that as $|g_n| \leq |g_{n+1}|$ for all $n \in \mathbb{N}$ and as $(g_n)_{n\geq 1}$ converges to g pointwise almost everywhere, the Monotone Convergence Theorem (Theorem 3.2.1) implies that

$$\left(\int_X |g|^q \, d\mu\right)^{\frac{1}{q}} = \lim_{n\to\infty} \left(\int_X |g_n|^q \, d\mu\right)^{\frac{1}{q}} \leq \|\psi\| < \infty.$$

Hence $g \in L_q(X, \mu)$.

Finally, to see that

$$\psi(f) = \int_X f g \, d\mu$$

for all $f \in L_p(X,\mu)$, let $f \in L_p(X,\mu)$ be arbitrary and for each $n \in \mathbb{N}$ let $f_n = f\chi_{X_n}$. Then

$$|f_n - f|^p \leq |f|^p$$

and $(|f_n - f|^p)_{n\geq 1}$ converges to zero almost everywhere. Therefore, since $|f|^p \in L_1(X,\mu)$, the Dominated Convergence Theorem (Theorem 3.6.1) implies that $\lim_{n\to\infty} \|f - f_n\|_p = 0$. Since ψ is continuous and since $f_n g_n = f_n g$ for all $n \in \mathbb{N}$, we have that

$$\psi(f) = \lim_{n\to\infty} \psi(f_n) = \lim_{n\to\infty} \psi_n(f_n) = \lim_{n\to\infty} \int_X f_n g_n \, d\mu = \lim_{n\to\infty} \int_X f_n g \, d\mu.$$

However, since $(f_n g)_{n\geq 1}$ converges pointwise to fg and since $|f_n g| \leq |fg| \in L_1(X,\mu)$ by Hölder's Inequality (Theorems 3.7.8 and 3.7.8), the Dominated Convergence Theorem (Theorem 3.6.1) implies that

$$\psi(f) = \lim_{n\to\infty} \int_X f_n g \, d\mu = \int_X fg \, d\mu$$

as desired. ■

To conclude this section, we formally show (using only the Riesz Representation Theorem (Theorem 7.2.1)) that a function can be verified to be in $L_q(X,\mu)$ via only integrals against $L_p(X,\mu)$ functions.

Corollary 7.2.5. *Let $(X,\mathcal{A},\mu)$ be a σ-finite measure space and let $p,q \in [1,\infty]$ be such that $\frac{1}{p} + \frac{1}{q} = 1$ and $q \neq 1$. If*

$$\sup\left\{ \left| \int_X fg \, d\mu \right| \,\middle|\, f \in L_p(X,\mu), \|f\|_p \leq 1 \right\} < \infty,$$

then $g \in L_q(X,\mu)$.

Proof. Define $\Psi : L_p(X,\mu) \to \mathbb{R}$ by

$$\Psi(f) = \int_X fg \, d\mu$$

for all $f \in L_p(X,\mu)$. By the assumptions in the statement, we easily see that Ψ is a well-defined continuous linear functional on $L_p(X,\mu)$. Therefore, by the Riesz Representation Theorem there exists an unique function $h \in L_q(X,\mu)$ such that

$$\Psi(f) = \int_X fh \, d\mu$$

for all $f \in L_p(X,\mu)$. In particular, for all $f \in L_p(X,\mu)$ and $A \in \mathcal{A}$ we see that

$$\int_A fg \, d\mu = \int_X (f\chi_A) g \, d\mu = \Psi(f\chi_A) = \int_X (f\chi_A) h \, d\mu = \int_A fh \, d\mu.$$

Therefore, since μ is σ-finite, by the Radon-Nikodym Theorem (Theorem 5.5.5) we obtain that $fg = fh$ for all $f \in L_p(X, \mu)$.

Since μ is σ-finite, there exists $\{X_n\}_{n=1}^\infty \subseteq \mathcal{A}$ such that $X = \bigcup_{n=1}^\infty X_n$, $\mu(X_n) < \infty$ for all $n \in \mathbb{N}$, and $\{X_n\}_{n=1}^\infty$ are pairwise disjoint. Since $\mu(X_n) < \infty$, $\chi_{X_n} \in L_p(X, \mu)$ for all $n \in \mathbb{N}$. Hence the above implies that

$$g\chi_{X_n} = h\chi_{X_n}$$

for all $n \in \mathbb{N}$. Therefore, since $X = \bigcup_{n=1}^\infty X_n$, we obtain that $g = h \in L_q(X, \mu)$ as desired. ∎

7.3 Other Riesz Representation Theorems

To conclude our course, we mention there are several other versions of the Riesz Representation Theorem we could analyze in the context of measure theory. Here are two which describe the dual spaces of two very natural collections of functions seen in this course.

Theorem 7.3.1 (Riesz Representation Theorem for L_∞). *Let $(X, \mathcal{A}, \mu)$ be a σ-finite measure space. If $\Psi \in L_\infty(X, \mu)_\mathbb{R} \to \mathbb{R}$ is a continuous linear functional, then there exists a unique 'bounded, finitely additive' signed measure ν such that ν is absolutely continuous with respect to μ and*

$$\Psi(f) = \int_X f \, d\nu$$

for all $f \in L_\infty(X, \mu)$. Moreover $\|\Psi\| = |\nu|(X)$.

Theorem 7.3.2 (Riesz-Markov Theorem). *Let X be a locally compact Hausdorff space, let $C_c(X)$ denote the continuous complex-valued functions of compact support on X, and let $\Psi : C_c(X) \to \mathbb{C}$ be such that $\Psi(f) \geq 0$ whenever $f \in C_c(X)$ and $f \geq 0$. Then there exists a unique regular measure μ on the Borel σ-algebra associated to X such that $\mu(K) < \infty$ for all compact subsets K of X and*

$$\Psi(f) = \int_X f \, d\mu$$

for all $f \in C_c(X)$.

Remark 7.3.3. It is not difficult to demonstrate that if X is a locally compact Hausdorff space then every continuous linear functional on $C_c(X)$ is a linear combination of four linear functional satisfying the hypotheses of Theorem 7.3.2.

Appendix A

Review of the Riemann Integral

In this appendix chapter, we will recall the construction and properties of the Riemann integral presented in undergraduate analysis. 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

$$\sup\{L(f, \mathcal{P}) \mid \mathcal{P} \text{ a partition of } [0, 1]\} \neq \inf\{U(f, \mathcal{P}) \mid \mathcal{P} \text{ a partition of } [0, 1]\}.$$

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

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

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

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

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

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

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

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

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

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

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

Corollary 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

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

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 1.4.6 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

$$A \preceq B \iff A \subseteq B.$$

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 $\mathrm{span}(B) \neq V$. Thus there exists a non-zero vector $\vec{v} \in V \setminus \mathrm{span}(B)$. 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 $\mathrm{span}(B) = V$ 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

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

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

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

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

We desire to invoke Zorn's Lemma (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

Banach Spaces

In this appendix, we briefly review many important definitions, concepts, and results relating to Banach spaces that are important to this course.

C.1 Metric and Normed Linear Spaces

Definition C.1.1. Let X be a non-empty set. A *metric* on X is a function $d : X \times X \to [0, \infty)$ such that

1. for $x, y \in X$, $d(x, y) = 0$ if and only if $x = y$,
2. $d(x, y) = d(y, x)$ for all $x, y \in X$, and
3. (Triangle Inequality) $d(x, y) \leq d(x, z) + d(z, y)$ for all $x, y, z \in X$.

Definition C.1.2. A *metric space* is a pair $(\mathcal{X}, d)$ where $\mathcal{X}$ is a non-empty set and d is a metric on $\mathcal{X}$.

Example C.1.3. For any $c > 0$, define $d : \mathbb{R} \times \mathbb{R} \to [0, \infty)$ by $d(x, y) = c|x - y|$. Then $(\mathbb{R}, d)$ is a metric space. In particular, there are many metrics that may be placed on a given set.

Example C.1.4. The usual notion of measuring the distance between two points in a plane is a metric. Indeed define $d_2 : \mathbb{R}^2 \times \mathbb{R}^2 \to [0, \infty)$ by $d_2((x_1, y_1), (x_2, y_2)) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$. Then $(\mathbb{R}^2, d_2)$ is a metric space and the metric d_2 is called the *Euclidean metric*.

Example C.1.5. Define $d : \mathbb{C} \times \mathbb{C} \to [0, \infty)$ by $d(x, y) = |x - y|$. Then $(\mathbb{C}, d)$ is a metric space.

Example C.1.6. Given $n \in \mathbb{N}$, define $d_1 : \mathbb{C}^n \times \mathbb{C}^n \to [0, \infty)$ by

$$d_1((z_1, \ldots, z_n), (w_1, \ldots, w_n)) = \sum_{k=1}^{n} |z_k - w_k|$$

for all $(z_1, \ldots, z_n), (w_1, \ldots, w_n) \in \mathbb{C}^n$. Then it is easy to verify that $(\mathbb{C}^n, d_1)$ is a metric space.

In fact, given any set, it is possible to place a metric on the set.

Example C.1.7. Let $\mathcal{X}$ be a non-empty set. Define $d : \mathcal{X} \times \mathcal{X} \to [0, \infty)$ by

$$d(x, y) = \begin{cases} 0 & \text{if } x = y \\ 1 & \text{if } x \neq y \end{cases}.$$

It is elementary to verify that d is a metric, which is called the *discrete or trivial metric*.

Although the above gives several examples of metrics, not all metrics were created equal. In particular, we desire to study special types of metric spaces. These metric spaces come from specific functions on vector spaces that behave like the absolute value does on $\mathbb{R}$ and $\mathbb{C}$. Consequently, we will restrict to vector spaces where the scalars are either the real or the complex numbers. Consequently, it will be convenient to use $\mathbb{K}$ to denote either $\mathbb{R}$ or $\mathbb{C}$.

The following is our generalization of the absolute value to vector spaces.

Definition C.1.8. Let V be a vector space over $\mathbb{K}$. A *norm* on V is a function $\|\cdot\| : V \to [0, \infty)$ such that

1. for $\vec{v} \in V$, $\|\vec{v}\| = 0$ if and only if $\vec{v} = \vec{0}$,
2. $\|\alpha \vec{v}\| = |\alpha| \, \|\vec{v}\|$ for all $\alpha \in \mathbb{K}$ and $\vec{v} \in V$, and
3. (Triangle Inequality) $\|\vec{v} + \vec{w}\| \leq \|\vec{v}\| + \|\vec{w}\|$ for all $\vec{v}, \vec{w} \in V$.

Definition C.1.9. A *normed linear space* is a pair $(V, \|\cdot\|)$ where V is a vector space over $\mathbb{K}$ and $\|\cdot\|$ is a norm on V.

As our motivation for generalizing the absolute value was to induce a metric, we note the following.

Proposition C.1.10. *If $(V, \|\cdot\|)$ is a normed linear space, then V is a metric space with the metric $d : V \times V \to [0, \infty)$ defined by $d(\vec{v}, \vec{w}) = \|\vec{v} - \vec{w}\|$. We call d the metric induced by $\|\cdot\|$.*

Proof. It suffices to show that d is a metric. Clearly $d : V \times V \to [0, \infty)$. Furthermore notice $d(\vec{v}, \vec{w}) = 0$ if and only if $\|\vec{v} - \vec{w}\| = 0$ if and only if $\vec{v} - \vec{w} = \vec{0}$ if and only if $\vec{v} = \vec{w}$.

Next notice for all $\vec{v}, \vec{w} \in V$ that

$$d(\vec{v}, \vec{w}) = \|\vec{v} - \vec{w}\| = \|(-1)(\vec{w} - \vec{v})\| = |-1| \, \|\vec{w} - \vec{v}\| = d(\vec{w}, \vec{v}).$$

Finally, to see that d satisfies the triangle inequality, notice for all $\vec{v}, \vec{w}, \vec{z} \in V$ that

$$\begin{aligned} d(\vec{v}, \vec{z}) + d(\vec{z}, \vec{w}) &= \|\vec{v} - \vec{z}\| + \|\vec{z} - \vec{w}\| \\ &\geq \|(\vec{v} - \vec{z}) + (\vec{z} - \vec{w})\| \\ &= \|\vec{v} - \vec{w}\| = d(\vec{v}, \vec{w}). \end{aligned}$$

Hence d is a metric. ∎

Remark C.1.11. Notice in the proof of the triangle inequality in Proposition C.1.10 that using $\vec{w} = \vec{0}$ produced $\|\vec{v} - \vec{z}\| + \|\vec{z}\| \geq \|\vec{v}\|$ for all $\vec{v}, \vec{z} \in V$. Hence

$$\|\vec{v}\| - \|\vec{z}\| \leq \|\vec{v} - \vec{z}\|$$

for all $\vec{v}, \vec{z} \in V$. Thus, by interchanging $\vec{v}$ and $\vec{z}$, we obtain that

$$|\|\vec{v}\| - \|\vec{z}\|| \leq \|\vec{v} - \vec{z}\|$$

for all $\vec{v}, \vec{z} \in V$. This potentially useful inequality is often called the *reverse triangle inequality.*

Clearly the absolute value on $\mathbb{K}$ is an norm on $\mathbb{K}$. Furthermore, the metric induced by this norm is exactly the metric introduced in Examples C.1.3 and C.1.5. In fact, some of the other metrics we have seen come from norms.

Example C.1.12. For $n \in \mathbb{N}$, define $\|\cdot\|_1 : \mathbb{K}^n \to [0, \infty)$ by

$$\|(z_1, \ldots, z_n)\|_1 = \sum_{k=1}^{n} |z_k|$$

for all $(z_1, \ldots, z_n) \in \mathbb{K}^n$. It is elementary to verify that $\|\cdot\|_1$ is a norm on $\mathbb{K}^n$ that induces the metric d_1 as in Example C.1.6. We call $\|\cdot\|_1$ the 1-*norm.*

Example C.1.13. Define $\|\cdot\|_2 : \mathbb{R}^2 \to [0, \infty)$ by

$$\|(x, y)\|_2 = \sqrt{|x|^2 + |y|^2}$$

for all $(x, y) \in \mathbb{R}^2$. It is possible to verify that $\|\cdot\|_2$ is a norm on $\mathbb{R}^2$ that induces the Euclidean metric as in Example C.1.4. We call $\|\cdot\|_2$ the *Euclidean norm* or the 2-*norm.*

However, some of the metrics we have seen are not norms. For example, if V is a vector space over $\mathbb{K}$, the trivial metric cannot be induced by a norm since if a norm (and thus its induced metric) takes the value 1, then it takes all values in $[0, \infty)$.

There are many more useful norms. In fact, there are several norms we can place on $\mathbb{K}^n$.

Example C.1.14. For $n \in \mathbb{N}$, define $\|\cdot\|_\infty : \mathbb{K}^n \to [0, \infty)$ by

$$\|(z_1, \ldots, z_n)\|_\infty = \sup_{1 \leq k \leq n} |z_k|$$

for all $(z_1, \ldots, z_n) \in \mathbb{K}^n$. It is elementary to verify that $\|\cdot\|_\infty$ is a norm on $\mathbb{K}^n$. We call $\|\cdot\|_\infty$ the sup-*norm* or the ∞-*norm.*

Using the idea of the ∞-norm, we can develop a norm on vector spaces we have yet to consider.

Example C.1.15. Let $C[a,b]$ denote the set of all real-valued continuous functions on a closed interval $[a,b]$. Then $C[a,b]$ is a vector space over $\mathbb{R}$ under the operations of pointwise addition and scalar multiplication. Define $\|\cdot\|_\infty : C[a,b] \to [0,\infty)$ by

$$\|f\|_\infty = \sup_{x\in[a,b]} |f(x)|$$

for all $f \in C[a,b]$. Note $\|\cdot\|_\infty$ does take values in $[0,\infty)$ by the Extreme Value Theorem. It is elementary to see that $\|\cdot\|_\infty$ is a norm on $C[a,b]$. We call $\|\cdot\|_\infty$ the sup-*norm* or the ∞-*norm*.

Of course, the sup-norm works perfectly well if we restrict the set of continuous functions to, for example, the polynomials. In particular, this holds true in more generality.

Proposition C.1.16. *Let $(V, \|\cdot\|)$ be a normed linear space and let W be a subspace of V. The restriction of $\|\cdot\|$ to W is a norm on W.*

There are many more norms we can place on $\mathbb{K}^n$. In particular, we can generalize the Euclidean norm.

Example C.1.17. For $n \in \mathbb{N}$, define $\|\cdot\|_2 : \mathbb{K}^n \to [0,\infty)$ by

$$\|(z_1,\ldots,z_n)\|_2 = \left(\sup_{1\leq k\leq n} |z_k|^2 \right)^{\frac{1}{2}}$$

for all $(z_1,\ldots,z_n) \in \mathbb{K}^n$. Then $\|\cdot\|_\infty$ is a norm on $\mathbb{K}^n$ called the *Euclidean norm* or the 2-*norm*.

Example C.1.18. For $n \in \mathbb{N}$ and a fixed $p \in (1,\infty)$, define $\|\cdot\|_p : \mathbb{K}^n \to [0,\infty)$ by

$$\|(z_1,\ldots,z_n)\|_p = \left(\sum_{k=1}^n |z_k|^p \right)^{\frac{1}{p}}$$

for all $(z_1,\ldots,z_n) \in \mathbb{K}^n$. Then $\|\cdot\|_p$ is a norm on $\mathbb{K}^n$ called the *p-norm*.

C.2 Topology on Metric Spaces

In this section, we will analyze the notion of convergent sequences in metric spaces. Of course we could jump right in and define the convergence of a sequence using our distance function. However, in doing so we will not obtain too much information about the structure of our spaces and of the subsets of our spaces. Thus we will begin with another view of how to define a sequence to converge thereby permitting a deeper discussion of types and properties of subsets of metric spaces.

C.2.1 Open and Closed Sets

One way to interpret the notion of a convergence sequence of real numbers without a notion of distance is to say that $a_n \in (L - \epsilon, L + \epsilon)$ for all $n \geq N$. Thus for $(a_n)_{n\geq 1}$ to be 'close' to L means that each element in $(a_n)_{n\geq 1}$ must eventually be in any fixed open interval containing L. Thus if we can analyze the essential properties of open intervals and generalize these to metric spaces, we may generalize the notion of a convergent sequence. If fact, we want a concept slightly more general than an open interval.

Definition C.2.1. Let X be a non-empty set. A collection $\mathcal{T} \subseteq \mathcal{P}(X)$ is said to be a *topology* on X if

1. $\emptyset, X \in \mathcal{T}$,
2. if $\{U_\alpha \mid \alpha \in I\} \subseteq \mathcal{T}$, then $\bigcup_{\alpha \in I} U_\alpha \in \mathcal{T}$, and
3. if $n \in \mathbb{N}$ and $U_1, \ldots, U_n \in \mathcal{T}$, then $\bigcap_{i=1}^n U_i \in \mathcal{T}$.

The elements of $\mathcal{T}$ are called the *open sets* of the topology.

There are many examples of topologies we may place on a set.

Example C.2.2. Let X be a non-empty set. The set $\mathcal{T} = \{X, \emptyset\}$ is a topology on X known as the *trivial topology*.

Example C.2.3. Let X be a non-empty set. The set $\mathcal{T} = \mathcal{P}(X)$ is a topology on X known as the *discrete topology*.

Of course, the above topologies may not be the best topologies for a metric space as we desire a topology related to the metric. Thus we begin with the following definitions.

Definition C.2.4. Let $(\mathcal{X}, d)$ be a metric space. Given an $x \in \mathcal{X}$ and an $r > 0$, the *open ball* of radius r centred at x, denoted $B(x, r)$, is the set

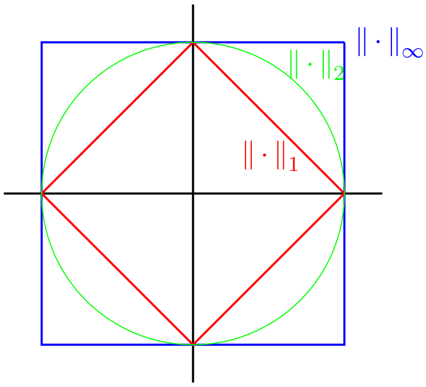
$$B(x, r) = \{y \in \mathcal{X} \mid d(x, y) < r\}.$$

Similarly, given an $x \in \mathcal{X}$ and an $r \geq 0$, the *closed ball* of radius r centred at x, denoted $B[x, r]$, is the set

$$B[x, r] = \{y \in \mathcal{X} \mid d(x, y) \leq r\}.$$

Example C.2.5. In $\mathbb{R}$ with the absolute value metric, $B(x, r) = (x - r, x + r)$ and $B[x, r] = [x - r, x + r]$ for all $x \in \mathbb{R}$ and $r > 0$.

Example C.2.6. For $\mathbb{R}^2$, the following diagram illustrates $B(0,1)$ for various p-norms:



Example C.2.7. Let $\mathcal{X}$ be a non-empty set and let d be the discrete metric on $\mathcal{X}$. Then, for all $x \in \mathcal{X}$,

$$\begin{aligned} B(x,r_1) = B[x,r_2] = \{x\} &\quad \text{if } r_1 \leq 1 \text{ and } r_2 < 1, \text{ and} \\ B(x,r_1) = B[x,r_2] = \mathcal{X} &\quad \text{if } r_1 > 1 \text{ and } r_2 \geq 1. \end{aligned}$$

Unsurprisingly, to obtain a desirably topology on a metric space, we will use our open balls to construct the open sets.

Theorem C.2.8. *Let $(\mathcal{X},d)$ be a metric space. Let $\mathcal{T}$ be the set of all subsets U of $\mathcal{X}$ such that for each $x \in U$ there exists an $\epsilon > 0$ such that $B(x,\epsilon) \subseteq U$. Then $\mathcal{T}$ is a topology on $\mathcal{X}$.*

Proof. To see that $\mathcal{T}$ is a topology, we must verify the three properties in Definition C.2.1. It is clear by definition that $\emptyset, \mathcal{X} \in \mathcal{T}$.

Suppose $\{U_\alpha\}_{\alpha\in I}$ is a set of elements of $\mathcal{T}$. To see that $\bigcup_{\alpha\in I} U_\alpha \in \mathcal{T}$, let $x \in \bigcup_{\alpha\in I} U_\alpha$ be arbitrary. Then there must be an $i \in I$ such that $x \in U_i$. Since $U_i \in \mathcal{T}$, there exists an $\epsilon > 0$ such that $B(x,\epsilon) \subseteq U_i$. Hence $B(x,\epsilon) \subseteq U_i \subseteq \bigcup_{\alpha\in I} U_\alpha$. As $x \in \bigcup_{\alpha\in I} U_\alpha$ was arbitrary, $\bigcup_{\alpha\in I} U_\alpha \in \mathcal{T}$.

Finally, suppose $U_1,\ldots,U_n \in \mathcal{T}$. To see that $\bigcap_{i=1}^n U_i \in \mathcal{T}$, suppose $x \in \bigcap_{i=1}^n U_i$ be arbitrary. Hence $x \in U_i$ for all $i \in \{1,\ldots,n\}$. Since each $U_i \in \mathcal{T}$, there exists an $\epsilon_i > 0$ such that $B(x,\epsilon_i) \subseteq U_i$ for all $i \in \{1,\ldots,n\}$. Let $\epsilon = \min_{1\leq i\leq n} \epsilon_i > 0$. Notice for each $i \in \{1,\ldots,n\}$ that

$$B(x,\epsilon) \subseteq B(x,\epsilon_i) \subseteq U_i.$$

Hence $B(x,\epsilon) \subseteq \bigcap_{i=1}^n U_i$. As $x \in \bigcap_{i=1}^n U_i$ was arbitrary, $\bigcap_{i=1}^n U_i \in \mathcal{T}$ as desired. ∎

Definition C.2.9. Let $(\mathcal{X},d)$ be a metric space. The topology $\mathcal{T}$ from Theorem C.2.8 is called the *metric space topology* on $(\mathcal{X},d)$. Unless otherwise specified, given a metric space $(\mathcal{X},d)$ the topology on $\mathcal{X}$ will always be the metric space topology and the elements of $\mathcal{T}$ will be referred to as open sets.

Of course, it is useful to be able to determine which sets are open. It should not be a surprise that our open balls are indeed open sets. In fact, it is not difficult to see that the metric topology is the smallest topology where every open ball is an open set.

Proposition C.2.10. *Let $(\mathcal{X}, d)$ be a metric space. Every open ball in $\mathcal{X}$ is an open set.*

Proof. Consider the open ball $B(x, \epsilon)$ for some $x \in \mathcal{X}$ and $\epsilon > 0$. To see that $B(x, \epsilon)$ is open, let $y \in B(x, \epsilon)$ be arbitrary. Thus $d(x, y) < \epsilon$.

Let $\delta = \epsilon - d(x, y) > 0$. We claim that $B(y, \delta) \subseteq B(x, \epsilon)$. To see this, let $z \in B(y, \delta)$ be arbitrary. Then $d(z, y) < \delta$ so, by the triangle inequality,

$$d(z, x) \leq d(z, y) + d(y, x) < \delta + d(y, x) = \epsilon.$$

Therefore, since $z \in B(y, \delta)$ was arbitrary, $B(y, \delta) \subseteq B(x, \epsilon)$. Hence $B(x, \epsilon)$ is open as $y \in B(x, \epsilon)$ was arbitrary. ∎

We also note the following complete description of open subsets of $\mathbb{R}$.

Proposition C.2.11. *Every open subset of $\mathbb{R}$ is a countable union of open intervals*

Proof. Let U be an arbitrary non-empty open subset of $\mathbb{R}$. Define a relation $\sim$ on U by $x \sim y$ if and only if whenever $x < z < y$ or $y < z < x$ then $z \in U$. We claim that $\sim$ is an equivalence relation on U. To see this first notice that if $x \in U$, then $x \sim x$ trivially. Furthermore, clearly if $x \sim y$ then $z \in U$ whenever $x < z < y$ or $y < z < x$, and thus $y \sim x$. Finally, suppose $x, y, w \in U$ are such that $x \sim y$ and $y \sim w$. To see that $x \sim w$, we divide the discussion into five cases:

Case 1: $x \leq y \leq w$. In this case, we have $x < z < y$ implies $z \in U$ and $y < z < w$ implies $z \in U$. If z is such that $x < z < w$, then either $x < z < y$, $y < z < w$, or $y = z$. As all of these imply $z \in U$, we have $x \sim w$ in this case.

Case 2: $w \leq y \leq x$. This case follows from Case 1 by interchanging x and w.

Case 3: $y \leq x \leq w$. In this case, we have $y < z < w$ implies $z \in U$. Thus if $x < z < w$ then $y < z < w$ so $z \in U$. Hence $z \sim x$ in this case.

Case 4: $y \leq w \leq x$. This case follows from Case 3 by interchanging x and w.

Case 5: $x \leq w \leq y$ or $w \leq x \leq y$. This case follows from Cases 3 and 4 by reversing the inequalities.

Thus, in any case $x \sim w$. Thus $\sim$ is an equivalence relation.

Next we claim that each equivalence class is an open interval. To see this let $x \in U$ be arbitrary and let E_x denote the equivalence class of x with respect to $\sim$. To see that E_x is an open interval, let

$$\alpha_x = \inf(E_x) \qquad \text{and} \qquad \beta_x = \sup(E_x).$$

We claim that $E_x = (\alpha_x, \beta_x)$.

First, we claim that $\alpha_x < \beta_x$. To see this, notice that $x \in E_x \subseteq U$. Hence, as U is open, there exists an $\epsilon > 0$ such that $(x - \epsilon, x + \epsilon) \subseteq U$. Clearly $y \sim x$ for all $y \in (x - \epsilon, x + \epsilon)$ so

$$\alpha_x \leq x - \epsilon < x + \epsilon \leq \beta_x.$$

To see that $(\alpha_x, \beta_x) \subseteq E_x$, let $y \in (\alpha_x, \beta_x)$ be arbitrary. Since $\alpha_x < y < \beta_x$, by the definition of inf and sup there exists $z_1, z_2 \in E_x$ such that

$$\alpha_x \leq z_1 < y < z_2 \leq \beta_x.$$

Since $z_1, z_2 \in E_x$, we have $z_1 \sim x$ and $z_2 \sim x$. Thus $z_1 \sim z_2$ so $[z_1, z_2] \subseteq U$. Hence $y \in [z_1, z_2] \subseteq U$. Therefore, as $y \in (\alpha_x, \beta_x)$ was arbitrary, $(\alpha_x, \beta_x) \subseteq E_x$.

To see that $E_x \subseteq (\alpha_x, \beta_x)$, note that $E_x \subseteq (\alpha_x, \beta_x) \cup \{\alpha_x, \beta_x\}$ by the definition of α_x and β_x. Thus it suffices to show that $\alpha_x, \beta_x \notin E_x$. Suppose $\beta_x \in E_x$ (this implies $\beta_x \neq \infty$). Then $\beta_x \in U$ so there exists an $\epsilon > 0$ so that $(\beta_x - \epsilon, \beta_x + \epsilon) \subseteq U$. Hence $\beta_x + \frac{1}{2}\epsilon \sim \beta_x \sim x$ (as $\beta_x \in E_x$). Hence $\beta_x + \frac{1}{2}\epsilon \in E_x$. However $\beta_x + \frac{1}{2}\epsilon > \beta_x$ so $\beta_x + \frac{1}{2}\epsilon \in E_x$ contradicts the fact that $\beta_x = \sup(E_x)$. Hence we have obtained a contradiction so $\beta_x \notin E_x$. Similar arguments show that $\alpha_x \notin E_x$. Hence $E_x = (\alpha_x, \beta_x)$ as desired.

To complete the proof, first notice that clearly

$$U = \bigcup_{x \in U} E_x$$

so U is a union of open intervals. It remains to be verify that the above union can be made countable. Since each E_x is an open interval, $E_x \cap \mathbb{Q} \neq \emptyset$. Hence, as each $E_x \cap \mathbb{Q}$ is non-empty, by the Axiom of Choice there exists a function $f : \{E_x \mid x \in U\} \to \mathbb{Q}$ such that $f(E_x) \in E_x$ for all $x \in U$. Hence, as $E_x \cap E_y = \emptyset$ if $E_x \neq E_y$, f is an injective function. Hence $\{E_x \mid x \in U\}$ is countable. Thus the union $U = \bigcup_{x \in U} E_x$ can be made into a countable union by choosing one representative from each equivalence class (or, alternatively, $U = \bigcup_{q \in \mathbb{Q}} f^{-1}(\{q\})$). ∎

Remark C.2.12. Note that Definition C.2.1 only requires that a finite intersection of open sets is open. To see why this is required, note that in $\mathbb{R}$ that $U_n = (-\frac{1}{n}, \frac{1}{n})$ is an open subsets of $\mathbb{R}$ for all $n \in \mathbb{N}$ yet $\bigcap_{n=1}^{\infty} U_n = \{0\}$ is not an open set.

Remark C.2.13. Although we have many norms on $\mathbb{K}^n$, each metric space topology we have seen on $\mathbb{K}^n$ agrees. To see this, fix $n \in \mathbb{N}$. If $p \in [1, \infty]$ let $\mathcal{T}_p$ denoted the topology on $\mathbb{K}^n$ induced by the p-norm.

To see that $\mathcal{T}_p = \mathcal{T}_\infty$ for all $p \in [1,\infty)$ (and thus $\mathcal{T}_p = \mathcal{T}_q$ for all $p,q \in [1,\infty]$), first notice for an arbitrary $\vec{x} = (x_1, \ldots, x_n) \in \mathbb{K}^n$ that

$$\begin{aligned}\|\vec{x}\|_\infty^p &= \sup\{|x_k|^p \mid 1 \leq k \leq n\} \\ &\leq \sum_{k=1}^n |x_k|^p \\ &= \|\vec{x}\|^p \\ &\leq \sum_{k=1}^n \|\vec{x}\|_\infty^p \\ &= n \|\vec{x}\|_\infty^p .\end{aligned}$$

Hence $\|\vec{x}\|_\infty \leq \|\vec{x}\|_p \leq n^{\frac{1}{p}} \|\vec{x}\|_\infty$ for all $\vec{x} \in \mathbb{K}^n$.

To show that $\mathcal{T}_p = \mathcal{T}_\infty$ we must show that every open subset of $\mathbb{K}^n$ with respect to either norm is open with respect to the other norm. For notational simplicity, we will use $B^p(\vec{x}, r)$ to denote the open ball centred at $\vec{x}$ of radius r with respect to the p-norm and we will use $B^\infty(\vec{x}, r)$ to denote the open ball centred at $\vec{x}$ of radius r with respect to the ∞-norm

To begin, let $U \in \mathcal{T}_p$ be arbitrary. To see that $U \in \mathcal{T}_\infty$, let $x \in U$ be arbitrary. Since $U \in \mathcal{T}_p$ there exists an $r > 0$ such that $B^p(\vec{x}, r) \subseteq U$. As $B^\infty\left(\vec{x}, \frac{1}{n^{\frac{1}{p}}} r\right) \subseteq B^p(\vec{x}, r) \subseteq U$ by the above norm estimates, and as $x \in U$ was arbitrary, we obtain that $U \in \mathcal{T}_\infty$. Hence $\mathcal{T}_p \subseteq \mathcal{T}_\infty$.

For the other inclusion, let $U \in \mathcal{T}_\infty$ be arbitrary. To see that $U \in \mathcal{T}_p$, let $x \in U$ be arbitrary. Since $U \in \mathcal{T}_\infty$ there exists an $r > 0$ such that $B^\infty(\vec{x}, r) \subseteq U$. As $B^p(\vec{x}, r) \subseteq B^\infty(\vec{x}, r) \subseteq U$ by the above norm estimates, we obtain that $U \in \mathcal{T}_p$. Hence $\mathcal{T}_\infty \subseteq \mathcal{T}_p$. Thus $\mathcal{T}_\infty = \mathcal{T}_p$ as desired.

Although we have been interested in open sets in relation to convergent sequences, the complements of open sets will be of incredibly interest.

Definition C.2.14. Let $\mathcal{T}$ be a topology on a set X. A subset $F \subseteq X$ is said to be *closed* if F^c is open.

Example C.2.15. Let $(\mathcal{X}, d)$ be a metric space. Then $\emptyset$ and $\mathcal{X}$ are both closed and open sets.

Example C.2.16. In $\mathbb{R}$ with the absolute value metric, $(a, b]$ is neither open nor closed. Indeed $(a, b]$ is not open as there is no open ball around b contained in $(a, b]$, and $(a, b]$ is not closed as $(a, b]^c = (-\infty, a] \cup (b, \infty)$ is not open since there is no open ball around a contained in $(a, b]^c$.

Example C.2.17. In $\mathbb{R}$ with the absolute value metric, $[a, b]$ is closed for all $a, b \in \mathbb{R}$ since $[a, b]^c = (-\infty, a) \cup (b, \infty)$ is a union of open sets and thus open.

Proposition C.2.18. *Every closed ball in a metric space $(\mathcal{X}, d)$ is a closed set.*

Proof. Let $x \in \mathcal{X}$ and $r > 0$. We claim that $B[x, r]^c$ is open. To see this, let $y \in B[x, r]^c$ be arbitrary. Then $d(x, y) > r$. Let $\epsilon = d(x, y) - r > 0$. Notice if $z \in B(y, \epsilon)$ then

$$d(x, y) \leq d(x, z) + d(z, y) < d(x, z) + \epsilon = d(x, z) + d(x, y) - r$$

which implies $r < d(x, z)$. Hence $B(y, \epsilon) \subseteq B[x, r]^c$. Therefore, as $y \in B[x, r]^c$ was arbitrary, $B[x, r]^c$ is an open set. Hence $B[x, r]$ is closed. ∎

Example C.2.19. Let d be the discrete metric on a non-empty set $\mathcal{X}$. Then every open ball is closed and every closed ball is open.

Like with open sets, there are set operations we may perform on closed sets.

Proposition C.2.20. *Let $\mathcal{T}$ be a topology on a set X, let I be an non-empty set, and for each $\alpha \in I$ let F_α be a closed subset of X. Then*

- $\bigcap_{\alpha \in I} F_\alpha$ *is closed in X, and*
- $\bigcup_{\alpha \in I} F_\alpha$ *is open in X provided I has a finite number of element.*

Proof. Since De Morgan's Laws imply

$$\left(\bigcap_{\alpha \in I} F_\alpha\right)^c = \bigcup_{\alpha \in I} F_\alpha^c \qquad \text{and} \qquad \left(\bigcup_{\alpha \in I} F_\alpha\right)^c = \bigcap_{\alpha \in I} F_\alpha^c,$$

the result follows by the definition of a closed set along with the definition of a topology. ∎

Remark C.2.21. Complementing the fact that a countable intersection of open sets need not be open, a countable union of closed sets need not be closed. Indeed $A = \bigcup_{n=1}^{\infty} \{\frac{1}{n}\}$ is a countable union of closed sets in $\mathbb{R}$ that is not closed since $0 \in A^c$ yet $(-\epsilon, \epsilon) \nsubseteq A^c$ for all $\epsilon > 0$ (we will see later that $A \cup \{0\}$ is a closed set). Furthermore, there exist closed subsets of $\mathbb{R}$ that are not countable unions of closed intervals.

Of course, the most important property of a closed set is related to convergent sequences (see Theorem C.2.32).

C.2.2 Convergence of Sequences

By modelling the notion of a convergent sequence in $\mathbb{R}$, we have finally arrived at defining when a sequence in a metric spaces converges.

Definition C.2.22. Let $(\mathcal{X}, d)$ be a metric space and let $(x_n)_{n\geq 1}$ be a sequence in $\mathcal{X}$. The sequence $(x_n)_{n\geq 1}$ is said to *converge* in $\mathcal{X}$ to an element $x_0 \in \mathcal{X}$ if for all $\epsilon > 0$ there exists an $N \in \mathbb{N}$ such that $d(x_n, x_0) < \epsilon$ for all $n \geq N$. In this case x_0 is said to be a *limit* of the sequence $(x_n)_{n\geq 1}$ and we write $x_0 = \lim_{n\to\infty} x_n$.

Of course, like in previous courses, the '$< \epsilon$' in Definition C.2.22 can be replaced with '$\leq \epsilon$' without changing the definition. Furthermore, as we have seen examples of convergent sequences in $\mathbb{R}$ in previous courses, we will examine some more exotic examples.

Example C.2.23. Let d be the discrete metric on a non-empty set $\mathcal{X}$. If $(x_n)_{n\geq 1}$ is a sequence in $\mathcal{X}$, then $(x_n)_{n\geq 1}$ converges to a point $x_0 \in \mathcal{X}$ if and only if there exists an $N \in \mathbb{N}$ such that $x_n = x_0$ for all $n \geq N$; that is, the sequence is eventually constant.

Example C.2.24. Let $m \in \mathbb{N}$ and $p \in [1, \infty]$. For each $n \in \mathbb{N}$, let $\vec{x}_n = (z_{1,n}, \ldots, z_{m,n}) \in \mathbb{K}^m$. Given $\vec{x} = (z_1, \ldots, z_m) \in \mathbb{K}^m$, the following are equivalent:

1. $(\vec{x}_n)_{n\geq 1}$ converges to $\vec{x}$ with respect to the p-norm.
2. $\lim_{n\to\infty} |z_{k,n} - z_k| = 0$ for all $k \in \{1, \ldots, m\}$

Indeed notice $|z_{k,n} - z_k| \leq \|\vec{x}_n - \vec{x}\|_p$. Thus (1) implies (2). For the other direction, notice if $|z_{k,n} - z_k| < \epsilon$ for all $k \in \{1, \ldots, m\}$ then

$$\|\vec{x}_n - \vec{x}\|_p = \left(\sum_{k=1}^m |z_{k,n} - z_k|^p\right)^{\frac{1}{p}} \leq \left(\sum_{k=1}^m \epsilon^p\right)^{\frac{1}{p}} = m^{\frac{1}{p}}\epsilon.$$

As m is fixed, $m^{\frac{1}{p}}\epsilon$ may be made as small as desired thereby completing the equivalence.

Example C.2.25. Repeating the same arguments in Example C.2.24 for $\mathbb{K} = \mathbb{R}$ and $p = 2$, if $(z_n)_{n\geq 1}$ is a sequence in $\mathbb{C}$, $z \in \mathbb{C}$, and $a_n, b_n, a, b \in \mathbb{R}$ are such that $z = a + bi$ and $z_n = a_n + b_n i$ for all $n \in \mathbb{N}$, then $z = \lim_{n\to\infty} z_n$ if and only if $a = \lim_{n\to\infty} a_n$ and $b = \lim_{n\to\infty} b_n$.

Example C.2.26. Given a sequence $(f_n)_{n\geq 1}$ of elements of $C[a, b]$, notice that $(f_n)_{n\geq 1}$ converges to an element $f \in C[a, b]$ with respect to $\|\cdot\|_\infty$ if and only if for all $\epsilon > 0$ there exists an $N \in \mathbb{N}$ such that $|f_n(x) - f(x)| < \epsilon$ for all $x \in [a, b]$ and $n \geq N$. This is precisely the notion of *uniform convergence of functions* discussed in previous analysis courses.

In the case of normed linear spaces, the notion of convergent sequences behaves well with respect to vector space operations.

Proposition C.2.27. *Let $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ be a normed linear space over $\mathbb{K}$. If $(\vec{x}_n)_{n\geq 1}$ and $(\vec{y}_n)_{n\geq 1}$ are sequences that converge to $\vec{x}$ and $\vec{y}$ respectively, then*

- *$(\vec{x}_n + \vec{y}_n)_{n\geq 1}$ converges to $\vec{x} + \vec{y}$, and*
- *$(\alpha\vec{x}_n)_{n\geq 1}$ converges to $\alpha\vec{x}$ for all $\alpha \in \mathbb{K}$.*

Proof. Let $\epsilon > 0$. Since

$$\begin{gathered} \|(\vec{x}_n + \vec{y}_n) - (\vec{x} - \vec{y})\| \leq \|\vec{x}_n - \vec{x}\| + \|\vec{y}_n - \vec{y}\| \text{ and} \\ \|\alpha\vec{x}_n - \alpha\vec{x}\| \leq |\alpha|\,\|\vec{x}_n - \vec{x}\| \end{gathered}$$

for all n and since we may chose N sufficiently large so that the right-hand sides of both inequalities is less than ϵ, the result follows. ∎

As the statement "$d(x_n, x_0) < \epsilon$" is equivalent to saying that $x_n \in B(x_0, \epsilon)$, we directly have a connection between convergence of sequences and topology.

Proposition C.2.28. *Let $(\mathcal{X}, d)$ be a metric space. A sequence $(x_n)_{n\geq 1}$ converges to an element $x_0 \in \mathcal{X}$ if and only if for every open set U of $\mathcal{X}$ such that $x_0 \in U$ there exists an $N \in \mathbb{N}$ such that $x_n \in U$ for all $n \geq N$.*

For general topological spaces (i.e. a space with a topology), the notion of convergence is defined via Proposition C.2.28. One thinks of each open set as a 'neighbourhood' around a point and for a sequence to converge to a point, it must eventually inside every open set. This becomes a problem for a general topological space due to examples like the trivial topology for which two or more points are contained in the same open sets and thus are both limits of the same sequences. In metric spaces, we do not have this problem.

Proposition C.2.29. *Let $(\mathcal{X}, d)$ be a metric space and let $(x_n)_{n\geq 1}$ be a sequence in $\mathcal{X}$. If $x_0 = \lim_{n\to\infty} x_n$ and $y_0 = \lim_{n\to\infty} x_n$, then $x_0 = y_0$.*

Proof. Suppose $x_0 = \lim_{n\to\infty} x_n$ and $y_0 = \lim_{n\to\infty} x_n$. Let $\epsilon > 0$ be arbitrary. Since $x_0 = \lim_{n\to\infty} x_n$ there exists an $N_1 \in \mathbb{N}$ such that $d(x_n, x_0) < \epsilon$ for all $n \geq N_1$. Similarly, since $y_0 = \lim_{n\to\infty} x_n$ there exists an $N_2 \in \mathbb{N}$ such that $d(x_n, y_0) < \epsilon$ for all $n \geq N_2$. Therefore, if $N = \max\{N_1, N_2\}$, we obtain that

$$0 \leq d(x_0, y_0) \leq d(x_0, x_N) + d(x_N, y_0) < 2\epsilon.$$

Since the above inequality holds for all $\epsilon > 0$, we obtain that $d(x_0, y_0) = 0$. Hence $x_0 = y_0$ by property (1) of Definition C.1.1 ∎

Given a sequence, it is often useful to be able to construct other sequences by removing elements. This leads to the following notion.

Definition C.2.30. Let $(\mathcal{X}, d)$ be a metric space. A *subsequence* of a sequence $(x_n)_{n\geq 1}$ of elements of $\mathcal{X}$ is any sequence $(y_n)_{n\geq 1}$ such that there exists an increasing sequence of natural numbers $(k_n)_{n\geq 1}$ so that $y_n = x_{k_n}$ for all $n \in \mathbb{N}$.

Unsurprisingly, if a sequence converges to a point, so does every subsequence.

Proposition C.2.31. *Let $(\mathcal{X}, d)$ be a metric space and let $(x_n)_{n\geq 1}$ be a sequence that converges to $x \in \mathcal{X}$. Every subsequence of $(x_n)_{n\geq 1}$ converges to x.*

Proof. Let $(x_{k_n})_{n\geq 1}$ be a subsequence of $(x_n)_{n\geq 1}$. Let $\epsilon > 0$. Since $x = \lim_{n\to\infty} x_n$, there exists an $N \in \mathbb{N}$ such that $d(x_n, x) < \epsilon$ for all $n \geq N$. Since $(k_n)_{n\geq 1}$ is an increasing sequence of natural numbers, there exists an $N_0 \in \mathbb{N}$ such that $k_n \geq N$ for all $n \geq N_0$. Hence $d(x_{k_n}, x) < \epsilon$ for all $n \geq N_0$. Therefore, as $\epsilon > 0$ was arbitrary, we obtain that $\lim_{n\to\infty} x_{k_n} = x$ by the definition of the limit. ∎

Of course, convergent sequences can be used to characterize closed sets.

Theorem C.2.32. *Let $(\mathcal{X}, d)$ be a metric space and let $A \subseteq \mathcal{X}$. Then A is a closed set if and only if whenever $(a_n)_{n\geq 1}$ is a sequence of elements of A that converge to an element $x \in \mathcal{X}$, then $x \in A$.*

Proof. Suppose that A is a closed set. Suppose to the contrary that there exists a sequence $(a_n)_{n\geq 1}$ of elements from A such that $x = \lim_{n\to\infty} a_n$ and $x \in A^c$. Since A is closed, A^c is open so there exists an $\epsilon > 0$ such that $B(x, \epsilon) \subseteq A^c$. However, since $x = \lim_{n\to\infty} a_n$, there exists an $N \in \mathbb{N}$ such that $a_n \in B(x, \epsilon) \subseteq A^c$ for all $n \geq N$. Notice this is a contradiction as $a_n \in A$ for all $n \in \mathbb{N}$. Therefore one direction is complete.

To see the converse, suppose that whenever $(a_n)_{n\geq 1}$ is a sequence of elements of A that converge to an element $x \in \mathcal{X}$, then $x \in A$. Suppose to the contrary that A is not closed. Therefore A^c is not open. Thus there exists an $x \in A^c$ such that $B(x, \epsilon) \cap A \neq \emptyset$ for all $\epsilon > 0$. For each $n \in \mathbb{N}$ choose $a_n \in B(x, \frac{1}{n}) \cap A$. Clearly $(a_n)_{n\geq 1}$ is a sequence of elements of A that converges to x so, by assumption, $x \in A$. As this contradicts the fact that $x \in A^c$, the proof is complete. ∎

C.3 Continuity

In this section we will focus on studying continuous functions as these are functions that are very well-behaved with respect to the topological properties of metric spaces. In particular, continuous functions are those that preserve convergent sequences. Furthermore, continuous functions interact in a very specific and useful way with the metric space topology.

C.3.1 Continuity and Topology

To generalize the notion of a continuous function on $\mathbb{R}$ to a function between metric spaces, we begin by recalling the ϵ-δ notion of continuity.

Definition C.3.1. Recall a function $f : \mathbb{R} \to \mathbb{R}$ is said to be continuous at $a \in \mathbb{R}$ if for all $\epsilon > 0$ there exists a $\delta > 0$ such that $|x - a| < \delta$ implies $|f(x) - f(x)| < \epsilon$.

It is clear that to generalize the notion of continuity to functions between metric spaces, we need only insert the appropriate notion of distance.

Definition C.3.2. Let $(\mathcal{X}, d_\mathcal{X})$ and $(\mathcal{Y}, d_\mathcal{Y})$ be metric spaces. It is said that a function $f : \mathcal{X} \to \mathcal{Y}$ is *continuous* at a point $x_0 \in \mathcal{X}$ if for all $\epsilon > 0$ there exists a $\delta > 0$ such that if $d_\mathcal{X}(x, x_0) < \delta$ then $d_\mathcal{Y}(f(x), f(x_0)) < \epsilon$. Otherwise it is said that f is *discontinuous* at x_0.

The set of continuous functions from $\mathcal{X}$ to $\mathcal{Y}$ is denoted $\mathcal{C}(\mathcal{X}, \mathcal{Y})$.

Remark C.3.3. Note that the '$<$' in both the '$< \delta$' and '$< \epsilon$' portions of Definition C.3.2 may be replaced by '$\leq$'. Indeed this follows since for all $x \in \mathcal{X}$ and $r > 0$,

$$B\left(x, \frac{1}{2}r\right) \subseteq B\left[x, \frac{1}{2}r\right] \subseteq B\left(x, r\right).$$

Definition C.3.4. Let $(\mathcal{X}, d_\mathcal{X})$ and $(\mathcal{Y}, d_\mathcal{Y})$ be metric spaces. It is said that a function $f : \mathcal{X} \to \mathcal{Y}$ is *continuous* (on $\mathcal{X}$) if f is continuous at each point in $\mathcal{X}$.

We have already seen several continuous functions on $\mathbb{R}$ in previous courses (e.g. polynomials, trigonometric functions, exponentials, etc.). Here are some more unusual examples.

Example C.3.5. Let $(\mathcal{X}, d_\mathcal{X})$ and $(\mathcal{Y}, d_\mathcal{Y})$ be metric spaces. If $d_\mathcal{X}$ is the discrete metric, then any function $f : \mathcal{X} \to \mathcal{Y}$ is continuous. If $d_\mathcal{Y}$ is the discrete metric, then a function $f : \mathcal{X} \to \mathcal{Y}$ is continuous at x_0 if and only if there exists a neighbourhood U of x_0 such that f is constant on U. In particular, if $\mathcal{X} = \mathbb{R}$ and $d_\mathcal{Y}$ is the discrete metric, $f : \mathcal{X} \to \mathcal{Y}$ is continuous if and only if f is constant.

As with continuous functions on $\mathbb{R}$, continuity of functions between metric spaces may be characterized via preservation of convergent sequences. Furthermore, continuity can also be characterized using topological properties.

Theorem C.3.6. *Let $(\mathcal{X}, d_\mathcal{X})$ and $(\mathcal{Y}, d_\mathcal{Y})$ be metric spaces, let $f : \mathcal{X} \to \mathcal{Y}$, and let $x_0 \in \mathcal{X}$. The following are equivalent:*

(1) f is continuous at x_0.

(2) For every sequence $(x_n)_{n\geq 1}$ in $\mathcal{X}$ that converges to x_0, the sequence $(f(x_n))_{n\geq 1}$ converges to $f(x_0)$.

(3) For every neighbourhood V of $f(x_0)$, $f^{-1}(V)$ is a neighbourhood of x_0.

Proof. To see that (1) implies (2), suppose f is continuous at x_0 and that $(x_n)_{n\geq 1}$ is a sequence in $\mathcal{X}$ that converges to x_0. To see that $(f(x_n))_{n\geq 1}$ converges to $f(x_0)$, let $\epsilon > 0$. Since f is continuous at x_0, there exists a $\delta > 0$ such that if $d_{\mathcal{X}}(x, x_0) < \delta$ then $d_{\mathcal{Y}}(f(x), f(x_0)) < \epsilon$. Since $x_0 = \lim_{n\to\infty} x_n$, there exists an $N \in \mathbb{N}$ such that $d(x_n, x_0) < \delta$ for all $n \geq N$. Hence $d(f(x_n), f(x_0)) < \epsilon$ for all $n \geq N$. Since $\epsilon > 0$ was arbitrary, we obtain that $f(x_0) = \lim_{n\to\infty} f(x_n)$ as desired.

To see that (2) implies (3), suppose to the contrary that there exists a neighbourhood V of $f(x_0)$ such that $f^{-1}(V)$ is not a neighbourhood of x_0. Since $x_0 \in f^{-1}(V)$ this implies that $B\left(x_0, \frac{1}{n}\right) \cap (f^{-1}(V))^c \neq \emptyset$ for all $n \in \mathbb{N}$. For each $n \in \mathbb{N}$ choose an element

$$x_n \in B\left(x_0, \frac{1}{n}\right) \cap (f^{-1}(V))^c.$$

Hence $(x_n)_{n\geq 1}$ converges to x_0. So, by the assumption of (2), we obtain that $f(x_0) = \lim_{n\to\infty} f(x_n)$. Since V is a neighbourhood of $f(x_0)$, this implies $f(x_n) \in V$ for some $n \in \mathbb{N}$ which implies $x_n \in f^{-1}(V)$. As $x_n \in (f^{-1}(V))^c$, we have obtained a contradiction. Hence (2) implies (3).

To see that (3) implies (1), let $\epsilon > 0$ be arbitrary. Since $B(f(x_0), \epsilon)$ is a neighbourhood of $f(x_0)$, $f^{-1}(B(f(x_0), \epsilon))$ is a neighbourhood of x_0 by the assumption of (3). Hence there exists a $\delta > 0$ such that

$$B(x_0, \delta) \subseteq f^{-1}(B(f(x_0), \epsilon)).$$

Thus, if $d(x, x_0) < \delta$ then

$$x \in B(x_0, \delta) \subseteq f^{-1}(B(f(x_0), \epsilon))$$

so $f(x) \in B(f(x_0), \epsilon)$ and thus $d(f(x), f(x_0)) < \epsilon$. Hence f is continuous by definition. ∎

In addition to the above we obtain the following characterization of continuity using open sets. As the following characterization makes no use of the metric, one may generalize this result to obtain a definition of continuous functions between any two sets with given topologies.

Proposition C.3.7. *Let $(\mathcal{X}, d_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}})$ be metric spaces. A function $f : \mathcal{X} \to \mathcal{Y}$ is continuous if and only if $f^{-1}(V)$ is open (in $\mathcal{X}$) for every open subset V of $\mathcal{Y}$.*

Proof. If $f^{-1}(V)$ is open (in $\mathcal{X}$) for every open subset V of $\mathcal{Y}$, then the fact that f is continuous at each point in $\mathcal{X}$ follows from the proof of (3) implies (1) in Theorem C.3.6.

Suppose $f : \mathcal{X} \to \mathcal{Y}$ is continuous and let V be an open subset of $\mathcal{Y}$. Let $U = f^{-1}(V)$ and let $x \in U$ be arbitrary. Since $f(x) \in V$ and since V is open, there exists an $\epsilon > 0$ such that $B(f(x), \epsilon) \subseteq V$. Since f is continuous, there exists a $\delta > 0$ such that if $y \in B(x, \delta)$ then $f(y) \in B(f(x), \epsilon) \subseteq V$. Hence $B(x, \delta) \subseteq U$. Therefore, since x was arbitrary U is open as desired. ∎

Corollary C.3.8. *Let $(\mathcal{X}, d_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}})$ be metric spaces. A function $f : \mathcal{X} \to \mathcal{Y}$ is continuous if and only if $f^{-1}(F)$ is closed (in $\mathcal{X}$) for every closed subset F of $\mathcal{Y}$.*

Proof. Since $f^{-1}(A^c) = (f^{-1}(A))^c$, the result follows from Proposition C.3.7. ∎

As with continuous functions on $\mathbb{R}$, composition continuous functions preserves continuity.

Proposition C.3.9. *Let $(\mathcal{X}, d_{\mathcal{X}})$, $(\mathcal{Y}, d_{\mathcal{Y}})$, and $(\mathcal{Z}, d_{\mathcal{Z}})$ be metric spaces, and let $f : \mathcal{X} \to \mathcal{Y}$ and $g : \mathcal{Y} \to \mathcal{Z}$ be continuous functions. Then $g \circ f : \mathcal{X} \to \mathcal{Z}$ is continuous.*

Proof. Let $x \in \mathcal{X}$ be arbitrary. To see that $g \circ f$ is continuous at x, let $(x_n)_{n\geq 1}$ be a sequence in $\mathcal{X}$ that converges to x. Therefore, by Theorem C.3.6, $f(x) = \lim_{n\to\infty} f(x_n)$ as f is continuous and thus $g(f(x)) = \lim_{n\to\infty} g(f(x_n))$ as g is continuous. Hence $g \circ f$ is continuous at x by Theorem C.3.6. As $x \in \mathcal{X}$ was arbitrary, the proof is complete. ∎

C.3.2 Useful Continuous Functions

In this section we will describe some useful continuous functions and the existence of certain continuous functions on metric spaces. All of these results we be related to the following notion.

Definition C.3.10. Let $(\mathcal{X}, d)$ be a metric space and let $A \subseteq \mathcal{X}$ be a non-empty set. Given $x \in \mathcal{X}$, the *distance from x to A*, denoted $\mathrm{dist}(x, A)$, is defined to be

$$\mathrm{dist}(x, A) = \inf\{d(x, a) \mid a \in A\}.$$

Example C.3.11. If $A = \{a\}$ then clearly $d(x, A) = d(x, a)$. Furthermore, if $\mathcal{X}$ is a normed linear space, then $d(x, A) = \|x - a\|$.

As a further example and to exhibit some important properties of $\mathrm{dist}(x, A)$, we note the following.

Lemma C.3.12. *Let $(\mathcal{X}, d)$ be a metric space and let $A \subseteq \mathcal{X}$ be a non-empty set. For each $x \in \mathcal{X}$, $\mathrm{dist}(x, A) = 0$ if and only if $x \in \overline{A}$ (the set of all points that are limits of sequences of points from A). Consequently $\mathrm{dist}(x, A) = \mathrm{dist}(x, \overline{A})$ for all $x \in \mathcal{X}$.*

Proof. First suppose $x \in X$ is such that $\mathrm{dist}(x, A) = 0$. Therefore for all $n \in \mathbb{N}$ there exists an $a_n \in A$ such that $d(x, a_n) < \frac{1}{n}$. Hence $x = \lim_{n\to\infty} a_n$ so $x \in \overline{A}$.

Conversely, suppose $x \in \overline{A}$. Hence there exists a sequence $(a_n)_{n\geq 1}$ of elements of A such that $x = \lim_{n\to\infty} a_n$. Thus $\lim_{n\to\infty} d(x, a_n) = 0$ so $\mathrm{dist}(x, A) = 0$.

For the second part, note clearly $A \subseteq \overline{A}$ so $\mathrm{dist}(x, \overline{A}) \leq \mathrm{dist}(x, A)$ for all $x \in \mathcal{X}$. To see the other inequality, fix $x \in \mathcal{X}$. Let $\epsilon > 0$ be arbitrary. By the definition of the distance, there exists an $y \in \overline{A}$ such that $d(x, y) \leq \mathrm{dist}(x, \overline{A}) + \epsilon$. However, since $y \in \overline{A}$ there exists an $a \in A$ such that $d(y, a) < \epsilon$. Hence

$$d(x, a) \leq d(x, y) + d(y, a) \leq \mathrm{dist}(x, \overline{A}) + 2\epsilon.$$

Hence, as $a \in A$,

$$\mathrm{dist}(x, A) \leq \mathrm{dist}(x, \overline{A}) + 2\epsilon.$$

Therefore, as $\epsilon > 0$ was arbitrary, $\mathrm{dist}(x, A) \leq \mathrm{dist}(x, \overline{A})$ thereby completing the proof. ∎

Next we demonstrate the continuity of the distance function to a set. In particular, by applying the following to the examples contained in Example C.3.11, we obtain that the distance to a point in any metric space and the norm in any normed linear space are continuous functions.

Theorem C.3.13. *Let $(\mathcal{X}, d)$ be a metric space and let $A \subseteq \mathcal{X}$ be a non-empty set. The function $F : \mathcal{X} \to \mathbb{R}$ defined by $F(x) = \mathrm{dist}(x, A)$ for all $x \in X$ is continuous.*

Proof. To see that F is continuous, let $x, y \in \mathcal{X}$ be arbitrary. By the definition of the distance function, given any $\delta > 0$ there exists an $a \in A$ such that $d(x, a) \leq \mathrm{dist}(x, A) + \delta$. Therefore

$$\mathrm{dist}(y, A) \leq \mathrm{dist}(y, a) \leq d(x, y) + d(x, a) \leq d(x, y) + \mathrm{dist}(x, A) + \delta.$$

As the above inequality holds for all $\delta > 0$, we obtain that $F(y) \leq F(x) + d(x, y)$. By reversing the roles of x and y, we obtain that $F(x) \leq F(y) + d(x, y)$ and hence $|F(x) - F(y)| \leq d(x, y)$.

To see now that F is continuous, fix $x_0 \in \mathcal{X}$ and let $\epsilon > 0$. Let $\delta = \epsilon > 0$. Therefore, if $y \in \mathcal{X}$ is such that $d(x_0, y) < \delta$ then $|F(x_0) - F(y)| \leq d(x_0, y) < \delta = \epsilon$. Hence F is continuous at x_0. Therefore, as x_0 was arbitrary, F is continuous as desired. ∎

To conclude this section, we note the following result that 'separates' closed subsets of a metric space using continuous functions.

Theorem C.3.14 (Urysohn's Lemma). *Let $(\mathcal{X}, d)$ be a metric space and let B and C be two non-empty disjoint closed subsets of $\mathcal{X}$. There exists a function $f : \mathcal{X} \to [0, 1]$ such that $f(x) = 0$ if $x \in B$, $f(x) = 1$ if $x \in C$, and $0 < f(x) < 1$ if $x \notin B \cup C$.*

Proof. Consider the function $f : \mathcal{X} \to \mathbb{R}$ defined by

$$f(x) = \frac{\text{dist}(x, B)}{\text{dist}(x, B) + \text{dist}(x, C)}.$$

for all $x \in \mathcal{X}$. We claim that f is well-defined; that is, the denominator never vanishes. To see this, suppose there exists an $x \in \mathcal{X}$ such that $\text{dist}(x, B) + \text{dist}(x, C) = 0$. Thus $\text{dist}(x, B) = \text{dist}(x, C) = 0$ so by Lemma C.3.12, $x \in \overline{B} = B$ and $x \in \overline{C} = C$ as B and C are closed. Therefore, as $B \cap C = \emptyset$, we have obtained a contradiction. Hence f is well-defined.

Clearly $f(x) \geq 0$ for all $x \in \mathcal{X}$. Since

$$0 \leq \text{dist}(x, B) \leq \text{dist}(x, B) + \text{dist}(x, C)$$

we see that $f : \mathcal{X} \to [0, 1]$. Furthermore, by Theorem C.3.13 and elementary properties of continuous functions, f is continuous (i.e. use part (2) of Theorem C.3.6 together with Proposition C.2.27 to show the sum of continuous functions is continuous).

To complete the proof, first notice that $f(x) = 0$ if and only if $\text{dist}(x, B) = 0$ if and only if $x \in \overline{B} = B$ by Lemma C.3.12. Similarly $f(x) = 1$ if and only if $\text{dist}(x, B) = \text{dist}(x, B) + \text{dist}(x, C)$ if and only if $\text{dist}(x, C) = 0$ if and only if $x \in \overline{C} = C$ by Lemma C.3.12. Since $f : \mathcal{X} \to [0, 1]$, we obtain that $0 < f(x) < 1$ for all $x \notin B \cup C$ thereby completing the proof. ∎

C.3.3 Metric Spaces of Continuous Functions

Unfortunately, the set of continuous functions between two metric spaces need not be a 'nice' metric space. Of course we may place the discrete metric on any set, but for continuous functions we would like a non-trivial metric such that the distance between two functions is related to the pointwise distance between the functions. To do this, we will need to restrict the collection of continuous functions. To do so, we define the following.

Definition C.3.15. Let $(\mathcal{X}, d)$ be a metric space and let $A \subseteq \mathcal{X}$. It is said that A is *bounded* if there exists an $x \in \mathcal{X}$ such that

$$\sup\{d(x, a) \mid a \in A\} < \infty.$$

Remark C.3.16. Since for all $y \in \mathcal{X}$ we have

$$d(y,a) \leq d(y,x) + d(x,a),$$

the choice of x does not matter in Definition C.3.15. Hence, if $\mathcal{X}$ is a normed linear space, we may choose $x = \vec{0}$ to obtain that A is bounded if and only if

$$\sup\{\|a\|_{\mathcal{X}} \mid a \in A\} < \infty.$$

Definition C.3.17. Let $(\mathcal{X}, d_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}})$ be metric spaces. A function $f : \mathcal{X} \to \mathcal{Y}$ is said to be *bounded* if $f(\mathcal{X})$ is a bounded set in $\mathcal{Y}$. The set of all bounded continuous functions $f : \mathcal{X} \to \mathcal{Y}$ is denoted $\mathcal{C}_b(\mathcal{X}, \mathcal{Y})$.

Example C.3.18. If $\mathcal{X} = \mathbb{N}$ and $\mathcal{Y} = \mathbb{K}$, then $\mathcal{C}_b(\mathcal{X}, \mathcal{Y}) = \ell_\infty(\mathbb{N}, \mathbb{K})$.

Theorem C.3.19. *Let $(\mathcal{X}, d_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}})$ be metric spaces. Then $\mathcal{C}_b(\mathcal{X}, \mathcal{Y})$ is a metric space with the metric*

$$d(f,g) = \sup\{d_{\mathcal{Y}}(f(x), g(x)) \mid x \in \mathcal{X}\}.$$

Proof. First, given $f, g \in \mathcal{C}_b(\mathcal{X}, \mathcal{Y})$, to see that $d(f,g) < \infty$, we note there exists an $a \in y$ such that

$$\sup\{d_{\mathcal{Y}}(f(x), a) \mid x \in \mathcal{X}\} < \infty \quad \text{and} \quad \sup\{d_{\mathcal{Y}}(g(x), a) \mid x \in \mathcal{X}\} < \infty.$$

From this it clearly follows from the triangle inequality on $d_{\mathcal{Y}}$ that $d(f,g) < \infty$. The remaining properties of a metric are trivial to verify. ∎

Of course, with continuous functions on $\mathbb{R}$, the sum of continuous functions is continuous and a scalar multiple of continuous functions is continuous. This means that continuous functions on $\mathbb{R}$ are a vector space. To repeat these ideas for $\mathcal{C}_b(\mathcal{X}, \mathcal{Y})$ is only possible if $\mathcal{Y}$ is a normed linear space. This yields the following thereby generalizing the sup norm on $C[a,b]$.

Theorem C.3.20. *Let $(\mathcal{X}, d_{\mathcal{X}})$ be a metric space and let $(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}})$ be a normed linear space over $\mathbb{K}$. Then $\mathcal{C}_b(\mathcal{X}, \mathcal{Y})$ is a normed linear space over $\mathbb{K}$ with the operations of pointwise addition and scalar multiplication, and the norm*

$$\|f\|_\infty = \sup\{\|f(x)\|_{\mathcal{Y}} \mid x \in \mathcal{X}\}.$$

The norm $\|\cdot\|_\infty$ is called the supremum norm *or the* infinity norm.

Proof. If $f, g : \mathcal{X} \to \mathcal{Y}$ are continuous functions, then one can verify that $f + g$ and αf are continuous for all $\alpha \in \mathbb{K}$ by using part (2) of Theorem C.3.6 together with Proposition C.2.27. If f and g are bounded, the properties of $\|\cdot\|_{\mathcal{Y}}$ easily imply that $f + g$ and αf are bounded. Hence $\mathcal{C}_b(\mathcal{X}, \mathcal{Y})$ is a vector space over $\mathbb{F}$. The fact that $\|\cdot\|_\infty$ is a norm easily follows (with the proof that it is finite following as in Theorem C.3.19). ∎

C.3.4 Continuous Linear Maps

To complete this section, we desire to analyze continuity in the context of normed linear spaces. In particular, the 'nice' maps between vector spaces are the linear maps as these are precisely the functions that preserve the vector space operations. Thus we desire to study when a linear map between normed linear spaces is continuous. To do this, as linear maps will clearly not be bounded as defined above, we modify the definition of boundedness.

Definition C.3.21. Let $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ and $(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}})$ be normed linear spaces over $\mathbb{K}$. A linear map $T : \mathcal{X} \to \mathcal{Y}$ is said to be *bounded* if

$$\sup\{\|T(\vec{x})\|_{\mathcal{Y}} \mid \vec{x} \in \mathcal{X}, \|\vec{x}\|_{\mathcal{X}} \leq 1\} < \infty.$$

If T is bounded, then we write

$$\|T\| = \sup\{\|T(\vec{x})\|_{\mathcal{Y}} \mid \vec{x} \in \mathcal{X}, \|\vec{x}\|_{\mathcal{X}} \leq 1\}.$$

The quantity $\|T\|$ is called the *operator norm* of T. Furthermore, the set of bounded linear maps from $\mathcal{X}$ to $\mathcal{Y}$ is denoted $\mathcal{B}(\mathcal{X}, \mathcal{Y})$.

To see that the operator norm is indeed a norm, we note that the only non-trivial property of Definition C.1.8 to verify is that if $\|T\| = 0$, then T is the zero linear map. Note the following lemma yields the result.

Lemma C.3.22. *Let $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ and $(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}})$ be normed linear spaces over $\mathbb{K}$ and let $T \in \mathcal{B}(\mathcal{X}, \mathcal{Y})$. Then*

$$\|T(\vec{x})\|_{\mathcal{Y}} \leq \|T\| \, \|\vec{x}\|_{\mathcal{X}}$$

for all $\vec{x} \in \mathcal{X}$.

Proof. Since $\left\|T(\vec{0})\right\|_{\mathcal{Y}} = \left\|\vec{0}\right\|_{\mathcal{X}} = 0$, the result holds when $\vec{x} = \vec{0}$. If $\vec{x} \neq \vec{0}$, then $\|\vec{x}\|_{\mathcal{X}} \neq 0$. Consequently, as

$$\left\|\frac{1}{\|\vec{x}\|_{\mathcal{X}}}\vec{x}\right\|_{\mathcal{X}} = \frac{1}{\|\vec{x}\|_{\mathcal{X}}} \|\vec{x}\|_{\mathcal{X}} = 1,$$

we obtain from the definition of the operator norm that

$$\frac{1}{\|\vec{x}\|_{\mathcal{X}}} \|T(\vec{x})\|_{\mathcal{Y}} = \left\|\frac{1}{\|\vec{x}\|_{\mathcal{X}}} T(\vec{x})\right\|_{\mathcal{Y}} = \left\|T\left(\frac{1}{\|\vec{x}\|_{\mathcal{X}}}\vec{x}\right)\right\|_{\mathcal{Y}} \leq \|T\|.$$

Therefore $\|T(\vec{x})\|_{\mathcal{Y}} \leq \|T\| \, \|\vec{x}\|_{\mathcal{X}}$ as desired. ∎

Theorem C.3.23. *Let $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ and $(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}})$ be normed linear spaces over $\mathbb{K}$. Then $\mathcal{B}(\mathcal{X}, \mathcal{Y})$ is a normed linear space over $\mathbb{K}$ with the operator norm as defined in Definition C.3.21.*

The reason we have been analyzing bounded linear maps in reference to continuous function is that $\mathcal{B}(\mathcal{X}, \mathcal{Y})$ is all continuous linear functions from $\mathcal{X}$ to $\mathcal{Y}$.

Theorem C.3.24. *Let $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ and $(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}})$ be normed linear spaces over $\mathbb{K}$ and let $T : \mathcal{X} \to \mathcal{Y}$ be linear. The following are equivalent:*

(1) T is continuous.

(2) T is continuous at 0.

(3) T is bounded.

Proof. Clearly (1) implies (2). To see that (2) implies (3), let $\epsilon = 1$. Since T is continuous at 0, there exists a $\delta > 0$ such that if $\|\vec{x}\| \leq \delta$ then $\|T(\vec{x})\| \leq 1$. Therefore, if $\vec{x} \in \mathcal{X}$ is such that $\|\vec{x}\| \leq 1$, then $\|\delta\vec{x}\| \leq \delta$ so

$$\delta \|T(\vec{x})\| = \|\delta T(\vec{x})\| = \|T(\delta\vec{x})\| \leq 1.$$

Hence $\|\vec{x}\| \leq 1$ implies $\|T(\vec{x})\| \leq \delta^{-1}$ so T is bounded by definition.

To see that (3) implies (1), let $\vec{x}_0 \in \mathcal{X}$ be arbitrary. To see that T is continuous at x, let $\epsilon > 0$. Let $\delta = \frac{\epsilon}{\|T\|+1} > 0$. If $\vec{x} \in \mathcal{X}$ is such that $\|\vec{x} - \vec{x}_0\| < \delta$, then Lemma C.3.22 implies that

$$\|T(\vec{x}) - T(\vec{x}_0)\| = \|T(\vec{x} - \vec{x}_0)\| \leq \|T\| \|\vec{x} - \vec{x}_0\| < \|T\| \frac{\epsilon}{\|T\| + 1} < \epsilon.$$

Therefore T is continuous at $\vec{x}_0$ as $\epsilon > 0$ was arbitrary. Therefore, as $\vec{x}_0 \in \mathcal{X}$ was arbitrary, T is continuous on $\mathcal{X}$. ∎

C.4 Cauchy Sequences

In order for a sequence to converge, given any $\epsilon > 0$ all the elements of the sequence must be within ϵ of their limit. In particular, this means that the terms in the sequence must eventually be within 2ϵ of each other. This leads us to the following concept previously seen for sequences in $\mathbb{R}$.

Definition C.4.1. Let $(\mathcal{X}, d)$ be a metric space. A sequence $(x_n)_{n\geq 1}$ is said to be *Cauchy* if for all $\epsilon > 0$ there exists an $N \in \mathbb{N}$ such that $d(x_n, x_m) < \epsilon$ for all $n, m \geq N$.

Remark C.4.2. There exists sequences $(x_n)_{n\geq 1}$ such that

$$\lim_{n\to\infty} d(x_n, x_{n+1}) = 0$$

that are not Cauchy. Indeed let $x_n = \sum_{k=1}^{n} \frac{1}{k}$ for all $n \in \mathbb{N}$. Clearly $d(x_n, x_{n+1}) = \frac{1}{n+1}$ yet $(x_n)_{n\geq 1}$ is not Cauchy as for all $m \in \mathbb{N}$

$$\sup_{m\to\infty} d(x_n, x_m) = \sup_{m\to\infty} \sum_{k=n}^{m} \frac{1}{k} = \infty.$$

There are immediately sequences we can deduce are not Cauchy.

Lemma C.4.3. *Every Cauchy Sequence in a metric space is bounded.*

Proof. Let $(x_n)_{n\geq 1}$ be a Cauchy sequence in a metric space $(\mathcal{X}, d)$. Since $(x_n)_{n\geq 1}$ is Cauchy, there exists an $N \in \mathbb{N}$ such that $d(x_n, x_m) < 1$ for all $n, m \geq N$.

Let $M = \max\{d(x_1, x_N), \ldots, d(x_{N-1}, x_N), 1\}$. Using the above paragraph, we see that $d(x_n, x_N) \leq M$ for all $n \in \mathbb{N}$. Hence $(x_n)_{n\geq 1}$ is bounded. ∎

Furthermore, we have already seen several examples of Cauchy sequences.

Lemma C.4.4. *Every convergent sequence in a metric space is Cauchy.*

Proof. Let $(x_n)_{n\geq 1}$ be a convergent sequence in a metric space $(\mathcal{X}, d)$. Let $x_0 = \lim_{n\to\infty} x_n$. To see that $(x_n)_{n\geq 1}$ is Cauchy, let $\epsilon > 0$ be arbitrary. Since $x_0 = \lim_{n\to\infty} x_n$, there exists an $N \in \mathbb{N}$ such that $d(x_n, x_0) < \frac{\epsilon}{2}$ for all $n \geq N$. Therefore, for all $n, m \geq N$,

$$d(x_n, x_m) \leq d(x_n, x_0) + d(x_0, x_m) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Thus, as $\epsilon > 0$ was arbitrary, $(x_n)_{n\geq 1}$ is Cauchy by definition. ∎

Corollary C.4.5. *Every convergent sequence in a metric spaces is bounded.*

Of course, it would be nice if the converse Lemma C.4.4 were true as this would enable us to deduce the convergence of a sequence by checking it is Cauchy without any knowledge of the limit. Thus we make the following definition.

Definition C.4.6. A metric space $(\mathcal{X}, d)$ is said to be *complete* if every Cauchy sequence converges.

Any metric space with the discrete metric is complete as any Cauchy sequence with respect to the discrete metric is eventually constant. Furthermore $\mathbb{R}$ is complete. We will quickly recall the proof that $\mathbb{R}$ is complete by beginning with the following result which holds in any metric space.

Lemma C.4.7. *Let $(x_n)_{n\geq 1}$ be a Cauchy sequence in a metric space $(\mathcal{X}, d)$. If a subsequence of $(x_n)_{n\geq 1}$ converges, then $(x_n)_{n\geq 1}$ converges.*

Proof. Let $(x_n)_{n\geq 1}$ be a Cauchy sequence with a convergent subsequence $(x_{k_n})_{n\geq 1}$ and let $x_0 = \lim_{n\to\infty} x_{k_n}$. We claim that $\lim_{n\to\infty} x_n = x_0$. To see this, let $\epsilon > 0$ be arbitrary. Since $(x_n)_{n\geq 1}$ is Cauchy, there exists an $N \in \mathbb{N}$ such that $d(x_n, x_m) < \frac{\epsilon}{2}$ for all $n, m \geq N$. Furthermore, since $x_0 = \lim_{n\to\infty} x_{k_n}$, there exists an $k_j \geq N$ such that $d(x_{k_j}, x_0) < \frac{\epsilon}{2}$. Hence, if $n \geq N$, then

$$d(x_n, x_0) \leq d(x_n, x_{k_j}) + d(x_{k_j}, x_0) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Thus, as $\epsilon > 0$ was arbitrary, $(x_n)_{n\geq 1}$ is converges to x_0 by definition. ∎

In addition, recall the following theorem.

Theorem C.4.8 (Bolzano-Weierstrass Theorem). *Every bounded sequence of real numbers has a convergent sequence.*

Theorem C.4.9 (Completeness of the Real Numbers). *Every Cauchy sequence of real numbers converges.*

Proof. Let $(x_n)_{n\geq 1}$ be a Cauchy sequence of real numbers. Thus $(x_n)_{n\geq 1}$ is bounded by Lemma C.4.3. Therefore $(x_n)_{n\geq 1}$ has a convergent sequence by the Bolzano-Weierstrass Theorem. Hence $(x_n)_{n\geq 1}$ converges by Lemma C.4.7. ∎

For other examples of complete metric spaces, we turn to the following.

Corollary C.4.10. *For every $p \in [1, \infty]$ and $n \in \mathbb{N}$, $(\mathbb{K}^n, \|\cdot\|_p)$ is complete.*

Proof. To see that $(\mathbb{R}^n, \|\cdot\|_p)$ is complete, let $(\vec{x}_k)_{k\geq 1}$ be an arbitrary Cauchy sequence in $(\mathbb{R}^n, \|\cdot\|_p)$. Write $\vec{x}_k = (x_{k,1}, \ldots, x_{k,n})$. Since for all $k, m \in \mathbb{N}$ we have

$$|x_{k,j} - x_{m,j}| \leq \|\vec{x}_k - \vec{x}_m\|_p\,,$$

it is elementary to see that $(x_{k,j})_{k\geq 1}$ is a Cauchy sequence in $\mathbb{R}$ for all $j \in \{1, \ldots, n\}$. Since $\mathbb{R}$ is complete, for each $j \in \{1, \ldots, n\}$ there exists an $x_j \in \mathbb{R}$ such that $x_j = \lim_{k\to\infty} x_{k,j}$. If $\vec{x} = (x_1, \ldots, x_n)$, then $\vec{x} = \lim_{k\to\infty} \vec{x}_k$ in $(\mathbb{R}^n, \|\cdot\|_p)$ by Example C.2.24. Therefore, as $(\vec{x}_k)_{k\geq 1}$ was arbitrary, $(\mathbb{R}^n, \|\cdot\|_p)$ is complete.

To see that $(\mathbb{C}^n, \|\cdot\|_p)$, it suffices by the same arguments to show that $(\mathbb{C}, |\cdot|)$ is complete. To see that $(\mathbb{C}, |\cdot|)$ is complete, let $(z_k)_{k\geq 1}$ be an arbitrary Cauchy sequence in $\mathbb{C}$. For each k, write $z_k = a_k + ib_k$ where $a_k, b_k \in \mathbb{R}$. Since for all $k, m \in \mathbb{N}$ we have

$$|a_k - a_m|, |b_k - b_m| \leq |z_k - z_m|,$$

it is elementary to see that $(a_k)_{k\geq 1}$ and $(b_k)_{k\geq 1}$ are Cauchy sequences in $\mathbb{R}$. Since $\mathbb{R}$ is complete, $a = \lim_{k\to\infty} a_k$ and $b = \lim_{k\to\infty} b_k$ exist. Hence $z = a + bi$, then $z = \lim_{k\to\infty} z_k$ by Example C.2.25. Hence, as $(z_k)_{k\geq 1}$ was arbitrary, $(\mathbb{C}, |\cdot|)$ is complete. ∎

Using our knowledge of complete metric spaces, we can construct additional examples.

Theorem C.4.11. *Let $(\mathcal{X}, d)$ be a complete metric space and let $A \subseteq \mathcal{X}$ be non-empty. Then (A, d) is complete if and only if A is closed in $\mathcal{X}$.*

Proof. Suppose (A, d) is complete. To see that A is closed, let $(a_n)_{n\geq 1}$ be an arbitrary sequence of elements from A that converges to some element $x \in \mathcal{X}$. Since $(a_n)_{n\geq 1}$ converges in $\mathcal{X}$, $(a_n)_{n\geq 1}$ is Cauchy in $\mathcal{X}$ by Lemma

C.4.4 and therefore is Cauchy in (A, d). Hence $(a_n)_{n\geq 1}$ converges in A to some element $a \in A$ as (A, d) is complete. Since limits in metric spaces are unique (Proposition C.2.29), $a = x$. Hence $x \in A$ so A is closed by Theorem C.2.32.

For the converse, suppose A is closed in $\mathcal{X}$. To see that (A, d) is complete, let $(a_n)_{n\geq 1}$ be an arbitrary Cauchy sequence in (A, d). Hence $(a_n)_{n\geq 1}$ is a Cauchy sequence in $(\mathcal{X}, d)$. Since $(\mathcal{X}, d)$ is complete, $(a_n)_{n\geq 1}$ converges to some element $x \in \mathcal{X}$. Since A is closed in $\mathcal{X}$, Theorem C.2.32 implies that $x \in A$. Hence as $(a_n)_{n\geq 1}$ was an arbitrary Cauchy sequence, (A, d) is complete. ∎

Corollary C.4.12. *Every closed subset of* $\mathbb{K}^n$ *is a complete metric space.*

Notice that one direction of the proof of Theorem C.4.11 did not require $(\mathcal{X}, d)$ to be complete. Thus we obtain the following.

Corollary C.4.13. *Let* $(\mathcal{X}, d)$ *be a complete metric space and let* $A \subseteq \mathcal{X}$ *be non-empty. If* (A, d) *is complete, then* A *is closed in* $\mathcal{X}$*.*

C.5 Banach Spaces

The above produced several examples of complete metric spaces including many that were not normed linear spaces. As complete normed linear spaces are incredibly nice and important for the remainder of the course, and as saying/typing complete normed linear spaces is rather cumbersome, we make the following definition.

Definition C.5.1. A *Banach space* is a complete normed linear space.

Corollary C.4.10 produced for us a collection of Banach spaces. For the remainder of this subsection, we will note several of the normed linear spaces we have seen previously are Banach spaces. Furthermore, via Theorem C.4.11, we obtain any closed vector subspace of these Banach spaces is also a Banach space (and any closed subset is a complete metric space).

As we go through the following, note there is a similar theme to the proofs.

Proposition C.5.2. *For each* $p \in [1, \infty]$, $(\ell_p(\mathbb{N}, \mathbb{K}), \|\cdot\|_p)$ *is a Banach space.*

Proof. Note the proof of this proposition is very similar to that of Proposition C.2.29 except for the complication that convergences entrywise need not imply convergence in $(\ell_p(\mathbb{N}, \mathbb{K}), \|\cdot\|_p)$. To bi-pass this problem, we will invoke a technique that will be used repeatedly in this section.

Fix $p \in [1, \infty]$ and let $(\vec{x}_n)_{n\geq 1}$ be an arbitrary Cauchy sequence in $(\ell_p(\mathbb{N}, \mathbb{K}), \|\cdot\|_p)$. For each $n \in \mathbb{N}$, write $\vec{x}_n = (x_{n,k})_{k\geq 1}$. Since for all $m, j, k \in \mathbb{N}$,

$$|x_{m,k} - x_{j,k}| \leq \|\vec{x}_m - \vec{x}_j\|_p \,,$$

we see that for each $k \in \mathbb{N}$ the sequence $(x_{n,k})_{n\geq 1}$ is Cauchy in $(\mathbb{K}, |\cdot|)$. Therefore, as $(\mathbb{K}, |\cdot|)$ is complete, $y_k = \lim_{n\to\infty} x_{n,k}$ exists in $(\mathbb{K}, |\cdot|)$ for each $k \in \mathbb{N}$.

Let $\vec{y} = (y_n)_{n\geq 1}$. To complete the proof, it suffices to verify two things: that $\vec{y} \in \ell_p(\mathbb{N}, \mathbb{K})$, and that $\lim_{n\to\infty} \|\vec{y} - \vec{x}_n\|_p = 0$. We will only discuss the case $p \neq \infty$ and the case $p = \infty$ is similar. For $p \neq \infty$ notice for all $m \in \mathbb{N}$ that

$$\left(\sum_{k=1}^{m} |y_k - x_{1,k}|^p\right)^{\frac{1}{p}} = \lim_{n\to\infty} \left(\sum_{k=1}^{m} |x_{n,k} - x_{1,k}|^p\right)^{\frac{1}{p}} \leq \limsup_{n\to\infty} \|\vec{x}_n - \vec{x}_1\|_p .$$

Since $(\vec{x}_n)_{n\geq 1}$ is Cauchy in $(\ell_p(\mathbb{N}, \mathbb{K}), \|\cdot\|_p)$, $(\vec{x}_n)_{n\geq 1}$ is bounded in $(\ell_p(\mathbb{N}, \mathbb{K}), \|\cdot\|_p)$ since Cauchy sequences are bounded. Hence $\limsup_{n\to\infty} \|\vec{x}_n - \vec{x}_1\|_p$ is finite. Therefore, by taking the limit as m tends to infinity, we obtain that

$$\left(\sum_{k=1}^{\infty} |y_k - x_{1,k}|^p\right)^{\frac{1}{p}} \leq \limsup_{n\to\infty} \|\vec{x}_n - \vec{x}_1\|_p .$$

Hence $\vec{z} = (y_k - x_{1,k})_{k\geq 1} \in \ell_p(\mathbb{N}, \mathbb{K})$. Therefore, as $\vec{y} = \vec{z} + \vec{x}_1$, we obtain that $\vec{y} \in \ell_p(\mathbb{N}, \mathbb{K})$ by the triangle inequality.

To see that $\lim_{n\to\infty} \|\vec{y} - \vec{x}_n\|_p = 0$, let $\epsilon > 0$ be arbitrary. Note the above proof also shows for all $j \in \mathbb{N}$ that

$$\|\vec{y} - \vec{x}_j\|_p \leq \limsup_{n\to\infty} \|\vec{x}_n - \vec{x}_j\|_p .$$

Since $(\vec{x}_n)_{n\geq 1}$ is Cauchy in $(\ell_p(\mathbb{N}, \mathbb{K}), \|\cdot\|_p)$, there exists an $N \in \mathbb{N}$ such that $\|\vec{x}_m - \vec{x}_j\|_p \leq \epsilon$ for all $m, j \geq N$. Hence if $j \geq N$, the above implies $\|\vec{y} - \vec{x}_j\|_p \leq \epsilon$. Therefore, as $\epsilon > 0$ was arbitrary, we obtain that $\lim_{n\to\infty} \|\vec{y} - \vec{x}_n\|_p = 0$. Hence $(\vec{x}_n)_{n\geq 1}$ converges in $(\ell_p(\mathbb{N}, \mathbb{K}), \|\cdot\|_p)$ so, as $(\vec{x}_n)_{n\geq 1}$ was arbitrary, $(\ell_p(\mathbb{N}, \mathbb{K}), \|\cdot\|_p)$ is complete. ∎

To discuss Banach spaces consisting of functions, we first note the following types of convergence and a lemma which guarantees certain limits are continuous. This lemma is the generalization to metric spaces of a result that is a cornerstone of any first course in analysis.

Definition C.5.3. Let $(\mathcal{X}, d_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}})$ be metric spaces. For each $n \in \mathbb{N}$ let $f_n : \mathcal{X} \to \mathcal{Y}$. Given $f : \mathcal{X} \to \mathcal{Y}$, it is said that the sequence $(f_n)_{n\geq 1}$

- *converges pointwise* to f if $\lim_{n\to\infty} f_n(x) = f(x)$ for all $x \in \mathcal{X}$.
- *converges uniformly* to f if $(f_n)_{n\geq 1}$ converges to f with respect to the uniform metric (provided it makes sense); that is, for all $\epsilon > 0$ there exists an $N \in \mathbb{N}$ such that $d_{\mathcal{Y}}(f(x), f_n(x)) < \epsilon$ for all $n \geq N$ and for all $x \in \mathcal{X}$.

Theorem C.5.4. *Let $(\mathcal{X}, d_\mathcal{X})$ and $(\mathcal{Y}, d_\mathcal{Y})$ be metric spaces and let $f : \mathcal{X} \to \mathcal{Y}$. If $(f_n)_{n\geq 1}$ is a sequence of continuous functions from $\mathcal{X}$ to $\mathcal{Y}$ that converge to f uniformly, then f is continuous.*

Proof. To see that f is continuous, let $x_0 \in \mathcal{X}$ be arbitrary. To see that f is continuous at x_0 let $\epsilon > 0$ be arbitrary. Since $(f_n)_{n\geq 1}$ converges to f uniformly, there exists an $N \in \mathbb{N}$ such that $d_\mathcal{Y}(f(x), f_N(x)) < \frac{\epsilon}{3}$ for all $x \in \mathcal{X}$. Since f_N is continuous at x_0, there exists a $\delta > 0$ such that if $d_\mathcal{X}(x, x_0) < \delta$ then $d_\mathcal{Y}(f_N(x), f_N(x_0)) < \frac{\epsilon}{3}$. Hence if $x \in \mathcal{X}$ and $d_\mathcal{X}(x, x_0) < \delta$, then, by the triangle inequality,

$$\begin{aligned} d_\mathcal{Y}(f(x), f(x_0)) &\leq d_\mathcal{Y}(f(x), f_N(x)) + d_\mathcal{Y}(f_N(x), f_N(x_0)) + d_\mathcal{Y}(f_N(x_0), f(x_0)) \\ &< \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon. \end{aligned}$$

Hence, as $\epsilon > 0$ was arbitrary, f is continuous at x_0. Thus, as x_0 was arbitrary, f is continuous on $\mathcal{X}$. ∎

Using the above, we obtain the following result for metric spaces.

Theorem C.5.5. *Let $(\mathcal{X}, d_\mathcal{X})$ and $(\mathcal{Y}, d_\mathcal{Y})$ be metric spaces. If $\mathcal{Y}$ is complete, then $(\mathcal{C}_b(\mathcal{X}, \mathcal{Y}), d_\infty)$ is a complete metric space.*

Proof. Let $(f_n)_{n\geq 1}$ be an arbitrary Cauchy sequence in $(\mathcal{C}_b(\mathcal{X}, \mathcal{Y}), d_\infty)$. For each $x \in \mathcal{X}$, notice

$$d_\mathcal{Y}(f_n(x), f_m(x)) \leq d_\infty(f_n, f_m)$$

for all $n, m \in \mathbb{N}$. Hence it is elementary to see that $(f_n(x))_{n\geq 1}$ is a Cauchy sequence in $\mathcal{Y}$ for all $x \in \mathcal{X}$. Therefore, since $\mathcal{Y}$ is complete, for each $x \in \mathcal{X}$ there exists an $f(x) \in \mathcal{Y}$ such that $f(x) = \lim_{n\to\infty} f_n(x)$. Clearly the function $x \mapsto f(x)$ defines a function $f : \mathcal{X} \to \mathcal{Y}$.

To complete the proof, it suffices to verify three things: that $f : \mathcal{X} \to \mathcal{Y}$ is continuous, that f is bounded, and that $\lim_{n\to\infty} d_\infty(f, f_n) = 0$. For the first, we claim that $(f_n)_{n\geq 1}$ converges to f uniformly on $\mathcal{X}$. To see this, first notice for all $x \in \mathcal{X}$ and $m \in \mathbb{N}$ that

$$d_\mathcal{Y}(f(x), f_m(x)) = \lim_{n\to\infty} d_\mathcal{Y}(f_n(x), f_m(x)) \leq \limsup_{n\to\infty} d_\infty(f_n, f_m).$$

Since $(f_n)_{n\geq 1}$ is Cauchy in $(\mathcal{C}_b(\mathcal{X}, \mathcal{Y}), d_\infty)$, $(f_n)_{n\geq 1}$ is bounded in $(\mathcal{C}_b(\mathcal{X}, \mathcal{Y}), d_\infty)$ since Cauchy sequences are bounded. Hence $\limsup_{n\to\infty} d_\infty(f_n, f_m)$ is finite. Therefore, by taking the supremum over all $x \in \mathcal{X}$, we obtain that

$$\sup\{d_\mathcal{Y}(f(x), f_m(x)) \mid x \in \mathcal{X}\} \leq \limsup_{n\to\infty} d_\infty(f_n, f_m)$$

for all $m \in \mathbb{N}$. Thus, by taking $m = 1$ and using the fact that f_1 is bounded, we easily see that f is bounded.

To see that f is continuous, we will show that $(f_n)_{n\geq 1}$ converges uniformly to f using the above. Thus let $\epsilon > 0$ be arbitrary. Since $(f_n)_{n\geq 1}$ is Cauchy in $(\mathcal{C}_b(\mathcal{X}, \mathcal{Y}), d_\infty)$, there exists an $N \in \mathbb{N}$ such that $d_\infty(f_j, f_m) \leq \epsilon$ for all $m, j \geq N$. Hence if $m \geq N$, the above implies

$$\sup\{d_{\mathcal{Y}}(f(x), f_m(x)) \mid x \in \mathcal{X}\} < \epsilon.$$

Thus $(f_n)_{n\geq 1}$ converges to f uniformly on $\mathcal{X}$. Hence f is continuous by Theorem C.5.4.

As the above shows that $\lim_{m\to\infty} d_\infty(f, f_m) = 0$, $(f_n)_{n\geq 1}$ converges to f in $(\mathcal{C}_b(\mathcal{X}, \mathcal{Y}), d_\infty)$. Thus, as $(f_n)_{n\geq 1}$ was an arbitrary Cauchy sequence, $(\mathcal{C}_b(\mathcal{X}, \mathcal{Y}), d_\infty)$ is complete. ∎

Since $\mathcal{C}_b(\mathcal{X}, \mathcal{Y})$ is a normed linear space provided $\mathcal{Y}$ is, we obtain the following.

Corollary C.5.6. *Let $(\mathcal{X}, d_{\mathcal{X}})$ be a metric space and let $(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}})$ be a Banach space. Then $(\mathcal{C}_b(\mathcal{X}, \mathcal{Y}), \|\cdot\|_\infty)$ is a Banach space.*

Corollary C.5.7. *Let $(\mathcal{X}, d_{\mathcal{X}})$ be a metric space. Then $(\mathcal{C}_b(\mathcal{X}, \mathbb{R}), \|\cdot\|_\infty)$ is a Banach space.*

Finally, returning to bounded linear maps between normed linear spaces, we obtain the following.

Theorem C.5.8. *Let $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ and $(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}})$ be normed linear spaces. If $\mathcal{Y}$ is a Banach space, then $(\mathcal{B}(\mathcal{X}, \mathcal{Y}), \|\cdot\|)$ is a Banach space (where $\|\cdot\|$ is the operator norm).*

Proof. Let $(T_n)_{n\geq 1}$ be an arbitrary Cauchy sequence in $(\mathcal{B}(\mathcal{X}, \mathcal{Y}), \|\cdot\|)$. For each $\vec{x} \in \mathcal{X}$, notice

$$\|T_n(\vec{x}) - T_m(\vec{x})\|_{\mathcal{Y}} \leq \|T_n - T_m\| \, \|\vec{x}\|_{\mathcal{X}}$$

for all $n, m \in \mathbb{N}$. Hence it is elementary to see that $(T_n(\vec{x}))_{n\geq 1}$ is a Cauchy sequence in $\mathcal{Y}$ for all $\vec{x} \in \mathcal{X}$. Therefore, since $\mathcal{Y}$ is complete, for each $\vec{x} \in \mathcal{X}$ there exists an $T(\vec{x}) \in \mathcal{Y}$ such that $T(\vec{x}) = \lim_{n\to\infty} T_n(\vec{x})$.

To complete the proof, it suffices to verify three things: that $T : \mathcal{X} \to \mathcal{Y}$ is linear, that T is bounded, and that $\lim_{n\to\infty} \|T - T_n\| = 0$. To see that T is linear, notice for all $\vec{x}_1, \vec{x}_2 \in \mathcal{X}$ and $\alpha \in \mathbb{K}$ that

$$T(\alpha\vec{x}_1 + \vec{x}_2) = \lim_{n\to\infty} T_n(\alpha\vec{x}_1 + \vec{x}_2) = \lim_{n\to\infty} \alpha T_n(\vec{x}_1) + T)n(\vec{x}_2) = \alpha T(\vec{x}_1) + T(\vec{x}_2).$$

Hence T is linear.

To see that T is bounded, notice for all $\vec{x} \in \mathcal{X}$ with $\|\vec{x}\|_{\mathcal{X}} \leq 1$ and $m \in \mathbb{N}$ that

$$\|T(\vec{x}) - T_m(\vec{x})\|_{\mathcal{Y}} = \lim_{n\to\infty} \|T_n(\vec{x}) - T_m(\vec{x})\|_{\mathcal{Y}} \leq \limsup_{n\to\infty} \|T_n - T_m\|$$

Since $(T_n)_{n\geq 1}$ is Cauchy in $(\mathcal{B}(\mathcal{X},\mathcal{Y}),\|\cdot\|)$, $(T_n)_{n\geq 1}$ is bounded in $(\mathcal{B}(\mathcal{X},\mathcal{Y}),\|\cdot\|)$ since Cauchy sequences are bounded. Hence $\limsup_{n\to\infty}\|T_n - T_m\|$ is finite. In particular, we obtain that there exists a constant K such that

$$\|T(\vec{x})\|_{\mathcal{Y}} \leq \|T_1(\vec{x})\|_{\mathcal{Y}} + K \leq \|T_1\| + K$$

for all $\vec{x}\in\mathcal{X}$ with $\|\vec{x}\|_{\mathcal{X}} \leq 1$. Hence T is bounded with $\|T\| \leq \|T_1\| + K$.

To see that $\lim_{n\to\infty}\|T - T_n\| = 0$, let $\epsilon > 0$ be arbitrary. Since $(T_n)_{n\geq 1}$ is Cauchy in $(\mathcal{B}(\mathcal{X},\mathcal{Y}),\|\cdot\|)$, there exists an $N\in\mathbb{N}$ such that $\|T_m - T_j\| \leq \epsilon$ for all $m, j \geq N$. Hence if $j \geq N$, the above implies $\|T(\vec{x}) - T_j(\vec{x})\| \leq \epsilon$ for all $\vec{x}\in\mathcal{X}$ with $\|\vec{x}\|_{\mathcal{X}} \leq 1$. Therefore, as $\epsilon > 0$ was arbitrary, we obtain that $\lim_{n\to\infty}\|T - T_n\| = 0$. Hence $(T_n)_{n\geq 1}$ converges in $(\mathcal{B}(\mathcal{X},\mathcal{Y}),\|\cdot\|)$ so, as $(T_n)_{n\geq 1}$ was arbitrary, $(\mathcal{B}(\mathcal{X},\mathcal{Y}),\|\cdot\|)$ is complete. ∎

C.6 Absolute Summability

The above has painstakingly demonstrated that several of the space we naturally desire to consider are Banach spaces. Thus, as we have several Banach spaces and complete metric spaces, it is nice to determine what additional properties these spaces have beyond the convergence of all Cauchy sequences. In this section, we will analyze one of these properties.

One important property of the real numbers is the convergence of specific types of series. In particular, every 'absolutely summable' series converges. We can generalize these concepts to metric spaces as follows.

Definition C.6.1. Let $(\mathcal{X},\|\cdot\|)$ be a normed linear space. A series $\sum_{n=1}^{\infty}\vec{x}_n$ is said to be *summable* if the sequence of partial sums $(s_n)_{n\geq 1}$ converges (where $s_n = \sum_{k=1}^{n}\vec{x}_k$).

A series $\sum_{n=1}^{\infty}\vec{x}_n$ is said to be *absolutely summable* if $\sum_{n=1}^{\infty}\|\vec{x}_n\| < \infty$.

Theorem C.6.2. *Let $(\mathcal{X},\|\cdot\|)$ be a normed linear space. Then $\mathcal{X}$ is complete (i.e. a Banach space) if and only if every absolutely summable series is summable.*

Proof. Suppose $(\mathcal{X},\|\cdot\|)$ is complete. Let $\sum_{n=1}^{\infty}\vec{x}_n$ be an absolutely summable series. To see that $\sum_{n=1}^{\infty}\vec{x}_n$ is summable, let $\epsilon > 0$ be arbitrary. Since $\sum_{n=1}^{\infty}\|\vec{x}_n\| < \infty$, there exists an $N\in\mathbb{N}$ such that $\sum_{n=N}^{\infty}\|\vec{x}_n\| < \epsilon$. Therefore, if $k, m \geq N$ and, without loss of generality, $m \geq k$, then

$$\begin{aligned}\|s_m - s_k\| &= \left\|\sum_{n=1}^{m} \vec{x}_n - \sum_{n=1}^{k} \vec{x}_n\right\| \\ &= \left\|\sum_{n=k+1}^{m} \vec{x}_n\right\| \\ &\leq \sum_{n=k+1}^{m} \|\vec{x}_n\| \\ &\leq \sum_{n=N}^{\infty} \|\vec{x}_n\| < \epsilon.\end{aligned}$$

Therefore, as $\epsilon > 0$ was arbitrary, the sequence of partial sums $(s_n)_{n\geq 1}$ is Cauchy. Hence $(s_n)_{n\geq 1}$ converges as $\mathcal{X}$ is complete. Thus, as $\sum_{n=1}^{\infty} \vec{x}_n$ was arbitrary, every absolutely summable series in $\mathcal{X}$ is summable.

For the converse, suppose every absolutely summable sequence in $\mathcal{X}$ is summable. To see that $\mathcal{X}$ is complete, let $(\vec{x}_n)_{n\geq 1}$ be an arbitrary Cauchy sequence. Since $(\vec{x}_n)_{n\geq 1}$ is Cauchy, there exists an $n_1 \in \mathbb{N}$ such that $\|\vec{x}_m - \vec{x}_j\| < \frac{1}{2}$ for all $m, j \geq n_1$. Similarly, since $(\vec{x}_n)_{n\geq 1}$ is Cauchy, there exists an $n_2 \in \mathbb{N}$ such that $n_2 > n_1$ and $\|\vec{x}_m - \vec{x}_j\| < \frac{1}{2^2}$ for all $m, j \geq n_2$. By repeating the above process, for each $k \in \mathbb{N}$ there exists an $n_k \in \mathbb{N}$ such that $n_k < n_{k+1}$ for all k and $\|\vec{x}_m - \vec{x}_j\| < \frac{1}{2^k}$ for all $m, j \geq n_k$.

For each $k \in \mathbb{N}$ let $\vec{y}_k = \vec{x}_{n_{k+1}} - \vec{x}_{n_k}$. By the above paragraph, we see that

$$\sum_{k=1}^{\infty} \|\vec{y}_k\| \leq \sum_{k=1}^{\infty} \frac{1}{2^k} < \infty.$$

Hence $\sum_{k=1}^{\infty} \vec{y}_k$ is an absolutely summable series in $\mathcal{X}$. Therefore, by the assumptions on $\mathcal{X}$, $\sum_{k=1}^{\infty} \vec{y}_k$ is summable in $\mathcal{X}$.

Let $\vec{x} = \vec{x}_{n_1} + \sum_{k=1}^{\infty} \vec{y}_k$. We claim that $(\vec{x}_{n_k})_{k\geq 1}$ converges to $\vec{x}$. To see this, let $\epsilon > 0$ be arbitrary. Then there exists a $M \in \mathbb{N}$ such that if $m \geq M$ then

$$\left\|\sum_{k=1}^{\infty} \vec{y}_k - \sum_{k=1}^{m} \vec{y}_k\right\| < \epsilon.$$

Therefore, if $m \geq M$,

$$\begin{aligned}\|\vec{x} - \vec{x}_{n_{m+1}}\| &\leq \left\|\sum_{k=1}^{\infty} \vec{y}_k - \sum_{k=1}^{m} \vec{y}_k\right\| + \left\|\vec{x}_{n_1} - \vec{x}_{n_{m+1}} + \sum_{k=1}^{m} \vec{y}_k\right\| \\ &< \epsilon + \left\|\vec{x}_{n_1} - \vec{x}_{n_{m+1}} + \sum_{k=1}^{m} \vec{x}_{n_{k+1}} - \vec{x}_{n_k}\right\| \\ &= \epsilon.\end{aligned}$$

Therefore, as $\epsilon > 0$ was arbitrary, $(\vec{x}_{n_k})_{k\geq 1}$ converges to $\vec{x}$. Hence $(\vec{x}_n)_{n\geq 1}$ converges to $\vec{x}$ by Lemma C.4.7. Therefore, as $(\vec{x}_n)_{n\geq 1}$ was an arbitrary Cauchy sequence, $\mathcal{X}$ is complete. ∎

As an immediate corollary, we obtain the following result pertaining to convergence of series of continuous functions.

Corollary C.6.3 (Weierstrass M-Test). *Let $(\mathcal{X}, d)$ be a metric space and let $(f_n)_{n\geq 1}$ be a sequence of functions from $\mathcal{C}_b(\mathcal{X}, \mathbb{R})$. Suppose there exists an $M \in \mathbb{R}$ such that $\sum_{n=1}^{\infty} \|f_n\|_\infty < M$. Then $\sum_{n=1}^{\infty} f_n$ converges uniformly on $\mathcal{X}$ to a continuous function.*

Index