Analysis I and Analysis II

Lecture Notes

Michael Strayer

As of April 26, 2014.
Important Note

This document represents the notes I have taken during the lectures of Dr. Archil Gulisashvili for Analysis I and Analysis II during the Fall 2013 and Spring 2014 terms at Ohio University, respectively. As such, while I have done my best, I make no guarantees (or even implications) that everything contained herein is accurate or complete. While the material is obviously not my original work, this document is and as such I reserve the copyright to it. The intended purpose of this document is to help students (primarily myself) prepare for the course exams and the comprehensive exam. Anyone who wishes to use this document to study may do so without my permission, but please do not distribute it in any official capacity without my permission. To request permission or to recommend changes (or, especially, to point out errors), please email me at ms770213@ohio.edu.† The first twelve sections correspond to Analysis I and the remaining sections correspond to Analysis II.

†As of May 2014, you may email me at mcs80@live.unc.edu.
ACKNOWLEDGEMENTS

I would like to thank my friends who have given me the encouragement to continue writing this file. In particular, I would like to thank Rebin Muhammad, who has contributed many valuable insights and has willingly proofread any section I have asked of him, as well as letting me borrow his class notes from days I missed. Thank you, my friend.
## CONTENTS

Important Note i  
Acknowledgements ii  
Introduction v  
1. The Real Numbers and Completeness 1  
2. Countable and Uncountable Sets; Open, Closed, and Borel Sets 4  
   2.1. Countable and Uncountable Sets 4  
   2.2. Topological Set Concepts 5  
3. Sequences of Real Numbers 8  
4. Lebesgue Outer Measure 9  
5. Carathéodory Contraction (\(\sigma\)-algebra) 11  
6. Regularity and Continuity Properties of \(m\) 16  
   6.1. Regularity Properties 16  
   6.2. The Continuity Properties of Measure 17  
   6.3. The Borel-Cantelli Lemma 19  
7. Non-Measurable Sets 21  
8. The Cantor Set and the Devil’s Staircase 23  
   8.1. Construction and Properties of the Cantor Set 23  
   8.2. The Cantor-Lebesgue Function 24  
9. Lebesgue Measurable Functions 27  
10. Limits and Modes of Convergence 32  
11. Egorov’s Theorem, Simple Approximation, and Lusin’s Theorem 37  
   11.1. Egorov’s Theorem 37  
   11.2. Simple Approximation 38  
   11.3. Lusin’s Theorem and the Standard Representation of a Simple Function 40  
12. Lebesgue Integration 42  
   12.1. Simple Functions 43  
   12.2. Bounded Measurable Functions Supported on a Set of Finite Measure 45  
   12.3. Non-Negative Measurable Functions 50  
   12.4. Integrable Functions: \(L^1\) 51  
13. More Integration and Convergence Theorems 53  
   13.1. Convergence Theorems 53  
   13.2. More Properties of Integration and \(L^1\) Convergence 56  
14. \(n\)-Dimensional Lebesgue Measure 58  
   14.1. Topology of \(\mathbb{R}^n\) 58  
   14.2. Measure in \(\mathbb{R}^n\) 59  
   14.3. Continuity and Regularity properties for \(m_n\) 64  
15. Lebesgue Measurable Functions and Integration for \(\mathbb{R}^n\) 66  
   15.1. Measurable Functions 66  
   15.2. Lebesgue Integration 67  
   15.3. Convergence Theorems, Chebychev’s Inequality, and Modes of Convergence 69  
16. Differentiation 71
INTRODUCTION

This document is a reasonably self-contained treatment of Real Analysis, as presented at Ohio University. The purpose of this document is two-fold. My first and primary reason for writing this document is to help myself understand the course content to the extent that I would feel comfortable explaining it to others. In order to achieve this goal, I needed to make sure I carefully recreated every argument that Dr. Gulisashvili gave in lectures. Additionally, there were times when Dr. Gulisashvili left a piece or two of a proof to us as an exercise. In order to achieve a thorough understanding of the material, I have tried in most circumstances to “fill in the gaps” left by Dr. Gulisashvili.

The second purpose of this document is to help my friends study for the comprehensive exam in Analysis at Ohio University.† With this goal in mind, one may often find the arguments I present here to be annoyingly detailed. This simultaneously serves both of my goals for this document, as it ensures I personally understand the argument and allows anyone reading to see the fine details, if needed. It is my hope that this does not lead to a pattern of memorization while studying for the comprehensive exam, but rather encourages my friends to also work out the details for themselves independently of what they read in this document. I welcome recommendations to improve this document, especially of the form “There is an easier argument for...” or “You did not need to include this detail here because...” or even “There is an error here...” All such comments are welcome, and even encouraged.

†I would have this goal for myself as well, but as of March 2014, it was confirmed that I will be transferring to a different school starting in Fall 2014.
Analysis I

1. The Real Numbers and Completeness

Definition 1.1. The real numbers are an ordered field with the completeness axiom.

Definition 1.2. A field is a set (we concern ourselves with \( \mathbb{R} \)) with two operations, + and \( \cdot \), such that the following hold.

1. \( a + b = b + a \) (Commutativity)
2. \( (a + b) + c = a + (b + c) \) (Associativity)
3. There exists an element \( 0 \in \mathbb{R} \) such that \( a + 0 = a \) (Neutral Element)
4. For all \( a \in \mathbb{R} \), there exists an element \( b = -a \) such that \( a + (-a) = 0 \) (Additive Inverse)
5. \( ab = ba \) (Commutativity)
6. \( (ab)c = a(bc) \) (Associativity)
7. There exists an element \( 1 \in \mathbb{R} \) such that \( 1a = a \) (Neutral Element)
8. For all \( a \neq 0 \), there exists an element \( b = \frac{1}{a} \) such that \( a \cdot \frac{1}{a} = 1 \) (Multiplicative Inverse)
9. \( 1 \neq 0 \) (Non-Triviality)
10. \( a(b + c) = ab + ac \) (Distributivity)

Definition 1.3. Subtraction: \( a - b = a + (-b) \), and Division: \( \frac{a}{b} = a \cdot \frac{1}{b} \), where \( b \neq 0 \).

Axiomatically, there exists a special subset of \( \mathbb{R} \) called the set of positive numbers, denoted \( \mathbb{R}^+ \), such that

1. \( a, b \in \mathbb{R}^+ \) implies \( a + b \in \mathbb{R}^+ \), \( ab \in \mathbb{R}^+ \), and
2. For all \( a \in \mathbb{R} \), one and only one condition amont \( a \in \mathbb{R}^+ \), \( -a \in \mathbb{R}^+ \), or \( a = 0 \) is true.

We define an ordering by the following rule: \( a < b \) if and only if \( b - a \in \mathbb{R}^+ \). This allows us to define intervals:

\[ (a, b) = \{ x : a < x < b \} \]
\[ [a, b] = \{ x : a \leq x \leq b \} \]

Definition 1.4. For a subset \( A \subset \mathbb{R} \), we say \( A \) is bounded above if there is a \( b \in \mathbb{R} \) such that \( a \leq b \) for all \( a \in A \). Similarly, we say \( A \) is bounded below if there is a \( m \in \mathbb{R} \) such that \( m \leq a \) for all \( a \in A \). A set that is bounded above and below is simply called bounded.

Note 1.5. The completeness axiom states that any subset of \( \mathbb{R} \) that is bounded from above has a least upper bound.

Definition 1.6. The least upper bound (supremum) of \( A \subset \mathbb{R} \), denoted sup \( A \), is the number \( \alpha \) such that

1. \( \alpha \) is an upper bound for \( A \), and
2. For every upper bound \( b \) of \( A \), we have \( \alpha \leq b \).

Example 1.7. Let \( A = \{ a \in \mathbb{Q}^+ : a^2 < 2 \} \). Then \( A \subset \mathbb{Q} \), but there is not least upper bound for \( A \) in \( \mathbb{Q} \). Thus \( \mathbb{Q} \) is not complete.

Corollary 1.8 (to completeness axiom). If \( B \) is bounded from below, then there exists the greatest lower bound \( \beta \) such that
(1) $\beta$ is a lower bound for $B$, and 
(2) $c \leq \beta$ for any other lower bound $c$ of $B$.

**Proof.** The idea is to take negatives and use completeness axiom to derive this result. $\square$

**Definition 1.9.** We define the **absolute value** as follows:

$$|a| = \begin{cases} 
a & \text{if } a \geq 0 \\
-a & \text{if } a < 0. \end{cases}$$

**Note 1.10.** The absolute value has the following properties:

1. $|a| = 0$ if and only if $a = 0$,
2. $|\alpha a| = |\alpha| \cdot |a|$, and
3. $|a + b| \leq |a| + |b|$.

Properties (1), (2), and (3) combine to make a **normed space**. This gives us a distance metric defined by $d(a, b) = |b - a|$.

**Theorem 1.11** (The Archimedean Property of Real Numbers). Suppose $a, b > 0$. Then there exists a positive integer $n$ such that $b < na$.

**Proof.** We show $\frac{b}{a} < n$; that is, we show the existence of $n$. Suppose to the contrary that no integer $n$ exists. Then $n \leq \frac{b}{a}$ for all $n \in \mathbb{N}$. Then it follows that $\mathbb{N}$ is bounded from above by $\frac{b}{a}$. Hence, by the completeness axiom, let $c = \sup\{\mathbb{N}\}$. Consider $c - 1$, which is not an upper bound for $\mathbb{N}$. Thus there exists an integer $m \in \mathbb{N}$ such that $c - 1 < m$, so $c < m + 1 \in \mathbb{N}$, a contradiction. $\square$

**Theorem 1.12.** Let $E \subset \mathbb{N}$. Then $E$ has a smallest element.

**Proof.** We know $E$ is bounded below by 1 since $\mathbb{N}$ is bounded below by 1. Hence $c = \inf\{E\}$ exists. Then $c + 1$ is not a lower bound for $E$, so there exists an integer $m \in E$ such that $m < c + 1$. Let us prove that $m$ is the smallest element in $E$. If $m$ is not the smallest element of $E$, then there exists an $n \in E$ such that $n < m$. Note that $c \leq n < m < c + 1$, which implies that $m - n < 1$ a contradiction since $m$ and $n$ are distinct integers. $\square$

**Theorem 1.13.** $\mathbb{Q}$ is dense in $\mathbb{R}$.

**Proof.** Our goal is to find $m, n \in \mathbb{Z}$ such that $a < \frac{m}{n} < b$, or equivalently, $na < m < nb$, where $a, b \in \mathbb{R}$ and $a < b$. Note that $b - a > 0$. By the Archimedean Property, there exists an $n \in \mathbb{Z}$ such that $n(b - a) > 1$. That is,

$$nb > 1 + na. \quad (1)$$

Also by the Archimedean Property, there exists an $m_1$ such that $na < m_1$ (using positive numbers $na$ and 1 in the hypotheses of the A.P.). Now $0 < na < m_1$. Consider

$$S = \{k \in \mathbb{N} : na < k\},$$

which is non-empty since $[m_1] \in S$. We also know that $S$ is bounded from below since $0 < s$ for all $s \in S$. Thus let $m = \inf\{S\} \in \mathbb{N}$. Then $m - 1 \leq na < m$, or

$$m < na + 1 < m + 1. \quad (2)$$
Combining (1) and (2) we get

\[ na < m \leq 1 + na < nb, \]

so \( na < m < nb, \) or \( a < \frac{m}{n} < b, \) establishing the result. \( \square \)
2. Countable and Uncountable Sets; Open, Closed, and Borel Sets

2.1. Countable and Uncountable Sets.

Definition 2.1. We make several definitions for the set $S$:

1. We say $S$ is **finite** if there is a one-to-one correspondence between $S$ and $\{1, \ldots, n\}$.
2. We say $S$ is **countably infinite** if there is a one-to-one correspondence between $S$ and $\mathbb{N}$.
3. We say $S$ is **countable** if $S$ is either finite or countably infinite.
4. We say $S$ is **uncountable** if it is not countable.

Note 2.2. We formulate two ideas to determine the countability of a set.

1. Try to find a counting procedure.
   
   Example 2.3. Let $S = \mathbb{N} \times \mathbb{N} \times \cdots \times \mathbb{N} = \mathbb{N}^n$. Show that $S$ is countably infinite.
   
   Proof. We only do it for $\mathbb{N} \times \mathbb{N}$. We do the array trick where we put every point of $\mathbb{N} \times \mathbb{N}$ on a grid and draw diagonal lines originating at the point $(0,0)$. See notebook for the picture. □

2. We also have the following result.

   Theorem 2.4. Any subset of a countable set is countable.
   
   Proof. Let $M$ be a countable set.
   
   Case 1: $M$ is finite. Then $M \longleftrightarrow \{1, \ldots, n\}$. Let $S \subset M$. Let $f : \{1, \ldots, n\} \to M$ describe this correspondence. Define $g$ as follows:
   
   \[ g(1) = \min\{j : f(j) \in S\}, \quad g(2) = \min\{j : f(j) \in S, j \neq g(1)\}, \]
   
   and so on. This provides the numbering and by simple induction, shows $S$ is finite.
   
   Case 2: Suppose $M$ is countably infinite so that there is a bijection $f : \mathbb{N} \to M$.
   
   Let $S \subset M$. Define $g$ similarly as above. If the numbering procedure stops, then $S$ is finite. Suppose it never stops. Then $g : \mathbb{N} \to \tilde{S} \subset S$. Want to show $f \circ g(\mathbb{N}) = S$. Take $s \in S$. Induction gives $g(j) \geq j$. Now $s = f(m)$ and we have $g(m) \geq m$. There has to be an easier way.........look in the book. □

Corollary 2.5. The set $\mathbb{Q}$ is countable.

Proof. Let $r \in \mathbb{Q}$. Then $r = \frac{m}{n} = (m,n) \in \mathbb{Z} \times \mathbb{Z}$. In fact, since rational representation is not unique, we only need a subset of $\mathbb{Z} \times \mathbb{Z}$. But the set $\mathbb{Z} \times \mathbb{Z}$ is countable. □

Corollary 2.6. Suppose $\{S_n\}_{n \in \mathbb{N}}$ is a countable family of countable sets. Then $\bigcup_{n \in \mathbb{N}} S_n$ is countable.

Proof. He drew a picture and called it rigorous. □

Theorem 2.7. Let $I = [a,b]$ where $a < b$. Then the set $I$ is uncountable.
Note 2.8. A corollary of this is that $\mathbb{R}$ and any interval (e.g. $(a,b)$) is uncountable. Another corollary is that $\mathbb{R} \setminus \mathbb{Q}$ is uncountable.

Proof of Theorem 2.7. Assume to the contrary that $[a,b]$ is countable. Then we may enumerate as follows: $[a,b] = \{x_n\}_{n \in \mathbb{N}}$. The denseness of $\mathbb{Q}$ guarantees that we can find $[a_1, b_1] \subset [a, b]$ with $x_1 \notin [a_1, b_1]$. Similarly, there exists $[a_2, b_2] \subset [a_1, b_1]$ such that $x_2 \notin [a_2, b_2]$. Continue in this fashion. Thus there exists $[a_n, b_n] \subset [a_{n-1}, b_{n-1}]$ with $x_n \notin [a_n, b_n]$. Consider $\{a_n\}_{n \in \mathbb{N}}$. This set is bounded above by $b$; in fact, this set is bounded above by every $b_n$. Thus the completeness axiom says we can find $x^* = \sup \{a_n : n \in \mathbb{N}\}$. It follows from the previous comment that $a_n \leq x^* \leq b_n$ for all $n \in \mathbb{N}$. Now $x^* \in [a, b]$, so $x^* = x_{n_0}$ for some $n_0 \in \mathbb{N}$. Hence, we have $a_{n_0} \leq x^* = x_{n_0} \leq b_{n_0}$. But by construction, $x_{n_0} \notin [a_{n_0}, b_{n_0}]$. This contradiction establishes our result. \hfill \(\square\)

2.2. Topological Set Concepts.

Definition 2.9. A set $O \subset \mathbb{R}$ is called an open set if for every $x \in O$ there exists $(a, b)$ such that $x \in (a, b) \subset O$. The family of all open subsets of $\mathbb{R}$ is a topology.

Note 2.10. Properties of open sets.

1. Any union of open sets is open.
2. Any intersection of a finite family of open sets is an open set.

Theorem 2.11 (Structure Theorem for Open Sets). A non-empty set $O \subset \mathbb{R}$ is open if and only if it can be represented as a unique countable union of disjoint open intervals.

Proof. $\Leftarrow$ Obvious. $\Rightarrow$ Let $x \in O$ so that $(x - \epsilon, x + \epsilon) \subset O$. Let $S = \{y \in O : (x, y) \subset O\}$ and $T = \{y \in O : (y, x) \subset O\}$. Clearly $S \neq \emptyset \neq T$. Let $a_x = \inf T$ and $b_x = \sup S$. Let $I_x = (a_x, b_x)$. It is clear that for $x_1, x_2 \in O$ that $I_{x_1} = I_{x_2}$ or $I_{x_1} \cap I_{x_2} = \emptyset$. Now $O \subset \bigcup_{x \in O} I_x$, so this proves the theorem after we throw away repeated sets in this union. (We don’t prove uniqueness or countability of the union, although this latter part is simple.) \hfill \(\square\)

Definition 2.12. We make some definitions about closed sets.

1. Suppose $S \subset \mathbb{R}$. A point $x \in \mathbb{R}$ is called a point of closure for $S$ if for every open interval $I$ such that $x \in I$, there exists $y \in S$ such that $y \in I$.
2. The closure of $S$ is $\overline{S} = \{\text{the family of all points of closure for } S\}$.
3. The set $S$ is called a closed set if $S = \overline{S}$.

Lemma 2.13. The following are equivalent:

1. $S$ is a closed set,
2. For every sequence $\{x_n\} \subset S$ such that $x_n \to x \in \mathbb{R}$, we have $x \in S$, and
3. The complement $\mathbb{R} \setminus S$ is an open set.


1. $\emptyset$ is closed.
2. If $\{S_\alpha\}_{\alpha \in A}$ is a collection of closed sets, then $\bigcap_{\alpha \in A} S_\alpha$ is closed, and
3. If $\{S_1, \ldots, S_n\}$ is a finite collection of closed sets, then $\bigcup_{i=1}^n S_i$ is closed.
Definition 2.15. A set \( S \) is a **compact set** if for every open cover of \( S \), i.e.
\[
S \subset \bigcup_{\alpha} \mathcal{O}_\alpha,
\]
there exists a finite subcover \( S \subset \bigcup_{\alpha=1}^{n} \mathcal{O}_{\alpha_i} \).

**Theorem 2.16** (Heine-Borel Theorem). Suppose a set \( S \subset \mathbb{R} \) is bounded and closed. Then \( S \) is compact.

**Note 2.17.** The converse is also true. We will prove both.

**Proof of Heine-Borel Theorem.** \( \Rightarrow \) Let \( E \) be compact in \( \mathbb{R} \). Suppose \( E \) is unbounded. Note that \( E \subset \bigcup_{n=1}^{\infty} (-n, n) \), so \( \{(-n, n)\}_{n \in \mathbb{N}} \) is an open cover of \( E \). By compactness, there is a finite subcover \((-n_1, n_1) \cup \cdots \cup (-n_k, n_k) \). This contradicts \( E \) being unbounded, though, so we see that \( E \) is indeed bounded.

Now suppose \( E \) is not closed. Thus there is an \( x \in \mathbb{R} \) such that \( x \) is the limit of some sequence \( \{x_n\}_{n \in \mathbb{N}} \subset E \) and \( x \notin E \). For each \( y \in E \), let \( \epsilon_y < \frac{1}{2}|x-y| \). Then \( E \subset \bigcup \{(y-\epsilon_y, y+\epsilon_y)\}_{y \in E} \). But this cover has no finite subcover, since any finite subcover would not cover \( \{x_n\} \). Thus \( E \) is closed.

\( \Leftarrow \) Suppose \( E \) is closed and bounded. We begin with a special case. Suppose \( E = [a, b] \) where \( a < b \). Let \( \mathcal{F} = \{\mathcal{O}_\alpha\}_{\alpha \in A} \) be such that \( [a, b] \subset \bigcup_{\alpha} \mathcal{O}_\alpha \). Let
\[
F = \{x \in E : [a, x] \text{ can be covered by a finite family of } \mathcal{F}\}
\]
Note that \( a \in F \), so \( F \neq \emptyset \). Not also that \( F \) is bounded above by \( b \), so we may apply the completeness axiom. Let \( c = \sup F \). Note that \( c \in E \), so \( c \in \mathcal{O}_{\alpha_0} \) for some \( \alpha_0 \in A \). Thus \( (c-\epsilon, c+\epsilon) \subset \mathcal{O}_{\alpha_0} \). Now note that there is a \( y \) such that \( c - \epsilon < y \in F \) (otherwise \( c \) would not be a least upper bound of \( F \)). Thus \( [a, y] \) can be covered by \( \mathcal{O}_{\alpha_1}, \ldots, \mathcal{O}_{\alpha_m} \). Consider \( \{\mathcal{O}_{\alpha_1}, \ldots, \mathcal{O}_{\alpha_m}, \mathcal{O}_{\alpha_0}\} \), and note that if \( \epsilon' < \epsilon \), then this finite family covers \([a, c + \epsilon']\). Hence, if \( c < b \), then we have a contradiction. Thus \( c = b \). But we don’t know yet that \( c \in F \). But this follows from essentially the same argument, so \( [a, b] \) is compact.

Now, take any closed and bounded set \( S \). Thus there exist \( a, b \) such that \( S \subset [a, b] \). Then \( S^\sim = \mathbb{R} \setminus S \) is open. Take an open cover of \( S \), \( S \subset \bigcup_{\alpha \in A} \mathcal{O}_\alpha \), and consider the family \( \{\mathcal{O}_\alpha \cup S^\sim : \alpha \in A\} \). Note that
\[
\mathbb{R} = \left( \bigcup_{\alpha \in A} \mathcal{O}_\alpha \right) \cup S^\sim,
\]
and so \([a, b] \subset \bigcup_{\alpha \in A} (\mathcal{O}_\alpha \cup S^\sim) \). Thus, since \([a, b] \) is compact, there exists a finite subcover \((\mathcal{O}_{\alpha_1} \cup S^\sim) \cup \cdots \cup (\mathcal{O}_{\alpha_n} \cup S^\sim) \), so
\[
S \subset \mathcal{O}_{\alpha_1} \cup \cdots \cup \mathcal{O}_{\alpha_n},
\]
as desired to complete the proof. \( \square \)

**Corollary 2.18** (Nested Set Theorem). Let \( \{F_n\}_{n \in \mathbb{N}} \) be a sequence of subsets of \( \mathbb{R} \) such that

1. \( F_{n+1} \subset F_n \) for all \( n \geq 1 \). That is, \( F_n \downarrow \).
2. \( F_n \) is a closed set for all \( n \geq 1 \).
3. \( F_1 \) is a bounded set.

Then \( \bigcap_{n=1}^{\infty} F_n \neq \emptyset \).
Proof. Assume to the contrary that $\bigcap_{n=1}^{\infty} F_n = \emptyset$. Then for every $x \in \mathbb{R}$ there exists $n \in \mathbb{N}$ such that $x \notin F_n$. Let $\mathcal{O}_n \subset \mathbb{R} \setminus F_n$ for all $n$, and note that $\mathcal{O}_n$ is open. Hence $x \in \mathcal{O}_n$, so $\bigcup_{n \in \mathbb{N}} \mathcal{O}_n = \mathbb{R}$. Thus $F_1 \subset \bigcup_{n=1}^{\infty} \mathcal{O}_n$, an open cover of $F_1$. Since $F_1$ is closed and bounded (and hence compact by Heine-Borel), we can write

$$F_1 \subset \bigcup_{i=1}^{k} \mathcal{O}_i,$$

for some $k \in \mathbb{N}$. Let $N = \max_{1 \leq j \leq k} \{n_j\}$. Then $\mathcal{O}_n \not\supset \mathcal{O}_N$. Thus $F_1 \subset \mathcal{O}_N = \mathbb{R} \setminus F_N$, which implies that $F_1 \cap F_N = \emptyset$. This is a contradiction since $F_N \subset F_1$. $\square$

**Definition 2.19.** Let $S$ be a set. A non-empty family $\mathcal{F}$ of subsets of $S$ is a $\sigma$-algebra if

1. $\emptyset \in \mathcal{F}$,
2. If $E \in \mathcal{F}$, then $S \setminus E \in \mathcal{F}$, and
3. If $\{E_n\}$ is a countable family of subsets of $S$ such that $E_n \in \mathcal{F}$ for all $n$, then $\bigcup_{n \geq 1} E_n \in \mathcal{F}$.

Elements of $\mathcal{F}$ are called measurable sets with respect to $\mathcal{F}$.

**Example 2.20.** We have several examples of $\sigma$-algebras.

1. The trivial $\sigma$-algebra: $S$, $\mathcal{F} = \{S, \emptyset\}$.
2. $\mathcal{P}(S) = \{\tilde{S} : \tilde{S} \subset S\}$, the power set of $S$. Both of these first two examples are extremes.
3. Borel $\sigma$-algebra of a topological space $\mathcal{B}(S)$ is the $\sigma$-algebra generated by the family $\{\mathcal{O}_\alpha\}$ of all open sets.

**Definition 2.21.** Let $S$ be a set, $G = \{S_\alpha\}$ a family of subsets of $S$. The $\sigma$-algebra $\sigma(G)$ generated by $G$ is the intersection of all $\sigma$-algebras containing $G$. In other words, $\sigma(G)$ is the smallest $\sigma$-algebra containing $G$.

**Definition 2.22.** The smallest $\sigma$-algebra $\tilde{\mathcal{F}}$ containing $G$ is defined as follows:

1. $G \subset \tilde{\mathcal{F}}$, and
2. For every $\sigma$-algebra $\tilde{\mathcal{G}}$ such that $G \subset \tilde{\mathcal{G}}$, we have $\tilde{\mathcal{F}} \subset \tilde{\mathcal{G}}$.

**Lemma 2.23.** In accordance with the notation of the previous two definitions, $\tilde{\mathcal{F}} = \sigma(G)$.

Proof. Since $\tilde{\mathcal{F}}$ is a $\sigma$-algebra, and since $\tilde{\mathcal{F}} \supset G$, it is immediate that $\sigma(G) = \bigcap_{G \subset \tilde{\mathcal{G}}} \tilde{\mathcal{G}} \subset \tilde{\mathcal{F}}$. It remains to show the reverse inclusion.

Note that $\sigma(G) = \bigcap_{G \subset \tilde{\mathcal{G}}} \tilde{\mathcal{G}}$ is a $\sigma$-algebra because the intersection of $\sigma$-algebras is a $\sigma$-algebra. Note also that $G \subset \sigma(G)$. Thus $G \subset \tilde{\mathcal{F}} \subset \sigma(G)$ by the second axiom in Definition 2.22 $\square$

Let $S$ be a topological space. Then $\mathcal{B}(S)$ is the Borel $\sigma$-algebra. There are two important sub-families of $\mathcal{B}(S)$:

1. $F_\sigma$-sets: $F = \bigcup_{n=1}^{\infty} E_n$ where $E_n$ are closed sets.
2. $G_\delta$-sets: $E = \bigcap_{n=1}^{\infty} \mathcal{O}_n$, where $\mathcal{O}_n$ are open.

**Lemma 2.24.** Any $F_\sigma$ or $G_\delta$ set is a Borel set.

Proof. It is clear that every closed set is a Borel set since $\sigma$-algebras are closed under taking complements and every closed set is the complement of an open (Borel) set. Thus $F_\sigma$ sets are Borel sets by the definition. By taking complements, every $G_\delta$ set is also a Borel set. $\square$
3. Sequences of Real Numbers

Definition 3.1. A sequence is a function \( f : \mathbb{N} \to \mathbb{R} \) where \( f(n) = x_n \in \mathbb{R} \). Often written \( \{x_n\} \).

Note 3.2. There are many types of sequences.

1. i) Bounded: \( |x_n| < M \) for some \( M \) and some \( n \in \mathbb{N} \).
   ii) Unbounded.
2. Monotone i) increasing, ii) decreasing.
3. Convergent sequences, divergent sequences.

Lemma 3.3. The following hold:

1. Increasing sequences bounded above are convergent.
2. Decreasing sequences bounded below are convergent.

Proof. Too easy... \( \square \)

Theorem 3.4 (Bolzano-Weierstrass Theorem). Suppose \( \{x_n\} \) is a bounded sequence in \( \mathbb{R} \). Then \( \{x_n\} \) has a convergent subsequence.

Proof. Use nested set theorem with the sets \( E_m = \{x_n : n \geq m\} \). This gives \( x \in \bigcap_{m \geq 1} E_m \). We want to show that there exists a subsequence \( x_{n_j} \to x \). Consider the intervals \( [x - \frac{1}{j}, x + \frac{1}{j}] \). Since \( x \in E_m \) for all \( m \), and since \( E_m \) is closed, there exists a sequence in \( E_m \) that converges to \( x \). But \( E_m \) is made up of \( \{x_n\} \)...not sure how he finishes...I like the proof I’ve seen before better. \( \square \)

Theorem 3.5. Let \( E \subset \mathbb{R} \). Then the following are equivalent.

1. \( E \) is compact.
2. For every sequence \( \{x_n\} \subset E \) there exists a subsequence \( \{x_{n_j}\} \) and \( x \in E \) such that \( x_{n_j} \to x \) as \( j \to \infty \).

Proof. To see (1) \( \Rightarrow \) (2), assume \( E \) is compact. Then Heine-Borel + Bolzano-Weierstrass = result.

Now we work toward showing (2) \( \Rightarrow \) (1). Assume to the contrary that (2) holds but (1) doesn’t. By the Heine-Borel theorem, either \( E \) is not closed or not bounded. Suppose \( E \) is not closed. Then there exists a sequence \( \{x_n\} \subset E \) such that \( x_n \to x \) but \( x \notin E \). Then (2) for this sequence because any subsequence converges to \( x \).

Now assume \( E \) is unbounded. Then there exists a sequence \( \{x_n\} \) such that either \( x_n \to \infty \) or \( x_n \to -\infty \). In either case, every subsequence tends to \( \infty \) or \( -\infty \), violating (2). These contradictions establish (1). \( \square \)
4. Lebesgue Outer Measure

How do we measure the “length” of a “complicated” subset of \( \mathbb{R} \)? For example, \( \ell([a,b]) = b - a \), \( \ell((a,b)) = b - a \), etc. As another example, it would make sense if \( E = \bigcup_{i=1}^{\infty} [a_i, b_i] \) (disjoint), then \( m(E) = \sum_{i=1}^{\infty} (b_i - a_i) \).

Our goal is to assign to some sets of \( \mathbb{R} \) (say, \( E \)) a number, \( m(E) \), called the measure of \( E \), such that \( m \) is a set function \( m : \mathcal{F} \to [0, \infty] \), where \( \mathcal{F} \) is some \( \sigma \)-algebra.

**Note 4.1** (Axioms we want for our Measure). Let us denote by \( \mathcal{L} \) the \( \sigma \)-algebra of Lebesgue measurable sets. Then we want to construct a set function \( m : \mathcal{L} \to [0, \infty] \) such that the following axioms are satisfied.

1. \( m(\emptyset) = 0 \).
2. \( m(I) = \ell(I) \) (i.e. the length of the interval).
3. (MOST IMPORTANT) For every disjoint countable family \( E_n \in \mathcal{L} \), we have \( m(\bigcup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} m(E_n) \). This is called countable additivity.
4. \( m(E + h) = m(E) \) for all \( E \in \mathcal{L} \) and \( h \in \mathbb{R} \). (Note that \( E + h = \{ x + h : x \in E \} \in \mathcal{L} \) for our \( \sigma \)-algebra \( \mathcal{L} \).) This is called the translation invariance of \( m \).

We modify the third axiom as follows:

3* \( m^*(\bigcup_{n=1}^{\infty} E_n) \leq \sum_{n=1}^{\infty} m^*(E_n) \). This is called countable subadditivity. The outer measure, defined on \( \mathcal{P}(\mathbb{R}) \) satisfies (1), (2), (3*), and (4).

**Definition 4.2.** The outer measure is a set function \( m^* : \mathcal{P}(\mathbb{R}) \to [0, \infty] \) defined by

\[
m^*(E) = \inf \left\{ \sum_{n=1}^{\infty} (b_n - a_n) : E \subset \bigcup_{n=1}^{\infty} (a_n, b_n) \right\}.
\]

**Note 4.3** (Properties of \( m^* \)). The following are properties of the outer measure.

1. \( m^*(\emptyset) = 0 \).
   *Proof.\( \emptyset \subset (0, \epsilon) \) for all \( \epsilon > 0 \), so \( m^*(\emptyset) \leq \epsilon \) for all \( \epsilon > 0 \). Thus \( m^*(\emptyset) = 0 \). \( \square \)

2. (Monotonicity.) Let \( A, B \subset \mathbb{R} \) such that \( A \subset B \). Then \( m^*(A) \leq m^*(B) \).
   *Proof. The set of families covering \( A \) contain the set of families covering \( B \). \( \square \)

3. If \( A \) is a countable set, then \( m^*(A) = 0 \).
   *Proof. Suppose \( A \) is countable. Then we can write \( A = \{ x_n \}_{n \in \mathbb{N}} \). Let \( \epsilon > 0 \). Note that \( x_n \in (x_n - \frac{\epsilon}{2^n}, x_n + \frac{\epsilon}{2^n}) = I_n \). Hence,

\[
A \subset \bigcup_{n \in \mathbb{N}} I_n,
\]

which implies \( m^*(A) \leq \sum_{n=1}^{\infty} \ell(I_n) = \sum_{n=1}^{\infty} 2 \cdot \frac{\epsilon}{2^n} = 2\epsilon \).

Letting \( \epsilon \to 0 \), we get \( m^*(A) = 0 \). \( \square \)

**Corollary 4.4.** \( m^*(\mathbb{Q}) = 0 \).

4. \( m^*([a,b]) = b - a \).
Proof. Note that \([a, b] \subset (a - \epsilon, b + \epsilon)\), which by definition of \(m^*\) implies that

\[ m^*([a, b]) \leq \inf_{\epsilon} \{b - a + 2\epsilon\}, \]

which implies \(m^*([a, b]) \leq b - a\).

Conversely, we want to show that \(b - a \leq m^*([a, b])\), which happens if and only if \(b - a \leq \sum_{n=1}^{\infty} \ell(I_n)\) for any covering \([a, b] \subset \bigcup_{n=1}^{\infty} I_n\). By the compactness of \([a, b]\), it is enough to prove \(b - a \leq \sum_{n=1}^{m} \ell(I_n)\), where \([a, b] \subset \bigcup_{n=1}^{m} I_n\) and \(\ell(I_n) < \infty\). This follows from an easy argument. \(\square\)

(5) \(m^*((a, b)) = b - a\).

Proof. We have \((a, b) \subset [a, b]\), and so by monotonicity, \(m^*((a, b)) \leq m^*([a, b]) = b - a\).

Conversely, for \(\epsilon > 0\), we have \([a + \epsilon, b - \epsilon] \subset (a, b)\), and so by monotonicity,

\[ m^*([a + \epsilon, b - \epsilon]) \leq m^*((a, b)), \]

which implies that \(b - a - 2\epsilon \leq m^*((a, b))\).

Letting \(\epsilon \to 0\), we have \(b - a \leq m^*((a, b))\), as desired to complete the proof. \(\square\)

(6) Let \(A \subset \mathbb{R}, y \in \mathbb{R}\). Then \(m^*(A + y) = m^*(A)\).

Proof. Very easy. Translate the covering by \(y\) and the result follows (since interval lengths are translation invariant). \(\square\)

(7) (Countable Subadditivity.) Let \(\{E_n\}_{n \in \mathbb{N}} \subset \mathbb{R}\). Then

\[ m^* \left( \bigcup_{n=1}^{\infty} E_n \right) \leq \sum_{n=1}^{\infty} m^*(E_n). \]

Proof. If \(m^*(E_k) = \infty\) for any \(k\), then the result is trivial. So assume that \(m^*(E_n) < \infty\) for all \(n\). Take \(\epsilon > 0\). Find a cover \(E_n \subset \bigcup_{k=1}^{\infty} I_k^{(n)}\) such that \(\sum_{k=1}^{\infty} |I_k^{(n)}| \leq m^*(E_n) + \frac{\epsilon}{2^n}\).

Now \(\bigcup_{n=1}^{\infty} E_n \subset \bigcup_{n,k=1}^{\infty} I_k^{(n)}\). Hence

\[ m^* \left( \bigcup_{n=1}^{\infty} E_n \right) \leq \sum_{n,k=1}^{\infty} |I_k^{(n)}| = \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} |I_k^{(n)}| \leq \sum_{n=1}^{\infty} \left( m^*(E_n) + \frac{\epsilon}{2^n} \right) = \sum_{n=1}^{\infty} m^*(E_n) + \epsilon. \]

Letting \(\epsilon \to 0\), we have our result. \(\square\)
5. C
deathéodory Construction (σ-algebra)

**Definition 5.1.** A set $E$ in $\mathbb{R}$ is called **measurable** provided for any subset $A \subset \mathbb{R}$ we have

$$m^*(A) = m^*(A \cap E) + m^*(A \cap E^C),$$

where $E^C = \mathbb{R} \setminus E$.

**Note 5.2.** We will denote by $\mathcal{M}$ the family of measurable sets. We have the following two goals.

1. Prove that $\mathcal{M}$ is a σ-algebra, and
2. Show $m^*|_{\mathcal{M}} : \mathcal{M} \rightarrow [0, \infty]$ is countably additive.

**Note 5.3 (Observations about $\mathcal{M}$).** We make the following observations about $\mathcal{M}$.

1. When checking for measurability, it suffices to check $m^*(A) \geq m^*(A \cap E) + m^*(A \cap E^C)$, because subadditivity gives us the other inequality for free.
2. $\emptyset \in \mathcal{M}$.
   
   *Proof.* $m^*(A) \geq m^*(A \cap \emptyset) + m^*(A \cap \emptyset^C) = m^*(\emptyset) + m^*(A \cap \mathbb{R}) = m^*(A)$. □
3. If $E \subset \mathcal{M}$, then $E^C \subset \mathcal{M}$.
   
   *Proof.* The symmetry of the measurability criterion establishes this result for free. □
4. If $m^*(E) = 0$, then $E \in \mathcal{M}$.
   
   *Proof.* We have $A \cap E \subset E$, so monotonicity yields $m^*(A \cap E) \leq m^*(E) = 0$. We also have $A \cap E^C \subset A$, which tells us again by monotonicity that $m^*(A \cap E^C) \leq m^*(A)$. Combining these two statements, we have $E \in \mathcal{M}$. □
5. Let $\{E_n\}_{n=1}^m$ be a finite family of sets. Then if $E_n \in \mathcal{M}$ for all $n$, then $\bigcup_{n=1}^m E_n \in \mathcal{M}$.
   
   *Proof.* It suffices to prove for $m = 2$, since it will then extend to all finite unions. Let $E_1, E_2 \in \mathcal{M}$. Since $E_1$ and $E_2$ are measurable, we have

$$m^*(A) = m^*(A \cap E_1) + m^*(A \cap E_1^C) = m^*(A \cap E_1) + m^*((A \cap E_1^C) \cap E_2) + m^*((A \cap E_1^C) \cap E_2^C).$$

It suffices to show

$$m^*(A \cap E_1) + m^*((A \cap E_1^C) \cap E_2) \geq m^*(A \cap (E_1 \cup E_2)),$$

as this will imply

$$m^*(A) \geq m^*(A \cap (E_1 \cup E_2)) + m^*(A \cap (E_1 \cup E_2)^C).$$

Note the obvious containment $E_1 \cup E_2 \subset E_1 \cup (E_1^C \cap E_2)$. Taking the intersection of everything with $A$ gives

$$A \cap (E_1 \cup E_2) \subset (A \cap E_1) \cup (A \cap E_1^C \cap E_2),$$

which implies (3) by monotonicity. □
Definition 5.4. A pre-measure \( \mu \) on an algebra \( \mathcal{M} \) is a set function such that \( \mu : \mathcal{M} \to [0, \infty] \) and the following hold.

(a) \( \mu(\emptyset) = 0 \),
(b) For \( \{E_k\}_{k=1}^n \) disjoint, we have \( \mu(\bigcup_{k=1}^n E_k) = \sum_{k=1}^n \mu(E_k) \).

\( m^* \) is a pre-measure on \( \mathcal{M} \).

Proof. All we must prove is finite additivity. We will actually prove a stronger statement. For any \( A \subset \mathbb{R} \) we have
\[
m^* \left( A \cap \left( \bigcup_{k=1}^n E_k \right) \right) = \sum_{k=1}^n m^*(A \cap E_k).
\]
Taking \( A = \mathbb{R} \) will yield the desired result. We proceed by induction on \( n \). For the case \( n = 1 \), the result is clear. Now assume the statement holds for \( n - 1 \). We use the fact that \( E_n \in \mathcal{M} \). Since \( E_n \in \mathcal{M} \), we have
\[
m^* \left( A \cap \left( \bigcup_{k=1}^n E_k \right) \right) = m^* \left( A \cap \left( \bigcup_{k=1}^n E_k \right) \cap E_n \right) + m^* \left( A \cap \left( \bigcup_{k=1}^{n-1} E_k \right) \right)
\]
\[
= m^*(A \cap E_n) + \sum_{k=1}^{n-1} m^*(A \cap E_k)
\]
\[
= \sum_{k=1}^n m^*(A \cap E_k),
\]
where we used the fact that \( \{E_n\} \) is a disjoint family for the second equality and the induction hypothesis for the third equality. By induction, we are done. \( \square \)

(7) If \( \{E_k\}_{k=1}^\infty \) is a collection of sets in \( \mathcal{M} \), then \( E = \bigcup_{k=1}^\infty E_k \in \mathcal{M} \).

Proof. Can we find \( \{F_k\} \subset \mathcal{M} \) such that \( \{F_k\} \) is a disjoint family and \( \bigcup_{k=1}^\infty F_k = \bigcup_{k=1}^\infty E_k \)? The answer is yes. Let \( F_1 = E_1 \). Then let \( F_2 = E_2 \setminus E_1 \), \( F_3 = E_3 \setminus (E_1 \cup E_2) \), etc. This family clearly establishes the properties we are after. Note that \( F_k \subset E_k \) and \( F_k \in \mathcal{M} \) (why?). Thus
\[
E = \bigcup_{k=1}^\infty E_k = \bigcup_{k=1}^\infty F_k,
\]
where the second union is disjoint. Let \( E_n' = \bigcup_{k=1}^n F_k \). Now \( E_n' \) is measurable since \( \mathcal{M} \) is an algebra. Hence we have
\[
m^*(A) = m^*(A \cap E'_n) + m^*(A \cap (E'_n)^C)
\]
\[
\geq m^*(A \cap E'_n) + m^*(A \cap E^C) \quad \text{(since } E^C \subset (E'_n)^C \text{)}
\]
\[
= m^* \left( A \cap \left( \bigcup_{k=1}^n F_k \right) \right) + m^*(A \cap E^C)
\]
\[ \sum_{k=1}^{n} m^*(A \cap F_k) + m^*(A \cap E^C) \quad \text{(Property 6)}. \]

Since this holds for all \( n \in \mathbb{N} \), this implies

\[ m^*(A) \geq \sum_{k=1}^{\infty} m^*(A \cap F_k) + m^*(A \cap E^C) \quad \text{(Countable subadditivity)} \]

\[ = m^*\left( A \cap \left( \bigcup_{k=1}^{\infty} F_k \right) \right) + m^*(A \cap E^C) \]

\[ = m^*(A \cap E) + m^*(A \cap E^C), \]

which finishes the proof. \( \Box \)

Note that we now have proved the following result.

**Theorem 5.5.** \( \mathcal{M} \) is a \( \sigma \)-algebra. In fact, we call \( \mathcal{M} \) the Carathéodory \( \sigma \)-algebra.

(8) \( m^* \) is a countably additive set function on \( \mathcal{M} \).

**Proof.** Let \( \{E_n\}_{n=1}^{\infty} \) be a disjoint family of sets in \( \mathcal{M} \). Then

\[ m^*\left( \bigcup_{n=1}^{\infty} E_n \right) \leq \sum_{n=1}^{\infty} m^*(E_n) \quad \text{(4)} \]

by countable subadditivity. From Property 6, \( m^*\left( \bigcup_{n=1}^{m} E_n \right) = \sum_{n=1}^{m} m^*(E_n) \). Also, by monotonicity we have

\[ m^*\left( \bigcup_{n=1}^{\infty} E_n \right) \geq m^*\left( \bigcup_{n=1}^{m} E_n \right) = \sum_{n=1}^{m} m^*(E_n), \]

and this holds for all \( m \). Thus, letting \( m \to \infty \), we have

\[ m^*\left( \bigcup_{n=1}^{\infty} E_n \right) \geq \sum_{n=1}^{\infty} m^*(E_n). \quad \text{(5)} \]

Combining (4) and (5) gives us our result. \( \Box \)

**Definition 5.6.** Let \( S \) be a set and let \( G \) be a \( \sigma \)-algebra of its subsets. A set function \( \mu : G \to [0, \infty] \) is called a **measure** provided

(a) \( \mu(\emptyset) = 0 \) and

(b) For any disjoint family \( \{E_n\} \subset G, n \geq 1, \) we have

\[ \mu\left( \bigcup_{n=1}^{\infty} E_n \right) = \sum_{n=1}^{\infty} \mu(E_n). \]

In this scenario, we call \((S, G)\) a **measurable space** and \((S, G, \mu)\) a **measure space**.

**Example 5.7.** We have (tediously) shown that \((\mathbb{R}, \mathcal{M}, m^*|_{\mathcal{M}})\) is a measure space. This is called the **Lebesgue measure space**. From now on, we will denote this space as follows:
We return to the discussion of the Borel
Note 5.8. We return to the discussion of the Borel 
σ-algebra \( B(\mathbb{R}) \). Recall that \( B \) is 
generated by the family of all open subsets of \( \mathbb{R} \). By relatively easy arguments, it is also 
generated by 
(a) \( \{(a, b) : a, b \in \mathbb{R}\} \), (b) \( \{[a, b] : a, b \in \mathbb{R}\} \), and 
(c) \( \{(a, \infty) : a \in \mathbb{R}\} \).

**Lemma 5.9.** \( B \subset \mathcal{L} \).

**Proof.** Let \( (\ast) \) be the statement \( (a, \infty) \in \mathcal{L} \) for all \( a \in \mathbb{R} \). Assuming we know \( (\ast) \), we know that
\[
\sigma(\{(a, \infty)\}) \subset \mathcal{L}, \text{ i.e., since } \sigma(\{(a, b)\}) = B.
\]
Thus we must only show \( (\ast) \). We must therefore show
\[
m^*(A) \geq m^*(A \cap (a, \infty)) + m^*(A \cap (a, \infty)^c).
\]
For simplicity, let \( A_1 = A \cap (a, \infty) \) and \( A_2 = A \cap (a, \infty)^c \), so we must show \( m^*(A) \geq m^*(A_1) + m^*(A_2) \). Without loss of generality, we may assume \( a \notin A \) (why?). Let \( A \subset \bigcup_{i=1}^{\infty} I_i \). We will prove \( m^*(A_1) + m^*(A_2) \leq \sum_{i=1}^{\infty} \ell(I_i) \). Let
\[
I'_i = I_i \cap (a, \infty) \text{ and } I''_i = I_i \cap (-\infty, a).
\]
These sets are all open or empty. Note that \( I'_i \cap I''_i = \emptyset \). Also, \( \bigcup I'_i \supset A_1 \) and \( \bigcup I''_i \supset A_2 \) (here we are using the assumption that \( a \notin A \)). Thus
\[
\sum \ell(I'_i) \geq m^*(A_1) \text{ and } \sum \ell(I''_i) \geq m^*(A_2).
\]
Adding both sides of these inequalities yields
\[
\sum \left[ \ell(I'_i) + \ell(I''_i) \right] \geq m^*(A_1) + m^*(A_2). \tag{6}
\]
Since \( I'_i \cap I''_i = \emptyset \), we have \( m^*(I'_i \cup I''_i) = m^*(I'_i) + m^*(I''_i) = \ell(I'_i) + \ell(I''_i) \). Using this identity on the left hand side of (6), we get
\[
\sum_{i=1}^{\infty} m^*(I'_i \cup I''_i) \geq m^*(A_1) + m^*(A_2).
\]
We also note that \( I'_i \cup I''_i \subset I_i \), and hence
\[
\sum_{i=1}^{\infty} \ell(I_i) = \sum_{i=1}^{\infty} m^*(I_i) \geq \sum_{i=1}^{\infty} m^*(I'_i \cup I''_i) \geq m^*(A_1) + m^*(A_2).
\]
What we have shown is that \( m^*(A_1) + m^*(A_2) \) is majorized by \( \sum_{i=1}^{\infty} \ell(I_i) \) for any cover \( A \subset \bigcup_{i=1}^{\infty} I_i \). Thus \( m^*(A_1) + m^*(A_2) \) is a lower bound for the set
\[
\left\{ \sum_{i=1}^{\infty} \ell(I_i) : A \subset \bigcup_{i=1}^{\infty} I_i \right\},
\]
and hence by definition \( m^*(A) \geq m^*(A_1) + m^*(A_2) \). \( \square \)
(9) (Translation Invariance.) Let $E \in \mathcal{M}$ and $y \in \mathbb{R}$. Then $E + y \in \mathcal{M}$ and $m(E + y) = m(E)$.

Proof. If we show that $E + y \in \mathcal{M}$, then we are done because of the translation invariance of the outer measure. Thus we need to prove

$m^*(A) = m^*(A \cap (E + y)) + m^*(A \cap (E + y)^C)$.

Note that

$m^*(A) = m^*(A - y) = m^*([A - y] \cap E) + m^*([A - y] \cap E^C)$

$= m^*(A \cap [E + y]) + m^*(A \cap [E + y]^C)$.

The first equality holds because of the translation invariance of outer measure. The second equality holds because $E$ is measurable. The final equality holds via set theory. Therefore $E + y \in \mathcal{M}$. □
6. Regularity and Continuity Properties of \( m \)

6.1. Regularity Properties. Given any measurable set \( E \), we seek to approximate \( E \) by Borel sets.

**Theorem 6.1** (Outer (Exterior) Regularity). The following statements approximate \( E \in \mathcal{M} \) by open sets.

1. Let \( E \in \mathcal{M} \). Then for every \( \epsilon > 0 \) there exists an open set \( O_\epsilon \) such that \( E \subseteq O_\epsilon \) and \( m(O_\epsilon \setminus E) < \epsilon \).
2. Let \( E \in \mathcal{M} \). Then there exists an \( G_\delta \)-set \( G \) such that \( E \subseteq G \) and \( m(G \setminus E) = 0 \).

**Theorem 6.2** (Inner (Interior) Regularity). The following statements approximate \( F \in \mathcal{M} \) by closed sets.

1. Let \( F \in \mathcal{M} \). Then for every \( \epsilon > 0 \) there exists a closed set \( C_\epsilon \) such that \( C_\epsilon \subseteq F \) and \( m(F \setminus C_\epsilon) < \epsilon \).
2. Let \( F \in \mathcal{M} \). Then there exists an \( F_\sigma \)-set \( U \) such that \( U \subseteq F \) and \( m(F \setminus U) = 0 \).

**Note 6.3.** These two theorems are important because they show that every measurable set differs from a Borel set by a set of 0 measure! Schwell...

**Proof of Outer Regularity.** (1) First assume that \( m(E) < \infty \). Recall

\[
m(E) = m^*(E) = \inf \left\{ \sum_{k=1}^{\infty} \ell(I_k) : E \subseteq \bigcup_{k=1}^{\infty} I_k \right\}.
\]

Find a cover \( E \subseteq \bigcup I_k \) such that \( \sum_{k=1}^{\infty} \ell(I_k) \leq m^*(E) + \epsilon \). Denote \( O_\epsilon = \bigcup_{k=1}^{\infty} I_k \), which is an open set. Note that \( E \subseteq O_\epsilon \) and

\[
m(O_\epsilon \setminus E) = m(O_\epsilon) - m(E) \leq \sum_{k=1}^{\infty} \ell(I_k) - m(E) \leq m(E) + \epsilon - m(E) = \epsilon,
\]

where the first equality follows from the excision property since we are assuming \( m(E) < \infty \). This establishes the case for when \( m(E) < \infty \).

Now suppose \( m(E) = \infty \). Let \( E_n = E \cap (-n, n) \) for all \( n \in \mathbb{N} \). Then \( E = \bigcup_{n=1}^{\infty} E_n \) and \( m(E_n) \leq 2n < \infty \). By the first case, there exists an open set \( O_n \) such that \( E_n \subseteq O_n \) and \( m(O_n \setminus E) < \frac{\epsilon}{2^n} \). Take \( O = \bigcup_{n=1}^{\infty} O_n \), which is an open set. Also, \( E \subseteq O \) since \( \bigcup_{n=1}^{\infty} E_n \subseteq \bigcup_{n=1}^{\infty} O_n \).

Now \( O \setminus E \subseteq \bigcup_{n=1}^{\infty} (O_n \setminus E_n) \), which implies

\[
m(O \setminus E) \leq m \left( \bigcup_{n=1}^{\infty} (O_n \setminus E_n) \right) \leq \sum_{n=1}^{\infty} m(O_n \setminus E_n) \leq \sum_{n=1}^{\infty} \frac{\epsilon}{2^n} = \epsilon.
\]

(2) Let \( \epsilon = \frac{1}{n} \) and by part (1) find an open set \( O_n \) such that \( E \subseteq O_n \) and \( m(O_n \setminus E) < \frac{1}{n} \). Set \( G = \bigcap_{n=1}^{\infty} O_n \), which is by definition of type \( G_\delta \). Moreover, \( E \subseteq G \) since \( E \subseteq O_n \) for all \( n \). Now \( G \setminus E \subseteq O_n \setminus E \) for all \( n \), so

\[
m(G \setminus E) \leq m(O_n \setminus E) < \frac{1}{n},
\]

and this holds for all \( n \in \mathbb{N} \), so \( m(G \setminus E) = 0 \).

**Proof of Inner Regularity.** This proof follows from Outer Regularity by taking complements.
(1) Consider $E^C$. Applying outer regularity to $E^C$, we see that there exists an open set $O_\epsilon$ such that $E^C \subset O_\epsilon$ and $m(O_\epsilon \setminus E^C) < \epsilon$. Now let $C_\epsilon = O_\epsilon^C$, which is closed. Note that $C_\epsilon \subset E$ because $E^C \subset C_\epsilon^C = O_\epsilon$. Also note that $E \setminus C_\epsilon = O_\epsilon \setminus E^C$, so $m(E \setminus C_\epsilon) = m(O_\epsilon \setminus E^C) < \epsilon$.

(2) Consider $E^C$ and apply outer regularity. Then there exists a $G_\delta$-set $G$ such that $E^C \subset G$ and $m(G \setminus E^C) = 0$. Now let $F = G^C \subset E$, and so $F$ is $F_\sigma$. Since $E \setminus F = E \setminus G^C = G \setminus E^C$, we have

$$m(E \setminus F) = m(G \setminus E^C) = 0.$$ This completes the proof. □

**Definition 6.4.** We define $E \triangle G = (G \setminus E) \cup (E \setminus G)$ to be the symmetric difference of sets.

**Note 6.5.** The set $G$ approximates $E$ well (with no inclusion condition, as was needed in the regularity properties) when $m(G \Delta E) < \epsilon$.

**Theorem 6.6.** Let $\epsilon > 0$ be given. Let $E \subset \mathcal{M}$ be such that $m(E) < \infty$. Then there exists a set $G = \bigcup_{k=1}^n I_k$ (which is a disjoint union of open intervals) such that $m(G \Delta E) < \epsilon$.

**Proof.** Outer approximation gives that there exists an open set $O$ such that $E \subset O$ and $m(O \setminus E) < \frac{\epsilon}{2}$. Note this implies $O < \infty$. By the structure theorem, we have

$$O = \bigcup_{k=1}^\infty I_k,$$

where these sets $I_k$ are disjoint open intervals. Also, by countable additivity of $m$, we have

$$m(O) = \sum_{k=1}^\infty \ell(I_k) < \infty.$$ Since this series converges, we can find $n \geq 1$ such that $\sum_{k=n+1}^\infty \ell(I_k) < \frac{\epsilon}{2}$. Set $G = \bigcup_{k=1}^n I_k$. Now note

$$m(G \Delta E) = m(G \setminus E) + m(E \setminus G) \leq m(O \setminus E) + m(O \setminus G) = m(O \setminus \bigcup_{k=n+1}^\infty I_k) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$ This completes the proof. □

### 6.2. The Continuity Properties of Measure.

**Theorem 6.7** (Continuity Properties of $m$). The following are known as continuity from below and continuity from above, respectively.

1. Let $E_n \in \mathcal{M}$ and $E_n \subset E_{n+1}$ for all $n$. Then $\lim_{n \to \infty} m(E_n) = m(\bigcup_{n=1}^\infty E_n)$.
2. Let $E_n \in \mathcal{M}$ and $E_n \supset E_{n+1}$ and $m(E_1) < \infty$. Then $\lim_{n \to \infty} m(E_n) = m(\bigcap_{n=1}^\infty E_n)$.

**Proof of continuity from below.** If $m(E_{n_0}) = \infty$ for at least one $n_0 \in \mathbb{N}$, then the result holds trivially. Thus without loss of generality we may assume $m(E_n) < \infty$ for all $n \in \mathcal{N}$. Let $C_1 = E_1$ and $C_n = E_n \setminus E_{n-1}$ for all other $n \in \mathbb{N}$. Clearly $\bigcup_{n=1}^\infty E_n = \bigcup_{n=1}^\infty C_n$, and $\{C_n\}$ is a disjoint family. Hence, $m\left(\bigcup_{n=1}^\infty E_n\right) = m\left(\bigcup_{n=1}^\infty C_n\right) = m\left(E_1 \cup \bigcup_{n=2}^\infty (E_n \setminus E_{n-1})\right)$.
\[
= m(E_1) + \sum_{n=2}^\infty m(E_n \setminus E_{n-1}) \\
= m(E_1) + \sum_{n=2}^\infty (m(E_n) - m(E_{n-1})) \\
= \lim_{j \to \infty} \left[ m(E_1) + \sum_{n=2}^{j} (m(E_n) - m(E_{n-1})) \right] \\
= \lim_{j \to \infty} m(E_j).
\]

This establishes continuity from below. \qed

**Proof of continuity from above.** Let \( D_n = E_1 \setminus E_n \) for each \( n \in \mathbb{N} \). Note that \( D_n \uparrow \) since \( E_n \downarrow \).

Thus we may use continuity from below to get

\[
\lim_{n \to \infty} m(D_n) = m \left( \bigcup_{n=1}^\infty D_n \right). \tag{7}
\]

Note that

\[
\bigcup_{n=1}^\infty D_n = \bigcup_{n=1}^\infty (E_1 \setminus E_n) = E_1 \setminus \left( \bigcap_{n=1}^\infty E_n \right). \tag{8}
\]

Note that since \( m(E_1) < \infty \), we have \( m(D_n) = m(E_1 \setminus E_n) = m(E_1) - m(E_n) \) by the Excision Property. Combining (7) and (8) and using this fact, we have

\[
m(E_1) - \lim_{n \to \infty} m(E_n) = \lim_{n \to \infty} [m(E_1) - m(E_n)] = \lim_{n \to \infty} m(D_n) \\
= m \left( \bigcup_{n=1}^\infty D_n \right) \\
= m \left( E_1 \setminus \left( \bigcap_{n=1}^\infty E_n \right) \right) \\
= m(E_1) - m \left( \bigcap_{n=1}^\infty E_n \right)
\]

This gives \( \lim_{n \to \infty} m(E_n) = m(\bigcap_{n=1}^\infty E_n) \), as desired to complete the proof. \qed

**Definition 6.8.** We say that a some statement holds **almost everywhere** if the statements holds for all \( x \in \mathbb{R} \in \mathcal{M} \) such that \( m(\mathbb{R} \setminus \mathcal{R}) = 0 \).

**Definition 6.9.** Let \( E = \bigcup E_n \). We say \( E \) is **locally finite** if for every \( x \in E \), there exist only finite number of \( E_n \)'s containing \( x \). Or, for every \( x \in \mathbb{R} \), there exist no more than a finite number of \( E_n \) containing \( x \) (if we are allowing this number to be zero).

**Note 6.10.** Recall that \( \limsup_{n \to \infty} a_n = \inf_{n \geq k} \sup_{n \geq k} a_n \) and \( \liminf_{n \to \infty} a_n = \sup_{n \geq k} \inf_{n \geq k} a_n \). We are now going to formulate an analogous definition for sequences of sets.
Definition 6.11. Let \( \{ E_n \} \) be a family of \( S \). Then
\[
\limsup_{n \to \infty} E_n = \bigcap_{k \geq 1} \bigcup_{n \geq k} E_n \quad \text{and} \quad \liminf_{n \to \infty} E_n = \bigcup_{k \geq 1} \bigcap_{n \geq k} E_n.
\]

Note 6.12. We make some observations about \( \liminf \) and \( \limsup \).

1. If \( \{ E_n \} \subseteq M \), then \( \limsup_{n \to \infty} E_n \in M \) and \( \liminf_{n \to \infty} E_n \in M \).
2. If \( E_n \) are all open (closed), then \( \liminf_{n \to \infty} E_n \subseteq B \) and \( \limsup_{n \to \infty} E_n \in B \).
3. Suppose \( E_n \uparrow \). Then \( \limsup_{n \to \infty} E_n = \bigcup_{k=1}^\infty E_k \) and \( \liminf_{n \to \infty} E_n = \bigcup_{k=1}^\infty E_k \).
4. Suppose \( E_n \downarrow \). Then \( \limsup_{n \to \infty} E_n = \bigcap_{k=1}^\infty E_k \) and \( \liminf_{n \to \infty} E_n = \bigcap_{k=1}^\infty E_k \).

6.3. The Borel-Cantelli Lemma.

Theorem 6.13 (Borel-Cantelli Lemma). Suppose \( E_n \in M \), \( n \geq 1 \), is a sequence of sets such that
\[
\sum_{n=1}^\infty m(E_n) < \infty.
\]
Then the following equivalent statements are true:

1. \( m\left(\limsup_{n \to \infty} E_n\right) = 0 \), and
2. The family \( \{ E_n \} \) is locally finite almost everywhere.

Proof. Step 1: Show the equivalence of the two conclusions.

Note 6.14. It follows from DeMorgan that \( (\limsup_{n \to \infty} E_n)^C = \liminf_{n \to \infty} E_n^C \) and similarly \( (\liminf_{n \to \infty} F_n)^C = \limsup_{n \to \infty} F_n^C \).

Using this note, we see that
\[
m\left(\limsup_{n \to \infty} E_n\right) = m\left(\left[\liminf_{n \to \infty} E_n^C\right]^C\right) = m\left(\mathbb{R} \setminus \liminf_{n \to \infty} E_n^C\right).
\]
Condition (1) in the statement of the theorem states that
\[
m\left(\limsup_{n \to \infty} E_n\right) = m\left(\mathbb{R} \setminus \liminf_{n \to \infty} E_n^C\right) = 0,
\]
so we see that almost all \( x \in \mathbb{R} \) belong to \( \liminf_{n \to \infty} E_n^C \), so almost all \( x \in \mathbb{R} \) belongs to \( \bigcup_{k \geq 1} \bigcap_{n \geq k} E_k^C \). Now suppose \( x \in \bigcup_{k \geq 1} \bigcap_{n \geq k} E_k^C \). This means \( x \) is in all but finitely many \( E_k^C \), or equivalently that \( x \) belongs to at most a finite number of \( E_n \). Hence, we have shown that for almost all \( x \in \mathbb{R} \), it is true that \( x \) belongs to at most a finite number of \( E_n \). That is, \( \{ E_n \} \) is locally finite almost everywhere. Thus (1) \( \iff \) (2).

Step 2: We prove
\[
m\left(\limsup_{n \to \infty} E_n\right) = m\left(\bigcap_{k \geq 1} \bigcup_{n \geq k} E_n\right) = 0.
\]
Let \( F_n = \bigcup_{n \geq k} E_n \). Note that

1. \( m(F_k) \leq \sum_{n=k}^\infty m(E_n) < \infty \), and
2. \( F_k \downarrow \).

Hence we may apply continuity from above to state that
\[
m\left(\bigcap_{k \geq 1} F_k\right) = \lim_{k \to \infty} m(F_k).
\]

(9)
Now we observe that
\[ m(F_k) \leq \sum_{n=k}^{\infty} m(E_n) \to 0 \]
since it is the tail of a convergent series. Hence
\[ \lim_{k \to \infty} m(F_k) = 0. \]  
(10)
Combining (9) and (10) we have
\[ m \left( \limsup_{n \to \infty} E_n \right) = m \left( \bigcap_{k \geq 1} \bigcup_{n \geq k} E_n \right) = m \left( \bigcap_{k \geq 1} F_k \right) = \lim_{k \to \infty} m(F_k) = 0, \]
as desired to complete the proof.  \[ \square \]
7. Non-Measurable Sets

**Theorem 7.1.** There exists a proper subset $\mathcal{N}$ of $\mathbb{R}$ such that $\mathcal{N} \notin \mathcal{M}$.

The following theorem is stronger than Theorem 7.1, but we do not include the proof.

**Theorem 7.2.** Let $A \in \mathcal{M}$ such that $m(A) > 0$. Then there exists a set $\mathcal{N}_A \subset \mathbb{R}$ such that $\mathcal{N}_A \subset A$ and $\mathcal{N}_A \notin \mathcal{M}$.

**Proof of Theorem 7.1.** Consider $I = [0, 1]$. Let us introduce a relation $\sim$ on $I \times I$ as follows: $x \sim y$ if and only if $x - y \in \mathbb{Q}$. It is not hard to see that this is an equivalence relation. Consider the equivalence classes. For $x \in I$, let $I_x = \{y \in I : y \sim x\}$. Since equivalence classes partition $I$, we have $I = \bigcup I_{x_0}$, which is a disjoint union. Using the axiom of choice, we find a choice set $\mathcal{N}$ of representatives of each equivalence class. This $\mathcal{N}$ will be our non-measurable set.

We make two very important observations.

1. For any two distinct $x, y \in \mathcal{N}$, $x - y \notin \mathbb{Q}$.
2. For each $x \in I$, there exists a unique $c \in \mathcal{N}$ such that $x = c + q$, where $q \in \mathbb{Q}$.

Consider the family of sets $\{\mathcal{N} + r\}_{r \in [-1, 1] \cap \mathbb{Q}}$. What can we say about this family?

First of all, this is a countable family. This is obvious because the sets are indexed by rational numbers.

Second of all, this is a disjoint family. To see this, suppose for a moment that the family were not disjoint. Then there would be distinct $r_1, r_2 \in [-1, 1] \cap \mathbb{Q}$ such that $x \in (\mathcal{N} + r_1) \cap (\mathcal{N} + r_2)$. That is, $x - r_1 \in \mathcal{N}$ and $x - r_2 \in \mathcal{N}$. Thus we would have $(x - r_1) - (x - r_2) = r_1 - r_2 \in \mathbb{Q}$, which contradicts property (1) above. Thus $\{\mathcal{N} + r\}_{r \in [-1, 1] \cap \mathbb{Q}}$ is a disjoint family.

Lastly, we have

$$[0, 1] \subset \bigcup_{r \in [-1, 1] \cap \mathbb{Q}} (\mathcal{N} + r) \subset [-1, 2]. \tag{11}$$

To see this first inclusion, let $y \in [0, 1]$. Then by property (2) above, there exists a $c \in \mathcal{N}$ such that $y = c + q$ for some $q \in \mathbb{Q}$. So certainly it seems like $y \in \mathcal{N} + q$, but we need to make sure $q \in [-1, 1]$; otherwise, $\mathcal{N} + q$ is not in our family of sets. Note that $q = y - c$. Thus $q$ takes its minimum value when $y = 0$ and $c = 1$; that is, the minimum value $q$ can take is $-1$. The maximum value $q$ can take occurs when $y = 1$ and $c = 0$; that is, the maximum value $q$ can take is 1. Thus, $q \in [-1, 1] \cap \mathbb{Q}$. We have now shown the first inclusion.

The second inclusion is more or less obvious. The minimum value from $\mathcal{N}$ added to the minimum value from $[-1, 1] \cap \mathbb{Q}$ is $-1$, and the maximum value from $\mathcal{N}$ added to the maximum value from $[-1, 1] \cap \mathbb{Q}$ is 2, so the second inclusion follows.

We are now ready to complete the proof. Suppose to derive a contradiction that $\mathcal{N}$ is measurable. Note that

$$0 \leq m(\mathcal{N}) \leq 1,$$

and since $\mathcal{N}$ is measurable, $\mathcal{N} + r$ is measurable for all $r \in [-1, 1] \cap \mathbb{Q}$; furthermore, $m(\mathcal{N}) = m(\mathcal{N} + r)$ by the translation invariance of measure.
We have two possible cases. First, suppose $m(\mathcal{N}) = 0$. Then we have by (11) that
\[
1 = m([0,1]) \leq m \left( \bigcup_{r \in [-1,1] \cap \mathbb{Q}} \{ \mathcal{N} + r \} \right) = \sum_{r \in [-1,1] \cap \mathbb{Q}} m(\mathcal{N} + r) = \sum_{r \in [-1,1] \cap \mathbb{Q}} 0 = 0,
\]
where the inequality follows from monotonicity and the second equality (following the inequality) follows from countable additivity. This is clearly a contradiction.

Consider now the possibility that $m(\mathcal{N}) = \alpha > 0$. Then using (11) we have
\[
3 = m([-1,2]) \geq m \left( \bigcup_{r \in [-1,1] \cap \mathbb{Q}} \{ \mathcal{N} + r \} \right) = \sum_{r \in [-1,1] \cap \mathbb{Q}} \alpha = \infty,
\]
where the inequality once again follows from monotonicity and the second equality follows from countable additivity. This is also clearly a contradiction. Thus it must be that $\mathcal{N}$ is not measurable. 
\[\square\]
8. The Cantor Set and the Devil’s Staircase

8.1. Construction and Properties of the Cantor Set. In this section, we describe the Cantor Set \( C \), which is a subset of the unit interval.

**Definition 8.1.** Let \( C_1 = \left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right] \), \( C_2 = \left[0, \frac{1}{9}\right] \cup \left[\frac{2}{9}, \frac{1}{3}\right] \cup \left[\frac{2}{3}, \frac{7}{9}\right] \cup \left[\frac{8}{9}, 1\right] \), etc. We continue removing the “middle third” (which is why the Cantor Set is sometimes called the “middle third set”) from each remaining piece in each step. Thus each stage \( C_n \) consists of \( 2^n \) disjoint closed intervals of length \( 3^{-n} \). Then the Cantor set is defined to be

\[
C = \bigcap_{n=1}^{\infty} C_n.
\]

**Note 8.2.** We will show several properties of the Cantor set.

1. \( C \) is compact. (Obviously \( C \) is bounded; what we really care to point out here is that \( C \) is closed.)
2. \( C \) contains no isolated points.

**Note 8.3.** Properties (1) and (2) make the Cantor set a perfect set. Specifically, a perfect set is a closed set containing no isolated points.

3. \( C \) is uncountable, and
4. \( m(C) = 0 \).

**Proof of Properties.** We prove the properties listed above in Note 8.2.

1. \( C \) is clearly closed because it is the intersection of closed sets. \( \square \)
2. To show \( C \) has no isolated points, split \( C \) into two groups: the endpoints, and everything else. See the special homework assignment for proof of why “everything else” exists. Now we have two cases to consider.

   **Case 1:** Let \( x \in C \) be an endpoint. Then within the interval using \( x \) as an endpoint, there is a sequence of endpoints (from later removed stages) which tends to \( x \) and are different from \( x \). Hence, \( x \) is not isolated.

   **Case 2:** Let \( x \) be a “pure Cantor point,” that is, not an endpoint. Again, \( x \) is the limit of a sequence of endpoints of intervals of diameter tending to 0. All points are different from \( x \) because they are all endpoints and \( x \) is not. Hence, \( x \) is not isolated. \( \square \)

3. We show \( C \) is uncountable. To do this, we construct a one-to-one correspondence between \( C \leftrightarrow [0, 1] \). We use the dyadic representation of \([0, 1]\). Let \( tx \in [0, 1] \), and represent

\[
x = \sum_{n=0}^{\infty} \frac{a_n}{2^n} \quad \text{where} \quad a_n = \begin{cases} 0 \\ 1 \end{cases}
\]

This is not always unique. For instance, \( 1 = 1 \) (where \( a_0 = 1, a_n = 0 \) for all \( n \geq 1 \)), and \( 1 = \sum_{n=1}^{\infty} \frac{1}{2^n} \) (where \( a_0 = 0 \) and \( a_n = 1 \) for all \( n \geq 1 \)). In fact, all dyadic rationals have two representations.

**Definition 8.4.** A dyadic rational is a rational \( r \) such that \( r = \sum_{n=1}^{m} \frac{a_n}{2^n} \), where \( a_n = 0 \) or 1 for all \( n \in \mathbb{N} \).
To see that all dyadic rationals have two representations, observe without loss of generality that \( a_m = 1 \) and
\[
r = \sum_{n=0}^{m-1} \frac{a_n}{2^n} + \sum_{n=m+1}^{\infty} \frac{1}{2^n}.
\]
All other numbers only have one representation, which is the infinite representation. Hence, let us limit ourselves to thinking only of the infinite representations for all numbers \( x \in [0, 1] \), whether or not it is a dyadic rational. (That is, if we have a dyadic rational, take the infinite representations instead of the finite representation.) Thus to each number we may associate an infinite string of 0s and 1s.

We can do essentially the same construction with triadic rationals.

**Definition 8.5.** A **triadic rational** is a rational \( r \) such that \( r = \sum_{n=1}^{m} \frac{b_n}{3^n} \), where \( b_n = 0 \) or 1 or 2 for all \( n \in \mathbb{N} \).

In particular, for a triadic rational \( s \), we have
\[
s = \sum_{n=1}^{m} \frac{b_n}{3^n} = \sum_{n=1}^{m-1} \frac{b_n}{3^n} + b_m \begin{cases} \sum_{n=1}^{m-1} \frac{b_n}{3^n} + \sum_{n=m+1}^{\infty} \frac{2}{3^n} & \text{if } b_m = 1 \\
\sum_{n=1}^{m-1} \frac{b_n}{3^n} + \frac{1}{3^m} + \sum_{n=m+1}^{\infty} \frac{2}{3^n} & \text{if } b_m = 2. \end{cases}
\]

So we may again only consider when we have infinite representations. Thus, for \( y \in C \), we can form the infinite triadic representation
\[
y = \sum_{n=1}^{\infty} \frac{b_n}{3^n} \text{ where } b_n = \begin{cases} 0 \\ 2. \end{cases}
\]

Let \( S_1 = [0, 1] \setminus \{ \text{diadic rationals} \} \) and \( S_2 = C \setminus \{ \text{triadic rationals} \} \). We claim that there is a one-to-one correspondence between \( S_1 \) and \( S_2 \). Let \( y \in S_2 \). We will illustrate the rule for this one-to-one correspondence.
\[
y = \sum_{n=1}^{\infty} \frac{b_n}{3^n} \leftrightarrow x = \sum_{n=1}^{\infty} \frac{a_n}{2^n} \in S_1, \text{ where } a_n = \frac{b_n}{2}.
\]

Clearly \( S_1 \) is uncountable, so \( S_2 \) is uncountable as well. Since \( C \supset S_2 \), it follows that \( C \) is uncountable. □

(4) We show \( m(C) = 0 \). Note that \( C \supset C_n \) for all \( n \in \mathbb{N} \), so \( m(C) \leq m(C_n) \) for all \( n \in \mathbb{N} \). Since \( C_n \) consists of \( 2^n \) disjoint intervals of length \( 3^{-n} \), we have
\[
m(C_n) = 2^n \left( \frac{1}{3^n} \right) = \left( \frac{2}{3} \right)^n,
\]
which goes to 0 as \( n \to \infty \). Hence, \( m(C) = 0 \). □

We have now proved all of the properties listed, and so we are done. □

8.2. **The Cantor-Lebesgue Function.** We now begin working toward showing that \( B \neq M \).

**Example 8.6** (The Cantor-Lebesgue Function (aka Devil’s Staircase)). We consider a function \( \varphi : [0, 1] \to [0, 1] \), which we call the **Cantor-Lebesgue function**. Let us describe how to construct
this function. First of all, \( \varphi(0) = 0 \) and \( \varphi(1) = 1 \). Now suppose \( x \in \left( \frac{1}{3}, \frac{2}{3} \right) \). Then \( \varphi(x) = \frac{1}{2} \). This is the first segment of \([0, 1]\) removed in the construction of the Cantor set. The next two segments removed are \( \left( \frac{1}{9}, \frac{2}{9} \right) \) and \( \left( \frac{7}{9}, \frac{8}{9} \right) \). For \( x \in \left( \frac{1}{9}, \frac{2}{9} \right) \) we define \( \varphi(x) = \frac{1}{4} \) and for \( x \in \left( \frac{7}{9}, \frac{8}{9} \right) \) we define \( \varphi(x) = \frac{3}{4} \). We continue this construction at each stage of the construction of the Cantor set. See Figure 1. By induction, we have defined \( \varphi(x) \) for all \( x \in [0, 1] \). (Note that \([0, 1] = C \cup O \) and \( C \cap O = \emptyset \), we have \( m([0, 1]) = m(C) + m(O) \), which implies \( m(O) = 1 \).)

Now for an arbitrary \( y \in (0, 1] \), we define \( \varphi(y) = \sup \{ \varphi(x) : x \in O \cap [0, y) \} \).

**Figure 1.** The first three stages in the construction of the Cantor-Lebesgue function.

**Properties of \( \varphi \):**

1. \( \varphi \) is continuous. We omit the proof.
2. \( \varphi \) is non-decreasing by definition.
3. \( \varphi'(x) = 0 \) almost everywhere, since the only place \( \varphi \) increases is on elements of \( C \), which is a set of measure 0.

**Example 8.7.** We now consider another function related to the Cantor-Lebesgue function \( \varphi \). Consider \( \psi : [0, 1] \to [0, 2] \) where

\[ \psi(x) = \varphi(x) + x. \]

**Properties of \( \psi \):**

1. \( \psi \) is continuous, since it is the sum of two continuous functions.
2. \( \psi \) is strictly increasing since it is the sum of a non-decreasing function and a strictly increasing function.
3. \( \psi'(x) = 1 \) almost everywhere since \( \varphi'(x) = 0 \) almost everywhere and \( (x)' = 1 \) everywhere.

Now note that \( \psi^{-1} : [0, 2] \to [0, 1] \) is continuous, so \( \psi \) is a homeomorphism. Thus \( \psi \) maps closed (open) sets to closed (open) sets. Hence, we may say a few more things about \( \psi \):

4. \( \psi(C) \) is a closed set and is hence measurable.
5. \( \psi(O) \) (where \( O = [0, 1] \setminus C \)) is open and \( m(\psi(O)) = 1 \).

**Proof.** Since \( O \) is open, we can write it as \( O = \bigcup_{n} I_n \) for disjoint intervals \( I_n \) (removed in the construction of \( C \)). Note, then that \( \psi(O) = \bigcup_{n} \psi(I_n) \), where this is a disjoint union, since \( I_n \) are disjoint and \( \psi \) is a bijection. Also, \( m(I_n) = m(\psi(I_n)) \) (think about why this
is...it is because of the fact that \( \varphi'(x) = 1 \) on \( I_n \). Hence, we have

\[
m(\psi(O)) = m \left( \bigcup_n \psi(I_n) \right) = \sum_n m(\psi(I_n)) = \sum_n m(I_n) = m(O) = 1.
\]

This completes the proof. \( \square \)

(6) \( m(\psi(C)) = 1 \).

Proof. Since \( \psi([0,1]) = [0,2] \) and since \( \psi([0,1]) = \psi(O) \cup \psi(C) \) and \( \psi(O) \cap \psi(C) = \emptyset \), and since by (5) we have \( m(\psi(O)) = 1 \), it follows that \( m(\psi(C)) = 1 \). \( \square \)

**Theorem 8.8** \( (B \neq \mathcal{M}) \). There exists a Lebesgue measurable set \( F \) that is not a Borel set.

Recall the following result stated earlier but not proved.

**Lemma 8.9.** Let \( A \in \mathcal{M} \) such that \( m(A) > 0 \). Then there exists a set \( N_A \subset A \) such that \( N_A \notin \mathcal{M} \).

**Proof of Theorem 8.8.** Recall \( \psi(C) \) is measurable and \( m(\psi(C)) > 0 \). By Lemma 8.9, there exists a non-measurable set \( S \) inside \( \psi(C) \). Furthermore, \( \psi^{-1}(S) \subset C \). Since \( m(C) = 0 \), it follows that \( \psi^{-1}(S) \in \mathcal{M} \) and \( m(\psi^{-1}(S)) = 0 \). Let \( F = \psi^{-1}(S) \). Then, by what we have just shown, \( F \in \mathcal{M} \), but we claim \( F \) is not a Borel set.

Suppose to the contrary that \( F \) is Borel. Then since homeomorphisms map Borel sets to Borel sets, we have \( \psi(F) = \psi(\psi^{-1}(S)) = S \) is Borel and hence measurable, which is a contradiction. \( \square \)
Suppose \( f \) is a function defined on \( E \in \mathcal{M} \), where \( f : E \to (-\infty, \infty) \) (and we sometimes allow \( \infty \) to be in the range).

**Definition 9.1.** The function \( f \) is **Lebesgue measurable** if for every \( c \in \mathbb{R} \), the set 
\[ \{ x \in E : f(x) > c \} \]
is measurable. We will denote the set of all measurable functions by \( \mathcal{M} \) and the set of all measurable functions defined on \( E \in \mathcal{M} \) by \( \mathcal{M}_E \).

**Note 9.2.** The set \( \{ x \in E : f(x) > c \} \) is called a **level set**. We will (ab)use the following notation:
\[ \{ x \in E : f(x) > c \} = \{ f > c \} . \]

**Theorem 9.3 (Equivalent Conditions for \( f \in \mathcal{M} \)).** Let \( f \) be a function defined on \( E \subset \mathcal{M} \). Then the following are equivalent (and hence serve simultaneously as the definition of \( f \in \mathcal{M} \)):

1. \( \{ f > c \} \) for all \( c \in \mathbb{R} \),
2. \( \{ f \geq c \} \) for all \( c \in \mathbb{R} \),
3. \( \{ f < c \} \) for all \( c \in \mathbb{R} \), and
4. \( \{ f \leq c \} \) for all \( c \in \mathbb{R} \).

**Proof.** Note that \( \{ f > c \}^C = \{ f \leq c \} \) and \( \{ f < c \}^C = \{ f \geq c \} \), and so if \( \{ f > c \} \in \mathcal{M} \), then \( \{ f \leq c \} \in \mathcal{M} \), and also if \( \{ f < c \} \in \mathcal{M} \), then \( \{ f \geq c \} \in \mathcal{M} \). Hence we must only show the equivalence of (1) and (2).

Now assume (1). Then we have 
\[ \{ f \geq c \} = \bigcap_{n=1}^{\infty} \left\{ f > c - \frac{1}{n} \right\} . \]

Now everything on the right-hand side is measurable by assumption, and hence the left-hand side is measurable, implying (2).

Now assume (2). Then we have 
\[ \{ f > c \} = \bigcup_{n=1}^{\infty} \left\{ f \geq c + \frac{1}{n} \right\} . \]

Everything on the right-hand side is measurable by assumption, and hence the left-hand side is measurable, implying (1). Hence (1)-(4) are equivalent. \( \square \)

**Definition 9.4.** For \( A \subset \mathbb{R} \), we define
\[ \chi_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A. \end{cases} \]

Then \( \chi_A \) is called the **indicator function** or **characteristic function** for \( A \).

**Note 9.5.** There exists a non-measurable function. In particular, for a non-measurable set \( \mathcal{N} \), the indicator function \( \chi_{\mathcal{N}} \) is not measurable because \( \{ \chi_{\mathcal{N}} > \frac{1}{2} \} = \mathcal{N} \notin \mathcal{M} \).

**Lemma 9.6.** If \( f \in \mathcal{M} \), then \( \{ f = c \} \in \mathcal{M} \) for all \( c \in \mathbb{R} \).
We will give two proofs.

\textbf{Proof 1.} Let \( c \in \mathbb{R} \). Then \( \{ f = c \} = \bigcup_{c \in M} \{ f \geq c \} \cup \{ f \leq c \} \in M. \)

\textbf{Proof 2.} Let \( c \in \mathbb{R} \). Then it is clear that \( \{ f = c \} = \bigcap_{n \geq 1} \left\{ c - \frac{1}{n} < f(x) < c + \frac{1}{n} \right\} \). Now \( \left\{ c - \frac{1}{n} \right\} \in M \) by definition, but we do not assume yet that \( \{ f < c + \frac{1}{n} \} \in M \). So we must make an argument to show that this is the case. We have

\[
\begin{aligned}
\left\{ f < c + \frac{1}{n} \right\} &= \mathbb{R} \setminus \left\{ f \geq c + \frac{1}{n} \right\} = \mathbb{R} \setminus \left( \bigcap_{m > n} \left\{ f > c + \frac{1}{n} - \frac{1}{m} \right\} \right), \\
&\in M \\
&\in M \\
&\in M,
\end{aligned}
\]

which shows that \( \{ f < c + \frac{1}{n} \} \in M \), as desired. Hence it follows that \( \{ f = c \} \in M \) from what we showed above. \qedhere

\textbf{Note 9.7.} A useful assertion is that inverse images of functions preserve all set-theoretical operations (e.g. unions, intersections, complements, etc).

\textbf{Lemma 9.8.} For a function \( f \), we have \( f \in \mathcal{M} \) if and only if \( f^{-1}((c, \infty)) \in M \) for all \( c \in \mathbb{R} \).

\textbf{Proof.} \( \{ f > c \} = \{ x : f(x) \in (c, \infty) \} = f^{-1}((c, \infty)) \). \qedhere

\textbf{Lemma 9.9.} For a function \( f \), we have \( f \in \mathcal{M} \) if and only if \( f^{-1}(O) \in M \) for all open sets \( O \subset \mathbb{R} \).

\textbf{Proof.} First, suppose \( f^{-1}(O) \in M \) for all open \( O \subset \mathbb{R} \). Then in particular, \( f^{-1}((c, \infty)) \in M \) for all \( c \in \mathbb{R} \), and by Lemma 9.8 we have \( f \in \mathcal{M} \).

Conversely, suppose \( f \in \mathcal{M} \). Then \( f^{-1}((c, \infty)) \in M \) for all \( c \in \mathbb{R} \), which implies the equivalent conditions (Theorem 9.3) that \( f^{-1}((a, b)) \in M \) for all \( b \in \mathbb{R} \). This implies

\[
\begin{aligned}
f^{-1}((a, b)) &= f^{-1}((\infty, b) \cap (a, \infty)) \\
&= \bigcup_{E \in M} f^{-1}((\infty, b) \cap (a, \infty) \cap E) \\
&\in M,
\end{aligned}
\]

where the second equality follows from Note 9.7.

Now take an open set \( O \subset \mathbb{R} \). By the Structure Theorem, \( O = \bigcup_{i=1}^{\infty} (a_i, b_i) \), so

\[
f^{-1}(O) = f^{-1}\left( \bigcup_{i=1}^{\infty} (a_i, b_i) \right) = \bigcup_{i=1}^{\infty} f^{-1}((a_i, b_i)) \in M,
\]

as desired. \qedhere

\textbf{Corollary 9.10.} Continuous functions on \( \mathbb{R} \) are measurable.

\textbf{Proof.} Suppose \( f \) is continuous. Then for every open set \( O \), the set \( f^{-1}(O) \) is open. But open sets are measurable; hence, \( f^{-1}(O) \in M \). Hence, by Lemma 9.9, \( f \in \mathcal{M} \). \qedhere
Note 9.11. Does the following equivalence hold?

\[ f \in \mathcal{M} \iff f^{-1}(E) \in \mathcal{M} \text{ for all } E \in \mathcal{M}. \]

The answer is no! Refer to Homework 5, Exercise 37. In particular, \( \psi, \psi^{-1} \in \mathcal{M} \) and \( (\psi^{-1})^{-1}(A) = \psi(A) \notin \mathcal{M} \), despite \( A \in \mathcal{M} \).

**Theorem 9.12 (Measurability and Operations).** The following hold for functions \( f \) and \( g \).

1. If \( f, g \in \mathcal{M} \), then \( f + g \in \mathcal{M} \).
2. If \( f \in \mathcal{M} \) and \( \alpha \in \mathbb{R} \), then \( \alpha f \in \mathcal{M} \).
3. If \( f, g \in \mathcal{M} \), then \( f \cdot g \in \mathcal{M} \).
4. If \( f, g \in \mathcal{M} \) and \( g(x) \neq 0 \) for all \( x \in \mathbb{R} \), then \( \frac{f}{g} \in \mathcal{M} \).
5. There are measurable functions \( f \) and \( g \) for which \( f \circ g \notin \mathcal{M} \).

Before we prove this theorem, we need two lemmas.

**Lemma 9.13.** If \( f \in \mathcal{M} \), then \( f^2 \in \mathcal{M} \).

*Proof.* Consider \( \{ f^2 \geq c \} \). Without loss of generality, we may assume \( c \geq 0 \), since if \( c < 0 \) then \( (f(x))^2 \geq c \) for all \( x \in E \), and so \( \{ f^2 \geq c \} = E \in \mathcal{M} \). Hence assume \( c \geq 0 \). Then

\[
\{ f^2 \geq c \} = \{|f| \geq \sqrt{c} \} = \bigcup_{\in \mathcal{M}} \{ f \geq \sqrt{c} \} \cup \bigcup_{\in \mathcal{M}} \{ f \leq -\sqrt{c} \} \in \mathcal{M},
\]
as desired. \( \square \)

**Lemma 9.14.** Suppose \( g \in \mathcal{M} \) and \( g(x) \neq 0 \) for all \( x \in E \). Then \( \frac{1}{g} \in \mathcal{M} \).

*Proof.* Note that

\[
\left\{ \frac{1}{g} > c \right\} = \left\{ g > 0 \text{ and } \frac{1}{g} > c \right\} \cup \left\{ g < 0 \text{ and } \frac{1}{g} > c \right\} = \left( \{ g > 0 \} \cap \left\{ \frac{1}{g} > c \right\} \right) \cup \left( \{ g < 0 \} \cap \left\{ \frac{1}{g} > c \right\} \right).
\]

We consider cases.

**Case 1:** Suppose \( c > 0 \). Then by the beginning set equality, we have

\[
\left\{ \frac{1}{g} > c \right\} = \{ g > 0 \} \cap \left\{ \frac{1}{g} > c \right\} = \{ g > 0 \} \cap \left\{ \frac{1}{g} \in \mathcal{M} \right\}
\]

since \( g \in \mathcal{M} \). The first equality follows because the other piece from the initial set equality is empty. This completes Case 1.

**Case 2:** Suppose \( c = 0 \). Then

\[
\left\{ \frac{1}{g} > c \right\} = \{ g > 0 \} \cap \left\{ \frac{1}{g} > 0 \right\} = \{ g > 0 \} \in \mathcal{M}.
\]

This completes Case 2.

**Case 3:** Suppose \( c < 0 \). Then

\[
\left\{ \frac{1}{g} > c \right\} = \{ g > 0 \} \cup \left\{ g < 0 \right\} \cup \left( \{ g < 0 \} \cap \left\{ g < \frac{1}{c} \right\} \right) \in \mathcal{M},
\]

\( \square \)
which completes the proof.

\[ \square \]

**Proof of Theorem 9.12.** We prove these properties out of order.

(2) Let \( \alpha \in \mathbb{R} \) and \( f, g \in \mathcal{M} \). If \( \alpha = 0 \), then \( \alpha f = 0 \in \mathcal{M} \). If \( \alpha > 0 \), then we have \( \{ \alpha f > c \} = \{ f > \frac{c}{\alpha} \} \in \mathcal{M} \). If \( \alpha < 0 \), then \( \{ \alpha f > c \} = \{ f < \frac{c}{\alpha} \} \in \mathcal{M} \). □

(1) Let \( f, g \in \mathcal{M} \). Consider \( \{ f + g < c \} \) for \( c \in \mathbb{R} \). We seek to prove

\[
\{ f + g < c \} = \bigcup_{r \in \mathbb{Q}} \{ f < r \} \cap \{ g < c - r \},
\]

for if we know this equality, then the result follows.

Let \( x \in \bigcup_{r \in \mathbb{Q}} \{ f < r \} \cap \{ g < c - r \} \), so there exists an \( r \in \mathbb{Q} \) such that \( x \in \{ f < r \} \cap \{ g < c - r \} \), so \( f(x) < r \) and \( g(x) < c - r \). This implies \( f(x) + g(x) < r + c - r = c \), so \( x \in \{ f + g < c \} \), giving us the first inclusion.

Now let \( x \in \{ f + g < c \} \), so that \( f(x) + g(x) < c \), and hence \( f(x) < c - g(x) \). By the denseness of \( \mathbb{Q} \), we can find an \( r \) such that \( f(x) < r < c - g(x) \). Thus, \( f(x) < r \) and \( g(x) < c - r \), meaning that \( x \in \{ f < r \} \cap \{ g < c - r \} \), hence \( x \in \bigcup_{r \in \mathbb{Q}} \{ f < r \} \cap \{ g < c - r \} \). This reverse inclusion, combined with the former inclusion, yields (12). □

(3) Let \( f, g \in \mathcal{M} \). Note that

\[
fg = \frac{(f + g)^2 - (f - g)^2}{4},
\]

and this function is measurable because \((f + g)^2\) and \((f - g)^2\) are measurable by Lemma 9.13 and parts (1) and (2), and the rest also follows by parts (1) and (2) of this theorem. □

(4) By Lemma 9.14, we have that \( \frac{1}{g} \in \mathcal{M} \), and by part (3) of this theorem, we have \( \frac{f}{g} \in \mathcal{M} \). □

(5) Recall that there exists a set \( A \in \mathcal{M} \) such that \( \psi(A) \notin \mathcal{M} \), where \( \psi \) is given in Example 8.7. Let \( f = \chi_A \), the indicator function for \( A \), and let \( g = \psi^{-1} \). Now clearly \( \chi_A \in \mathcal{M} \), and \( g \) is continuous (and hence measurable by Corollary 9.10) because \( \psi \) is a strictly increasing function defined on an interval (See Homework 5).

Now we claim that \( f \circ g = \chi_A \circ \psi^{-1} \) is not measurable. Let \( I = (\frac{1}{2}, \frac{3}{2}) \). Then

\[
(f \circ g)^{-1}(I) = \psi(\chi_A^{-1}(I)) = \psi(A) \notin \mathcal{M},
\]

which implies that \( f \circ g \notin \mathcal{M} \) by Lemma 9.9. □

**Theorem 9.15.** If \( f \) is continuous and \( g \) is measurable, then \( f \circ g \) is measurable.

**Proof.** Note that \( (f \circ g)^{-1}(A) = g^{-1}(f^{-1}(A)) \) for all sets \( A \). Let \( A = \mathcal{O} \) be an open set in \( \mathbb{R} \). Then \( f^{-1}(\mathcal{O}) \) is an open set by the continuity of \( f \). We also have \( g \in \mathcal{M} \), so since \( f^{-1}(\mathcal{O}) \) is open, by Lemma 9.9, we have

\[
g^{-1}(f^{-1}(\mathcal{O})) \in \mathcal{M},
\]

implying that \( (f \circ g)^{-1}(\mathcal{O}) \in \mathcal{M} \). Since \( \mathcal{O} \) was picked as an arbitrary open set, we have by Lemma 9.9 that \( f \circ g \in \mathcal{M} \). □

**Theorem 9.16.** Suppose \( f_1, \ldots, f_n \in \mathcal{M} \). Let \( f = \max\{f_1, \ldots, f_n\} \) and \( g = \min\{f_1, \ldots, f_n\} \). Then \( f \) and \( g \) are measurable.

30
Proof. We check that \( \{ f > c \} \in \mathcal{M} \) for all \( c \in \mathbb{R} \). It suffices to prove that
\[
\{ f > c \} = \bigcup_{k=1}^{n} \{ f_k > c \}
\] (13)
since the right hand side is assumed to be a measurable set.

Let \( x \in \bigcup_{k=1}^{n} \{ f_k > c \} \). Then there exists an \( m \in \{1, \ldots, k\} \) such that \( f_m(x) > c \). Since \( f \geq f_k \), it follows that \( f(x) > c \), which implies \( x \in \{ f > c \} \). Thus the first inclusion is established.

Conversely, suppose \( x \in \{ f > c \} \), or that \( f(x) > c \). Then there is some \( k \) for which \( f_k(x) > c \). Thus \( x \in \bigcup_{k=1}^{n} \{ f_k > c \} \), establishing the reverse inclusion and hence establishing (13).

A similar argument proves the result for \( g \). \( \square \)

Definition 9.17. Let \( f \) be a function, not necessarily measurable. Then \( f^+(x) = \max\{f(x), 0\} \) is called the positive part of \( f \) and \( f^-(x) = \max\{-f(x), 0\} \) is called the negative part of \( f \).

Note 9.18. We state a few properties of the positive and negative parts of \( f \).

1. If \( f \in \mathcal{M} \), then \( f^+, f^- \in \mathcal{M} \).
2. \( f(x) = f^+(x) - f^-(x) \).
3. \( |f(x)| = f^+(x) + f^-(x) \).
4. \( f^+ \) and \( f^- \) are disjointly supported. (The support of a function is the subset of the domain for which the function is non-zero.)
5. \( f^+ \geq 0 \) and \( f^- \geq 0 \).
10. LIMITS AND MODES OF CONVERGENCE

Note 10.1. This section is very important for exams, both in-course and comprehensive.

Definition 10.2 (Modes of Convergence). Let \( \{f_n\}_{n \in \mathbb{N}} \) be a sequence of measurable functions defined on \( E \subseteq \mathcal{M} \).

1. The sequence \( \{f_n\} \) is called **point-wise convergent** on \( E \) if the limit \( \lim_{n \to \infty} f_n(x) = f(x) \) exists for every \( x \in E \).
2. The sequence \( \{f_n\} \) is called **convergent almost everywhere** on \( E \) if \( \lim_{n \to \infty} f_n(x) = f(x) \) exists for all \( x \in E \setminus E_0 \) where \( m(E_0) = 0 \).
3. The sequence \( \{f_n\} \) converges **uniformly** on \( E \) if
   \[
   \lim_{n \to \infty} \left[ \sup_{x \in E} |f(x) - f_n(x)| \right] = 0.
   \]
4. The sequence \( \{f_n\} \) converges **almost uniformly** on \( E \) provided for all \( \epsilon > 0 \) there exists a set \( E_\epsilon \subset E \) such that \( E_\epsilon \in \mathcal{M} \) and such that \( f_n \to f \) uniformly on \( E_\epsilon \) and \( m(E \setminus E_\epsilon) < \epsilon \).
5. The sequence \( \{f_n\} \) converges **in measure** if for every \( \epsilon > 0 \) we have
   \[
   \lim_{n \to \infty} m\left( \{x \in E : |f(x) - f_n(x)| > \epsilon \} \right) = 0.
   \]

Lemma 10.3. Suppose \( f_n \in \mathcal{M}_E \) and \( f_n \to f \) pointwise on \( E \). Then \( f \in \mathcal{M}_E \).

Proof. Let \( c \in \mathbb{R} \) and consider the fact that
\[
\{f < c\} = \bigcup_{k=1}^{\infty} \left[ \cap_{n=1}^{\infty} \left\{ f_j < c - \frac{1}{n} \right\} \right].
\]
(14)

Showing (14) will yield our result since the right-hand side is clearly measurable by the properties of a \( \sigma \)-algebra. With this in mind, let \( x \) be in the right-hand side of (14). Then \( x \in \bigcap_{j=k}^{\infty} \left\{ f_j < c - \frac{1}{n} \right\} \) for some positive integers \( k \) and \( n \). This means that \( f_j(x) < c - \frac{1}{n} \) for all \( j \geq k \). Sending \( j \to \infty \), we see that \( f(x) \leq c - \frac{1}{n} < c \), so \( x \in \{f < c\} \), as desired. Hence,
\[
\{f < c\} \supset \bigcup_{k=1}^{\infty} \left[ \cap_{n=1}^{\infty} \left\{ f_j < c - \frac{1}{n} \right\} \right].
\]
(15)

Now suppose \( x \in \{f < c\} \). Then \( f(x) < c \). Since \( \lim_{j \to \infty} f_j(x) = f(x) < c \), we may choose \( n \) large enough so that \( f(x) + \frac{1}{n} < c - \frac{1}{n} \). Then, also due to the fact that \( \lim_{j \to \infty} f_j(x) = f(x) \), we can find a \( k \in \mathbb{N} \) such that \( f_j(x) < f(x) + \frac{1}{n} < c - \frac{1}{n} \) for all \( j \geq k \). What we have shown is that for all \( n \) satisfying \( f(x) + \frac{1}{n} < c - \frac{1}{n} \), there exists a \( k \) such that for all \( j \geq k \) we have \( f_j(x) < c - \frac{1}{n} \). This is precisely what it means for \( x \) to be in the right-hand side of (14). Hence,
\[
\{f < c\} \subset \bigcup_{k=1}^{\infty} \left[ \cap_{n=1}^{\infty} \left\{ f_j < c - \frac{1}{n} \right\} \right].
\]
(16)

Combining (15) and (16) gives us (14) and finishes the proof. \( \square \)
Note 10.4. We now begin the long and arduous task of drawing comparisons and implications between the various modes of convergence. Refer to the following diagrams; the modes of convergence are labeled as they are in Definition 10.2.

Here, the * beside the implications (1) ⇒* (4) and (1) ⇒* (5) imply that this result holds only if \( m(E) < \infty \). We will provide (later) a counter-example if \( m(E) = \infty \).

Once again, we say that ⇒* means that the implication holds for \( m(E) < \infty \).

Worth Noting: We make another comment on the relation between (2) and (5): Let \( m(E) < \infty \). If \( f_n \to f \) in measure on \( E \), then there exists a subsequence \( \{f_{n_k}\} \) such that \( f_{n_k} \to f \) almost everywhere on \( E \). Hence, (5) ⇒ (2) on subsequences if \( m(E) < \infty \).

We combine the remaining cases into a single figure.
Once again, ⇒* means the implication holds if \( m(E) < \infty \). This finishes the summary of implications. We now make justifications and comments, where necessary. For the more obvious cases (e.g. (1) ⇒ (2)), we do not make any justification or comment.

**Relation | Comments**

(2) ⇓ (1)  |  The almost everywhere convergence property of \( f_n \) is not affected if we change the values of \( f_n(x_0) \) at an arbitrary point \( x_0 \).

(1) ⇒* (4)  |  This follows from Egorov’s Theorem (to follow).

Here is a counterexample where \( m(E) = \infty \). Let \( E = \mathbb{R} \) and suppose

\[
f_n(x) = \begin{cases} 
1 & \text{if } x > n \\
0 & \text{if } x \leq n.
\end{cases}
\]

It is clear that \( \lim_{n \to \infty} f_n(x) = 0 \) for all \( x \in \mathbb{R} \). However, we show that \( f_n \not\to 0 \) almost uniformly. Suppose to the contrary that \( f_n \to 0 \) almost uniformly. Let \( \epsilon = \frac{1}{2} \). Then we can find a set \( \tilde{E} \subset \mathbb{R} \) such that

\[
m(\mathbb{R} \setminus \tilde{E}) < \frac{1}{2} \quad \text{and} \quad \lim_{n \to \infty} \left[ \sup_{x \in \tilde{E}} |f_n(x) - 0| \right] = \lim_{n \to \infty} \left[ \sup_{x \in \tilde{E}} |f_n(x)| \right] = 0.
\]

Now since \( m(\mathbb{R} \setminus \tilde{E}) < \frac{1}{2} \), for all \( k \in \mathbb{N} \) there exists an \( x_k \in \tilde{E} \) where \( x_k > k \), and consequently, where \( f_k(x_k) = 1 \). Hence, for all \( n \in \mathbb{N} \), we have \( \sup_{x \in \tilde{E}} |f_n(x)| = 1 \), which means that

\[
\lim_{n \to \infty} \left[ \sup_{x \in \tilde{E}} |f_n(x)| \right] = 1,
\]

a contradiction from our assumption. Hence \( f_n \not\to 0 \) almost uniformly on \( E \).

(4) ⇓ (1)  |  This follows from the same reason as (2) ⇓ (1).

This implication does not hold when \( m(E) = \infty \). The same counterexample as in the case (1) ⇓ (4) works here as well. For, if \( \epsilon = \frac{1}{2} \), then

\[
m \left( \left\{ x : |f_n(x)| \geq \frac{1}{2} \right\} \right) = \infty \not\to 0.
\]

as \( n \to \infty \).

(5) ⇓ (1)  |  This follows from the “sliding hump” counter-example (to follow).
<table>
<thead>
<tr>
<th>Relation</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>(2) ⇒* (4)</td>
<td>This follows from Egorov’s Theorem.</td>
</tr>
<tr>
<td>(2) ⇒* (5)</td>
<td>This follows from Egorov’s Theorem.</td>
</tr>
<tr>
<td>(4) ⇒* (2)</td>
<td>Proof. Recall we assume $m(E) &lt; \infty$. Suppose $f_n \to f$ on $E$ almost uniformly. Let $\epsilon = \frac{1}{m}$ and let $A_m \subset E$ such that $m(E \setminus A_m) &lt; \frac{1}{m}$ and $f_n \to f$ uniformly on $A_m$. Now let $B_k = \bigcup_{m=1}^{k} A_m$. Now it follows that ${B_k}<em>{k=1}^{\infty}$ is an increasing family (making ${E \setminus B_k}</em>{k=1}^{\infty}$ a decreasing family) and $m(E \setminus B_k) \leq m(E \setminus A_k) &lt; \frac{1}{k}$. Also, it follows that $f_n \to f$ uniformly on $B_k$. To see this, note that $\lim_{n \to \infty} \left[ \sup_{x \in B_k}</td>
</tr>
<tr>
<td>(5) \not\Rightarrow (2)</td>
<td>This follows from the “sliding hump” counter-example.</td>
</tr>
<tr>
<td>(3) ⇒* (5)</td>
<td>Proof. Recall we assume $m(E) &lt; \infty$. Suppose $f_n \to f$ uniformly on $E$. Then $\sup_{x \in E} {</td>
</tr>
<tr>
<td>(5) \not\Rightarrow (3)</td>
<td>This follows from the “sliding hump” counter-example.</td>
</tr>
<tr>
<td>(4) ⇒* (5)</td>
<td>This follows from Egorov’s Theorem.</td>
</tr>
</tbody>
</table>
Example 10.5 ("Sliding Hump"). First divide the unit interval into two equal segments. Then define
\[ f_1(x) = \begin{cases} 
1 & \text{if } x \in [0, \frac{1}{2}] \\
0 & \text{otherwise}
\end{cases} \quad \text{and} \quad f_2(x) = \begin{cases} 
1 & \text{if } x \in (\frac{1}{2}, 1] \\
0 & \text{otherwise}
\end{cases} \]
Next, divide the unit interval into 4 equal segments and define
\[ f_3(x) = \begin{cases} 
1 & \text{if } x \in [0, \frac{1}{4}] \\
0 & \text{otherwise}
\end{cases} \quad \text{and} \quad f_4(x) = \begin{cases} 
1 & \text{if } x \in (\frac{1}{4}, \frac{1}{2}] \\
0 & \text{otherwise}
\end{cases} \]
\[ f_5(x) = \begin{cases} 
1 & \text{if } x \in (\frac{1}{2}, \frac{3}{4}] \\
0 & \text{otherwise}
\end{cases} \quad \text{and} \quad f_6(x) = \begin{cases} 
1 & \text{if } x \in (\frac{3}{4}, 1] \\
0 & \text{otherwise}
\end{cases} \]
See Figure 2. We continue this construction for all \( n \in \mathbb{N} \), always repeating the cycles on smaller and smaller subdivisions of the unit interval. Now it is clear that \( f_n \) diverges at every point, since for all \( N \in \mathbb{N} \), there exist \( n, m > N \) such that \( f_n(x) = 0 \) and \( f_m(x) = 1 \). However, it is also clear that \( f_n \to 0 \) in measure.

Figure 2. The first six functions in the “sliding hump” counter-example.

We also state Egorov’s Theorem here, but we defer the proof to the next section.

**Theorem 10.6** (Egorov’s Theorem). Suppose \( m(E) < \infty \) and \( f_n \to f \) almost everywhere (or point wise) on \( E \). Then for all \( \epsilon > 0 \), there exists a closed set \( C_\epsilon \subset E \) such that \( m(E \setminus C_\epsilon) < \epsilon \) and \( f_n \to f \) uniformly on \( C_\epsilon \).

**Corollary 10.7.** Suppose \( E \in \mathcal{M} \) with \( m(E) < \infty \). Then
1. Pointwise convergence implies almost uniform convergence.
2. Almost everywhere convergence implies almost uniform convergence.

This corollary, along with the “sliding hump” example, helps fill out our table.
11. Egorov’s Theorem, Simple Approximation, and Lusin’s Theorem

11.1. Egorov’s Theorem. We restate Egorov’s Theorem for easy reference. It should also be noted that Dr. Gulisashvili stressed the importance of Egorov’s Theorem as one of the foundational theorems that could potentially show up on a comprehensive exam. We must know the statement and proof.

**Theorem 11.1** (Egorov’s Theorem). Suppose \( m(E) < \infty \) and \( f_n \to f \) almost everywhere (or point wise) on \( E \). Then for all \( \varepsilon > 0 \), there exists a closed set \( C_\varepsilon \) such that \( C_\varepsilon \subset E \) and \( m(E \setminus C_\varepsilon) < \varepsilon \) and \( f_n \to f \) uniformly on \( C_\varepsilon \).

**Proof.** First, assume that \( f_n \to f \) point-wise on \( E \). Let

\[
E_k^n = \left\{ x \in E : |f(x) - f_j(x)| < \frac{1}{n} \text{ for all } j \geq k \right\}.
\]

Fix \( n \). Then \( E_k^n \subset E_k^{n+1} \) (i.e. if it is true for all \( j \geq k \), then it is true for all \( j \geq k+1 \)), so it follows that \( E_k^n \) is an increasing sequence in \( k \). Now \( \bigcup_{k=1}^\infty E_k^n = E \) because of point-wise convergence of \( \{f_n\} \). Thus, by continuity from below, we have

\[
m(E) = \lim_{k \to \infty} m(E_k^n),
\]

which implies by the excision property that \( m(E \setminus E_k^n) \to 0 \) as \( k \to \infty \). Thus, for all \( n \in \mathbb{N} \), find a \( k_n \in \mathbb{N} \) such that \( m(E \setminus E_{k_n}^n) < \frac{1}{2^n} \). Fix \( \varepsilon > 0 \). Find \( N \) such that \( \sum_{n=N}^\infty \frac{1}{2^n} < \frac{\varepsilon}{2} \), and note that \( N \) depends only on \( \varepsilon \). Then, let \( \tilde{C}_\varepsilon = \bigcap_{n=N}^\infty E_{k_n}^n \). We perform the following estimate:

\[
m' \left( E \setminus \tilde{C}_\varepsilon \right) = m \left( E \setminus \bigcap_{n=N}^\infty E_{k_n}^n \right) = m \left( \bigcup_{n=N}^\infty \left( E \setminus E_{k_n}^n \right) \right) \leq \sum_{n=N}^\infty m \left( E \setminus E_{k_n}^n \right) < \sum_{n=N}^\infty \frac{1}{2^n} < \frac{\varepsilon}{2}.
\]

Now for all \( \delta > 0 \), take \( n \in \mathbb{N} \) such that \( n > \max \left\{ N, \frac{1}{\delta} \right\} \), so that \( \frac{1}{n} < \delta \). Note that if \( x \in \tilde{C}_\varepsilon \), then we have \( x \in E_{k_n}^n \) by how \( C_\varepsilon \) was defined. This implies

\[
|f(x) - f_j(x)| < \frac{1}{n} < \delta \text{ for all } j \geq k_n
\]

by how the set \( E_{k_n}^n \) is defined. Thus we have

\[
\lim_{j \to \infty} \left[ \sup_{x \in \tilde{C}_\varepsilon} \{|f(x) - f_j(x)|\} \right] = 0,
\]

which implies \( f_n \to f \) uniformly on \( \tilde{C}_\varepsilon \).

So far we have proved that there exists a set \( \tilde{C}_\varepsilon \in \mathcal{M} \) such that \( m(\tilde{E} \setminus \tilde{C}_\varepsilon) < \frac{\varepsilon}{2} \) and \( f_n \to f \) uniformly on \( \tilde{C}_\varepsilon \). By inner regularity, there exists a closed set \( C_\varepsilon \subset \tilde{C}_\varepsilon \) such that \( m(\tilde{C}_\varepsilon \setminus C_\varepsilon) < \frac{\varepsilon}{2} \). Thus

\[
m(E \setminus C_\varepsilon) = m(E \setminus \tilde{C}_\varepsilon) + m(\tilde{C}_\varepsilon \setminus C_\varepsilon) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,
\]

and since \( C_\varepsilon \subset \tilde{C}_\varepsilon \), we know \( f_n \to f \) uniformly on \( C_\varepsilon \). This completes the point-wise case.

Now suppose \( m(E) < \infty \) and \( f_n \to f \) almost everywhere on \( E \). Then there exists a set \( E_0 \subset E \) such that \( f_n \to f \) point-wise on \( E_0 \) where \( m(E_0) \leq m(E) < \infty \), and where \( m(E \setminus E_0) = 0 \). Let \( \varepsilon > 0 \). By what we have just shown, it follows that there exists a closed set \( C_\varepsilon \) such that \( C_\varepsilon \subset E_0 \subset E \) such that \( m(E_0 \setminus C_\varepsilon) < \varepsilon \) and \( f_n \to f \) uniformly on \( C_\varepsilon \). Hence, it remains to show
that \( m(E \setminus C_\epsilon) < \epsilon \). We have

\[
m(E \setminus C_\epsilon) = m(E \setminus E_0) + m(E_0 \setminus C_\epsilon) < 0 + \epsilon = \epsilon,
\]

which completes the proof for the almost everywhere case. \( \square \)

Before we consider Lusin’s Theorem, we discuss simple approximation.

11.2. Simple Approximation.

Definition 11.2. Let \( f \in \mathcal{M} \). Then \( f \) is a simple function if it can be written as

\[
f(x) = \sum_{n=1}^{m} a_n \chi_{E_n}(x),
\]

where \( m \geq 1 \) is finite, \( a_n \in \mathbb{R}, E_n \in \mathcal{M} \).

If \( f \in \mathcal{M}_E \), then \( f \) is simple provided that the same conditions hold as above, this time with \( E_n \in \mathcal{M} \) and \( E_n \subseteq E \in \mathcal{M} \).

Lemma 11.3 (Approximation Lemma). Suppose \( f \in \mathcal{M}_E \) where \( E \in \mathcal{M} \), and further suppose that \( f \) is bounded. Then for all \( \epsilon > 0 \) there exists simple functions \( \varphi_\epsilon \) and \( \psi_\epsilon \) with domain \( E \) such that

1. \( \varphi_\epsilon(x) \leq f(x) \leq \psi_\epsilon(x) \) and
2. \( \psi_\epsilon(x) - \varphi_\epsilon(x) < \epsilon \) for all \( x \in E \).

Proof. Since \( f \) is bounded, we have \( |f| \leq M \) for some \( M > 0 \). Thus, \( f(E) \subseteq [-M, M] \). Let \( \epsilon > 0 \). Choose \( m \in \mathbb{N} \) such that \( \frac{2M}{m} < \epsilon \). We define two partitions \( P_1 \) and \( P_2 \), where \( P_1 \) partitions \( [-M, M - \frac{2M}{m}] \) and \( P_2 \) partitions \( [-M + \frac{2M}{m}, M] \). We have

\[
P_1 = \left\{-M = a_1 < a_2 < \cdots < a_m = M - \frac{2M}{m} : a_i = a_{i-1} + \frac{2M}{m} \text{ for } 2 \leq i \leq m \right\}
\]

and

\[
P_2 = \left\{-M + \frac{2M}{m} = b_1 < b_2 < \cdots < b_m = M : b_i = b_{i-1} + \frac{2M}{m} \text{ for } 2 \leq i \leq m \right\}.
\]

Now that we have these partitions constructed, we observe that \( a_i = b_{i-1} \) for all \( i \in \{2, \ldots, m\} \). Let \( I_k = [a_k, b_k] \), and note from our previous observation that \( \ell(I_k) = b_k - a_k = b_k - b_{k-1} = \frac{2M}{m} < \epsilon \).

Now let \( E_k = f^{-1}(I_k) \), which is a measurable set since inverse images of Borel sets are Borel sets. Let

\[
\varphi_\epsilon(x) = \sum_{k=1}^{m} a_k \chi_{E_k}(x) \quad \text{and} \quad \psi_\epsilon(x) = \sum_{k=1}^{m} b_k \chi_{E_k}(x),
\]

where \( a_k \) and \( b_k \) are the lower and upper endpoints of \( I_k \), respectively, as defined above. Both properties stated in the Approximation Lemma are now clear. See Figure 3. \( \square \)

Corollary 11.4. Suppose \( f \in \mathcal{M}_E \) where \( E \in \mathcal{M} \), and further suppose that \( f \) is bounded. Then there exist sequences \( \{\varphi_n\} \) and \( \{\psi_n\} \) of simple functions on \( E \) such that

1. \( \varphi_n(x) \leq f(x) \) for all \( x \in E \),
2. \( \psi_n(x) \geq f(x) \) for all \( x \in E \), and
3. \( \varphi_n \to f \) and \( \psi_n \to f \) (both) uniformly on \( E \).
Proof. By the Approximation Lemma, for all \( n \in \mathbb{N} \) there exist \( \varphi_n \) and \( \psi_n \) such that \( \varphi_n(x) \leq f(x) \leq \psi_n(x) \) and \( \psi_n(x) - \varphi_n(x) < \frac{1}{n} \) for all \( x \in E \). Now \( |f(x) - \varphi_n(x)| < \frac{1}{n} \) and \( |\psi_n(x) - f(x)| < \frac{1}{n} \) for all \( x \in E \). Both conditions imply uniform convergence. \( \square \)

Lemma 11.5. The following statements hold for simple functions defined on a set \( E \in \mathcal{M} \).

1. If \( s_1 \) and \( s_2 \) are simple, then \( s_1 + s_2 \), \( s_1 - s_2 \), and \( s_1 \cdot s_2 \) are simple.
2. If \( s_1, \ldots, s_n \) are simple, then \( \max\{s_1, \ldots, s_n\} \) is simple.

Proof. A function is simple if and only if it has finite range. Let \( s_1(E) \) and \( s_2(E) \) denote the ranges of \( s_1 \) and \( s_2 \), respectively. Suppose \( |s_1(E)| = n \) and \( |s_2(E)| = m \). Then for any \( x \in E \), there are \( n \) possibilities for \( s_1(x) \) and \( m \) possibilities for \( s_2(x) \), and so there are \( nm \) possibilities for \( s_1(x) + s_2(x) \). This same argument holds for subtraction and multiplication as well; hence, \( nm \) forms an upper bound for \( |(s_1 + s_2)(E)| \), \( |(s_1 - s_2)(E)| \), and \( |(s_1 \cdot s_2)(E)| \), which proves the first statement.

For the second statement, let \( s = \max\{s_1, \ldots, s_n\} \) and let \( R = \bigcup_{k=1}^n s_k(E) \), the union of all the respective ranges. Since this is a finite union of finite sets, it is finite. Let \( x \in E \) be arbitrary. Then

\[
s(x) = \max\{s_1(x), \ldots, s_n(x)\} = s_j(x) \in s_j(E) \subset R,
\]

where this holds for some \( j \in \{1, \ldots, n\} \). In any case, we have shown that \( s(x) \subset R \) for any arbitrary \( x \in E \), and so \( s(E) \subset R \). Since \( R \) is finite, this shows \( s(E) \) is finite and hence \( s \) is a simple function. \( \square \)

Note 11.6. Note that Lemma 11.5 implies that if \( s_1, \ldots, s_n \) are simple functions, then \( -s_1, \ldots, -s_n \) are simple as well, and also that \( -\max\{-s_1, \ldots, -s_n\} = \min\{s_1, \ldots, s_n\} \) is simple.

Theorem 11.7 (The Simple Approximation Theorem). Let \( f \) be a measurable function on \( E \).

1. Suppose \( f(x) \geq 0 \) for all \( x \in E \). Then there exists a sequence of simple function \( \{s_k\} \) defined on \( E \) such that
(a) $s_k(x) \geq 0$ for all $k \in \mathbb{N}$ and for all $x \in E$,
(b) $s_k(x) \leq s_{k+1}(x)$ for all $k \in \mathbb{N}$ and for all $x \in E$, and
(c) $\lim_{k \to \infty} s_k(x) = f(x)$.

(2) There exists a sequence $\{s_k\}$ of simple functions defined on $E$ such that
(a) $|s_k(x)| \leq |f(x)|$ for all $k \in \mathbb{N}$ and all $x \in E$, and
(b) $\lim_{k \to \infty} s_k(x) = f(x)$.

Proof. We first argue that (1) $\Rightarrow$ (2). Recall that $f = f^+ - f^-$. Let $E_1 \subset E$ be such that $f^+ > 0$ on $E_1$ and let $E_2 \subset E$ be such that $f^- > 0$ on $E_2$. Then by (1) we can approximate $f^+$ by $\{s_k^{(1)}\}$ and $f^-$ by $\{s_k^{(2)}\}$. Let $s_k = s_k^{(1)} - s_k^{(2)}$, and note that $s_k(x) = 0$ for any $x \notin E_1 \cup E_2$. Now clearly $s_k \to f$ point-wise. Now we have

$$|s_k| = \left|s_k^{(1)}\right| + \left|s_k^{(2)}\right| \leq f^+ + f^- = |f|,$$

where the first equality follows because $E_1 \cap E_2 = \emptyset$. This proves (2); hence we must prove (1).

To prove (1), let $E_n = \{f < n\}$. Consider $f$ on $E_n$ and note that since $f \geq 0$, it follows that $f$ is bounded on $E_n$. By the Approximation Lemma, there exists two function $\varphi_n$ and $\psi_n$ defined on $E_n$, both simple, such that $\varphi_n(x) \leq f(x) \leq \psi_n(x)$ and $\psi_n(x) - \varphi_n(x) < \frac{1}{n}$ for all $x \in E_n$. Define

$$\tilde{\varphi}_n(x) = \begin{cases} \max\{\varphi_n(x), 0\} & \text{if } x \in E_n \\ n & \text{if } x \in E \setminus E_n, \end{cases}$$

which is a simple function by Lemma 11.5. Note that $0 \leq \tilde{\varphi}_n(x) \leq f(x)$ on $E$. To see this, we note that for $x \in E_n$, this is clear, and for $x \in E \setminus E_n$, we have $f(x) \geq n = \tilde{\varphi}_n(x)$ since if $f(x) < n$, this would contradict $x \notin E_n$. We wish to prove that $\lim_{n \to \infty} \tilde{\varphi}_n(x) = f(x)$. Take $x \in E$ and find $N \in \mathbb{N}$ such that $N > f(x)$. Now take $n \in \mathbb{N}$ such that $n > N$. Then by definition, we have $x \in E_N \subset E_n$. Hence, for all $n \geq N$, we have

$$f(x) - \tilde{\varphi}_n(x) = f(x) - \varphi_n(x) \leq \psi_n(x) - \varphi_n(x) < \frac{1}{n}.$$ 

Hence, $f(x) - \tilde{\varphi}_n(x) < \frac{1}{n}$ for all $n > N$, and so $\lim_{n \to \infty} \tilde{\varphi}_n(x) = f(x)$, as claimed.

Now define $s_n = \max\{\tilde{\varphi}_i : 1 \leq i \leq n\}$, which is simple because of Lemma 11.5. Now $\tilde{\varphi}_n \leq s_n \leq f$, so $s_n \to f$ point-wise on $E$. All the properties listed in part (1) are now clear. \qed

11.3. Lusin’s Theorem and the Standard Representation of a Simple Function. The next result, Lusin’s Theorem, is also a very important theorem, but it lies in the shadow of Egorov’s Theorem, and as such, the proof will be skipped. According to Dr. Gulisashvili, the proof of Lusin’s theorem will not appear on any exam, whether it be in-class or comprehensive.

**Theorem 11.8 (Lusin’s Theorem).** Let $E \in \mathcal{M}$ and $m(E) < \infty$. Let $f \in \mathcal{M}_E$. Then for all $\epsilon > 0$, there exists a closed set $F \subset E$ such that

1. $m(E \setminus F) < \epsilon$, and
2. $f|_F$ is continuous on $F$.

We continue briefly our discussion of simple functions.

**Definition 11.9.** The **standard representation**, or **standard form**, of a simple function is $s(x) = \sum_{n=1}^{k} b_n \chi_{F_n}(x)$ where $b_{n_1} \neq b_{n_2}$ if $n_1 \neq n_2$ and where $\{F_n\}$ is a disjoint family.
We give two proofs of the next result.

**Lemma 11.10.** Let \( s \) be a simple function. Then \( s \) has a unique standard representation.

**Proof 1.** Recall that a function is simple if and only if it has a finite range. Suppose \( s \) is a simple function and let \( \{b_1, \ldots, b_r\} \) be its range. Let \( F_m = s^{-1}(b_m) \) for all \( m \in \{1, \ldots, r\} \). Then

\[
s(x) = \sum_{m=1}^{r} b_m \chi_{F_m}(x),
\]

which is in standard form since clearly \( \{F_m\}_{m=1}^{r} \) is a disjoint set and \( b_n \neq b_m \) if \( n \neq m \). This completes the existence of a standard representation. Furthermore, uniqueness follows from the way we have constructed our sets \( F_m \). \( \square \)

**Proof 2.** Let \( s(x) = \sum_{n=1}^{k} a_n \chi_{E_n}(x) \) be any representation of the simple function \( s \). First, we show that \( s \) can be written as

\[
s(x) = \sum_{m=1}^{p} c_m \chi_{G_m}(x)
\]

where \( \{G_m\} \) is a disjoint family. Let \( D = \{1, \ldots, k\} \) and consider \( \mathcal{P}(D) \). Let \( q_j \in \mathcal{P}(D) \) for \( 1 \leq j \leq 2^k - 1 \) (where we exclude the empty set). Let

\[
G_j = \bigcap_{m \in q_j} E_m \cap \bigcap_{n \notin q_j} E_n^C.
\]

These sets \( \{G_m\} \) are disjoint. To see this, suppose \( x \in G_\alpha \cap G_\beta \) for \( \alpha, \beta \in \{1, \ldots, 2^k - 1\} \), where \( \alpha \neq \beta \). Then \( q_\alpha \neq q_\beta \), and so without loss of generality there must exist \( b \in q_\alpha \) such that \( b \notin q_\beta \). Since \( x \in G_\alpha \cap G_\beta \), this means that \( x \in E_b \) and \( x \notin E_b^C \), a clear contradiction. Hence, \( \{G_m\} \) is a disjoint family, as claimed.

The coefficients \( c_m \) are defined by \( c_m = \sum_{a \in q_m} a \). If any of the numbers \( c_m \) coincide, then we take the union of the corresponding sets, which is still a disjoint family, to make the coefficients unique. We have provided a constructive proof of the existence (and clearly uniqueness by this construction) of the standard representation of the simple function \( s \). \( \square \)
12. Lebesgue Integration

**Note 12.1.** Our goal is to define the integral $\int_{\mathbb{R}} f \, dx = \int f \, dx$ and $\int_E f \, dx$ for $E \in \mathcal{M}$. The questions we would like to answer are

1. What functions can be integrated?
2. How can we define the integral?
3. Can we define the integral in a way where 
   \[ \int (\alpha f + \beta g) \, dx = \alpha \int f \, dx + \beta \int g \, dx; \]
   i.e. where the integral is a linear operation, as we expect it should be?

**Note 12.2.** All integrals in this section are understood to be Lebesgue integrals. As such, we may just say “integral” instead of “Lebesgue integral.” Sometimes, if the context is not clear, we may write $\int f \, m(dx)$ or $\int f \, dm(x)$ to indicate we are integrating with respect to Lebesgue measure.

Our first step is to define the integral for simple functions defined on a set of finite measure.

**Definition 12.3.** Let $s$ be a simple function with standard representation $s(x) = \sum_{k=1}^{m} c_k \chi_{F_k}(x)$. Then we define the integral of $s$ to be 
\[
\int_E s \, dx = \sum_{k=1}^{m} c_k m(F_k).
\]

This is a very natural definition of the integral for simple functions. However, we would like something more general since this definition relies exclusively on using the standard representation. We eventually show this definition can be extended to an arbitrary representation of the standard function.

**Lemma 12.4.** Suppose $s(x) = \sum_{n=1}^{m} a_n \chi_{E_n}(x)$ on a set of finite measure $E$ such that $\{E_n\}$ is a disjoint family. Then 
\[
\int_E s \, dx = \sum_{n=1}^{m} a_n m(E_n).
\]

**Proof.** The proof follows easily from the fact that if any of the coefficients overlap, we simply group the corresponding sets together to form a new family of disjoint sets. This new family of disjoint sets corresponds to the standard form of $s$ to which we apply the definition of the integral given in Definition 12.3. The fact that both sums give the same value follows from distributivity property of real numbers as well as the additivity property of $m$ for disjoint families. \qed

**Note 12.5.** Lemma 12.4 provides a more general notion of the integral, not relying on the simple function being in standard form, but includes the case when the simple function is in standard form. It is still not general enough, though, since there are still restrictions on the representation of our simple function.

**Definition 12.6.** A linear function class $K$ is any set of functions such that if $f, g \in K$, then $\alpha f + \beta g \in K$ for all $\alpha, \beta \in \mathbb{R}$. This essentially splits into two conditions.

1. $\alpha f \in K$ for all $f \in K$ and $\alpha \in \mathbb{R}$, and
2. $f + g \in K$ for all $f, g \in K$. 
Note 12.7. To be a linear function class does not mean to be a class of linear functions. Be careful of this distinction in wording. For instance, we note that for all $\alpha \in \mathbb{R}$ we have that

$$f_{\alpha}(x) = \begin{cases} \alpha \sin(x) & \text{if } x \in [-\pi, \pi) \\ 0 & \text{otherwise} \end{cases}$$

clearly forms a linear function class. However, none of the functions in $\{f_{\alpha}\}$ is a linear function. This class is actually a subset of $K_2$, which we discuss later, since these functions are all bounded, continuous (and hence measurable) functions supported on a set of finite measure.

Definition 12.8 (Properties of Integration.). Let $K$ be a linear function class. The integral

$$\int : K \to (-\infty, \infty]$$

must satisfy the following axioms for all $f, g \in K$ and $\alpha, \beta \in \mathbb{R}$:

1. (Linearity.) $\int_E (\alpha f + \beta g) \, dx = \alpha \int_E f \, dx + \beta \int_E g \, dx$, that is,
   (a) $\int_E \alpha f \, dx = \alpha \int_E f \, dx$, and
   (b) $\int_E (f + g) \, dx = \int_E f \, dx + \int_E g \, dx$,

2. (Monotonicity.) If $f(x) \leq g(x)$ for all $x \in E$, then $\int_E f \, dx \leq \int_E g \, dx$, and

3. (Positivity.) If $f(x) \geq 0$ for all $x \in E$, then $\int_E f \, dx \geq 0$.

Lemma 12.9. Assume positivity and linearity hold. Then $\int_E 0 = 0$.

Proof. Let $f(x) = 0$ for all $x \in E$. Then $f(x) \geq 0$ for all $x \in E$, and by positivity, it follows that

$$\int_E f \, dx \geq 0. \quad (17)$$

On the other hand, since $f(x) \leq 0$ for all $x \in E$, we have $-f(x) \geq 0$ for all $x \in E$. Thus, positivity again yields $\int_E -f \, dx \geq 0$, which gives by linearity

$$\int_E f \, dx \leq 0. \quad (18)$$

Combining inequalities (17) and (18) gives our result. \qed

Note 12.10. The properties in Definition 12.8 are not independent of each other. In particular, if we know the linearity and positivity properties, then monotonicity follows as a result.

Proof. Let $f, g \in K$ for which $f(x) \leq g(x)$ for all $x \in E$. Then $h = g - f \in K$ is a function for which $h(x) \geq 0$ for all $x \in E$, and so positivity gives $\int_E h \, dx \geq 0$; it follows by linearity that

$$\int_E g \, dx - \int_E f \, dx = \int_E (g - f) \, dx = \int_E h \, dx \geq 0,$$

completing the proof. \qed

12.1. Simple Functions.

Example 12.11 (Linear function class of simple functions). Let $K_1 = \{\text{simple functions on } \mathbb{R}\}$. First of all, it is clear that if $s$ is a simple function, then so is $\alpha s$ for any $\alpha \in \mathbb{R}$. Furthermore, Lemma 11.5 completes the argument that $K_1$ is a linear function class. We prove that the definition for integration given in Definition 12.3 satisfies the properties of integration.
Proof. (3) Suppose $s$ is a simple function with standard representation

$$s(x) = \sum_{n=1}^{m} c_n \chi_{F_n}(x)$$

for which $s(x) \geq 0$ for all $x \in \mathbb{R}$. We must show $\int_{\mathbb{R}} s \, dx \geq 0$. Since $s(x) \geq 0$ for all $x \in \mathbb{R}$, it follows by necessity that $c_n \geq 0$ for all $n \in \{1, \ldots, m\}$. Hence, we clearly have $c_n m(F_n) \geq 0$ for all $n \in \{1, \ldots, m\}$, so

$$\int_{\mathbb{R}} s \, dx = \sum_{n=1}^{m} c_n m(F_n) \geq 0.$$

This completes the proof of (3).

(1) Suppose $s_1, s_2 \in K_1$, and suppose

$$s_1(x) = \sum_{n=1}^{m} a_n \chi_{E_n}(x) \quad \text{and} \quad s_2(x) = \sum_{j=1}^{p} b_j \chi_{F_j}(x)$$

are the standard representations of $s_1$ and $s_2$, respectively. Let $G_{n,j} = E_n \cap F_j$. Then $\{G_{n,j}\}$ is a finite family of disjoint sets, since $\{E_n\}$ and $\{F_j\}$ are both finite families of disjoint sets. Thus we can write

$$(s_1 + s_2)(x) = \sum_{n,j} (a_n + b_j) \chi_{G_{n,j}}(x),$$

which is “almost” in standard form (meaning that the only possibility for it to not be in standard form is for coefficients to overlap). By Lemma 12.4, we have

$$\int_{\mathbb{R}} (s_1 + s_2) \, dx = \sum_{n,j} (a_n + b_j) m(E_n \cap F_j)$$

$$= \sum_{n,j} a_n m(E_n \cap F_j) + \sum_{n,j} b_j m(E_n \cap F_j)$$

$$= \sum_{n} a_n \sum_{j} m(E_n \cap F_j) + \sum_{j} b_j \sum_{n} m(E_n \cap F_j)$$

$$= \sum_{n} a_n m(E_n) + \sum_{j} b_j m(F_j)$$

$$= \int_{\mathbb{R}} s_1 \, dx + \int_{\mathbb{R}} s_2 \, dx,$$

as desired. Now we must also show that the integral is linear with respect to multiplication by a constant. Suppose

$$s(x) = \sum_{n=1}^{m} a_n \chi_{E_n}(x)$$
is a simple function in standard form. Then

\[ \alpha s(x) = \sum_{n=1}^{m} \alpha a_n \chi_{E_n}(x) \]

is the standard form of \( \alpha s \). Hence

\[ \int \alpha s \, dx = \sum_{n=1}^{m} \alpha a_n m(E_n) = \alpha \sum_{n=1}^{m} a_n m(E_n) = \alpha \int s \, dx, \]

as desired to complete the proof of (1).

(2) This proof is unnecessary because of Note 12.10, but we prove anyway. Let \( s_1, s_2 \in K_1 \) be such that \( s_1(x) \leq s_2(x) \) for all \( x \in \mathbb{R} \). Then by Lemma 11.5 we know \( s = s_2 - s_1 \in K_1 \) and \( s(x) \geq 0 \) for all \( x \in \mathbb{R} \). Then by positivity, we have \( \int s \, dx \geq 0 \), and by linearity, we have

\[ \int s \, dx = \int (s_2 - s_1) \, dx = \int s_2 \, dx - \int s_1 \, dx \geq 0, \]

which implies \( \int s_2 \, dx \geq \int s_1 \, dx \), as desired to complete the proof of (2). \( \square \)

Note 12.12. The difficulty that arises in this proof, particularly in the proof of (1), is that the sum of standard representations is not a standard representation.

We are now finally in a position to establish the most general form of integration for simple functions that does not rely on the representation chosen. We will see that linearity takes care of the non-uniqueness problem.

Lemma 12.13. Let \( s(x) = \sum_{n=1}^{m} a_n \chi_{E_n}(x) \) be a simple function (with no restriction on the representation). Then

\[ \int s \, dx = \sum_{n=1}^{m} a_n m(E_n). \]

Proof. Note that \( s = s_1 + s_2 + \cdots + s_m \) where \( s_n = a_n \chi_{E_n} \), which is in standard form for all \( n \in \{1, \ldots, m\} \). Then by linearity of integration established in the preceding example, we have

\[
\int s \, dx = \int s_1 \, dx + \int s_2 \, dx + \cdots + \int s_m \, dx
= \int a_1 \chi_{E_1} \, dx + \int a_2 \chi_{E_2} \, dx + \cdots + \int a_m \chi_{E_m} \, dx
= a_1 m(E_1) + a_2 m(E_2) + \cdots + a_m m(E_m),
\]

as desired to complete the proof. \( \square \)

12.2. Bounded Measurable Functions Supported on a Set of Finite Measure.

Definition 12.14. Let \( f \) be a function. Then the support of \( f \) is \( \text{supp}(f) = \{ x \in \mathbb{R} : f(x) \neq 0 \} \).

Example 12.15 (Linear function class of bounded measurable functions supported on a set of finite measure). Define the class

\[ K_2 = \{ \text{all bounded measurable functions supported on a set of finite measure} \}. \]

That is, we have two conditions for \( f \in K_2 \):
(1) \(|f(x)| < M_f|\) for some \(M_f \in \mathbb{R}\) and for all \(x \in \mathbb{R}\), and
(2) \(m(\text{supp}(f)) < \infty\).

It is easy to see that this is a linear class. In particular, if \(f\) is bounded by \(M_f\), then \(\alpha f\) is bounded by \(\alpha M_f\), and if \(f\) and \(g\) are supported on the sets \(U\) and \(V\), respectively, then \(f + g\) is supported on the set \(U \cup V\), which has measure at most \(m(U) + m(V)\), which is finite.

**Definition 12.16** (The Lebesgue Integral on \(K_2\)). Suppose \(f \in K_2\). Then we define the lower Lebesgue integral as
\[
\int f^- dx = \sup_{\varphi \in K_1, \varphi \leq f} \left\{ \int \varphi dx \right\}
\]
and the upper Lebesgue integral as
\[
\int f^+ dx = \inf_{\psi \in K_1, \psi \geq f} \left\{ \int \psi dx \right\}.
\]
A function \(f \in K_2\) is integrable if \(\int f^- dx = \int f^+ dx\), and we define
\[
\int f dx = \int f^- dx = \int f^+ dx.
\]

**Lemma 12.17.** Every function \(f \in K_2\) is integrable.

*Proof.* Let \(f \in K_2\) and let \(E = \text{supp}(f)\). By the Simple Approximation Lemma (Lemma 11.3), there exists \(\varphi_n, \psi_n \in K_1\) supported on \(E\) such that \(\varphi_n \leq f \leq \psi_n\) and \(\psi_n - \varphi_n < \frac{1}{n}\). Note, then, that is means
\[
\psi_n(x) - \varphi_n(x) \leq \frac{1}{n} \chi_E(x).
\]
Thus we have
\[
0 \leq \int f^+ dx - \int f^- dx \leq \int \psi_n dx - \int \varphi_n dx = \int (\psi_n - \varphi_n) dx \leq \frac{1}{n} \int \chi_E dx = \frac{1}{n} m(E).
\]
Letting \(n \to \infty\), we have \(\int f^+ dx = \int f^- dx\). \(\square\)

**Definition 12.18.** Let \(f \in K_2\) and \(E \in M\). Then we define
\[
\int_E f dx = \int f \chi_E dx,
\]
noting that \(f \chi_E \in K_2\) since \(f\) and \(\chi_E\) are both bounded measurable functions, and since \(\text{supp}(f \chi_E) \subset \text{supp}(f)\).

**Theorem 12.19.** The definition of the Lebesgue integral given in Definition 12.16 satisfies the following properties.

1. **(Linearity.)** The integral is a linear operation on \(K_2\). That is,
   (a) For all \(\alpha \in \mathbb{R}\) and \(f \in K_2\), we have \(\int \alpha f dx = \alpha \int f dx\), and
   (b) For all \(f, g \in K_2\), we have \(\int (f + g) dx = \int f dx + \int g dx\).
2. **(Monotonicity.)** Suppose \(f, g \in K_2\) with \(f \leq g\). Then \(\int f dx \leq \int g dx\).
3. **(Positivity.)** Suppose \(f \in K_2\) such that \(f(x) \geq 0\) for all \(x \in \mathbb{R}\). Then \(\int f dx \geq 0\).
(4) (Additivity with respect to sets.) If \( E, F \in \mathcal{M} \) and \( E \cap F = \emptyset \), then
\[
\int_E f \, dx + \int_F f \, dx = \int_{E \cup F} f \, dx.
\]

Note 12.20. We also ask the following question: Can we pass to the limit under the integral sign if the sequence of functions converges almost everywhere? That is, suppose \( \{f_n\} \subset \mathcal{M} \) and \( f \in \mathcal{M} \) and \( \lim_{n \to \infty} f_n(x) = f(x) \) almost everywhere. Suppose also that \( f_n \in K_2 \) for all \( n \in \mathbb{N} \) and \( f \in K_2 \). Is it true that
\[
\lim_{n \to \infty} \int f_n \, dx = \int \left( \lim_{n \to \infty} f_n \right) \, dx = \int f \, dx.
\]

The answer is no, at least with this much generality. We provide two counter-examples.

First, consider the functions \( f_n \colon [0, 1] \to [0, \infty) \) given by
\[
f_n(x) = \begin{cases} 
n & \text{if } x \in \left[0, \frac{1}{n}\right) \\
0 & \text{if } x \in \left[\frac{1}{n}, 1\right]
\end{cases}
\]

We note that
\[
\lim_{n \to \infty} f_n(x) = \begin{cases} 
\infty & \text{if } x = 0 \\
0 & \text{if } x \in (0, 1]
\end{cases}
\]

Now \( \{f_n\} \subset K_2 \) and \( f_n \to 0 \) almost everywhere. We also have \( \int f_n \, dx = 1 \) for all \( n \in \mathbb{N} \). However, \( \int 0 \, dx = 0 \) but \( 1 \not\to 0 \). This completes the first counter-example. The problem here was that the functions \( \{f_n\} \) were not uniformly pointwise bounded, though they were all uniformly supported on a set of finite measure, namely \([0, 1]\).

Next, consider \( g_n \colon [0, \infty) \to [0, 1] \) given by
\[
g_n(x) = \begin{cases} 
1 & \text{if } x \in \left[n, n+1\right) \\
0 & \text{otherwise}
\end{cases}
\]

Note that \( \{g_n\} \subset K_2 \) and also \( \lim_{n \to \infty} g_n(x) = 0 \) for all \( x \in [0, \infty) \), so \( g_n \to 0 \) point-wise. Now note also that \( \int g_n \, dx = 1 \) for all \( n \in \mathbb{N} \), but once again, \( 1 \not\to 0 \). Hence, being uniformly bounded alone is not enough. What failed here is that they functions were not uniformly supported on a set of finite measure. Both of these examples violate the Bounded Convergence Theorem, which we will discuss later.

Proof of Properties Given in Theorem 12.19. (3) Suppose \( f \in K_2 \) such that \( f \geq 0 \) for all \( x \in \mathbb{R} \). Note that \( \int f \, dx = \int^+ f \, dx \) since \( f \in K_2 \) and hence is integrable. Hence we have
\[
\int f \, dx = \int^+ f \, dx = \inf_{\psi \geq f} \int_{\psi \in K_1} \psi \, dx \geq \inf_{\psi \geq 0} \int_{\psi \in K_1} \psi \, dx \geq 0
\]

since it is the infimum of non-negative numbers. This establishes the positivity property. \(\square\)

(1a) Let \( \alpha \in \mathbb{R} \) and \( f \in K_2 \).

Case 1: \( \alpha = 0 \); this case is trivial.

Case 2: \( \alpha > 0 \). Note \( \alpha f \in K_2 \) as well. Thus
\[
\int \alpha f \, dx = \int^+ \alpha f \, dx = \inf_{\psi \geq \alpha f} \left\{ \int_{\psi \in K_1} \psi \, dx \right\} = \alpha \inf_{\psi \geq 0} \left\{ \int_{\psi \alpha \in K_1} \frac{\psi}{\alpha} \, dx \right\}
\]
\[
\int_0^\infty f(x) \, dx = \int_0^\infty g(x) \, dx,
\]
where \(\xi = \frac{\psi}{\alpha} \in K_1\) is a simple function.

**Case 3:** \(\alpha < 0\). Note once again that \(f \in K_2\). Hence we have

\[
\int_0^\infty f(x) \, dx = \int_0^\infty g(x) \, dx = \int_0^\infty h(x) \, dx,
\]
where \(\xi = \frac{\psi}{\alpha} \in K_1\), as desired to complete the proof of (1a).

(1b) Let \(f, g \in K_2\). Let \(\epsilon > 0\) be given. Then choose \(\varphi_1 \in K_1\) such that \(\varphi_1 \leq f\) and such that \(\int \varphi_1 \, dx \geq \int f \, dx \geq \frac{\epsilon}{2}\), which is guaranteed to exist by the definition of a supremum. Likewise, choose \(\varphi_2 \in K_1\) such that \(\varphi_2 \leq g\) and \(\int \varphi_2 \, dx \geq \int g \, dx \geq \frac{\epsilon}{2}\). Then \(\varphi_1 + \varphi_2 \leq f + g\) and we have

\[
\int (f + g) \, dx = \int (\varphi_1 + \varphi_2) \, dx \geq \int f \, dx + \int g \, dx - \epsilon.
\]

Letting \(\epsilon \to 0\), we have \(\int (f + g) \, dx \geq \int f \, dx + \int g \, dx\).

Now we make a similar argument for the other case. Choose \(\psi_1 \in K_1\) such that \(\psi_1 \geq f\) and such that \(\int \psi_1 \, dx \leq \int f \, dx + \frac{\epsilon}{2}\), which is guaranteed to exist by the definition of an infimum. Likewise, choose \(\psi_2 \in K_1\) such that \(\psi_2 \geq g\) and \(\int \psi_2 \, dx \leq \int g \, dx + \frac{\epsilon}{2}\). Then \(\psi_1 + \psi_2 \geq f + g\) and we have

\[
\int (f + g) \, dx = \int (\psi_1 + \psi_2) \, dx \leq \int f \, dx + \int g \, dx + \epsilon.
\]

Letting \(\epsilon \to 0\), we have \(\int (f + g) \, dx \leq \int f \, dx + \int g \, dx\), and hence we have equality.

(2) Linearity and positivity imply monotonicity by Note 12.10.

(4) We note that

\[
\int_E f \, dx + \int_F f \, dx = \int f \chi_E \, dx + \int f \chi_F \, dx = \int f (\chi_E + \chi_F) \, dx = \int f \chi_{E \cup F} \, dx = \int_{E \cup F} f \, dx,
\]

where the first and fourth equalities follow from Definition 12.18 and where the second equality follows from linearity and the third equality follows from the fact that \(E\) and \(F\) are disjoint.

**Lemma 12.21.** Let \(f \in K_2\). Then \(|f| \in K_2\) and \(|\int f \, dx| \leq \int |f| \, dx\).
Proof. Recall that $f = f^+ - f^-$ and $|f| = f^+ + f^-$. Note also that $f^+ = f \chi_{\{f > 0\}}$ and $f^- = -f \chi_{\{f < 0\}}$, from which it follows that $f^+, f^- \in K_2$. Hence $f^+ + f^- = |f| \in K_2$. Now we have

\[
\left| \int f \, dx \right| = \left| \int (f^+ - f^-) \, dx \right|
\]

\[
= \left| \int f^+ \, dx - \int f^- \, dx \right|
\]

\[
\leq \int f^+ \, dx + \int f^- \, dx
\]

\[
= \int (f^+ + f^-) \, dx
\]

\[
= \int |f| \, dx,
\]

which completes the proof. \qed

The next result is the first convergence theorem we consider.

**Theorem 12.22 (Bounded Convergence Theorem).** Let $E \in \mathcal{M}$ with $m(E) < \infty$. Also, let \( \{f_n\} \subset K_2 \) be a family of functions defined on $E$. Assume that the family \( \{f_n\} \) is uniformly pointwise bounded on $E$. That is, there exists an $M > 0$ such that

\[
|f_n(x)| \leq M \quad \text{for all } x \in E \text{ and for all } n \in \mathbb{N}.
\]

Furthermore, assume that $\lim_{n \to \infty} f_n(x) = f(x)$ almost everywhere on $E$. Then $f \in K_2$ and

\[
\lim_{n \to \infty} \int_E f_n \, dx = \int_E f \, dx.
\]

Proof. It is clear that $f \in K_2$. We note from the previous lemma that

\[
\left| \int_E f \, dx - \int_E f_n \, dx \right| = \left| \int_E (f - f_n) \, dx \right| \leq \int_E |f - f_n| \, dx.
\]

Now take any set $A \subset E$ where $A \in \mathcal{M}$. Then we have

\[
\int_E |f - f_n| \, dx = \int_A |f - f_n| \, dx + \int_{E \setminus A} |f - f_n| \, dx
\]

\[
\leq \int_A |f - f_n| \, dx + \int_{E \setminus A} \underbrace{|f| \, dx}_{\leq M \cdot m(E \setminus A)} + \int_{E \setminus A} |f_n| \, dx
\]

\[
\leq \int_A |f - f_n| \, dx + 2M \cdot m(E \setminus A).
\]

Fix $\epsilon > 0$. By Egorov’s Theorem, there exists a set $A_\epsilon \in \mathcal{M}$ with $A_\epsilon \subset E$ and such that $m(E \setminus A_\epsilon) < \frac{\epsilon}{4M}$ and $f_n \to f$ uniformly on $A_\epsilon$. Hence by the previous observations we have

\[
\left| \int_E f \, dx - \int_E f_n \, dx \right| \leq \int_{A_\epsilon} |f - f_n| \, dx + \frac{\epsilon}{2} < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon
\]

if $n > n_0$ (where $n_0$ is defined such that $|f - f_n| < \frac{\epsilon}{2m(A_\epsilon)}$ for all $n > n_0$). This completes the proof. \qed
12.3. Non-Negative Measurable Functions.

**Example 12.23** (Class of all non-negative measurable functions). Define the class
\[ K_3 = \{ \text{all non-negative measurable functions} \} . \]

This class is not a linear class because of subtraction and multiplication by a negative scalar.

**Definition 12.24** (The Lebesgue Integral on \( K_3 \)). We define the integral for \( K_3 \) to be
\[ \int_E f \, dx = \sup_{g \in K_2, 0 \leq g \leq f} \left\{ \int_E g \, dx \right\} . \]

(We do not use the infimum in this definition, as in the class \( K_2 \), because this may be the infimum of an empty set, a scenario we would like to avoid.) Note the possibility of \( \int_E f \, dx = \infty \).

**Note 12.25.** The integral on the class \( K_3 \) also satisfies the properties of

1. Linearity,
2. Monotonicity,
3. Additivity with respect to sets, and
4. Positivity.

The proofs of such properties closely mimic the proofs for the class \( K_2 \), so we omit them (with the exception of the linearity property).

**Proof of Linearity Property.** We must show that \( \int (f + g) \, dx = \int f \, dx + \int g \, dx \). Let \( h_1, h_2 \in K_2 \) be such that \( 0 \leq h_1 \leq f \) and \( 0 \leq h_2 \leq g \). Then \( h_1 + h_2 \leq f + g \). Now
\[ \int (f + g) \, dx = \sup_{h \in K_2, 0 \leq h \leq f + g} \left\{ \int h \, dx \right\} \geq \int (h_1 + h_2) \, dx = \int h_1 \, dx + \int h_2 \, dx . \]

Now let \( \epsilon > 0 \) and suppose we found \( h_1, h_2 \in K_2 \) such that \( \int h_1 \, dx \geq \int f \, dx - \frac{\epsilon}{2} \) and \( \int h_2 \, dx \geq \int g \, dx - \frac{\epsilon}{2} \), which are guaranteed to exist by the definition of the integral for \( K_2 \) being a supremum. Then we would have
\[ \int h_1 \, dx + \int h_2 \, dx \geq \int f \, dx + \int g \, dx - \epsilon . \]

Letting \( \epsilon \to 0 \), we have
\[ \int (f + g) \, dx \geq \int f \, dx + \int g \, dx . \] (19)

Now take \( \ell \in K_2 \) with \( 0 \leq \ell \leq f + g \). Let \( h = \min \{ \ell, f \} \) and \( k = \ell - h \). Clearly \( h \in K_2 \) and hence \( k \in K_2 \). Furthermore, \( k(x) \geq 0 \) since \( 0 \leq h \leq \ell \). It is immediate that \( h \leq f \). We claim that \( k \leq g \). First, if \( \ell(x) \leq f(x) \), then \( h(x) = \ell(x) \), and so \( k(x) = 0 \), from which it follows that \( k \leq g \). Secondly, if \( \ell(x) > f(x) \), then \( h(x) = f(x) \), and so \( k(x) = \ell(x) - f(x) \leq g(x) \) by our original assumption that \( 0 \leq \ell \leq f + g \). We have established our claim that \( k \leq g \). Hence, we have
\[ \int f \, dx + \int g \, dx \geq \int h \, dx + \int k \, dx = \int (h + k) \, dx = \int \ell \, dx . \]
Since \( \int \ell \, dx \leq \int f \, dx + \int g \, dx \) for an arbitrary function \( \ell \in K_2 \) for which \( 0 \leq \ell \leq f + g \), it follows by the definition of \( \int (f + g) \, dx \) that
\[
\int (f + g) \, dx \leq \int f \, dx + \int g \, dx. \tag{20}
\]
Combining (19) and (20) finishes the proof. \( \square \)

**Theorem 12.26** (Chebychev’s Inequality). Suppose \( f \in K_3 \) and let \( \lambda > 0 \). Then
\[
m(\{ f > \lambda \}) \leq \frac{1}{\lambda} \int f \, dx.
\]

**Proof.** Note that
\[
\int f \, dx \geq \int_{\{f > \lambda\}} f \, dx \geq \int_{\{f > \lambda\}} \lambda \, dx = \lambda m(\{ f > \lambda \}),
\]
from which the result follows. \( \square \)

**Lemma 12.27.** Let \( f \in K_3 \). If \( \int f \, dx = 0 \), then \( f(x) = 0 \) almost everywhere.

**Proof.** By Chebyshev’s Inequality, we have \( m\{ f > \lambda \} \leq \frac{1}{\lambda} \int f \, dx \) for \( \lambda > 0 \). Hence \( m\{ f > \frac{1}{n} \} = 0 \) for all \( n \geq 1 \) by assumption since \( \int f \, dx = 0 \). Now consider that
\[
\{ f > 0 \} = \bigcup_{n=1}^{\infty} \left\{ f > \frac{1}{n} \right\},
\]
which implies \( m\{ f > 0 \} = 0 \), as desired. \( \square \)

**Definition 12.28.** A function \( f \in K_3 \) is called **Lebesgue integrable** if \( \int f \, dx < \infty \). (We will call this class \( L^1_+ \), which will be extended later.)

**Note 12.29.** Note that if \( f \in L^1_+ \) and \( \alpha \geq 0 \), then we have \( \alpha f \in L^1_+ \). Furthermore, if \( f, g \in L^1_+ \), then \( f + g \in L^1_+ \).

### 12.4. Integrable Functions: \( L^1 \).

**Example 12.30** (Linear function class of all integrable functions). Define the class
\[
K_4 = \{ \text{all integrable functions} \}.
\]
We say that \( f \) is **integrable** if \( f \) is measurable and \( \int |f| \, dx < \infty \), where the integral is as defined in the \( K_3 \) case (since \( |f| \in K_3 \)).

**Definition 12.31** (The Lebesgue Integral on \( K_4 \)). Recall that \( |f|, f^+, \) and \( f^- \) are all in \( K_3 \) and \( |f| \geq f^+ \) and \( |f| \geq f^- \), so by monotonicity we have \( 0 \leq \int f^+ \, dx \leq \int |f| \, dx < \infty \) and \( 0 \leq \int f^- \, dx \leq \int |f| \, dx < \infty \). Hence, we define the **integral** for \( f \in K_4 \) as
\[
\int f \, dx = \int f^+ \, dx - \int f^- \, dx.
\]

**Note 12.32** (Properties of the integral on \( K_4 \)). The family of all integrable functions is denoted by \( L^1 \) and has the following properties.

1. \( L^1 \) is a linear function class.
2. (Monotonicity.) For \( f, g \in L^1 \) with \( f \leq g \), we have \( \int f \leq \int g \).
(3) (Linearity.) Let $f, g \in L^1$ and $\alpha \in \mathbb{R}$.
(a) $\int \alpha f = \alpha \int f$, and
(b) $\int (f + g) = \int f + \int g$.

(4) (Additivity with respect to disjoint sets.) $\int_{E \cup F} f = \int_E f + \int_F f$ provided $E, F \in \mathcal{M}$ and $E \cap F = \emptyset$.

(5) (Positivity.) If $f \geq 0$ then $\int f \geq 0$. (This follows from $K_3$.)

(6) As is the case for $K_3$ (i.e. Lemma 12.21), we have $\left| \int f \, dx \right| \leq \int |f| \, dx$.

Proof. Most of these proofs are routine. We prove (3b), however, which is more difficult because the sum of positive parts of two functions is not the positive part of the sum of two functions. Note that $(f + g) = (f + g)^+ - (f + g)^-$ and also $(f + g) = f^+ - f^- + g^+ - g^-$; hence, we have

$$f^+ - f^- + g^+ - g^- = (f + g)^+ - (f + g)^-,$$

from which it follows that

$$\begin{align*}
\left( \underbrace{f^+ + g^+}_{\in K_3} \right) + (f + g)^- = (f + g)^+ + \left( \underbrace{f^- + g^-}_{\in K_3} \right).
\end{align*}$$

Hence, by the linearity of $K_3$ we have

$$\int f^+ + \int g^+ + \int (f + g)^- = \int (f + g)^+ + \int f^- + \int g^-.$$

Re-organizing, we get

$$\int f^+ - \int f^- + \int g^+ - \int g^- = \int (f + g)^+ - \int (f + g)^-.$$

from which it follows that

$$\int f + \int g = \int (f + g),$$

as desired. \qed

Analysis I material stops here.
13. More Integration and Convergence Theorems

We begin with a review of concepts covered at the end of Analysis I.

**Note 13.1.** Recall that \( K_4 \) is the linear function class of all Lebesgue integrable functions. In particular, \( f \in K_4 \) provided that \( f \in \mathcal{M} \) and \( \int |f| \, dx < \infty \). (Recall that \( |f| \in K_3 \), and so the integral is defined for this function.)

**Definition 13.2** (Integral for \( K_4 \)). Let \( f \in K_4 \). Then recall that we can write \( f = f^+ - f^- \) and \( |f| = f^+ + f^- \). The **Lebesgue integral** for \( f \) is defined by

\[
\int f \, dx = \int f^+ \, dx - \int f^- \, dx.
\]

Note that \( f^+ \leq |f| \) and \( f^- \leq |f| \), and so we have \( 0 \leq \int f^+ \, dx \leq \int |f| \, dx < \infty \) and \( 0 \leq \int f^- \, dx \leq \int |f| \, dx < \infty \).

**13.1. Convergence Theorems.**

**Note 13.3** (Convergence Theorems). We now state and prove four convergence theorems, and which linear function class each applies to. (Please see the Analysis I section for the proof of the Bounded Convergence Theorem.)

1. Bounded Convergence Theorem (\( K_2 \))
2. Fatou’s Lemma (\( K_3 \))
3. Monotone Convergence Theorem (\( K_3 \))
4. Lebesgue Dominated Convergence Theorem (\( K_4 \))

From this point forward, all functions are considered to be measurable unless otherwise stated.

**Note 13.4.** The convergence theorems seek to answer the following general question: Given \( \{f_n\}_{n=1}^{\infty} \) and \( f \) such that \( f_n \to f \) almost everywhere or point-wise, can we pass to the limit under the integral? That is, does the following equality hold?

\[
\lim_{n \to \infty} \int f_n \, dx = \int \lim_{n \to \infty} f_n \, dx = \int f \, dx
\]

We restate the Bounded Convergence Theorem for ease of reference.

**Theorem 13.5** (Bounded Convergence Theorem). Suppose \( \{f_n\} \subset K_2 \) and suppose there is a function \( f \) such that \( f_n \to f \) point-wise (or almost everywhere). Let \( E \) be a set of finite measure for which \( \{x : f_n(x) \neq 0 \text{ for some } n \in \mathbb{N}\} \subset E \)

(that is, the collection \( \{f_n\} \) is uniformly supported on a set of finite measure). Furthermore, suppose there is some \( M > 0 \) such that \( |f_n(x)| \leq M \) for all \( n \geq 1 \) and all \( x \in E \). Then

\[
\lim_{n \to \infty} \int f_n \, dx = \int f \, dx.
\]

The next result is of fundamental importance whose statement and proof may be asked on a comprehensive exam.
Lemma 13.6 (Fatou’s Lemma (aka Fatso’s Lemma)). Let \( \{f_n\}_{n=1}^{\infty} \subset K_3 \) and suppose \( f_n \to f \) point-wise. Then

\[
\int f \, dx \leq \liminf_{n \to \infty} \int f_n \, dx.
\]

Proof. First note that \( f \in K_3 \) since \( f_n \in K_3 \) for all \( n \in \mathbb{N} \) (cf. Lemma 10.3). Now

\[
\int f \, dx = \sup_{0 \leq h \leq f, h \in K_2} \left\{ \int h \, dx \right\}.
\]

Thus it suffices to show that \( \int h \, dx \leq \liminf_{n \to \infty} \int f_n \, dx \) for all \( h \in K_2 \) with \( 0 \leq h \leq f \). Thus, choose an arbitrary function \( h \in K_2 \) such that \( 0 \leq h \leq f \). Since \( h \in K_2 \) there exists an \( M > 0 \) such that \( 0 \leq h(x) \leq M \) for all \( x \). Furthermore, we have \( m(E_0) < \infty \) where \( E_0 = \{ x : h(x) \neq 0 \} = \text{supp}(h) \). Define \( h_n(x) = \min\{h(x), f_n(x)\} \) and note that \( 0 \leq h_n(x) \leq h(x) \leq M \). Note also that \( \text{supp}(h_n) \subset E_0 \) (since if \( h_n(x) \neq 0 \), then \( h(x) \neq 0 \) by how \( h_n \) is defined and since \( 0 \leq h \leq f \)). It follows that \( \{h_n\}_{n \in \mathbb{N}} \subset K_2 \).

We claim that \( \lim_{n \to \infty} h_n(x) = h(x) \). We must only consider when \( x \in E_0 \), since outside of \( E_0 \) everything is 0. We divide \( \mathbb{N} \) into two sets. Let

\[ N_1 = \{ n \in \mathbb{N} : h(x) \leq f_n(x) \} \quad \text{and} \quad N_2 = \{ n \in \mathbb{N} : h(x) > f_n(x) \}. \]

If \( |N_2| < \infty \), then there exists an \( n_0 \in \mathbb{N} \) such that \( h(x) \leq f_n(x) \) for all \( n > n_0 \); that is, \( h_n(x) = h(x) \) for all \( n > n_0 \), implying obviously that \( \lim_{n \to \infty} h_n(x) = h(x) \). Thus the only case we must consider is when \( |N_2| = \infty \). Let \( N_2 = \{ n_j : j \in \mathbb{N} \} \) be the corresponding subsequence indexing set (i.e. introducing an order to \( N_2 \) so that we may talk about the subsequence whose elements are indexed by \( N_2 \), and note that

\[ f(x) = \lim_{n \to \infty} f_n(x) = \lim_{j \to \infty} f_{n_j}(x) \]

since the limit of a sequence (if it exists, which in this case it does) is the same as the limit of any subsequence. Since \( h(x) > f_{n_j}(x) \) for all \( j \in \mathbb{N} \), we have \( h(x) \geq \lim_{j \to \infty} f_{n_j}(x) = f(x) \). By our original assumption, we also have \( h(x) \leq f(x) \). Thus we must have \( h(x) = f(x) \). Thus

\[ \lim_{n \to \infty} h_n(x) = \lim_{n \to \infty} f_n(x) = f(x) = h(x), \]

where the first equality follows because \( h_n(x) = h(x) = f(x) \) for any \( n \in N_1 \) and \( h_n(x) = f_n(x) \) for any \( n \in N_2 \). This completes our claim.

We are now in a position to use the Bounded Convergence Theorem for the sequence \( \{h_n\}_{n \in \mathbb{N}} \). Hence, we have

\[ \lim_{n \to \infty} \int h_n \, dx = \int h \, dx. \]

Note that \( h_n \leq f_n \), so by monotonicity we have \( \int h_n \, dx \leq \int f_n \, dx \). Hence,

\[ \liminf_{n \to \infty} \int h_n \, dx \leq \liminf_{n \to \infty} \int f_n \, dx. \]

Notice that \( \liminf_{n \to \infty} \int h_n \, dx = \lim_{n \to \infty} \int h_n \, dx = \int h \, dx \), and so we have

\[ \int h \, dx \leq \liminf_{n \to \infty} \int f_n \, dx, \]

which completes the proof. \( \square \)
Note 13.7. This is not the most general form of Fatou’s Lemma. See both Royden, et. al. and Wheeden, et. al. for more general forms.

Theorem 13.8 (Monotone Convergence Theorem). Let \( \{f_n\} \subset K_3 \) and suppose \( f_n \nearrow \), where \( \lim_{n \to \infty} f_n(x) = f(x) \) and \( f \) is finite almost everywhere. Then
\[
\lim_{n \to \infty} \int f_n \, dx = \int f \, dx.
\]

Proof. By Fatou’s Lemma, we have
\[
\int f \, dx \leq \liminf_{n \to \infty} \int f_n \, dx.
\]
Since \( f_n \leq f \) (since \( f_n \) is an increasing sequence), we have \( \int f_n \, dx \leq \int f \, dx \) for all \( n \in \mathbb{N} \), and so
\[
\limsup_{n \to \infty} \int f_n \, dx \leq \int f \, dx.
\]
Thus we have
\[
\int f \, dx \leq \liminf_{n \to \infty} \int f_n \, dx \leq \limsup_{n \to \infty} \int f_n \, dx \leq \int f \, dx,
\]
and hence this chain of inequalities is “tight.” Thus the limit exists and
\[
\lim_{n \to \infty} \int f_n \, dx = \int f \, dx,
\]
as desired to finish the proof. \( \square \)

Theorem 13.9 (Lebesgue’s Dominated Convergence Theorem). Suppose \( \{f_n\} \subset L^1 \) with \( \lim_{n \to \infty} f_n(x) = f(x) \). Also assume the dominating condition: there exists a function \( g \in L^1 \) such that \( |f_n(x)| \leq g(x) \). Then \( |f(x)| \leq g(x) \) and hence \( f \in L^1 \) and
\[
\lim_{n \to \infty} \int f_n \, dx = \int f \, dx.
\]

Proof. The first conclusion is immediate. Now consider \( h_n = g - f_n \). Note that \( h_n \geq 0 \) since \( g \geq f_n \). Also, note that \( \lim_{n \to \infty} h_n(x) = g(x) - f(x) \). Using Fatou’s Lemma, we have \( \int (g - f) \, dx \leq \liminf_{n \to \infty} \int (g - f_n) \, dx \), and so
\[
\int g \, dx - \int f \, dx = \int (g - f) \, dx
\leq \liminf_{n \to \infty} \int (g - f_n) \, dx
= \liminf_{n \to \infty} \left[ \int g \, dx - \int f_n \, dx \right]
= \int g \, dx + \liminf_{n \to \infty} \left[ - \int f_n \, dx \right]
= \int g \, dx - \limsup_{n \to \infty} \int f_n \, dx.
\]
From this it follows that \( - \int f \, dx \leq - \limsup_{n \to \infty} \int f_n \, dx \), and ultimately that
\[
\limsup_{n \to \infty} \int f_n \, dx \leq \int f \, dx. \tag{21}
\]
Now consider the function \( s_n = g + f_n \), which again has the property that \( s_n \geq 0 \) from the dominating condition. Furthermore, \( \lim_{n \to \infty} s_n(x) = g(x) + f(x) \). By Fatou’s Lemma, we have 
\[
\int (g + f) \leq \lim \inf_{n \to \infty} \int (g + f_n) \, dx,
\]
and so 
\[
\int g \, dx + \int f \, dx = \int (g + f) \, dx 
\leq \lim \inf_{n \to \infty} (g + f_n) \, dx 
= \lim \inf_{n \to \infty} \left[ \int g \, dx + \int f_n \, dx \right] 
= \int g \, dx + \lim \inf_{n \to \infty} \int f_n \, dx,
\]
from which we get 
\[
\int f \, dx \leq \lim \inf_{n \to \infty} \int f_n \, dx \tag{22}
\]
Combining (21) and (22) we get 
\[
\int f \, dx \leq \lim \inf_{n \to \infty} \int f_n \, dx \leq \lim \sup_{n \to \infty} \int f_n \, dx \leq \int f \, dx.
\]
Thus this chain of inequalities is tight, and so everything is equal. Hence the limit exists and the common value is 
\[
\lim_{n \to \infty} \int f_n \, dx = \int f \, dx,
\]
as desired to complete the proof. \( \square \)

13.2. More Properties of Integration and \( L^1 \) Convergence.

**Theorem 13.10** (Countable Additivity of Lebesgue Integral). Suppose \( f \in L^1 \) and \( \{E_n\}_{n \in \mathbb{N}} \subset \mathcal{M} \) is a disjoint family of measurable sets. For convenience of notation, write \( E = \bigcup_{n \in \mathbb{N}} E_n \). Then 
\[
\int_E f \, dx = \sum_{n=1}^{\infty} \int_{E_n} f \, dx.
\]

**Proof.** Define \( G_m = \bigcup_{n=1}^{m} E_n \) and let \( f_m = f \cdot \chi_{G_m} \). First of all, it is clear that \( |f_m(x)| \leq |f(x)| \), and so we have a domination condition. Secondly, it is clear that \( \lim_{m \to \infty} f_m(x) = f(x) \) for all \( x \in E \) by how we defined \( G_m \). Thus by the Dominated Convergence Theorem we have \( \lim_{m \to \infty} \int f_m \, dx = \int_E f \, dx \). Thus we have 
\[
\int f_m \, dx = \int f \cdot \chi_{G_m} \, dx 
= \int_{G_m} f \, dx 
= \int_{\bigcup_{n=1}^{m} E_n} f \, dx 
= \sum_{n=1}^{m} \int_{E_n} f \, dx.
\]

The second equality follows by definition (consider a generalization of Definition 12.18) and the fourth equality follows from the same result for the finite case (cf. Note 12.32(4)). Hence,
\[
\int_E f \, dx = \lim_{m \to \infty} \int f_m \, dx = \lim_{m \to \infty} \sum_{n=1}^{m} \int_{E_n} f \, dx = \sum_{n=1}^{\infty} \int_{E_n} f \, dx,
\]
as desired to complete the proof.

**Definition 13.11** (Another mode of convergence). Let \( \{f_n\} \subset L^1 \) and let \( f \in L^1 \). Then we say \( \{f_n\} \) **converges to** \( f \) in \( L^1 \), denoted \( f_n \overset{L^1}{\to} f \), if
\[
\int |f - f_n| \, dx \to 0
\]
as \( n \to \infty \).

For now, we only make one implication in terms of the modes of convergence already given.

**Lemma 13.12.** \( L^1 \)-convergence implies convergence in measure.

**Proof.** We use Chebychev’s inequality for the functions \( |f - f_n| \in K_3 \). Fix \( \epsilon > 0 \). Then by Chebychev’s inequality we have
\[
m\{|f - f_n| > \epsilon\} \leq \frac{1}{\epsilon} \int |f - f_n| \, dx.
\]
Now \( \int |f - f_n| \, dx \to 0 \) as \( n \to \infty \) by assumption since \( f_n \overset{L^1}{\to} f \). Thus \( n\{|f - f_n| > \epsilon\} \to 0 \) as \( n \to \infty \); this is the definition of convergence in measure.
14. n-Dimensional Lebesgue Measure

14.1. Topology of $\mathbb{R}^n$. We begin by describing some topological concepts. Our basic set we work with is

$$\mathbb{R}^n = \{(x_1, \ldots, x_n) : x_i \in \mathbb{R}\}.$$  

We may use vector notation $\vec{x}$ (but we may be lazy and assume the reader knows that it is an $n$-dimensional vector; actually, we will probably always be lazy; hopefully the context will always be clear). Addition and scalar multiplication are as usual:

$$x + y = (x_1 + y_1, \ldots, x_n + y_n)$$

and

$$\alpha x = (\alpha x_1, \ldots, \alpha x_n).$$

We define the norm on a vector $x$ as follows:

$$\|x\| = \sqrt{x_1^2 + \cdots + x_n^2}.$$  

We define the distance between $x$ and $y$ to be the following:

$$d(x, y) = \|x - y\|.$$  

An open ball $B(x, r)$ is given by

$$B(x, r) = \{y : \|x - y\| < r\}.$$  

An open set $\mathcal{O}$ has the following property: for every $x \in \mathcal{O}$, there exists a positive number $r_x > 0$ such that $B(x, r_x) \subset \mathcal{O}$. A set $C$ is closed if and only if $\mathbb{R}^n \setminus C$ is open. Another characterization of a closed set is the following: A set $C$ is closed if and only if for every sequence $\{x_k\} \subset C$ such that $x_k \to x$ as $k \to \infty$, we have $x \in C$.

**Definition 14.1.** A set $A \subset \mathbb{R}^n$ is called **compact** if for every open cover $\{\mathcal{O}_\alpha\}_{\alpha \in A}$ of $A$ there is a finite subcover.

**Theorem 14.2** (The Heine-Borel Theorem). A set $A \subset \mathbb{R}^n$ is compact if and only if $A$ is closed and bounded.

**Lemma 14.3.** A set $A \subset \mathbb{R}^n$ is compact if and only if for every sequence $\{x_n\} \subset A$ there exists a subsequence $\{x_{n_k}\}$ such that $\lim_{k \to \infty} x_{n_k} = x \in A$.

The emphasis in the preceding lemma was the necessity for the limit point $x$ to be an element of $A$. If the same conditions hold while dropping the restriction of the limit point being an element of $A$, the set is called pre-compact.

**Definition 14.4.** If for every sequence $\{x_n\} \subset A$ there exists a subsequence $\{x_{n_k}\}$ such that $\lim_{k \to \infty} x_{n_k} = x$, where $x$ is not necessarily in $A$, then $A$ is said to be **pre-compact**.

**Note 14.5.** The set $A$ is pre-compact if and only if $\overline{A}$ is compact.

We now seek to build the Lebesgue measure on $\mathbb{R}^n$ by repeating essentially the same construction we used for the 1-dimensional case. In particular, we have a measure space $(\mathbb{R}^n, \mathcal{M}, m)$, where $\mathcal{M}$ is a $\sigma$-algebra containing $\mathcal{B}_n$ (the Borel sets in $\mathbb{R}^n$) and $m$ is a countably-additive set function.

Recall that $\mathcal{B}_n$ is defined to be the smallest $\sigma$-algebra containing all open sets in $\mathbb{R}^n$.

The “simple sets” in $\mathbb{R}^n$ are “rectangles.” For instance, in $\mathbb{R}^2$, they are literally rectangles and in $\mathbb{R}^3$ they are rectangular prisms. These sets may be open (no boundary) or closed (with boundary).
For an analytical approach to these rectangles, let \( x \in \mathbb{R}^n \). Then we say that \( x = (x_1, \ldots, x_n) \in R = [a_1, b_1] \times \cdots \times [a_n, b_n] \) if

\[
a_1 \leq x_1 \leq b_1, \ldots, a_n \leq x_n \leq b_n.
\]

(23)

Here, we may replace this closed rectangle with an open rectangle, making the inequalities in (23) strict. We choose rectangles as basic sets because we can measure the volume of a rectangle. In particular, the volume is given by

\[
|R| = (b_1 - a_1)(b_2 - a_2) \cdots (b_n - a_n).
\]

Note that the volume is the same for open rectangles, closed rectangles, or anything in between.

We now continue the generalization with the notion of outer measure.

### 14.2. Measure in \( \mathbb{R}^n \).

**Definition 14.6.** The outer measure is a set function \( m^* : \mathcal{P}(\mathbb{R}^n) \rightarrow [0, \infty] \) where for \( A \subset \mathbb{R}^n \), we have

\[
m^*(A) = \inf \left\{ \sum_{k=1}^{\infty} |C_k| : A \subset \bigcup_{k=1}^{\infty} C_k \right\},
\]

where the infimum is taken of all families \( \{C_k\}_{k \in \mathbb{N}} \) of closed cubes that cover \( A \).

**Homework Exercise (not collected):** We could have defined the outer measure instead in terms of closed rectangles, open cubes, or open rectangles. Prove that these four definitions give the same value of \( m^*(A) \).

**Note 14.7** (Main Properties of Outer Measure). The outer measure for \( \mathbb{R}^n \) has the same fundamental properties it has in the 1-dimensional case.

1. \( m^*(\emptyset) = 0 \).
2. (Monotonicity.) If \( A \subset B \), then \( m^*(A) \leq m^*(B) \).
3. (Countable sub-additivity.) If \( \{A_k\} \subset \mathcal{P}(\mathbb{R}^n) \), then \( m^*(\bigcup_{k=1}^{\infty} A_k) \leq \sum_{k=1}^{\infty} m^*(A_k) \).

We also repeat the Carathéodory construction for measurable sets.

**Definition 14.8.** We call a set \( E \subset \mathbb{R}^n \) measurable if for every set \( A \subset \mathbb{R}^n \) we have

\[
m^*(A) = m^*(A \cap E) + m^*(A \cap E^C).
\]

**Note 14.9.** As in the 1-dimensional case, to show measurability, it suffices to show

\[
m^*(A) \geq m^*(A \cap E) + m^*(A \cap E^C).
\]

The other inequality follows from the countable sub-additivity property.

**Note 14.10.** As in the 1-dimensional case, \( \mathcal{M} \) is a \( \sigma \)-algebra. We call this the Carathéodory \( \sigma \)-algebra. (We do not call it the Lebesgue \( \sigma \)-algebra because the definition of measurable sets is not the same as Lebesgue used, though the definitions are equivalent.) If we need to specify the dimension, we will write \( \mathcal{M}_n \).

We state, sometimes with proof, sometimes without, several properties of measurable sets. These results for the most part parallel the results given in the 1-dimensional case. If a proof is omitted, it is because it follows similarly to the 1-dimensional case.
Lemma 14.11. If $m^*(E) = 0$, then $E \in \mathcal{M}$.

Proof. Since $0 \leq m^*(A \cap E) \leq m^*(E) = 0$ by monotonicity, it suffices to show $m^*(A) \geq m^*(A \cap E^C)$, but this also follows from monotonicity. \qed

Definition 14.12. We say that a family of sets $\{E_n\}$ is **almost disjoint** if $\text{Int}(E_j) \cap \text{Int}(E_k) = \emptyset$ whenever $j \neq k$.

Definition 14.13. Let $\{U_{\alpha}\}_{\alpha \in A}$ be a cover of a space $X$. Then $\{V_{\beta}\}_{\beta \in B}$ is a **refinement** of $\{U_{\alpha}\}_{\alpha \in A}$ if it is a cover of $X$ and for every $\beta \in B$, there is an $\alpha \in A$ such that $V_{\beta} \subset U_{\alpha}$.

Lemma 14.14. For any closed rectangle $R$, we have $m^*(R) = |R|$.

Proof. It is clear that $m^*(R) \leq |R|$ since $R$ is an admissible cover of $R$. Therefore, we must show $m^*(R) \geq |R|$. Let $\{C_n\}_{n=1}^\infty$ be an admissible cover of $R$ of open rectangles. We seek to show

$$\sum_{n=1}^\infty |C_n| \geq |R|.$$  \hspace{1cm} (24)

Since $\{C_n\}$ is an arbitrary cover of open rectangles, showing (24) will imply that

$$m^*(R) = \inf \left\{ \sum_{n=1}^\infty |C_n| : R \subset \bigcup_{n=1}^\infty C_n, \text{ C_n is open for all } n \in \mathbb{N} \right\} \geq |R|,$$

which will finish the proof. Since $R$ is compact, take a finite subcover $\{C_{n_k}\}_{k=1}^m$. Now extend the edges of each rectangle $C_{n_k}$ and $R$ to form a collection of open rectangles $\{I_j\}_{j=1}^l$ such that $\{T_j\}_{j=1}^l$ is almost disjoint and such that the only sets in $\{T_j\}_{j=1}^l$ are those that $\text{Int}(I_j) \subset \text{Int}(C_{n_k}) \cap \text{Int}(R)$ for some $k \in \{1, \ldots, m\}$. See Figure 4. Note also that $R \subset \bigcup_{j=1}^l T_j$, and furthermore, we have

|Figure 4. (a) A rectangle $R$ covered by open rectangles. (b) $R$ covered by smaller closed rectangles.|

$$|R| = \sum_{j=1}^l |T_j| = \sum_{j=1}^l |I_j|$$ since $|T_j| = |I_j|$. Since $I_j \subset C_{n_k}$ for some $k \in \{1, \ldots, m\}$ for every
\( j \in \{1, \ldots, l\} \), it follows that

\[
|R| = \sum_{j=1}^{l} |I_j| \leq \sum_{k=1}^{m} |C_{n_k}| \leq \sum_{n=1}^{\infty} |C_n|,
\]

as desired to finish the proof. \( \square \)

**Lemma 14.15.** Suppose \( R = [a_1, b_1] \times \cdots \times [a_n, b_n] \) is a rectangle (where we may replace closed intervals with open or half-open intervals). Then \( R \in \mathcal{M} \).

**Proof.** First, we prove that \( m^*(\partial R) = 0 \), where \( \partial R \) is the boundary of \( R \). In particular, we may cover \( \partial R \) with \( \epsilon \)-thin rectangles of specified length; that is, each edge of \( \partial R \) can be written as \( E_{k,i} = [a_1, b_1] \times \cdots \times [e_{k,i} - \epsilon/2, e_{k,i} + \epsilon/2] \times \cdots \times [a_n, b_n] \), and so we cover this edge with a rectangle \( R_{k,i} = [a_1, b_1] \times \cdots \times [e_{k,i} - \epsilon/2, e_{k,i} + \epsilon/2] \times \cdots \times [a_n, b_n] \) for \( i = 1, 2 \). This rectangle is an admissible cover of this edge, and has volume \( |R_{k,i}| = \epsilon(b_1 - a_1) \cdots (b_{k-1} - a_{k-1})(b_{k+1} - a_{k+1}) \cdots (b_n - a_n) = \epsilon \cdot \frac{|R|}{(b_k - a_k)} \), which can be made as small as we wish since it is controlled by \( \epsilon \). Hence, \( \{R_{k,i}\} \) is an admissible cover of \( \partial R \), and so

\[
m^*(\partial R) \leq \sum_{i=1}^{2} \sum_{k=1}^{m} |R_{k,i}| = \sum_{k=1}^{n} 2\epsilon \cdot \frac{|R|}{(b_k - a_k)} = \epsilon \sum_{k=1}^{n} \frac{2|R|}{(b_k - a_k)} \to 0 \text{ as } \epsilon \to 0.
\]

This shows that \( m^*(\partial R) = 0 \), as desired.

Since \( m^*(\partial R) = 0 \) and hence by Lemma 14.11 we have \( \partial R \in \mathcal{M} \), and since \( R = \text{Int}(R) \cup \partial R \) (or some subset of the boundary, also in \( \mathcal{M} \)), we may assume that \( R \) is open to begin with. That is, showing that \( \text{Int}(R) \in \mathcal{M} \) will suffice to show \( R \in \mathcal{M} \), so we assume without loss of generality that \( R \) is open.

Since every open rectangle is the intersection of open half-planes, we must show every open half-plane is measurable. Let \( E \) be an open half-plane and let \( A \subseteq \mathbb{R}^n \) be any set. We wish to show \( m^*(A) \geq m^*(A \cap E) + m^*(A \cap E^C) \). Let \( \{C_k\} \) be any admissible cover for \( A \) of open rectangles. We define

\[
C_k' = C_k \cap E \quad \text{and} \quad C_k'' = C_k \cap E^C,
\]

which, when considering only non-empty sets, is a collection of open rectangles and closed rectangles, respectively (because the intersection of a rectangle with a half-plane is a rectangle). Now note that

\[
A \cap E \subseteq \bigcup_{k=1}^{\infty} (C_k \cap E) = \bigcup_{k=1}^{\infty} C_k' \quad \text{and} \quad A \cap E^C \subseteq \bigcup_{k=1}^{\infty} C_k \cap E^C = \bigcup_{k=1}^{\infty} C_k''
\]

and hence

\[
m^*(A \cap E) \leq \sum_{k=1}^{\infty} |C_k'| \quad \text{and} \quad m^*(A \cap E^C) \leq \sum_{k=1}^{\infty} |C_k''|.
\]

Finally, note that since \( C_k = C_k' \cup C_k'' \) is a disjoint union, we have \( |C_k| = |C_k'| + |C_k''| \) and so

\[
\sum_{k=1}^{\infty} |C_k| = \sum_{k=1}^{\infty} |C_k'| + |C_k''| \geq m^*(A \cap E) + m^*(A \cap E^C),
\]

Taking the infimum gives us \( m^*(A) \geq m^*(A \cap E) + m^*(A \cap E^C) \), as desired to finish the proof. \( \square \)
**Note 14.16.** From Lemmas 14.14 and 14.15, we get that $m(R) = |R|$ in the measure space $(\mathbb{R}^n, \mathcal{M}, m)$, where $m = m^*|\mathcal{M}$. This is particularly nice and parallels the 1-dimensional case of $m(I) = \ell(I)$, the length of the interval.

**Theorem 14.17 (Countable Additivity).** For a pairwise disjoint family $\{E_k\}_{k=1}^\infty \subset \mathcal{M}$, we have

$$m \left( \bigcup_{k=1}^\infty E_k \right) = \sum_{k=1}^\infty m(E_k).$$

**Note 14.18.** For a set $E \subset \mathbb{R}^n$ and vector $h \in \mathbb{R}^n$, we define the translation of $E$ by $h$ similarly to the 1-dimensional case. That is, $E + h = \{v + h : v \in E\}$.

**Note 14.19.** We make use of the following fairly obvious set equalities of translations.

1. $[(B - h) \cap E] + h = B \cap (E + h)$.
2. $E^C + h = (E + h)^C$.
3. $[(B - h) \cap E^C] + h = B \cap (E + h)^C$.

**Lemma 14.20.** The outer measure for $n$-dimensional Euclidean space is translation invariant.

**Proof.** Let $E \subset \mathbb{R}^n$. We must show two things. First, we must show that $\{C_n\}$ is a covering of $E$ consisting of cubes if and only if $\{C_n + h\}$ is a covering of $E + h$ consisting of cubes, and secondly, we must show that $|C_n| = |C_n + h|$. That $m^*(E) = m^*(E + h)$ will follow from these two facts.

For the former, note that if $C_n$ is a cube, it is clear that $C_n + h$ is a cube, and vice versa. Suppose $\{C_n\}$ is a cover of $E$ consisting of cubes. Let $x \in E + h$. Then $x - h \in E$, and so there is an $n \in \mathbb{N}$ for which $x - h \in C_n$. This means that $x \in C_n + h$, which shows $\{C_n + h\}$ is a cover for $E + h$ consisting of cubes. A similar argument establishes that if $\{C_n + h\}$ is a cover for $E + h$ consisting of cubes, then $\{C_n\}$ is a cover for $E$ consisting of cubes.

For the latter, note that if $C_n = [a_1, b_1] \times \cdots \times [a_n, b_n]$, and if $h = (h_1, \ldots, h_n)$, then $C_n + h = [a_1 + h_1, b_1 + h_1] \times \cdots \times [a_n + h_n, b_n + h_n]$, from which it follows that $|C_n| = |C_n + h|$. \hfill \Box

**Lemma 14.21.** The $n$-dimensional Lebesgue measure is translation invariant; that is, for $E \in \mathcal{M}$ and $h \in \mathbb{R}^n$, we have $E + h \in \mathcal{M}$ and $m(E) = m(E + h)$.

**Proof.** To show $E + h \in \mathcal{M}$, we must show

$$m^*(B) = m^*(B \cap (E + h)) + m^*(B \cap (E + h)^C)$$

(25)

for every $B \subset \mathbb{R}^n$. Since $E \in \mathcal{M}$, we have $m^*(A) = m^*(A \cap E) + m^*(A \cap E^C)$ for every $A \subset \mathbb{R}^n$. Therefore, choose $A = B - h \subset \mathbb{R}^n$. Then

$$m^*(B - h) = m^*((B - h) \cap E) + m^*((B - h) \cap E^C).$$

But the outer measure is translation invariant, and so we may shift each of these sets by $h$ (making use of Note 14.19) to get

$$m^*(B) = m^*(B \cap (E + h)) + m^*(B \cap (E + h)^C),$$

as desired to show $E + h \in \mathcal{M}$. The remaining assertion follows immediately since

$$m(E) = m^*(E) = m^*(E + h) = m(E + h).$$

The proof is complete. \hfill \Box
Definition 14.22. The Borel $\sigma$-algebra, denoted $B_n$, is the smallest $\sigma$-algebra of subsets of $\mathbb{R}^n$ that contains all the open sets of $\mathbb{R}^n$. Formally,

$$B_n = \bigcap_{\alpha \in A} S_\alpha,$$

where $\{S_\alpha\}_{\alpha \in A}$ is the collection of all $\sigma$-algebras that contain all open sets.

Lemma 14.23 (The $n$-dimensional Counterpart to the Structure Theorem (Theorem 2.11)). Any open set $O \subset \mathbb{R}^n$ can be represented as an almost disjoint countable union of closed cubes.

Proof. Let $O \subset \mathbb{R}^n$. We form countably many grids. Let $G_1$ be the grid with closed cubes, each of which has sides of length 1. See Figure 5 for an illustration of the first grid $G_1$ for an open set $O \subset \mathbb{R}^2$. This will form an almost disjoint family of closed cubes. At this stage, we consider three types of cubes in the grid.

1. Accepted cubes: Those cubes for which $C \subset O$. These cubes are shaded blue in the figure.
2. Rejected cubes: Those cubes for which $C \subset O^C$. These cubes are left white in the figure.
3. Tentatively accepted cubes: Those cubes for which $C \cap O \neq \emptyset$ and $C \cap O^C \neq \emptyset$. These cubes are shaded yellow in the figure.

To pass from $G_1$ to $G_2$, we follow the following steps.

1. We keep the accepted cubes.
2. We discard the rejected cubes.
3. We form the next grid $G_2$ of cubes whose sides are length $\frac{1}{2}$, only inside the tentatively accepted cubes from the grid $G_1$.
4. We repeat this procedure inside the tentatively accepted cubes.

Thus, we generalize this procedure to pass from grid $G_k$ to grid $G_{k+1}$.

1. We keep the accepted cubes in $G_k$. 

![Figure 5](image-url)
(2) We discard the rejected cubes in $G_k$.

(3) We form the next grid $G_{k+1}$ of cubes whose sides are of length $\frac{1}{2^k}$, only inside the tentatively accepted cubes from the grid $G_k$.

(4) Lather, rinse, repeat.

Now denote by $A$ the set of all accepted cubes from every stage of the construction. First of all, $A$ is a countable almost disjoint family of closed cubes, so we wish to show $O = \bigcup A$. It is clear by our construction that $\bigcup A \subseteq O$. It remains to show $O \subseteq \bigcup A$.

We provide two proofs for the remainder of the proof. The first is given by Dr. Gulisashvili in class. The second is a simpler proof I came up with myself.

**Remainder of Proof (Version 1).** Let $x \in O$. Now there is some closed cube $C$ with dyadic endpoints such that $x \in C \subseteq O$. If $C \in A$, then we are done. If $C \notin A$, then there is an “ancestor” $\tilde{C}$ of $C$ (i.e. cube from an earlier stage in the construction) such that $\tilde{C} \in A$ or $\tilde{C}$ was tentatively accepted. (To see this, note that the “first ancestor” of $C$, namely the cube $x$ resided in from the first grid, is such a cube.) If $\tilde{C} \in A$, then $x \in C \subseteq \tilde{C}$, and so we are done. Hence, the only case to consider is when $\tilde{C}$ was tentatively accepted. Consider the smallest cube which is an ancestor of $C$ and was tentatively accepted and call it $\tilde{C}_k$. Then $\tilde{C}_{k+1}$ was accepted (since $x \in O$ and $x \in \tilde{C}_{k+1}$) and furthermore $x \in \tilde{C}_{k+1} \in A$. This completes the proof.

**Remainder of Proof (Version 2).** Along with $A$, consider the set $R$ of all rejected cubes and the set $T$ of all tentatively accepted cubes after all stages in the construction have been completed. Let $G$ be the set of ALL cubes from the construction, so that $G = A \cup R \cup T$. Let $I \in T$. Then $I$ was a tentatively accepted cube at some stage in the process, but this is a contradiction because $I$ would be split into smaller cubes in the next stage of the construction. Hence, $T = \emptyset$, so $G = A \cup R$, which is a disjoint union. Furthermore, $\bigcup G = \mathbb{R}^n$. Let $x \in O$. Then $x \in \mathbb{R}^n = \bigcup G$, and so $x \in C$ for some $C \in G$. But $C \notin R$ since $x \in O \cap C$, and so $x \in C \in A$, as required to finish the proof.

**Note 14.24.** Dr. Gulisashvili made the comment that Lemma 14.23 was asked to be proven on the most recent comprehensive exam.

**Theorem 14.25.** $B_n \subseteq M_n$.

**Proof.** Let $O \subset \mathbb{R}^n$. By Lemma 14.23, we can write $O = \bigcup_{k=1}^{\infty} R_k$, where $\{R_k\}$ is an almost disjoint family of closed cubes. Since $M_n$ is a $\sigma$-algebra, it follows that $O \subseteq M_n$. Since $M_n$ is a $\sigma$-algebra that contains all the open sets of $\mathbb{R}^n$, it follows that $B_n \subseteq M_n$, as desired.

**14.3. Continuity and Regularity properties for $m_n$.** We are almost done with $n$-dimensional Lebesgue measure. We state (without proof) the generalizations of a few more results from the 1-dimensional case.

**Theorem 14.26** (Continuity Properties of $m_n$). The following are the $n$-dimensional versions of continuity from below and continuity from above, respectively.

1. Let $\{E_k\} \subset M_n$ and $E_k \subseteq E_{k+1}$ for all $k$. Then $\lim_{k \to \infty} m_n(E_k) = m_n(\bigcup_{k=1}^{\infty} E_k)$.
2. Let $\{E_k\} \subset M_n$ and $E_k \supseteq E_{k+1}$ and $m_n(E_1) < \infty$. Then $\lim_{k \to \infty} m_n(E_k) = m_n(\bigcap_{k=1}^{\infty} E_k)$.

**Theorem 14.27** (Regularity Properties of $m_n$). The following are the $n$-dimensional versions of Exterior and Interior Regularity.

64
(1) Let $E \in \mathcal{M}_n$. Then for every $\epsilon > 0$ there exists an open set $O_\epsilon \in \mathbb{R}^n$ such that $E \subset O_\epsilon$ and $m_n(O_\epsilon \setminus E) < \epsilon$.

(2) Let $E \in \mathcal{M}_n$. Then for every $\epsilon > 0$ there exists a closed set $C_\epsilon \in \mathbb{R}^n$ such that $C_\epsilon \subset E$ and $m_n(E \setminus C_\epsilon) < \epsilon$.

**Theorem 14.28** (Approximating Measurable Sets by $G_\delta$- and $F_\sigma$-Sets). The following are the $n$-dimensional versions of the approximation of measurable sets by $G_\delta$- and $F_\sigma$-sets.

(1) Let $E \subset \mathcal{M}_n$. Then there exists a $G_\delta$-set $G \in \mathbb{R}^n$ such that $E \subset G$ and $m_n(G \setminus E) = 0$.

(2) Let $E \subset \mathcal{M}_n$. Then there exists an $F_\sigma$-set $K \in \mathbb{R}^n$ such that $K \subset E$ and $m_n(E \setminus K) = 0$.

This completes the description of the generalization to $n$-dimensional Lebesgue measure.
15. **Lebesgue Measurable Functions and Integration for \( \mathbb{R}^n \)**

15.1. **Measurable Functions.** We now generalize the notions of measurable functions and integration for \( \mathbb{R}^n \). The results parallel the 1-dimensional case, and so proofs are omitted in most cases.

From a notational standpoint, recall that \( \{ f > c \} = \{ x \in E : f(x) > c \} \), and likewise for the other relations \( \geq, <, \leq \).

**Definition 15.1.** A function \( f : \mathbb{R}^n \to \mathbb{R} \cup \{ \infty \} \) is called **measurable** if for every \( c \in \mathbb{R} \) we have \( \{ f > c \} \in \mathcal{M} \).

**Note 15.2.** Recall the equivalent definitions of measurability of a function in the 1-dimensional case. These also hold in the \( n \)-dimensional case. That is, \( f \in \mathcal{M} \) if any (and hence all) of the following hold for \( f \):

1. \( \{ f > c \} \in \mathcal{M} \).
2. \( \{ f \geq c \} \in \mathcal{M} \).
3. \( \{ f < c \} \in \mathcal{M} \).
4. \( \{ f \leq c \} \in \mathcal{M} \).

(We use the notation \( \mathcal{M} \) to denote the set of all measurable functions from \( \mathbb{R}^n \) into \( \mathbb{R} \cup \{ \infty \} \).

**Theorem 15.3 (Measurability and Operations).** The following hold for measurable functions.

1. If \( f, g \in \mathcal{M} \), then \( f \pm g \in \mathcal{M} \), \( f \cdot g \in \mathcal{M} \), and \( f/g \in \mathcal{M} \) (provided \( g(x) \neq 0 \)).
2. If \( f \in \mathcal{M} \), then \( \alpha f \in \mathcal{M} \) for all \( \alpha \in \mathbb{R} \).
3. If \( \{ f_k \}^m_{k=1} \subset \mathcal{M} \), then \( \max \{ f_1, \ldots, f_m \} \in \mathcal{M} \) and \( \min \{ f_1, \ldots, f_m \} \in \mathcal{M} \).
4. If \( \{ f_k \}^\infty_{k=1} \subset \mathcal{M} \), then
   
   (a) \( \limsup_{k \to \infty} f_k \in \mathcal{M} \) and \( \liminf_{k \to \infty} f_k \in \mathcal{M} \), and
   
   (b) If \( f_k \to f \), then \( f \in \mathcal{M} \).
5. If \( f \in \mathcal{M} \) and \( f = g \) almost everywhere, then \( g \in \mathcal{M} \).

**Definition 15.4.** A **simple function** has the form \( s(x) = \sum_{k=1}^m a_k \chi_{E_k}(x) \) where \( a_k \in \mathbb{R} \) and \( E_k \in \mathcal{M} \) for every \( k \geq 1 \).

**Note 15.5.** From the definition of a simple function, it follows that \( s \in \mathcal{M} \) for every simple function \( s \).

**Note 15.6.** As in the 1-dimensional case, there is a **standard representation** of every simple function \( s \). Namely,

\[
s = \sum_{k=1}^m a_k \chi_{E_k}
\]

is the standard representation if \( \{ E_k \}^m_{k=1} \) is a disjoint family and all coefficients \( a_k \) are distinct.

**Theorem 15.7 (Simple Approximation).** The following is the \( n \)-dimensional version of the Simple Approximation Theorem (Theorem 11.7). Let \( f \in \mathcal{M} \).

1. If \( f(x) \geq 0 \) for all \( x \in \mathbb{R}^n \), then there exists a sequence \( \{ s_k \}^\infty_{k=1} \) of simple functions satisfying the following properties:
   
   (a) \( s_k(x) \geq 0 \) for all \( k \in \mathbb{N} \) and for all \( x \in \mathbb{R}^n \),
(b) \( s_k(x) \leq s_{k+1} \) for all \( k \in \mathbb{N} \); that is, \( s_k \uparrow \), and
(c) \( \lim_{k \to \infty} s_k(x) = f(x) \).

(2) If \( f \) is finite almost everywhere then there exists a sequence \( \{ s_k \}_{k=1}^{\infty} \) of simple functions satisfying the following properties:
(a) \( |s_k(x)| \leq |s_{k+1}(x)| \) for all \( k \in \mathbb{N} \); that is, \( |s_k| \uparrow \), and
(b) \( \lim_{k \to \infty} s_k(x) = f(x) \).

**Note 15.8.** As a consequence of the Simple Approximation Theorem, it follows that if \( f \in \mathcal{M}_n \) is finite almost everywhere, then \( \lim_{k \to \infty} |s_k(x)| = |f(x)| \).

**Proof.** Note that \( |f| = |f - s_k + s_k| \leq |f - s_k| + |s_k| \), and so \( |f| - |s_k| \leq |f - s_k| \). Likewise, since \( |s_k| = |s_k - f + f| \leq |s_k - f| + |f| \), and so \( |s_k| - |f| \leq |s_k - f| = |f - s_k| \). Hence
\[
||f| - |s_k|| \leq |f - s_k| \to 0
\]
as \( k \to \infty \). This completes the proof. \( \square \)

The following are the \( n \)-dimensional versions of Egorov’s Theorem and Lusin’s Theorem.

**Theorem 15.9** (Egorov’s Theorem). Suppose \( E \in \mathcal{M}_n \) with \( m(E) < \infty \). Let \( \{ f_k \}_{k=1}^{\infty} \subset \mathcal{M}_n \) be such that \( f_k \to f \) almost everywhere on \( E \). Then given \( \epsilon > 0 \) there exists a closed set \( C_\epsilon \) such that
(1) \( C_\epsilon \subset E \) and \( m(E \setminus C_\epsilon) < \epsilon \) and
(2) \( f_n \to f \) uniformly on \( C_\epsilon \).

(Recall these two conditions give almost uniform convergence.)

**Theorem 15.10** (Lusin’s Theorem). Let \( E \subset \mathcal{M}_n \) with \( m(E) < \infty \). Let \( f \in \mathcal{M}_E \). Given \( \epsilon > 0 \) there exists a closed set \( \tilde{C}_\epsilon \) such that
(1) \( \tilde{C}_\epsilon \subset E \) and \( m(E \setminus \tilde{C}_\epsilon) < \epsilon \) and
(2) \( f|_{\tilde{C}_\epsilon} \) is a continuous function on \( \tilde{C}_\epsilon \).

15.2. Lebesgue Integration. We now begin the discussion of Lebesgue integration for \( \mathbb{R}^n \). Let \( f \in \mathcal{M}_n \). Our goal is to define
\[
\int f \, dx_1 dx_2 \cdots dx_n = \int f \, dx.
\]
We need to
(1) Define integrable functions.
(2) Define the integral.
We start by reviewing the cases \( K_1, K_2, K_3, \) and \( K_4 \) from the 1-dimensional case.

**Note 15.11** (Case 1). Recall that we defined
\[
K_1 = \left\{ \text{all simple functions } s(x) = \sum_{k=1}^{m} a_k \chi_{E_k}(x) \text{ such that } m(E_k) < \infty \text{ if } a_k < 0 \right\}.
\]
Then we define the integral as
\[
\int s \, dx = \sum_{k=1}^{m} a_k m(E_k).
\]
The restriction that \( m(E_k) < \infty \) is in place to avoid the case of \( \infty - \infty \). As in the 1-dimensional case, this definition is invariant of the representation of \( s \).

**Note 15.12** (Case 2). Recall that we defined

\[ K_2 = \{ \text{all bounded measurable functions } f \text{ such that } \text{supp}(f) \subset E \text{ with } m(E) < \infty \}. \]

The upper integral is defined to be

\[ \int^+ f \, dx = \inf_{\psi \in K_1 \atop f \leq \psi} \left\{ \int \psi \, dx \right\} \]

and the lower integral is defined to be

\[ \int^- f \, dx = \sup_{\varphi \in K_1 \atop f \geq \varphi} \left\{ \int \varphi \, dx \right\}. \]

Recall that \( f \) is integrable if \( \int^+ f \, dx = \int^- f \, dx \), and in this case

\[ \int f \, dx = \int^+ f \, dx = \int^- f \, dx. \]

**Note 15.13.** As in the 1-dimensional case, if \( f \in K_2 \), then \( f \) is integrable.

**Note 15.14.** If \( f \in K_2 \), then

\[ \left| \int f \, dx \right| \leq M \cdot m(E) < \infty, \]

where \( M > 0 \) is such that \( |f(x)| < M \) and where \( E \) is such that \( \text{supp}(f) \subset E \) with \( m(E) < \infty \).

**Note 15.15** (Case 3). Recall that we defined

\[ K_3 = \{ \text{non-negative measurable functions} \}. \]

The integral is defined as

\[ \int f \, dx = \sup_{0 \leq g \leq f \atop g \in K_2} \left\{ \int g \, dx \right\}. \]

Note that for \( f \in K_3 \) we have \( 0 \leq \int f \, dx \leq \infty \), so the integral may be infinite. For example, consider \( f(x) = 1 \) for all \( x \in \mathbb{R} \). Clearly \( f \in K_3 \), and the family \( \{g_k\}_{k=1}^{\infty} \) defined by

\[ g_k(x) = \begin{cases} 1 & \text{if } x \in [-k, k] \\ 0 & \text{otherwise} \end{cases} \]

is a family of functions in \( K_2 \) such that \( 0 \leq g_k \leq f \) for all \( k \in \mathbb{N} \). Hence \( \int f \, dx \geq \int g_k \, dx = 2k \) for all \( k \in \mathbb{N} \), and so \( \int f \, dx \) is unbounded. Thus \( \int f \, dx = \infty \).

**Definition 15.16.** A function \( f \in \mathcal{M}_n \) is **integrable** if

\[ \int |f| < \infty. \]

In this case, we write \( f \in L^1(\mathbb{R}^n) \). We usually abbreviate \( L^1(\mathbb{R}^n) \) by simply \( L^1 \) unless the context is not clear as to which dimension we are working with.
Note 15.17. The class $L^1(\mathbb{R}^n)$ is a linear function class, meaning that if $f, g \in L^1$ and $\alpha, \beta \in \mathbb{R}$, then $\alpha f + \beta g \in L^1$.

Note 15.18 (Case 4). Let $K_4 = L^1(\mathbb{R}^n)$. Recall that $|f| = f^+ + f^-$ and $f = f^+ - f^-$, where $f^+, f^- \in K_3$. In particular, $0 \leq f^+ \leq |f|$ and $0 \leq f^- \leq |f|$. Hence, if $f \in L^1$, we have by monotonicity of integration that $0 \leq \int f^+ \, dx \leq \int |f| \, dx < \infty$ and $0 \leq \int f^- \, dx \leq \int |f| \, dx < \infty$. Thus, there is no ambiguity (i.e. there is no situation where we have $\infty - \infty$) in defining the integral for $f \in L^1$ as

$$
\int f \, dx = \int f^+ \, dx - \int f^- \, dx.
$$

Note 15.19 (Properties of the Integral). We recall (without proof, since the proofs are similar) the following properties of the Lebesgue integral. We generally only consider functions from $K_3$ and $L^1$, as $K_1$ and $K_2$ were simply used to build the integral for these more interesting cases.

1. Linearity ($K_3$ and $L^1$);
   (a) For $f, g \in K_3$ ($\subset L^1$), we have $\int (f + g) \, dx = \int f \, dx + \int g \, dx$.
   (b) For $f \in K_3$ ($\subset L^1$) and $\alpha \in \mathbb{R}$, we have $\int \alpha f \, dx = \alpha \int f \, dx$.

2. Additivity;
   (a) For $f \in K_3$ ($\subset L^1$) and disjoint $E_1, E_2 \in \mathcal{M}_n$, we have

   $$
   \int_{E_1 \cup E_2} f \, dx = \int_{E_1} f \, dx + \int_{E_2} f \, dx.
   $$

   This extends to all finite disjoint unions of measurable sets.
   (b) If $f \in L^1$ and $\{E_k\}_{k=1}^\infty \subset \mathcal{M}_n$ is a disjoint family of measurable sets, then

   $$
   \int_{\bigcup_{k=1}^\infty E_k} f \, dx = \sum_{k=1}^\infty \int_{E_k} f \, dx.
   $$

   This property in particular is very important.

3. Monotonicity ($K_3$ and $L^1$);
   (a) For $f, g \in K_3$ ($\subset L^1$), if $f(x) \leq g(x)$ for all $x \in E$ then $\int_E f \, dx \leq \int_E g \, dx$.

4. Triangle Inequality in $L^1$;
   (a) For $f, g \in L^1$, we have $\int |f + g| \, dx \leq \int |f| \, dx + \int |g| \, dx$.

15.3. Convergence Theorems, Chebychev’s Inequality, and Modes of Convergence. Recall the following problem from the 1-dimensional case. Suppose we have a family of functions $\{f_k\} \subset K_i$ ($1 \leq i \leq 4$) and $f_k \to f$ point-wise or almost everywhere. Under what conditions do we have the equality

$$
\int f \, dx = \lim_{k \to \infty} \int_{k} f_k \, dx.
$$

That is, under what conditions can we “pass to the limit under the integral”? This question is the focus of the following convergence theorems.

Theorem 15.20 (Bounded Convergence Theorem for $K_2$). Suppose $\{f_k\} \subset K_2$ satisfies the following conditions.

1. There is an $M > 0$ such that $|f_k(x)| \leq M$ for all $k \in \mathbb{N}$ and all $x \in \mathbb{R}^n$.
2. $\text{supp}(f_k) \subset E \in \mathcal{M}_n$ for all $k \in \mathbb{N}$, where $m_n(E) < \infty$.


(3) \( f_k \to f \) point-wise or almost everywhere. Then \( \int f \, dx = \lim_{k \to \infty} \int f_k \, dx \).

**Theorem 15.21** (Monotone Convergence Theorem for \( K_3 \)). Suppose \( \{f_k\} \subset K_3 \) satisfies the following conditions.

1. \( f_k \to f \) point-wise or almost everywhere.
2. \( f(x) < \infty \) almost everywhere.
3. \( 0 \leq f_k(x) \leq f_{k+1}(x) \) for all \( k \in \mathbb{N} \) and all \( x \in \mathbb{R}^n \).

Then \( \int f \, dx = \lim_{k \to \infty} \int f_k \, dx \).

**Theorem 15.22** (Fatou’s Lemma for \( K_3 \)). Let \( \{f_k\} \subset K_3 \) and suppose \( f_k \to f \in K_3 \) point-wise or almost everywhere. Then \( \int f \, dx \leq \liminf_{k \to \infty} \int f_k \, dx \).

**Theorem 15.23** (Lebesgue’s Dominated Convergence Theorem for \( L^1 \)). Suppose \( \{f_k\} \subset L^1 \) and \( f_k \to f \) point-wise or almost everywhere. Also, assume there exists a function \( g \in L^1 \) such that \( |f_k| \leq g \) for all \( k \in \mathbb{N} \). (Note this implies \( f \in L^1 \).) Then \( \int f \, dx = \lim_{k \to \infty} \int f_k \, dx \).

We now note that Chebychev’s inequality holds in \( \mathbb{R}^n \). In fact, the proof is the same.

**Theorem 15.24** (Chebyshev’s Inequality). Let \( f \in L^1 \) and \( \lambda > 0 \). Then

\[
m_n \{x \in \mathbb{R}^n : |f(x)| > \lambda \} \leq \frac{1}{\lambda} \int |f| \, dx.
\]

We close this section with a discussion of two modes of convergence, one we have seen before in the 1-dimensional case and one we have not seen before. Note these are both global properties.

**Definition 15.25.** Let \( \{f_k\}_{k=1}^\infty \subset \mathcal{M}_n \) and let \( f \in \mathcal{M}_n \). We say that \( f_k \to f \) in measure if for every \( \delta > 0 \),

\[
m_n \{x \in \mathbb{R}^n : |f(x) - f_k(x)| > \delta \} \to 0 \quad \text{as} \quad k \to \infty.
\]

**Definition 15.26.** Let \( \{f_k\}_{k=1}^\infty \subset L^1 \) and let \( f \in L^1 \). We say that \( f_k \to f \) in \( L^1 \) if

\[
\|f - f_k\|_1 = \int |f - f_k| \, dx \to 0 \quad \text{as} \quad k \to \infty.
\]

Let us compare these two modes of convergence.

**Lemma 15.27.** \( L^1 \) convergence implies convergence in measure.

**Proof.** This follows immediately from Chebychev’s inequality. \( \square \)

**Note 15.28.** Neither convergence in measure nor convergence almost everywhere imply \( L^1 \) convergence. To see both cases, one example will suffice. Let

\[
f_k(x) = \begin{cases} 
  k & \text{if } x \in [0, \frac{1}{k}] \\
  0 & \text{otherwise}.
\end{cases}
\]

Then \( f_k \to f \equiv 0 \) almost everywhere and in measure. However, \( \int |f - f_k| \, dx = \int f_k \, dx = 1 \) for all \( k \in \mathbb{N} \), and so \( \int |f - f_k| \, dx \to 1 \neq f \) as \( k \to \infty \).
16. Differentiation

16.1. Increasing Functions. We return to the 1-dimensional case and seek to answer the question of which functions are differentiable almost everywhere. We restrict ourselves to closed intervals, but can “extend” to open intervals. In this section, $f$ will be defined everywhere on $[a,b]$ and will be assumed to be an increasing function on $[a,b]$.

**Definition 16.1.** Suppose $f$ is increasing on $[a,b]$. Let $x \in [a,b]$. Then we define the limits from above and below, denoted $f(x+)$ and $f(x-)$, respectively, as

$$f(x+) = \lim_{h \to 0^+} f(x + h) = \inf_{h > 0} f(x + h) \quad \text{and} \quad f(x-) = \lim_{h \to 0^+} f(x - h) = \sup_{h > 0} f(x - h).$$

The equivalence of the definition with the infimum and supremum versions is a result of the fact that $f$ is an increasing function.

**Definition 16.2.** Let $f$ be an increasing function of $[a,b]$. If $f(x+) - f(x-) > 0$, then we say there is a jump at $x$. We say that $x$ is a jump point, and the size of the jump is $J_f(x) = f(x+) - f(x-) > 0$. We denote the set of all jump points of a function $f$ by $S_f$. If $f(x+) - f(x-) = 0$, then we say $f$ is continuous at $x$.

**Lemma 16.3.** The set $S_f$ of all jump points of an increasing function $f$ is countable.

**Proof.** Let $x$ be a jump point and call $I_x = (f(x-), f(x+))$ the jump interval at $x$. The family of all jump intervals $\{I_x\}_{x \in S_f}$ is a disjoint family since $f$ is an increasing function. Furthermore, we can find a rational number inside each jump interval. These two observations tell us there can only be countably many jump intervals, and hence countably many jump points.

Since jumps are the only types of discontinuities for increasing functions, this previous theorem states that every increasing function may only have jump discontinuities, and no more than countably many of them. The next result shows the existence of some very wild increasing functions.

**Lemma 16.4.** Let $S$ be a countable subset of $(a,b)$. Then there exists an increasing function $f$ on $[a,b]$ such that $S_f = S$.

**Proof.** For the case that $S$ is finite, any increasing (although not strictly increasing) step function will do the job. Hence we may assume that $S$ is countably infinite. The we can write $S = \{x_j\}_{j=1}^{\infty}$. Note this indexing of $S$ does not assume an order. We want to construct an increasing function $f$ such that

1. $f$ has a jump at every point in $S$, and
2. $f$ is continuous at every point $x \in (a,b) \setminus S$.

With these goals in mind, consider the function $f : [a,b] \to \mathbb{R}$ defined by

$$f(x) = \sum_{k : x_k \leq x} \frac{1}{2^k}.$$

Now clearly $f$ is well-defined, bounded, and increasing. We show it satisfies both properties.

For property (1), take $x_j \in S$ and let $x < x_j$. Since $f(x) = \sum_{k : x_k \leq x} \frac{1}{2^k}$, we see that $\frac{1}{2^j}$ is not in this sum. But $\frac{1}{2^j}$ is in the sum $f(x_j) = \sum_{k : x_k \leq x_j} \frac{1}{2^k}$. Hence we have $f(x) < f(x) + \frac{1}{2^j} \leq f(x_j)$.
Since $x$ was taken as any number in $[a, b]$ strictly less than $x_j$, it follows that $f(x_j) \geq \frac{1}{2^j} + f(x_j -)$, and hence we have

$$f(x_j-) < f(x_j) \leq f(x_j+).$$

This implies there is a jump at $x_j$, and so (1) is satisfied.

For (2), take $x \in (a, b) \setminus S$. Fix some positive integer $n$ and consider the set $\{x_1, \ldots, x_n\}$. In particular, $x \not= x_i$ for any $i \in \{1, \ldots, n\}$. This is a finite collection, so there is some interval $I_n$ containing $x$ such that $x_i \not\in I_n$ for every $i \in \{1, \ldots, n\}$. Let $y \in I_n$. Now

$$f(x) - f(y) = \sum_{k: x_k \leq x} \frac{1}{2^k} - \sum_{k: x_k \leq y} \frac{1}{2^k}.$$

For each $i \in \{1, \ldots, n\}$, since $x_i \not\in I_n$, we have that either $\frac{1}{2^i}$ is a part of the sum for both $f(x)$ and $f(y)$ or for neither. In either case, this value is not considered in the difference of $f(x) - f(y)$. Therefore we may write

$$f(x) - f(y) = \sum_{k: x_k \leq x, k > n} \frac{1}{2^k} - \sum_{k: x_k \leq y, k > n} \frac{1}{2^k}.$$

It follows from the triangle inequality that

$$|f(x) - f(y)| \leq |f(x)| + |f(y)| \leq \sum_{k=n+1}^{\infty} \frac{1}{2^k} + \sum_{k=n+1}^{\infty} \frac{1}{2^k} = 2 \sum_{k=n+1}^{\infty} \frac{1}{2^k} = \frac{1}{2^{n-1}}.$$

Hence, given any $\epsilon > 0$, we can find a positive integer $n$ such that $\frac{1}{2^n} < \epsilon$, and what we have just shown is that we can place an interval around $x$ (hence providing our $\delta > 0$) such that $|f(x) - f(y)| < \epsilon$ whenever $|x - y| < \delta$ (i.e. whenever $y$ is in this interval). This is the definition of continuity at $x$, so (2) is satisfied. \hfill \Box

16.2. Bounded Variation. We now study bounded variation, which is related to the study of increasing functions. Note that we are not requiring our functions to be increasing anymore.

**Definition 16.5.** Let $f$ be a function on $[a, b]$ and let $[x_k, x_{k+1}] \subset [a, b]$. The **variation of $f$ over** $[x_k, x_{k+1}]$ is $|f(x_{k+1}) - f(x_k)|$. Consider a partition

$$P = \{a = x_0 < x_1 < \cdots < x_n = b\}.$$

The **variation of $f$ with respect to the partition** $P$, denoted $V(P; f)$, is

$$V(P; f) = \sum_{k=0}^{n-1} |f(x_{k+1}) - f(x_k)|.$$

The **total variation** of $f$ over $[a, b]$, denoted $TV_f([a, b])$, is

$$TV_f([a, b]) = \sup_P V(P; f),$$

where the supremum is taken over all possible partitions $P$ of $[a, b]$. It is said that $f$ is a **function of bounded variation** (written $f \in BV$) if $TV_f([a, b]) < \infty$.

**Note 16.6 (Properties).** The following are properties of variation.

(1) If $f$ is increasing or decreasing, then $f \in BV[a, b]$. 

72
Proof. Let $P = \{a = x_0 < \cdots < x_n = b\}$ be any partition of $[a, b]$. Then

$$V(P; f) = \sum_{k=0}^{n-1} |f(x_{k+1}) - f(x_k)| = \sum_{k=0}^{n-1} f(x_{k+1}) - f(x_k) = f(b) - f(a).$$

This is a telescoping sum. Notice we assumed here that $f$ is increasing. If $f$ were decreasing, then by a similar argument, we would have $V(P; f) = f(a) - f(b)$. In either case, since $P$ was an arbitrary partition, we have $TV_f([a, b]) < \infty$, so $f \in BV[a, b]$.

(2) $BV[a, b]$ is a linear class.

Proof. Let $\alpha, \beta \in \mathbb{R}$ and let $f, g \in BV$. Let $P = \{a = x_0 < \cdots < x_n = b\}$ be a partition of $[a, b]$. Then

$$V(P, \alpha f + \beta g) = \sum_{k=0}^{n-1} |\alpha f(x_{k+1}) + \beta g(x_{k+1}) - \alpha f(x_k) - \beta g(x_k)|$$

$$\leq \sum_{k=0}^{n-1} |\alpha f(x_{k+1}) - \alpha f(x_k)| + |\beta g(x_{k+1}) - \beta g(x_k)|$$

$$= |\alpha| \sum_{k=0}^{n-1} |f(x_{k+1}) - f(x_k)| + |\beta| \sum_{k=1}^{n-1} |g(x_{k+1}) - g(x_k)|$$

$$= |\alpha| V(P; f) + |\beta| V(P; g).$$

Hence we have

$$TV_{\alpha f + \beta g}([a, b]) = \sup_P V(P, \alpha f + \beta g) \leq \sup_P \{\alpha V(P; f) + |\beta| V(P; g)\}$$

$$\leq |\alpha| \sup_P V(P; f) + |\beta| \sup_P V(P; g) = |\alpha| TV_f([a, b]) + |\beta| TV_g([a, b]) < \infty,$$

and so $\alpha f + \beta g \in BV$, as desired.

(3) Let $P$ be a partition of $[a, b]$ and $P'$ be a refinement of $P$. Let $f$ be any function defined on $[a, b]$. Then $V(P; f) \leq V(P'; f)$.

Proof. It is enough to consider just one extra point $c$, since any refinement can be realized by adding one point at a time for a finite number of steps. Suppose $x_i < c < x_{i+1}$. Then

$$V(P; f) = \sum_{k=0}^{n-1} |f(x_{k+1}) - f(x_k)|$$

$$= \sum_{k=0}^{i-1} |f(x_{k+1}) - f(x_k)| + |f(x_{i+1}) - f(x_i)| + \sum_{k=i+1}^{n-1} |f(x_{k+1}) - f(x_k)|$$

$$= \sum_{k=0}^{i-1} |f(x_{k+1}) - f(x_k)| + |f(x_{i+1}) - f(c) + f(c) - f(x_i)| + \sum_{k=i+1}^{n-1} |f(x_{k+1}) - f(x_k)|$$

$$\leq \sum_{k=0}^{i-1} |f(x_{k+1}) - f(x_k)| + |f(x_{i+1}) - f(c)| + |f(c) - f(x_i)| + \sum_{k=i+1}^{n-1} |f(x_{k+1}) - f(x_k)|$$

$$= V(P'; f),$$

73
16.3. Lipschitz Functions, Absolutely Continuous Functions, and Functions of Bounded Variation. We now consider three classes of functions of a closed interval $[a,b]$. One is the class $BV$ of functions of bounded variation which we have been considering. The other two classes are the class Lip1 of Lipschitz functions and the class $AC$ of absolutely continuous functions.

**Definition 16.7.** A function $f$ belongs to Lip1 (i.e. Lipschitz 1) if there exists a constant $c > 0$ for which $|f(x) - f(y)| < c|x - y|$ for all $x, y \in [a,b]$.

**Note 16.8.** We note that Lip1 $\subset C[a,b]$, where $C[a,b]$ is the class of continuous functions on $[a,b]$. To see this, recall that a functions is continuous if, given any $\epsilon > 0$ there is a $\delta > 0$ such that $|f(x) - f(y)| < \epsilon$ whenever $|x - y| < \delta$. For a function $f \in$ Lip1, let $\delta = \frac{\epsilon}{c}$, where $c$ is the Lipschitz constant for $f$ guaranteed by the definition. Thus, whenever $|x - y| < \delta$, we have

$$|f(x) - f(y)| < c \cdot |x - y| < c \cdot \delta = c \cdot \frac{\epsilon}{c} = \epsilon,$$

which shows that $f$ is continuous. Actually, this shows $f$ is uniformly continuous on $[a,b]$ since our choice of $\delta$ only depended on $\epsilon$ and not on any point in the interval $[a,b]$. Hence Lip1 $\subset C[a,b]$.

**Definition 16.9.** A function $f$ is called absolutely continuous, written $f \in AC$, if for every $\epsilon > 0$ there is a $\delta > 0$ such that for all choices of finite families of almost disjoint closed intervals $\{(a_i, b_i)\}_{i=1}^k$ in $[a,b]$, we have

$$\sum_{i=1}^k |b_i - a_i| < \delta \text{ implies } \sum_{i=1}^k |f(b_i) - f(a_i)| < \epsilon.$$

**Note 16.10.** We note that $AC \subset C[a,b]$ as well. This is easily seen by taking just one interval $[a_1, b_1] = [x, y]$. That is, given $\epsilon > 0$ there is a $\delta > 0$ such that $|x - y| < \delta$ implies $|f(x) - f(y)| < \epsilon$. This is the definition of uniform continuity since it is valid for any interval $[x, y] \subset [a,b]$, provided $y - x < \delta$. Hence $AC \subset C[a,b]$.

**Example 16.11.** Suppose $f$ is a continuously differentiable function with $|f'(x)| \leq M$ for every $x \in [a,b]$. Then $f \in$ Lip1.

**Proof.** Without loss of generality, assume $x < y$. By the mean value theorem, there is a $\theta$ with $x < \theta < y$ such that $|f(x) - f(y)| = |f'(\theta)(x - y)| \leq M|x - y|$, where $M = f'(\theta)$. □

We now work toward showing Lip1 $\subset AC \subset BV$, and that these inclusions are proper.

**Lemma 16.12.** Lip1 $\subset AC$.

**Proof.** Assume $f \in$ Lip1. We estimate sums of the form $\sum_{i=1}^k |f(b_i) - f(a_i)|$ for almost disjoint $[a_i, b_i]_{1 \leq i \leq k}$. We have

$$\sum_{i=1}^k |f(b_i) - f(a_i)| \leq \sum_{i=1}^k M \cdot |b_i - a_i| = M \sum_{i=1}^k |b_i - a_i|,$$

where $M$ is the Lipschitz constant. Let $\epsilon > 0$. Choose $\delta = \frac{\epsilon}{M}$. Then if $\sum_{i=1}^k |b_i - a_i| < \delta$, we have

$$\sum_{i=1}^k |f(b_i) - f(a_i)| \leq M \sum_{i=1}^k |b_i - a_i| < M \cdot \frac{\epsilon}{M} = \epsilon,$$
which shows \( f \in AC \), as desired. \( \square \)

**Lemma 16.13.** \( AC \subset BV \).

**Proof.** Suppose \( f \in AC \). Take \( \epsilon = 1 \). Then there exists a \( \delta > 0 \) such that for every choice of almost disjoint families \([a_i, b_i]_{1 \leq i \leq k}\) we have

\[
\sum_{i=1}^{k} (b_i - a_i) < \delta \implies \sum_{i=1}^{k} |f(b_i) - f(a_i)| < 1.
\]

Let \( P \) and \( Q \) be a partitions of \([a, b]\) where

\[
P = \{ a = x_0 < \cdots < x_n = b \} \quad \text{and} \quad Q = \{ a = y_0 < \cdots < y_m = b \},
\]

where \( y_{j+1} - y_j < \delta \). Now consider the partition \( P' = P \cup Q \), so that \( P' \) is a refinement of \( P \). Fix an integer \( i \in \{1, \ldots, m\} \) and consider the reduction of the partition \( P' \) to \([y_{i-1}, y_i]\). This gives us some collection of almost disjoint closed intervals \([a_j, b_j]_{1 \leq j \leq t}\), where \( a_1 = y_{i-1} \), \( b_1 = y_i \), and \( a_j = b_{j-1} \) for each \( j \in \{2, \ldots, t\} \), and where each intermediate value is a value from the partition \( P \). Note then that

\[
\sum_{j=1}^{t} (b_j - a_j) = y_i - y_{i-1} < \delta
\]

since this is a telescoping sum. Hence by the absolute continuity of \( f \), we have

\[
\sum_{j=1}^{t} |f(b_j) - f(a_j)| < 1.
\]

We perform this same computation for each \( i \in \{1, \ldots, m\} \), which tells us that \( V(P'; f) \leq m \). Then, by Note 16.6(3) we have \( V(P; f) \leq V(P'; f) \leq m \). Taking the supremum over all partitions \( P \) of \([a, b]\) gives us \( TV_f([a, b]) \leq m \), which gives \( f \in BV \). \( \square \)

**Theorem 16.14** (The Absolute Continuity of the Lebesgue Integral). *Let \( f \in L^1(\mathbb{R}^n) \). For every \( \epsilon > 0 \) there is a \( \delta > 0 \) such that if \( m(E) < \delta \) then \( \int_E |f| \, dx < \epsilon \).*

**Proof.** Assume without loss of generality that \( f \geq 0 \). Denote \( E_k = \{ x \in \mathbb{R}^n : f(x) \leq k \} \). Let \( f_k(x) = f(x) \chi_{E_k}(x) \), and note that \( f_k \in \mathfrak{M} \). Note, also, that \( \{f_k\}_{k=1}^{\infty} \) is an increasing sequence of functions and \( \lim_{k \to \infty} f_k(x) = f(x) \). By the monotone convergence theorem, for every \( \epsilon > 0 \) there exists a \( k_0 \) such that

\[
\int (f - f_{k_0}) \, dx = \int f \, dx - \int f_{k_0} \, dx < \frac{\epsilon}{2}.
\]

Thus we have

\[
\int_E f \, dx = \int_E f \, dx - \int_E f_{k_0} \, dx + \int_E f_{k_0} \, dx
\]

\[
= \int_E (f - f_{k_0}) \, dx + \int_E f_{k_0} \, dx
\]

\[
\leq \int (f - f_{k_0}) \, dx + \int_E f_{k_0} \, dx
\]

\[
< \frac{\epsilon}{2} + k_0 \cdot m(E)
\]
Thus take \( \delta = \frac{\epsilon}{2k_0} \). Then if \( m(E) < \delta \), we have \( \int_E f \, dx \leq \frac{\epsilon}{2} + k_0 \cdot m(E) < \frac{\epsilon}{2} + k_0 \cdot \frac{\epsilon}{2k_0} = \epsilon \), as desired to complete the proof. Notice that \( k_0 \) depends only on \( \epsilon \), and so \( \delta \) depends only on \( \epsilon \). \( \square \)

**Note 16.15.** The absolute continuity of the Lebesgue integral does not depend on where the set \( E \) is located. So this is a global property in a sense.

We are able, with the following lemma, to produce a wide class of functions from \( AC \).

**Lemma 16.16.** Let \( f \in L^1[a,b] \). Define a new function \( g : [a,b] \rightarrow \mathbb{R} \) by \( g(x) = \int_a^x f(y) \, dy \). Then \( g \in AC \).

**Proof.** Take an almost disjoint family \([a_i, b_i]_{1 \leq i \leq k}\). Notice that
\[
\sum_{i=1}^k |g(b_i) - g(a_i)| = \sum_{i=1}^k \left| \int_{a_i}^{b_i} f \, dy - \int_{a}^{a_i} f \, dy \right| = \sum_{i=1}^k \left| \int_{a_i}^{b_i} f \, dy \right| \\
\leq \sum_{i=1}^k \int_{a_i}^{b_i} |f| \, dy = \int_{\bigcup_{i=1}^k I_i} |f| \, dy.
\]
The second equality and the fourth equality (using \( I_i = [a_i, b_i] \)) both follow from the additivity of the Lebesgue integral (cf. Note 15.19(2)), and the inequality follows from Lemma 12.21 (noting that exactly the same proof works for the case \( f \in L^1[a,b] \), since this lemma was proved for \( f \in K_2 \) in the 1-dimensional case).

Now let \( E = \bigcup_{i=1}^k I_i \), and note that \( m(E) = \sum_{i=1}^k (b_i - a_i) \) (noting that the “almost disjoint” assumption does not affect this measure). By the absolute continuity of the Lebesgue integral, for every \( \epsilon > 0 \) there is a \( \delta > 0 \) such that
\[
\int_E |f| \, dy < \epsilon \quad \text{provided} \quad m(E) < \delta.
\]
Fix an \( \epsilon > 0 \) and use \( E = \bigcup_{i=1}^k I_i \). Then if \( \sum_{i=1}^k (b_i - a_i) = m(E) < \delta \), it follows from our first observation that
\[
\sum_{i=1}^k |g(b_i) - g(a_i)| < \epsilon.
\]
This is the definition of absolute continuity. \( \square \)

As noted, this gives us a wide class of functions in \( AC \). We are now in a position to prove the inclusions \( \text{Lip}^1 \subset AC \subset BV \) are proper. We do this on the interval \([0,1]\), but note this can be extended to any bounded interval \([a,b]\).

**Example 16.17.** Let \( f = \sqrt{x} \) for \( x \in [0,1] \). Note that
\[
\int_0^1 \left| \frac{1}{2\sqrt{y}} \right| \, dy < \infty,
\]

Thus take \( \delta = \frac{\epsilon}{2k_0} \). Then if \( m(E) < \delta \), we have \( \int_E f \, dx \leq \frac{\epsilon}{2} + k_0 \cdot m(E) < \frac{\epsilon}{2} + k_0 \cdot \frac{\epsilon}{2k_0} = \epsilon \), as desired to complete the proof. Notice that \( k_0 \) depends only on \( \epsilon \), and so \( \delta \) depends only on \( \epsilon \). \( \square \)
and so \( \frac{1}{2\sqrt{y}} \in L^1[0,1] \). Hence, by Lemma 16.16, we have

\[
f(x) = \sqrt{x} = \int_0^x \frac{1}{2\sqrt{y}} \, dy \in AC.
\]

Hence it remains to show \( f \notin \text{Lip1} \). Suppose to the contrary that \( f \in \text{Lip1} \). Then \( |f(x) - f(y)| \leq M|x - y| \) for some Lipschitz constant \( M > 0 \). Take \( y = 0 \). Then we have \( f(x) = \sqrt{x} \leq Mx \) for every \( x \in [0,1] \), or equivalently, \( \frac{1}{\sqrt{x}} \leq M \) for every \( x \in [0,1] \), which is a contradiction because \( \frac{1}{\sqrt{x}} \) is unbounded as \( x \to 0 \) from the right. Hence \( f \notin \text{Lip1} \).

**Example 16.18.** Let \( f = \varphi \), the Cantor Lebesgue function on \([0,1]\). Now \( \varphi \in BV \) because \( \varphi \) is an increasing function (cf. Note 16.6(1)). We must therefore show \( \varphi \notin AC \). To do this, our goal is construct a special sequence of almost disjoint systems \( S_k = \{ [a_i^{(k)}, b_i^{(k)}] \}_{i=1}^{m_k} \) that has the properties we want. Let \( S_k \) be the collection of intervals remaining after the \( k \)th removal in the construction of the Cantor set. Recall, then, that

\[
\sum_{i=1}^{m_k} (b_i^{(k)} - a_i^{(k)}) = \left( \frac{2}{3} \right)^k
\]

is the total length of this collection \( S_k \). Hence, this total length tends to 0 as \( k \to \infty \). Note that \( \varphi(b_i) = \varphi(a_{i+1}) \) since \( \varphi \) does not increase on \([0,1] \setminus C_k \) (where \( C_k = \bigcup S_k \)).

Hence we have a telescoping sum

\[
\sum_{i=1}^{m_k} |\varphi(b_i) - \varphi(a_i)| = \sum_{i=1}^{m_k} (\varphi(b_i) - \varphi(a_i)) = -\varphi(a_1) + \sum_{i=1}^{m_k-1} (-\varphi(a_{i+1}) + \varphi(b_i)) + \varphi(b_{m_k})
\]

\[
= (\varphi(b_{m_k}) - \varphi(a_1)) + \sum_{i=1}^{m_k-1} (-\varphi(b_i) + \varphi(b_i)) = 0 = 1,
\]

and this holds for any \( k \in \mathbb{N} \).

Hence, suppose \( \varphi \in AC \). Then there exists a \( \delta > 0 \) such that for every almost disjoint family \( \{[a_i, b_i] \}_{i=1}^{m} \) with \( \sum_{i=1}^{m} (b_i - a_i) < \delta \), we have \( \sum_{i=1}^{m} (\varphi(b_i) - \varphi(a_i)) < \frac{1}{2} \). However, there exists a \( k \in \mathbb{N} \) such that \( \left( \frac{2}{3} \right)^k < \delta \), which is a contradiction. Hence \( \varphi \notin AC \).

**Note 16.19.** Let \( C = \{ \text{all continuous functions} \} \). Neither inclusion \( BV \subset C \) or \( C \subset BV \) hold. We provide counter-examples for each.

*Proof that \( BV \nsubset C \).* Any increasing step function will do. \( \square \)

*Proof that \( C \nsubset BV \).* Consider the function

\[
f(x) = \begin{cases} 
  x \cos \left( \frac{\pi}{2x} \right) & \text{if } 0 < x \leq 1 \\
  0 & \text{if } x = 0
\end{cases}
\]
This is clearly a continuous function on $[0, 1]$. We must show, however, that it is of unbounded variation. To do this, consider the sequence of partitions $P_n$ given by

$$P_n = \left\{ 0 < \frac{1}{2n} < \frac{1}{2n-1} < \frac{1}{2n-2} < \cdots < \frac{1}{2} < 1 \right\}.$$

We consider what value $f$ will take on these partition values. Note that if $k \in \mathbb{N}$ is even, we have

$$f \left( \frac{1}{k} \right) = \frac{1}{k} \cos \left( \frac{k\pi}{2} \right) = \pm \frac{1}{k},$$

and if $k$ is odd, we have

$$f \left( \frac{1}{k} \right) = \frac{1}{k} \cos \left( \frac{k\pi}{2} \right) = 0.$$

Hence, we have

$$V(P_n; f) = \left| f \left( \frac{1}{2n} \right) - f(0) \right| + \left| f \left( \frac{1}{2n-1} \right) - f \left( \frac{1}{2n} \right) \right| + \cdots + \left| f(1) - f \left( \frac{1}{2} \right) \right|$$

$$= \left( \pm \frac{1}{2n} - 0 \right) + \left( 0 - \frac{1}{2n} \right) + \cdots + \left( \frac{1}{2} - 0 \right)$$

$$= \frac{1}{n} \sum_{k=1}^{n} \frac{1}{k} \to \infty$$

as $n \to \infty$. Hence this sequence of partitions shows that $f$ is of unbounded variation.

We now begin building toward a fundamental result in analysis known as Jordan’s Theorem.

**Definition 16.20.** Let $f \in BV[a, b]$. Then the function given by

$$TV(f_{[a,x]}) = \sup_{P \text{ partitions } [a, x]} V(P, f) = \sup_{P=\{a=x_1<\cdots<x_k=x\}} \sum_{i=0}^{k-1} |f(x_{i+1}) - f(x_i)|$$

is called the **total variation function** for $f$.

**Note 16.21** (Properties of the Total Variation Function). The total variation function has the following properties.

1. $TV(f_{[a,x]})$ is an increasing function.
2. Let $c$ be such that $a < c < b$. Then the following **additivity property** holds:

$$TV(f_{[a,b]}) = TV(f_{[a,c]}) + TV(f_{[c,b]}).$$

**Proof.** We first show (2) ⇒ (1). Let $a \leq x_1 < x_2 \leq b$. Then by (2) we have

$$TV(f_{[a,x_2]}) = TV(f_{[a,x_1]}) + TV(f_{[x_1,x_2]}),$$

and thus $TV(f_{[a,x_1]}) \leq TV(f_{[a,x_2]})$, so (1) holds. Thus we must only show (2).

Note that by Note 16.6(3) the supremum does not change by adding one point $c$ to any partition, and so

$$TV(f_{[a,b]}) = \sup_P \sum_{i=1}^{k-1} |f(x_{i+1}) - f(x_i)| = \sup_{P_c} \sum_{i=1}^{k-1} |f(x_{x+1}) - f(x_i)|,$$

where $P_c$ is the partition obtained by adding the point $c$ to $P$.
where \( P_c \) is the collection of all partitions of \([a,b]\) including the point \( c \).

With this in mind, take a partition \( \pi \in P_c \) and let \( \pi' \) be the reduction of \( \pi \) to \([a,c]\). Likewise, let \( \pi'' \) be the reduction of \( \pi \) to \([c,b]\). Then

\[
V(\pi; f) = V(\pi'; f) + V(\pi''; f) \leq TV(f_{[a,c]}) + TV(f_{[c,b]}).
\]

This is for any partition \( \pi \in P_c \), so since the right-hand side doesn’t depend on the choice of the partition \( \pi \in P_c \), we have

\[
TV(f_{[a,b]}) \leq TV(f_{[a,c]}) + TV(f_{[c,b]}).
\]

Now fix \( \epsilon > 0 \). From the definition of \( TV(f_{[a,c]}) \) and \( TV(f_{[c,b]}) \) we may find partitions \( \pi' \) of \([a,c]\) and \( \pi'' \) of \([c,b]\) such that

\[
V(\pi'; f) \geq TV(f_{[a,c]}) - \frac{\epsilon}{2} \quad \text{and} \quad V(\pi''; f) \geq TV(f_{[c,b]}) - \frac{\epsilon}{2}.
\]

Now let \( \pi \) be the partition of \([a,b]\) by combining the partitions \( \pi' \) and \( \pi'' \). Then we have

\[
V(\pi; f) = V(\pi'; f) + V(\pi''; f) \geq TV(f_{[a,c]}) + TV(f_{[c,b]}) - \epsilon.
\]

Since \( \pi \) represents some partition of \([a,b]\), we may take the supremum and get

\[
TV(f_{[a,b]}) = \sup_{\pi \in P_c} V(\pi; f) \geq TV(f_{[a,c]}) + TV(f_{[c,b]}) - \epsilon.
\]

Sending \( \epsilon \to 0 \), we have

\[
TV(f_{[a,b]}) \geq TV(f_{[a,c]}) + TV(f_{[c,b]}).
\]

Combining (26) and (27), we are done. \( \square \)

16.4. **Jordan’s Theorem.** We are now ready to state and prove Jordan’s Theorem, which is the structure theorem for the class \( BV \).

**Theorem 16.22** (Jordan’s Theorem). A function \( f \) defined on \([a,b]\) is in \( BV \) if and only if there exists two increasing functions \( h \) and \( g \) defined on \([a,b]\) so that \( f = h - g \).

**Proof.** First, suppose there are increasing functions \( g \) and \( h \) for which \( f = g - h \). Then, because \( g, h \in BV \) and \( BV \) is a linear class (cf. Note 16.6(1) and (2)), it follows that \( f \in BV \).

Conversely, suppose \( f \in BV \). Take \( h(x) = f(x) + TV(f_{[a,x]}) \) and \( g(x) = TV(f_{[a,x]}) \). It is clear that \( f = h - g \), and furthermore we know \( g \) is an increasing function by Note 16.21(1). It remains to show \( h \) is an increasing function. With this in mind, let \( u, v \in [a,b] \) such that \( u < v \) and let \( \pi_0 \) be the trivial partition of \([u,v]\). Then we have

\[
f(u) - f(v) \leq |f(v) - f(u)| = V(\pi_0; f_{[a,v]}) \leq TV(f_{[a,v]}) = TV(f_{[a,v]}) - TV(f_{[a,u]}),
\]

where the last equality follows from additivity (cf. Note 16.21(2)). Re-arranging according to this inequality, we have

\[
f(u) + TV(f_{[a,u]}) \leq f(v) + TV(f_{[a,v]}),
\]

which is what we needed to show to establish that \( h \) is increasing. \( \square \)

**Note 16.23.** If \( f \in BV \), the representation \( f = h - g \) is not unique. In particular, you can add any increasing function \( s \) to both \( g \) and \( h \) to get \( f = (h + s) - (g + s) \). However, the
canonical representation is given in the proof of Theorem 16.22, where \( h(x) = f(x) + TV(f_{[a,x]}) \) and \( g(x) = TV(f_{[a,x]}) \). This is called the **Jordan decomposition** of \( f \).

Let \( f \in AC \). Then since \( AC \subset BV \), it follows that \( f \in BV \) and so Jordan’s theorem applies to \( f \). That is, we can write \( f = h - g \) for some increasing functions \( h \) and \( g \). A natural question to ask is whether \( h \) and \( g \) are in \( AC \). The next result answers this question.

**Theorem 16.24.** Let \( f \in AC[a,b] \). Let \( h = f + TV(f_{[a,i]}) \) and \( g = TV(f_{[a,i]}) \) be the increasing functions in the Jordan decomposition of \( f \). Then \( h \) and \( g \) are in \( AC[a,b] \).

**Proof.** Since \( AC \) is a linear functions class, it is enough to prove that \( TV(f_{[a,i]}) \in AC \). Fix \( \epsilon > 0. \) Since \( f \in AC \), there exists a \( \delta > 0 \) such that if \( [a_i, b_i] \) is an almost disjoint family of intervals with total length

\[
\sum_{i=1}^{k} (b_i - a_i) < \delta, \quad \text{then} \quad \sum_{i=1}^{k} |f(b_i) - f(a_i)| < \frac{\epsilon}{2}.
\]

Let \( [a_i, b_i] \) be a collection of almost disjoint intervals satisfying this total length requirement. Denote \( [a_i, b_i] = I_i \) for all \( i \in \{1, \ldots, k\} \). Let \( \pi_i \) be a partition of \( I_i \). Since the total length of each partition \( \pi_i \) combined is less than \( \delta \), we have

\[
\sum_{i=1}^{k} V(f, \pi_i) = \sum_{i=1}^{k} |f(b_i) - f(a_i)| < \frac{\epsilon}{2}.
\]  

(28)

Now for each \( \pi_i \) find a refinement \( \pi_i^* \) such that

\[
TV(f_{[a_i,b_i]}) \leq V(f, \pi_i^*) + \frac{\epsilon}{2k}.
\]

Then condition (28) still applies to the collection \( \{\pi_i^*\}_{i=1}^{k} \), so we have

\[
\sum_{i=1}^{k} TV(f_{[a_i,b_i]}) \leq \sum_{i=1}^{k} V(f, \pi_i^*) + \frac{\epsilon}{2} < \epsilon.
\]  

(29)

Now we want

\[
\sum_{i=1}^{k} |TV(f_{[a_i,b_i]}) - TV(f_{[a,b_i]})| < \epsilon,
\]

but by the additivity property and the fact that \( TV(f_{[a,i]}) \) is increasing, we have

\[
|TV(f_{[a,b_i]}) - TV(f_{[a,a_i]})| = TV(f_{[a,b_i]}),
\]

and so by (29), we are done.

The next result, Lebesgue’s Theorem, is of high importance in analysis, but for brevity we omit the proof and focus on some of its corollaries.

**Theorem 16.25** (Lebesgue’s Theorem). Let \( f \) be a monotone function on \([a,b]\). Then \( f \) is differentiable almost everywhere.

**Note 16.26.** Recall the derivative is defined as

\[
f'(x) = \lim_{y \to x} \frac{f(y) - f(x)}{y - x},
\]
if this limit exists. (If $x = a$ or $x = b$, we use one-sided limits.) Thus, a reformulation of Lebesgue's theorem is that if a function $f$ is monotone on $[a, b]$, then there exists a set $E_f \subset [a, b]$ such that $E_f \in \mathcal{M}$ and for which $m([a, b] \setminus E_f) = 0$ and $f'(x)$ exists for each $x \in E_f$.

**Corollary 16.27.** Suppose $f \in BV[a, b]$. Then $f$ is differentiable almost everywhere.

**Proof.** By Jordan’s theorem, take $f = h - g$ for increasing functions $h$ and $g$. By Lebesgue’s theorem, both $h$ and $g$ are differentiable almost everywhere, which implies $f = h - g$ is differentiable almost everywhere. For instance, if $E_h$ and $E_g$ represent the measurable sets on which $h$ and $g$ are differentiable, where $m([a, b] \setminus E_h) = 0 = m([a, b] \setminus E_g)$, then $E_f = E_h \cap E_g$, and

$$m([a, b] \setminus E_f) = m([a, b] \setminus (E_h \cap E_g)) = m(([a, b] \setminus E_h) \cup ([a, b] \setminus E_g)) \leq m([a, b] \setminus E_h) + m([a, b] \setminus E_g) = 0,$$

completing the proof. □

16.5. **The Fundamental Theorem of Calculus.** We first state (without proof, for the sake of brevity) Part One of The Fundamental Theorem of Calculus.

**Theorem 16.28** (Fundamental Theorem of Calculus, Part One). Let $f \in L^1[a, b]$ and let

$$g(x) = \int_a^x f \, dy,$$

where $x \in [a, b]$. Then $g$ is differentiable almost everywhere and $g'(x) = f(x)$.

**Proof.** Omitted. □

Now we will discuss Part Two of the Fundamental Theorem of Calculus. Recall in calculus, we assume that $f$ is continuously differentiable and we use the Riemann integral, stating that

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

We wish to consider this result in the context of the Lebesgue integral. The results do not hold in such generality, as the following example illustrates.

**Example 16.29.** Let $\varphi$ be the Cantor Lebesgue function. Now $\varphi$ is increasing, but recall we showed in Example 16.18 that $\varphi \notin AC$. Furthermore, note that $\varphi(1) - \varphi(0) = 1 - 0 = 1$. Also, $\varphi'(x) = 0$ if $x \in [0, 1] \setminus C$. It follows that $\int_0^1 \varphi' \, dx = 0$, which shows $\int_0^1 \varphi' \, dx < \varphi(1) - \varphi(0)$. This is a counter-example to the Fundamental Theorem of Calculus holding in full generality.

**Note 16.30.** As a matter of notation, let $\text{Diff}_h f(x) = \frac{f(x+h) - f(x)}{h}$ be the difference quotient for $f$, so $f'(x) = \lim_{h \to 0} \text{Diff}_h f(x)$.

**Lemma 16.31.** Let $f$ be an increasing function on $[a, b]$. Then $f' \in L^1[a, b]$ and

$$0 \leq \int_a^b f' \, dx \leq f(b) - f(a).$$

**Proof.** Let us consider the extension of $f$ to $[a, b + 1]$ by letting $f(x) = f(b)$ if $x \geq b$. Let $f_n(x) = \text{Diff}_{1/n} f(x)$. By Lebesgue’s Differentiation Theorem, we have that $f_n \to f'$ almost everywhere on $[a, b + 1]$. We also have $f_n(x) \geq 0$ for all $n \in \mathbb{N}$ and for all $x \in [a, b + 1]$. Finally, since $f$ is increasing it follows that $f_n$ is increasing for all $n \in \mathbb{N}$, and increasing functions are measurable.
Hence \( \{ f_n \}_{n=1}^{N} \subset K_3 \), and we may apply Fatou’s Lemma for this family of functions on \([a,b]\). Thus we have
\[
0 \leq \int_{a}^{b} f' \ dx \leq \liminf_{n \to \infty} \int_{a}^{b} \text{Diff}_1/n f(x) \ dx.
\]

Now we have
\[
\int_{a}^{b} \text{Diff}_1/n f(x) \ dx = n \int_{a}^{b} f(x + 1/n) - f(x) \ dx
\]
\[
\quad = n \left[ \int_{a}^{b+1/n} f(x) \ dx - \int_{a}^{b} f(x) \ dx \right]
\]
\[
\quad = n \left[ \int_{b}^{a+1/n} f(x) \ dx - \int_{a}^{a+1/n} f(x) \ dx \right]
\]
\[
\quad \leq n \left[ \frac{f(b)}{n} - \frac{f(a)}{n} \right]
\]
\[
\quad = f(b) - f(a).
\]

The second-to-last equality follows from the additivity property of the integral. The inequality follows from the fact that \( f \) is constant when \( x \geq b \), and also that \( f \) is increasing, so \( f(a) \leq f(a + \epsilon) \) for any \( \epsilon > 0 \), and hence we are subtracting a smaller quantity from \( f(b)/n \), thus making the expression larger. This completes the proof.

**Corollary 16.32.** If \( f \in BV[a,b] \), then \( f' \in L^1 \).

**Proof.** By the previous theorem, all increasing functions are in \( L^1 \). Thus, since \( f \in BV \), Jordan’s theorem allows us to write \( f = h - g \) for increasing functions \( h \) and \( g \). Since \( L^1 \) is a linear class, it follows that \( f = h - g \in L^1 \).

Note also the same proof shows that if \( f \in AC \), then \( f \in L^1 \) since \( AC \subset BV \). The class \( AC \) is actually intimately tied to the fundamental theorem of calculus. In fact, we will see that
\[
\int_{a}^{b} f' \ dx = f(b) - f(a)
\]
precisely when the function \( f \) is in \( AC \). This is the second part of the Fundamental Theorem of Calculus. To prove this, we first work toward the Vitali Covering Lemma.

**Definition 16.33.** Let \( E \subset \bigcup_{\alpha \in A} I_\alpha \). Then the open covering of intervals \( \mathcal{U} = \{ I_\alpha \}_{\alpha \in A} \) is called a **Vitali covering** if for every point \( x \in E \) and every \( \delta > 0 \) there exists an interval \( I_\alpha \in \mathcal{U} \) such that \( x \in I_\alpha \) and such that \( |I_\alpha| < \delta \).

**Lemma 16.34 (Vitali Covering Argument).** Suppose \( I_1, \ldots, I_N \) is a finite collection of open intervals. Then there exists a finite disjoint sub-collection \( I_{i_1}, \ldots, I_{i_k} \) such that
\[
m \left( \bigcup_{j=1}^{N} I_j \right) \leq 3 \sum_{n=1}^{k} |I_{i_n}|.
\]
Proof. We proceed in a finite number of stages. First, since \( \mathcal{U}_0 = \{ I_1, \ldots, I_N \} \) is a finite collection, we start by selecting (and re-indexing) an interval \( I_1 \in \mathcal{U}_0 \) of maximum length. Now let \( \mathcal{U}_1 \) be the collection formed by removing from \( \mathcal{U}_0 \) the set \( I_1 \), as well as any sets in \( \mathcal{U}_0 \) intersecting \( I_1 \). If \( \mathcal{U}_1 \) is empty, then the process terminates and we move to the final step, described below. If \( \mathcal{U}_1 \) is not empty, then we repeat the same process. That is, we select (and re-index) an interval \( I_2 \in \mathcal{U}_1 \) of maximum length, and then we define \( \mathcal{U}_2 \) to be the collection formed by removing from \( \mathcal{U}_1 \) the set \( I_2 \), as well as any sets in \( \mathcal{U}_1 \) intersecting \( I_2 \). We repeat this construction until we have exhausted all of the sets, which is guaranteed to happen in a finite number of steps since \( \mathcal{U}_0 \) was assumed to be finite.

Once this process has terminated, we may proceed to the final step. Let us first observe that the sets \( I_1, \ldots, I_k \) (where \( k \) is the largest integer for which \( \mathcal{U}_{k-1} \) is non-empty) form a disjoint family by construction, and also that \( k \) can potentially be any number between 1 and \( N \). In particular, \( k \) is finite. Now for each \( n \in \{ 1, \ldots, k \} \), let \( 3 \cdot I_n \) be the interval sharing a common midpoint with \( I_n \) and of 3 times the length of \( I_n \). We claim that

\[
\bigcup_{j=1}^N I_j \subseteq \bigcup_{n=1}^k 3 \cdot I_n.
\]

To see this, let \( x \in \bigcup_{j=1}^N I_j \). Then there is a \( j \in \{ 1, \ldots, N \} \) for which \( x \in I_j \). Suppose to the contrary that \( x \notin 3 \cdot I_n \) for any \( n \in \{ 1, \ldots, k \} \). By our construction, this implies that \( I_j \cap I_n = \emptyset \) for each \( n \in \{ 1, \ldots, k \} \), since we are working with open intervals and since \( I_n \) was chosen at each step to be an interval of maximal length at that stage in the construction. This is a contradiction, though, since this implies our process did not terminate when we are assuming it did. Hence (30) holds, and we consequently have

\[
m \left( \bigcup_{j=1}^N I_j \right) \leq m \left( \bigcup_{n=1}^k 3 \cdot I_n \right) \leq \sum_{n=1}^k m(3 \cdot I_n) = \sum_{n=1}^k 3|I_n| = 3 \sum_{n=1}^k |I_n|,
\]

as desired to finish the proof. \( \square \)

Lemma 16.35 (Vitali Covering Lemma). Let \( E \subseteq \mathcal{M} \) with \( m(E) < \infty \), and let \( \mathcal{I} = \{ I_\alpha \}_{\alpha \in A} \) be a Vitali covering of \( E \). Then for every \( \delta > 0 \) there exists a disjoint subfamily \( I_1, \ldots, I_N \) of \( \mathcal{I} \) such that \( \sum_{j=1}^N |I_j| \geq m(E) - \delta \).

Proof. This is another construction proof, similar in design to the proof of the Vitali Covering Argument. Accordingly, we proceed in stages that will terminate at some point. Before we begin, we note that if \( m(E) = 0 \), the result is trivial, so we assume that \( m(E) > 0 \). Let \( \delta > 0 \). Once again, the result is trivial if \( m(E) \leq \delta \), so assume that \( m(E) > \delta \). Finally, since \( m(E) < \infty \), we may assume all of the intervals in our Vitali covering \( \mathcal{I} \) are bounded, for if any such interval is not bounded, then we may remove it and still have a Vitali covering (by the definition of a Vitali covering).

First, since we assume \( E \) is measurable, we may use the interior regularity property to find a closed set \( F_1 \subseteq E \) such that \( m(F_1) > \delta \). Furthermore, we may choose \( F_1 \) in such a way that \( F_1 \) is compact (for if our original choice of \( F_1 \) was not compact, we may choose \( n \) large enough to replace \( F_1 \) with the set \( [-n,n] \cap F_1 \), where the measure of this new set is still greater than \( \delta \)).
Since \( F_1 \subset E \), we have \( F_1 \subset \bigcup_{\alpha \in A} I_\alpha \). Then, since \( F_1 \) is compact, we may choose a finite sub-cover \( \{ J_j \}_{j=1}^k \subset \{ I_\alpha \}_{\alpha \in A} \) such that \( F_1 \subset \bigcup_{j=1}^k J_j \). We now apply the Vitali Covering Argument (Lemma 16.34) to the collection \( \{ J_j \}_{j=1}^k \) to find a disjoint sub-family \( \{ I_i \}_{i=1}^{m_1} \subset \{ J_j \}_{j=1}^k \) such that

\[
m \left( \bigcup_{j=1}^k J_j \right) \leq 3 \sum_{i=1}^{m_1} |I_i|.
\]

Since \( \bigcup_{j=1}^k J_j \supset F_1 \) and \( m(F_1) > \delta \), this gives us

\[
\sum_{i=1}^{m_1} |I_i| > \frac{1}{3} m \left( \bigcup_{j=1}^k J_j \right) > \frac{1}{3} m(F_1) > \frac{1}{3} \delta.
\] (31)

This concludes the first step in the construction. Now if \( \sum_{i=1}^{m_1} |I_i| > m(E) - \delta \), then we are done as the conclusion is satisfied.

Suppose this is not the case. That is, suppose \( \sum_{i=1}^{m_1} |I_i| \leq m(E) - \delta \). Then let \( E_1 = E \setminus \bigcup_{i=1}^{m_1} T_i \). Since we know \( m(E) < \infty \) and \( m \left( \bigcup_{i=1}^{m_1} T_i \right) < \infty \), we may deduce

\[
m(E_1) > m(E) - m \left( \bigcup_{i=1}^{m_1} T_i \right) = m(E) - \sum_{i=1}^{m_1} |I_i| \geq m(E) - (m(E) - \delta) = \delta.
\]

Hence, we are almost in a position to repeat our first construction, but we need to make one modification first to address the following concern: we know that the sets \( I_1, \ldots, I_{m_1} \) are disjoint, but we do not know if they are pairwise disjoint with the remaining sets in \( \mathcal{I} \). To amend this, replace \( \mathcal{I} \) with \( \mathcal{I}_1 \) consisting of intervals from \( \mathcal{I} \) not intersecting \( \bigcup_{i=1}^{m_1} I_i \). We argue that \( \mathcal{I}_1 \) is a Vitali cover of \( E_1 \). Let \( x \in E_1 \) and let \( \gamma > 0 \). Then since \( E \setminus T_i = \emptyset \) for each \( i \in \{1, \ldots, m_1\} \), it follows that \( \epsilon = \min \{ \gamma, \text{dist}(x, I_1), \ldots, \text{dist}(x, I_{m_1}) \} > 0 \). Hence, since \( \mathcal{I} \) is a Vitali cover, choose an interval \( \tilde{I} \in \mathcal{I} \) such that \( x \in \tilde{I} \) and \( |\tilde{I}| < \epsilon \). This implies \( \tilde{I} \cap I_i = \emptyset \) for each \( i \in \{1, \ldots, m_1\} \), and so \( \tilde{I} \in \mathcal{I}_1 \), and since \( |\tilde{I}| < \gamma \), this proves that \( \mathcal{I}_1 \) is a Vitali cover of \( E_1 \).

Let us take a moment to collect ourselves and see what we have accomplished. We have first produced a finite sub-collection \( \{ I_i \}_{i=1}^{m_1} \subset \{ I_\alpha \}_{\alpha \in A} \) satisfying (31). If this further satisfies \( \sum_{i=1}^{m_2} |I_i| \geq m(E) - \delta \), then we are done. If not, then we have argued that we are in the same situation as before; namely, we can find a Vitali cover \( \mathcal{I}_2 \) of \( E_1 \), where the sets in \( \mathcal{I}_2 \) are pairwise disjoint with the sets \( I_1, \ldots, I_{m_1} \). Furthermore, we have shown that in this case \( m(E_1) > \delta \). Hence, we may use the exact same construction as before with \( E_1 \) and \( \mathcal{I}_2 \) to produce a set of pairwise disjoint sets \( \{ I_i \}_{i=m_1+1}^{m_2} \) (also pairwise disjoint from \( \{ I_i \}_{i=1}^{m_1} \)) such that

\[
\sum_{i=m_1+1}^{m_2} |I_i| > \frac{1}{3} \delta, \quad \text{which implies from (31) that} \quad \sum_{i=1}^{m_2} |I_i| > \frac{2}{3} \delta.
\]

If \( \sum_{i=1}^{m_2} |I_i| > \frac{2}{3} \delta > m(E) - \delta \), then we are done. If not, then we continue to repeat the construction, using the same arguments as before. Since \( m(E) \) is finite, we will eventually find an integer \( r > 0 \) such that

\[
\sum_{i=1}^{m_r} |I_i| > \frac{r}{3} \delta > m(E) - \delta,
\]

which completes the proof. \( \square \)
We are almost ready to conclude our work with the Fundamental Theorem of Calculus. Note that the following lemma does not hold for an arbitrary function \( f \in BV[a,b] \). The usual example is the Cantor-Lebesgue function, in which \( \varphi'(x) = 0 \) almost everywhere on \([0,1]\), but certainly \( \varphi \) is not constant on \([0.1]\); it is merely \textbf{piecewise constant} almost everywhere.

\textbf{Lemma 16.36.} Let \( f \in AC[a,b] \). Suppose \( f'(x) = 0 \) for almost all \( x \in [a,b] \). Then \( f(x) \) is constant on \([a,b]\).

\textbf{Proof.} We know \( f' \) exists almost everywhere by Corollary 16.27 since \( f \in BV[a,b] \). Let us denote by \( E \) the set of full measure (i.e. \( m(E) = b - a \)) where \( f' \) exists and where \( f'(x) = 0 \) for all \( x \in E \). (This set is guaranteed to exist; to see this, let \( A \) be the set of full measure on which \( f' \) exists, and let \( B \) be the set of full measure on which \( f'(x) = 0 \). Let \( E = A \cap B \). Then \( m([a,b] \setminus E) = m([a,b] \setminus (A \cap B)) \leq m([a,b] \setminus A) + m([a,b] \setminus B) = 0 \), as claimed.)

Furthermore, we may assume without loss of generality that \( a, b \notin E \). This does not change any of the properties of \( E \) which we need. Then we have

\[
\lim_{h \to 0} \left| \frac{f(x + h) - f(x)}{h} \right| = 0,
\]

and this holds because the limit without the absolute values tends to 0 by assumption when \( x \in E \).

Now fix \( \epsilon > 0 \) and also let \( \eta > 0 \) be some positive number. Then, by the \( \epsilon, \delta \) formulation of the limit, since \( \epsilon > 0 \) is fixed, we can find a \( \xi > 0 \) so that if \( |h| < \xi \), then \( \left| \frac{f(x + h) - f(x)}{h} \right| < \epsilon \), or equivalently,

\[
|f(x + h) - f(x)| < \epsilon \cdot |h|.
\]  

(We choose \( \xi \) instead of \( \delta \) because we have a special use in mind for \( \delta \) which will be explained later.) Let \((a_\eta, b_\eta)\) be an interval for which \( x \in (a_\eta, b_\eta) \) and \( x - a_\eta < \min\{(1/2)\eta, \xi\} \) and \( b_\eta - x < \min\{(1/2)\eta, \xi\} \). Then \( b_\eta - a_\eta < \eta \) by construction, and also by (32) we have

\[
|f(b_\eta) - f(x)| < \epsilon \cdot (b_\eta - x) \quad \text{and} \quad |f(a_\eta) - f(x)| < \epsilon \cdot (x - a_\eta).
\]

Using the triangle inequality, this gives us

\[
|f(b_\eta) - f(a_\eta)| \leq |f(b_\eta) - f(x)| + |f(x) - f(a_\eta)| < \epsilon \cdot [(b_\eta - x) + (x - a_\eta)] = \epsilon \cdot (b_\eta - a_\eta).
\]  

Note again that our choice of \( \eta > 0 \) was arbitrary. We perform this computation for every \( x \in E \), which is possible since we are assuming \( E \) does not contain its endpoints. For each \( x \in E \), let \((a_\eta, b_\eta, x)\) represent the interval in the construction we just described. We may assume without loss of generality that each of these intervals is contained entirely in \((a, b)\), for if it is not, then replace it by a smaller interval with the same properties. Then since \( \eta > 0 \) was arbitrary, \( \{(a_\eta, x, b_\eta, x)\} \) is by construction a Vitali covering of \( E \). Using the Vitali Covering Lemma 16.35, we obtain a disjoint subfamily \( I_1 = (a_1, b_1), \ldots, I_k = (a_k, b_k) \) such that

\[
\sum_{i=1}^{k} (b_i - a_i) > m(E) - \delta,
\]
where \( \delta > 0 \) is the \( \delta \) corresponding to the given \( \epsilon > 0 \) in the definition of absolute continuity (recalling we are assuming \( f \in AC[a,b] \)). By (33) we have

\[
\sum_{i=1}^{k} |f(b_i) - f(a_i)| \leq \sum_{i=1}^{k} \epsilon \cdot (b_i - a_i) \leq \epsilon \cdot (b - a).
\]

(35)

Now note that

\[ [a,b] \setminus \left( \bigcup_{i=1}^{k} I_i \right) = \bigcup_{i=1}^{k+1} [c_i, d_i] \]

where \( c_1 = a, d_{k+1} = b, d_i = a_i, \) and \( c_i+1 = b_i, \) which is a disjoint union. Notice also from (34) that

\[
\sum_{i=1}^{k+1} (d_i - a_i) = m(E) - \sum_{i=1}^{k} (b_i - a_i) < \delta,
\]

and so we may use the absolute continuity of \( f \) to establish that

\[
\sum_{i=1}^{k+1} |f(d_i) - f(c_i)| \leq \epsilon.
\]

(36)

Let \( \pi_\epsilon = \{a = c_1 < d_1 < c_2 < d_2 < \cdots < c_{k+1} < d_{k+1} = b\} \) be the partition we have formed here. Then combining (35) and (36) we have established that

\[
|f(b) - f(a)| = V(f, P) \leq V(f, \pi_\epsilon) < \epsilon \cdot (b - a) + \epsilon.
\]

The left-hand side of this expression does not depend on \( \epsilon, \) so let \( \epsilon \to 0 \) and conclude that \( |f(b) - f(a)| = 0, \) so \( f(b) = f(a). \) Thus \( f \) is constant on the endpoints of \( [a,b]. \) Now take any \( c \) for which \( a < c < b. \) The same argument holds for the interval \( [a,c], \) and so \( f(a) = f(c). \) Hence, we have that \( f \) is constant on \( [a,b], \) as desired to finish the proof. \( \Box \)

Finally, at long last we have

**Theorem 16.37** (Fundamental Theorem of Calculus, Part Two). Let \( f \in AC[a,b]. \) Then

\[
\int_{a}^{b} f' \, dx = f(b) - f(a).
\]

**Proof.** We already know from Lemma 16.13 and Corollaries 16.27 and 16.32 that \( f' \) exists almost everywhere and \( f' \in L^1[a,b]. \) Thus by Lemma 16.16 it follows that

\[
g = \int_{a}^{x} f' \, du \in AC[a,b].
\]

Consider the function \( h = f - g \in AC. \) Note that \( h'(x) = f'(x) - g'(x) = 0 \) almost everywhere by the Fundamental Theorem of Calculus, Part One (16.28). Thus by Lemma 16.36, it follows that \( h \) is constant on \( [a,b]. \) Hence

\[
 f(x) - g(x) = c
\]
for some c, which we now compute. Take $x = a$. Then

$$f(a) - g(a) = c \quad \text{and so} \quad c = f(a). \quad (37)$$

Now take $x = b$. Then we have from (37) that

$$f(b) - \int_a^b f'(u) \, du = f(a).$$

After re-arranging terms, the proof is complete. \qed

**Theorem 16.38** (Characterization of AC). Let $f$ be a function. Then $f \in AC[a,b]$ if and only if $f(x) = f(a) + \int_a^x g(y) \, dy$ for some $g \in L^1[a,b]$.

**Proof.** Suppose $f \in AC$. Then Theorem 16.37 gives us

$$\int_a^x f' \, dy = f(x) - f(a),$$

where we replace $b$ with $x$ for any $x \in (a,b]$. That is,

$$f(x) = f(a) + \int_a^x f' \, dy.$$

Taking $g = f' \in L^1$, we have established the first implication.

Conversely, assume

$$f(x) = f(a) + \int_a^x g \, dy$$

for some $g \in L^1$. By Lemma 16.16, it follows that $\int_a^x g \, dy \in AC$ since $g \in L^1$. We also have $f(a) \in AC$ since it is a constant. Since AC is a linear class, it follows that $f \in AC$. \qed
Suppose $X$ is a set. Let $\mathcal{M}$ be a $\sigma$-algebra of subsets of $X$.

**Note 17.1.** Recall that a $\sigma$-algebra $\mathcal{M}$ is a collection of subsets such that

1. $\emptyset \in \mathcal{M},$
2. If $A \in \mathcal{M}$, then $A^C \in \mathcal{M},$ and
3. If $\{A_i\}_{i=1}^\infty \subset \mathcal{M}$, then $\bigcup_{i=1}^\infty A_i \in \mathcal{M}$.

Then $(X, \mathcal{M})$ is called a **measurable space**. A **measure** $\mu : \mathcal{M} \to [0, \infty]$ is a set function on $\mathcal{M}$ such that

1. $\mu(\emptyset) = 0$ and
2. If $\{A_i\} \subset \mathcal{M}$ is a disjoint family, then $\mu\left(\bigcup_{i=1}^\infty A_i\right) = \sum_{i=1}^\infty \mu(A_i)$.

Then $(X, \mathcal{M}, \mu)$ is called a **measure space**.

**Note 17.2** (Properties). The following are properties of measure spaces.

1. (Finite Additivity.) Suppose $\{A_1, \ldots, A_k\} \subset \mathcal{M}$ are disjoint. Then

   $$\mu\left(\bigcup_{i=1}^k A_i\right) = \sum_{i=1}^k \mu(A_i).$$

   **Proof.** Define an infinite family of sets by $\{A_1, \ldots, A_k, \emptyset, \emptyset, \ldots\} \subset \mathcal{M}$, which is a disjoint family. Then by the first and second axiom of measure, we have

   $$\mu\left(\bigcup_{i=1}^\infty A_i\right) = \mu(A_1 \cup \cdots \cup A_k \cup \emptyset \cup \cdots) = \mu(A_1) + \cdots + \mu(A_k) + \mu(\emptyset) + \cdots = \sum_{i=1}^k \mu(A_i),$$

   as desired to finish the proof. \(\square\)

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

   **Proof.** We can write $B = A \cup (B \setminus A)$, and so by (1) we have

   $$\mu(B) = \mu(A) + \mu(B \setminus A) \geq \mu(A),$$

   as desired to finish the proof. \(\square\)

3. (Excision.) Let $A, B \in \mathcal{M}$, and let $A \subset B$ and $\mu(A) < \infty$. Then $\mu(B \setminus A) = \mu(B) - \mu(A)$.

   **Proof.** Following exactly the same procedure as in the previous proof, we have by the first equality in (38) that

   $$\mu(B) - \mu(A) = \mu(B \setminus A),$$

   where we are permitted to subtract $\mu(A)$ from both sides since $\mu(A) < \infty$. \(\square\)

4. (Countable Sub-Additivity.) Let $\{A_i\}_{i=1}^\infty \subset \mathcal{M}$. Then $\mu\left(\bigcup_{i=1}^\infty A_i\right) \leq \sum_{i=1}^\infty \mu(A_i)$.

   **Proof.** We construct a new disjoint collection of sets in $\mathcal{M}$ whose union is the same as $\bigcup_{i=1}^\infty A_i$. Let $C_1 = A_1$, $C_2 = A_2 \setminus A_1$, $C_3 = A_3 \setminus (A_1 \cup A_2)$, and in general,

   $$C_k = A_k \setminus \left(\bigcup_{i=1}^{k-1} A_i\right).$$
Then \( \{C_i\}_{i=1}^\infty \) is a disjoint family and \( \bigcup_{i=1}^\infty A_i = \bigcup_{i=1}^\infty C_i \). Thus we have
\[
\mu \left( \bigcup_{i=1}^\infty A_i \right) = \mu \left( \bigcup_{i=1}^\infty C_i \right) = \sum_{i=1}^\infty \mu(C_i) \leq \sum_{i=1}^\infty \mu(A_i),
\]
where the final inequality follows by monotonicity since \( C_i \subset A_i \) for all \( i \in \mathbb{N} \).

\[\square\]

**Theorem 17.3** (Continuity of General Measure). The following are the continuity properties of general measure.

1. (From Below.) Let \( \{A_i\}_{i=1}^\infty \subset \mathcal{M} \) with \( A_i \subset A_{i+1} \) for all \( i \in \mathbb{N} \). Then
\[
\mu \left( \bigcup_{i=1}^\infty A_i \right) = \lim_{i \to \infty} \mu(A_i).
\]
2. (From Above.) Let \( \{A_i\}_{i=1}^\infty \subset \mathcal{M} \) with \( A_i \supset A_{i+1} \) for all \( i \in \mathbb{N} \), and let \( \mu(A_1) < \infty \). Then
\[
\mu \left( \bigcap_{i=1}^\infty A_i \right) = \lim_{i \to \infty} \mu(A_i).
\]

**Proof.** Omitted since it follows from the same argument as the continuity for Lebesgue measure. \[\square\]

**Definition 17.4.** A measure \( \mu \) is called **finite** if \( \mu(X) < \infty \). A measure \( \mu \) is called **\( \sigma \)-finite** if there exists a representation \( X = \bigcup_{i=1}^\infty X_i \), where \( \mu(X_i) < \infty \) for all \( i \in \mathbb{N} \).

**Example 17.5.** The following are examples of measure spaces.

1. The Lebesgue measure on \( \mathbb{R}^n \): \( (\mathbb{R}^n, \mathcal{M}_n, m_n) \). We will not spend any time talking about this measure space; indeed, we have spent much time already learning about it. However, we will note that \( m_n \) is a \( \sigma \)-finite measure. To see this, consider
\[
\mathbb{R}^n = \bigcup_{k=1}^\infty B(0, k),
\]
where \( B(0, k) \) is the ball centered at the origin of radius \( k \). Note that \( m_n(B(0, k)) = \frac{4}{3}\pi k^3 \) \( < \infty \), showing that \( m_n \) is \( \sigma \)-finite.

2. The Borel Measure on \( \mathbb{R}^n \): \( (\mathbb{R}^n, \mathcal{B}_n, m_n) \). That is, the restriction of the Lebesgue measure to the \( \sigma \)-algebra of Borel sets. This is an important measure in probability theory.

3. The Counting Measure on \( X \). Let \( X \) be any set and \( \mathcal{M} = \mathcal{P}(X) \). Define \( \mu(A) = |A| \), the cardinality of \( A \). This measure is neither finite nor \( \sigma \)-finite if \( X \) is uncountable. This is because we cannot have an uncountable set \( X \) be the countable union of finite sets. Let us prove that it is a measure space. The first axiom is clear, so we must only check the second axiom. Let \( \{A_i\} \subset \mathcal{P}(X) \) be a disjoint family. We must show
\[
\mu \left( \bigcup_{i=1}^\infty A_i \right) = \left| \bigcup_{i=1}^\infty A_i \right| = \sum_{i=1}^\infty |A_i| = \sum_{i=1}^\infty \mu(A_i).
\]
This equality is clear, however, because for each \( x \in A_i \), we count \( x \) toward the cardinality exactly once in \( \bigcup_{i=1}^\infty A_i \) (since this is a disjoint union) and once in \( A_i \). Thus no element gets counted more (or less) than once on either side, and so we have equality.
(4) Let \( X \) be an uncountable set. Let \( \mathcal{M} \) be the family consisting of all countable subsets of \( X \), along with all subsets of \( X \) which have countable complement. Then define

\[
\mu(A) = \begin{cases} 
0 & \text{if } A \text{ is countable} \\
1 & \text{if } X \setminus A \text{ is countable.}
\end{cases}
\]

We claim that \((X, \mathcal{M}, \mu)\) is a measure space. We must show that \( \mathcal{M} \) is a \( \sigma \)-algebra and also that \( \mu \) is a measure.

**Proof.** First, we show that \( \mathcal{M} \) is a \( \sigma \)-algebra. We note the trivial statement that \( \emptyset \in \mathcal{M} \).

Now let \( E \in \mathcal{M} \). If \( X \setminus E \) is countable, then \( X \setminus E \in \mathcal{M} \). If \( E \) is countable, then \( X \setminus E \) has countable complement. In either case, \( X \setminus E \in \mathcal{M} \). Now let \( \{E_k\}_{k=1}^{\infty} \subset \mathcal{M} \). If all \( E_k \) are countable, then \( \bigcup_{k=1}^{\infty} E_k \) is countable. Note that for all \( j \in \mathbb{N} \), we have \( E_j \subset \bigcup_{k=1}^{\infty} E_k \), and so \( X \setminus (\bigcup_{k=1}^{\infty} E_k) \subset X \setminus E_j \). Hence, if there is even one \( j \in \mathbb{N} \) for which \( X \setminus E_j \) is countable; that is, \( E_j \) is uncountable, then we have \( X \setminus (\bigcup_{k=1}^{\infty} E_k) \) is countable. It follows that \( \bigcup_{k=1}^{\infty} E_k \in \mathcal{M} \). We have shown that \( \mathcal{M} \) is a \( \sigma \)-algebra.

Second, we show that \( \mu \) is a measure. We have \( \mu(\emptyset) = 0 \) trivially by the definition of \( \mu \).

We must therefore show countable additivity. Let \( \{E_k\}_{k=1}^{\infty} \subset \mathcal{M} \) be a disjoint family. Let \( E = \bigcup_{k=1}^{\infty} E_k \). We must show

\[
\mu(E) = \sum_{k=1}^{\infty} \mu(E_k). \tag{39}
\]

There are two cases. First, if all of the \( E_k \)’s are countable, then both the left-hand side and right-hand side of (39) are 0. For the second case, suppose there is a \( j \) for which \( X \setminus E_j \) is countable. Since the family is disjoint, it follows that \( E_k \subset X \setminus E_j \) whenever \( j \neq k \). It follows that \( E_k \) is countable for all such \( k \), and so the left-hand side and right-hand side of (39) is 1. This completes the proof that \( \mu \) is a measure. \( \square \)

(5) (Dirac Measure Space.) Let \( S \) be any set and let \( \mathcal{M} = \mathcal{P}(S) \). Let \( \mu = \delta_a \) where

\[
\delta_a(E) = \begin{cases} 
1 & \text{if } a \in E \\
0 & \text{otherwise.}
\end{cases}
\]

It is easy to see this is a measure.

**Definition 17.6.** A measure space \((X, \mathcal{M}, \mu)\) is **complete** if whenever \( A \in \mathcal{M} \) with \( \mu(A) = 0 \), it follows that if \( B \subset A \) then \( B \in \mathcal{M} \).

**Example 17.7.** As we have argued previously, the space \((\mathbb{R}^n, \mathcal{M}_n, m_n)\) is complete. However, the space \((\mathbb{R}^n, \mathcal{B}_n, m_n)\) is not complete. Our example to prove \( \mathcal{B} \neq \mathcal{M} \) involves a set inside the Cantor set \( C \), which has measure zero.

**Theorem 17.8** (Completion). Let \((X, \mathcal{M}, \mu)\) be an incomplete measure space. Define \( \mathcal{M}_0 \) to be the collection of sets in \( X \) of the form \( E = A \cup B \) where

\begin{itemize}
  \item[(1)] \( B \in \mathcal{M} \), and
  \item[(2)] \( A \subset C \), where \( C \in \mathcal{M} \) and \( \mu(C) = 0 \).
\end{itemize}

For such a set define \( \mu_0(E) = \mu(B) \). Then \( \mathcal{M}_0 \) is a \( \sigma \)-algebra, \( \mu_0 \) is a measure that extends to \( \mu \), and \((X, \mathcal{M}_0, \mu_0)\) is a complete measure space.
**Example 17.9.** The measure space \((\mathbb{R}^n, \mathcal{M}_n, m_n)\) is the completion of the measure space \((\mathbb{R}^n, \mathcal{B}_n, m_n)\).

*Proof.* Recall by inner regularity that for \(E \in \mathcal{M}_n\) there is an \(F_\sigma\)-set \(F\) such that \(F \subset E\) and \(m_n(E \setminus F) = 0\). (In fact, this is equivalent to the measurability of \(E\); cf. Theorem 11 on page 40 of Royden.) Hence \(E = F \cup (E \setminus F)\), so \(E\) has the correct form (given in Theorem 17.8) provided we can show \(F \in \mathcal{B}_n\) and \(E \setminus F \subset C\) where \(C \in \mathcal{B}_n\) with \(m_n(C) = 0\). Since \(F\) is an \(F_\sigma\)-set, it follows that \(F \in \mathcal{B}_n\). By outer regularity, there is a \(G_\delta\)-set \(G\) for which \(E \subset G\) and \(m_n(G \setminus E) = 0\). It follows that \(F \subset G\). Consider the set \(C = G \setminus F = G \cap F^C \in \mathcal{B}_n\). Since \(F \subset E \subset G\), it follows that \(E \setminus F \subset G \setminus F = C\). It remains to show \(m_n(C) = 0\). But since \(C = (G \setminus E) \cup (E \setminus F)\), which is a disjoint union, it follows that

\[
m_n(C) = m_n(G \setminus E) + m_n(E \setminus F) = 0,
\]

which completes the argument that every measurable set can be decomposed as in Theorem 17.8.

Now we show that every set of the form \(A \cup B\), where \(B \in \mathcal{B}_n\) and where \(A \subset C\), where \(C \in \mathcal{B}_n\) with \(m_n(C) = 0\), is measurable. But this is rather easily seen. In particular, since \(B\) is Borel, it is measurable. Furthermore, since \(C \in \mathcal{B}_n \subset \mathcal{M}_n\) and \(m_n(C) = 0\), any subset of \(C\) has outer measure zero (and is hence measurable), so \(A \in \mathcal{M}_n\). Hence \(A \cup B \in \mathcal{M}_n\). Thus we have shown that \(\mathcal{M}_n\) consists of exactly the sets that can be decomposed as in Theorem 17.8.

Finally, since we showed that every measurable set \(E \in \mathcal{M}_n\) has the decomposition into disjoint sets \(F \cup (E \setminus F)\) (where \(F\) is an \(F_\sigma\)-set), it follows that \(m_n(E) = m_n(F) + m_n(E \setminus F) = m_n(F)\), which completes the proof by applying Theorem 17.8. \(\Box\)

**17.1. Signed Measures.**

**Definition 17.10.** Let \((X, \mathcal{M})\) be a measurable space. Let \(\nu: \mathcal{M} \to [-\infty, \infty]\) be a set function for which

1. \(\nu\) omits one the values \(-\infty\) or \(\infty\) from its range,
2. \(\nu(\emptyset) = 0\), and
3. For all disjoint families \(\{E_i\}_{i=1}^\infty \subset \mathcal{M}\), we have

\[
\nu \left( \bigcup_{i=1}^\infty E_i \right) = \sum_{i=1}^\infty \nu(E_i).
\]

Then \(\nu\) is called a **signed measure**.

**Note 17.11.** We note that the finite additivity and excision properties carry over to signed measure with similar proofs as for general measure (with a modified assumption for excision that \(|\nu(A)| < \infty\)). The monotonicity, and consequently countable sub-additivity, properties do not carry over.

**Definition 17.12.** A **re-arrangement of** \(\mathbb{N}\) is a bijection \(\pi: \mathbb{N} \to \mathbb{N}\).

**Lemma 17.13.** Suppose \(\{a_i\}_{i=1}^\infty\) is a numerical sequence of finite numbers such that \(\sum_{i=1}^\infty a_{\pi(i)} < \infty\) for all re-arrangements \(\pi\). Then \(\sum_{i=1}^\infty |a_i| < \infty\).

*Proof.* Denote by \(\{a_k^+\}\) the sequence of positive terms in \(\{a_i\}\) and \(\{a_j^-\}\) the sequence of negative terms. Suppose to the contrary that \(\sum_{i=1}^\infty |a_i| = \infty\). Then at least one of \(\sum_{k=1}^\infty a_k^+\) or \(\sum_{j=1}^\infty (-a_j^-)\) is \(\infty\). Without loss of generality, assume \(\sum_{k=1}^\infty a_k^+ = \infty\). (Then the sequence \(\{a_j^-\}\) may be finite,
Let us consider the measure space \((\mathbb{R}^n, \mathcal{M}_n, \mu)\). Now since \(\sum_{k=1}^{\infty} a_k^+ = \infty\) and \(a_1^-\) is finite, we may find a positive integer \(n_1 \in \mathbb{N}\) for which
\[
a_1^+ + a_2^+ + \cdots + a_{n_1}^+ + a_1^- > 1.
\]
Likewise, we may find an integer \(n_2 \in \mathbb{N}\) such that
\[
a_{n_1+1}^+ + a_{n_1+2}^+ + \cdots + a_{n_2}^+ + a_2^- > 1.
\]
Indeed, for any \(k \in \mathbb{N}\), we may find an integer \(n_k \in \mathbb{N}\) for which
\[
a_{n_k-1}^+ + a_{n_k-2}^+ + \cdots + a_{n_k}^+ + a_k^- > 1.
\]
Then the re-arrangement \(\pi = \{a_1^+, \ldots, a_{n_1}^+, a_1^-, a_{n_1+1}^+, \ldots, a_{n_2}^+, a_2^-, \ldots\}\) gives us the series
\[
\underbrace{a_1^+ + a_2^+ + \cdots + a_{n_1}^+ + a_1^-}_{>1} + \underbrace{a_{n_1+1}^+ + a_{n_1+2}^+ + \cdots + a_{n_2}^+ + a_2^- + \cdots}_{>1} + \underbrace{a_{n_k-1}^+ + a_{n_k-2}^+ + \cdots + a_{n_k}^+ + a_k^- + \cdots}_{>1},
\]
which diverges to \(\infty\). This contradicts the assumption made that the series is finite for any given re-arrangement of \(\mathbb{N}\). Hence, it follows that the series converges absolutely, as required. \(\square\)

**Note 17.14.** As a consequence, if \(\{E_i\}_{i=1}^{\infty}\) is a disjoint family, and if \(|\nu(\bigcup_{i=1}^{\infty} E_i)| < \infty\), then \(\sum_{i=1}^{\infty} |\nu(E_i)| < \infty\).

**Proof.** Note that \(\nu(E_i)\) is finite for all \(i \in \mathbb{N}\), since otherwise it would follow that \(\nu(\bigcup_{i=1}^{\infty} E_i)\) would diverge to \(\infty\) or \(-\infty\). Let \(\pi\) be any re-arrangement of \(\mathbb{N}\). Then clearly \(\bigcup_{i=1}^{\infty} E_i = \bigcup_{i=1}^{\infty} E_{\pi(i)}\), so
\[
\sum_{i=1}^{\infty} \nu(E_{\pi(i)}) = \nu(\bigcup_{i=1}^{\infty} E_{\pi(i)}) = \nu(\bigcup_{i=1}^{\infty} E_i),
\]
which is, by assumption, finite. This applies to any re-arrangement of \(\mathbb{N}\). Thus we apply Lemma 17.13 to conclude that \(\sum_{i=1}^{\infty} |\nu(E_i)| < \infty\), as desired. \(\square\)

**Example 17.15.** Let us consider the measure space \((\mathbb{R}^n, \mathcal{M}_n, m_n)\). Let \(f \in \mathcal{M}_n\). Then \(f = f^+ - f^-\). Also, assume \(f^- \in L^1\). For \(E \in \mathcal{M}_n\), consider
\[
\nu(E) = \int_E f \, dx = \int_E f^+ \, dx - \int_E f^- \, dx.
\]
This is a signed measure not taking the value \(-\infty\) (since we assumed \(f^- \in L^1\)). For example, consider
\[
f(x) = \begin{cases} 
1 & \text{if } x > 0 \\
-1 & \text{if } x \in [-1, 0] \\
0 & \text{otherwise.}
\end{cases}
\]
Let \(E = [-1, 1]\). Then \(\nu(E) = 0\), and so we see that a set having zero signed measure does not tell us anything about its size.

**Definition 17.16.** Suppose \(\nu\) is a signed measure. We have the following types of sets:
Lemma 17.17. Let \( \nu \) be a signed measure on \((X, \mathcal{M})\). Let \( B \in \mathcal{M} \) and \( A \in \mathcal{M} \) with \( A \subset B \). If \(|\nu(B)| < \infty\), then \(|\nu(A)| < \infty\).

Proof. Note that \( B = A \cup (B \setminus A) \), which is a disjoint union, and so \( \nu(B) = \nu(A) + \nu(B \setminus A) \). Suppose to the contrary that \( \nu(B) = \infty \). Then since \(-\infty\) is excluded, we know \( \nu(B \setminus A) \neq -\infty \), and so \( \nu(B) = \nu(A) + \nu(B \setminus A) = \infty \), which is a contradiction from our assumption.

Now suppose \( \nu(A) = -\infty \). A similar contradiction to the first case holds. \( \square \)

Lemma 17.18. The following hold for a signed measure \( \nu \).

1. Let \( E \in \mathcal{M} \) be a positive set for \( \nu \). Let \( G \in \mathcal{M} \) with \( G \subset E \). Then \( G \) is a positive set.
2. Let \( \{E_i\}_{i=1}^{\infty} \subset \mathcal{M} \), where \( E_i \) is a positive set for all \( i \in \mathbb{N} \). Then \( E = \bigcup_{i=1}^{\infty} E_i \) is a positive set for \( \nu \).

Proof. (1) is immediate since any subset of \( G \) is also a subset of \( E \). For (2), take \( A \subset E \). Let \( A_1 = A \cap E_1 \), \( A_2 = (A \cap E_2) \setminus A_1 \), \( A_3 = (A \cap E_3) \setminus (A_1 \cup A_2) \), and in general,

\[
A_n = (A \cap E_n) \setminus \left( \bigcup_{i=1}^{n-1} A_i \right).
\]

Then \( \{A_i\}_{i=1}^{\infty} \) is a disjoint family of sets covering \( A \) with \( A = \bigcup_{i=1}^{\infty} A_i \), and so

\[
\nu(A) = \sum_{i=1}^{\infty} \nu(A_i).
\]

Observe also that \( A_i \subset E_i \), and since \( E_i \) is a positive set for all \( i \in \mathbb{N} \), we have \( \nu(A_i) \geq 0 \) for all \( i \in \mathbb{N} \). Hence, \( \nu(A) = \sum_{i=1}^{\infty} \nu(A_i) \geq 0 \). \( \square \)

Note 17.19. There are analogous statements (and virtually identical proofs) to Lemma 17.18 for negative sets and null sets.

Example 17.20. Let \( f \in L^1(\mathbb{R}) \) with \( \nu(A) = \int_A f \, dx \) for \( A \in \mathcal{M} \). Let \( A \) denote the union of all positive sets in \( \mathbb{R} \), let \( B \) denote the union of all negative sets in \( \mathbb{R} \), and let \( C \) denote the union of all null sets in \( \mathbb{R} \). Let \( A = A \) and let \( B = B \cup C \). Note that \( A \cap B = \emptyset \) and \( A \cup B = \mathbb{R} \). Notice that we have decomposed \( \mathbb{R} \) into two disjoint sets, one a positive set and one a negative set, whose union is all of \( \mathbb{R} \). It is also important to note that this decomposition is not necessarily unique, as we could have distributed the zeros of \( f \) differently. This example illustrates the Hahn Decomposition Theorem, which we build toward now.

Lemma 17.21 (Hahn’s Lemma). Let \((X, \mathcal{M})\) be a measurable space and \( \nu \) a signed measure on \( \mathcal{M} \). Suppose \( E \in \mathcal{M} \) such that \( 0 < \nu(E) < \infty \). Then there is an \( A \in \mathcal{M} \) such that

1. \( A \subset E \),
2. \( A \) is a positive set for \( \nu \), and
3. \( \nu(A) > 0 \).
Proof. If $E$ is a positive set, then we are done by taking $A = E$. Hence, assume that $E$ is not a positive set. Then there exists a set of negative $\nu$-measure inside of $E$. Thus we can find $E_1 \subset E$ and $m_1 \in \mathbb{N}$ such that
\[
\nu(E_1) < -\frac{1}{m_1},
\]
and in fact we may choose $m_1 \in \mathbb{N}$ such that $m_1$ is the smallest natural number for which such a set $E_1$ exists. Suppose $E \setminus E_1$ is a positive set. Then we claim $\nu(E \setminus E_1) > 0$. First of all, note that by Lemma 17.17 that $|\nu(E_1)| < \infty$. Since $E = E_1 \cup (E \setminus E_1)$, which is a disjoint union, we have (since $\nu(E_1)$ is finite)
\[
\nu(E) = \nu(E_1) + \nu(E \setminus E_1) \quad \text{and so} \quad \nu(E \setminus E_1) = \nu(E) - \nu(E_1) > 0,
\]
as desired, since $\nu(E) > 0$ and $\nu(E_1) < 0$. It follows that if $E \setminus E_1$ is a positive set, then the induction stops by taking $A = E \setminus E_1$.

If $E \setminus E_1$ is not a positive set, then $0 < \nu(E \setminus E_1) < \infty$, as we have already shown, and so we may use the same construction as before to find a set $E_2 \subset E \setminus E_1$ such that
\[
\nu(E_2) < -\frac{1}{m_2},
\]
where $m_2 \in \mathbb{N}$ is chosen to be the smallest natural number for which such a set $E_2$ exists. If $(E \setminus E_1) \setminus E_2 = E \setminus (E_1 \cup E_2)$ is a positive set, we are done. If not, then we repeat this construction. In particular, supposing the construction continues, at the $k$th step we may find a set $E_k \subset E \setminus \left( \bigcup_{i=1}^{k-1} E_i \right)$ where
\[
\nu(E_k) < -\frac{1}{m_k},
\]
where $m_k \in \mathbb{N}$ is chosen to be the smallest natural number for which such a set $E_k$ exists. This process may terminate after a finite number of steps, in which case we have satisfied the conclusion of the theorem with the set $A = E \setminus \left( \bigcup_{i=1}^{k-1} E_i \right)$ for some positive integer $k$.

Suppose this process does not terminate. Then put
\[
A = E \setminus \left( \bigcup_{i=1}^{\infty} E_k \right).
\]
Note that $A \cup E_1 \cup E_2 \cup \cdots = E$ is a disjoint union by our construction. Also note that
\[
-\infty < \nu \left( \bigcup_{k=1}^{\infty} E_k \right) = \sum_{k=1}^{\infty} \nu(E_k) < -\sum_{k=1}^{\infty} \frac{1}{m_k},
\]
where the first inequality follows from Lemma 17.17 since $\bigcup_{k=1}^{\infty} E_k \subset E$. This inequality then implies that $\sum_{k=1}^{\infty} \frac{1}{m_k} < \infty$. It follows from this that as $k \to \infty$, we also have $m_k \to \infty$ (in fact, at a much faster rate since the series $\sum_{k=1}^{\infty} \frac{1}{m_k}$ converges).

Since $A \subset E$, we must show two things: first, we must show that $A$ is a positive set, and second, we must show that $\nu(A) > 0$. For the first, let $B \in \mathcal{M}$ such that $B \subset A$. Therefore, for every $j \in \mathbb{N}$, we have
\[
B \subset A = E \setminus \left( \bigcup_{k=1}^{\infty} E_k \right) \subset E \setminus \left( \bigcup_{k=1}^{j-1} E_k \right).
\]
It follows directly from the minimality of choice of $m_j$ that $\nu(B) \geq -\frac{1}{m_j - 1}$, and this holds for all $j \in \mathbb{N}$. Hence, we have

$$\nu(B) \geq \lim_{j \to \infty} -\frac{1}{m_j - 1} = 0,$$

which shows that $A$ is a positive set.

To show $\nu(A) > 0$, we first remark that since $E \setminus A = \bigcup_{k=1}^{\infty} E_k$, which is a disjoint union, we have

$$\nu(E \setminus A) = \sum_{k=1}^{\infty} \nu(E_k) < -\sum_{k=1}^{\infty} \frac{1}{m_k} < 0,$$

and so by the excision property (using the fact that $0 < \nu(E) < \infty$), we have

$$\nu(A) = \nu(E) - \nu(E \setminus A) > 0,$$

as desired to finish the proof. $\square$

**Theorem 17.22** (Hahn’s Decomposition Theorem). Let $\nu$ be a signed measure on $(X,\mathcal{M})$. Then there exists a positive set $A$ and a negative set $B$ such that $A \cap B = \emptyset$ and $A \cup B = X$.

**Proof.** Without loss of generality, assume $\infty$ is the value omitted by $\nu$. Let

$$\mathcal{P} = \{ A : A \text{ is a positive set in } X \}.$$

Let $\lambda = \sup_{\mathcal{P}} \{ \nu(A) \}$, and note that $0 \leq \lambda \leq \infty$. In particular, $\emptyset \in \mathcal{P}$ gives us the inequality $0 \leq \lambda$. Now we may find a collection of sets $\{A_k\}_{k=1}^{\infty} \subset \mathcal{P}$ (not assumed to be distinct, since we cannot even assume that $\mathcal{P}$ is infinite) such that $\nu(A_k) \to \lambda$. Then let $A = \bigcup_{k=1}^{\infty} A_k$. Then by Lemma 17.18(2), it follows that $A \in \mathcal{P}$, and consequently $\nu(A) \leq \lambda$.

On the other hand, for each $k \in \mathbb{N}$, we have $A \setminus A_k \subset A$, and since $A$ is a positive set, it follows that $\nu(A \setminus A_k) \geq 0$. Hence we have

$$\nu(A) = \nu(A \setminus A_k) + \nu(A_k) \geq \nu(A_k).$$

Therefore, since the left-hand side of this inequality does not depend on $k$, we may send $k \to \infty$ to get the inequality

$$\nu(A) \geq \lim_{k \to \infty} \nu(A_k) = \lambda.$$

We have thus shown that $\nu(A) = \lambda$, and hence that $\lambda < \infty$ since $\nu$ omits $\infty$.

Now let $B = X \setminus A$. We must show that $B$ is a negative set. Suppose to the contrary that there exists a subset $E \subset B$ for which $\infty > \nu(E) > 0$. Then, using Hahn’s Lemma, it follows that there is a set $E' \subset E \subset B$ for which $E'$ is a positive set for $\nu$ and $\nu(E') > 0$. But then consider the set $A \cup E'$, which is a disjoint union of two positive sets and is hence positive, so it is a member of $\mathcal{P}$. Hence we have

$$\nu(A \cup E') = \nu(A) + \nu(E') > \nu(A) = \lambda,$$

which is a contradiction since $\lambda = \sup_{\mathcal{P}} \{ \nu(A) \}$. This contradiction establishes the theorem. $\square$

By excising a null set from $A$ and grafting it to $B$, we see that a Jordan decomposition is not unique. We now work toward the Jordan Decomposition Theorem.
Definition 17.23. Let $\mu_1$ and $\mu_2$ be measures on a measurable space $(X, \mathcal{M})$. Then $\mu_1$ and $\mu_2$ are called mutually singular if there exist disjoint sets $E_1, E_2 \in \mathcal{M}$ such that $E_1 \cup E_2 = X$, where $\mu_1(E_2) = 0$ and $\mu_2(E_1) = 0$. We denote mutually singular measures by $\mu_1 \perp \mu_2$.

Example 17.24. Let $f = f^+ - f^-$, where $f^- \in L^1$. Then $\nu(E) = \int_E f \, dx = \int_E f^+ \, dx - \int_E f^- \, dx$ is a signed measure for $(X, \mathcal{M})$, as discussed in Example 17.15. Then the measures $\nu_1(E) = \int_E f^+ \, dx$ and $\nu_2(E) = \int_E f^- \, dx$ are mutually singular measures for the disjoint sets $E_1 = \{x \in X : f(x) \geq 0\}$ and $E_2 = \{x \in X : f(x) < 0\}$.

Theorem 17.25 (Jordan’s Decomposition Theorem). Let $\nu$ be a signed measure on the measurable space $(X, \mathcal{M})$. Then there exists two measures $\nu_1$ and $\nu_2$ such that
\begin{enumerate}
    \item $\nu = \nu_1 - \nu_2$, and
    \item $\nu_1 \perp \nu_2$.
\end{enumerate}
Moreover, this decomposition is unique.

Proof. Let $\{A, B\}$ be a Hahn decomposition of $X$. Let $\nu_1$ and $\nu_2$ be given by
\[ \nu_1(E) = \nu(E \cap A) \quad \text{and} \quad \nu_2(E) = -\nu(E \cap B). \]
Then by construction both $\nu_1$ and $\nu_2$ are measures, and furthermore $\nu_1(B) = \nu(B \cap A) = 0$ and $\nu_2(A) = -\nu(A \cap B) = 0$, so $\nu_1 \perp \nu_2$. Furthermore, for $E \in \mathcal{M}$ we have
\[ \nu(E) = \nu((E \cap A) \cup (E \cap B)) = \nu(E \cap A) + \nu(E \cap B) = \nu_1(E) - \nu_2(E), \]
which completes the proof of existence. Dr. Gulisashvili skipped the proof of uniqueness. \qed

17.2. Outer Measure and the Carathéodory Construction for General Measures.

Definition 17.26. Let $S$ be a set and $\mathcal{P}(S)$ the power set of $S$. A set function $\mu^* : \mathcal{P}(S) \to [0, \infty]$ is called an outer measure provided that
\begin{enumerate}
    \item $\mu^*(\emptyset) = 0$,
    \item if $A \subseteq B$, then $\mu^*(A) \leq \mu^*(B)$, and
    \item if $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{P}(S)$, then
    \[ \mu^* \left( \bigcup_{i=1}^{\infty} A_i \right) \leq \sum_{i=1}^{\infty} \mu^*(A_i). \]
\end{enumerate}

We next describe the Carathéodory condition for measurable sets, which parallels the construction given in §5. In particular, we define the measurable sets by
\[ \mathcal{M} = \{ E \in \mathcal{P}(S) : \text{for all } A \in \mathcal{P}(S), \text{ we have } \mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap E^C) \}. \]
As in the Lebesgue case, it suffices to prove
\[ \mu^*(A) \geq \mu^*(A \cap E) + \mu^*(A \cap E^C) \]
for all $A$ with $\mu^*(A) < \infty$ (otherwise the result is trivial). Using this definition of $\mathcal{M}$, we have the following important theorem.

Theorem 17.27. Let $\mu^*$ be an outer measure on $S$. Then
\begin{enumerate}
    \item We have
\end{enumerate}
(a) If \( E \in \mathcal{M} \), then \( E^C \in \mathcal{M} \).
(b) If \( E_1, \ldots, E_n \in \mathcal{M} \), then \( \bigcup_{i=1}^n E_i \in \mathcal{M} \).
(c) If \( A \subset S \) and \( E_1, \ldots, E_n \in \mathcal{M} \) are disjoint, then
\[
\mu^*(A \cap \left[ \bigcup_{i=1}^n E_i \right]) = \sum_{i=1}^n \mu^*(A \cap E_i), \quad \text{and}
\]
(d) If \( \{E_i\}_{i=1}^\infty \subset \mathcal{M} \), then \( E = \bigcup_{i=1}^\infty E_i \in \mathcal{M} \).

In particular, \( \mathcal{M} \) is a \( \sigma \)-algebra. (That \( \emptyset \in \mathcal{M} \) is a triviality.)

(2) The set function \( \mu = \mu^*|_{\mathcal{M}} \) is a measure.

(3) The measure space \((S, \mathcal{M}, \mu)\) is complete.

**Proof.** (1)(a) Let \( E \in \mathcal{M} \). Then it is clear that \( E^C \in \mathcal{M} \) since the Carathéodory condition is symmetric to taking complements.

(b) Let \( E_1, \ldots, E_n \in \mathcal{M} \). By finite induction, it is enough to show the result for \( n = 2 \). That is, we must show \( E_1 \cup E_2 \in \mathcal{M} \). Using the measurability of \( E_1 \) and \( E_2 \), we have
\[
\mu^*(A) = \mu^*(A \cap E_1) + \mu^*(A \cap E_1^C)
\]
\[
= \mu^*(A \cap E_1) + \mu^*((A \cap E_1^C) \cap E_2) + \mu^*((A \cap E_1^C) \cap E_2^C) = \mu^*(A \cap (E_1 \cup E_2)^C)
\]

Hence, by this reasoning it suffices to show \( \mu^*(A \cap E_1) + \mu^*((A \cap E_1^C) \cap E_2) \geq \mu^*(A \cap (E_1 \cup E_2)) \).

But we note the obvious set equality \( A \cap E_1 \cup [A \cap E_1^C \cup E_2] = A \cap (E_1 \cup E_2) \), from which the desired inequality follows from finite sub-additivity (which we did not show for outer measure, but is obvious using the same tricks we used for the general measure). Hence \( E_1 \cup E_2 \in \mathcal{M} \).

(c) Again, by finite induction we show the result for \( n = 2 \). Using the fact that \( E_2 \in \mathcal{M} \), we have
\[
\mu^*(A \cap [E_1 \cup E_2]) = \mu^*(A \cap [E_1 \cup E_2] \cap E_2) + \mu^*(A \cap [E_1 \cup E_2] \cap E_2^C) = \mu^*(A \cap E_2) + \mu^*(A \cap E_2^C) = \mu^*(A \cap E_1).
\]

The equalities given in the underbraces follow because \( E_1 \) and \( E_2 \) are disjoint. This shows the result for \( n = 2 \), and consequently for any finite \( n \in \mathbb{N} \).

(d) Without loss of generality, assume \( \{E_i\}_{i=1}^\infty \) is a disjoint family. If not, we may replace this family with a disjoint family having the same union. Let \( F_n = \bigcup_{i=1}^n E_i \). By part (b), we know that \( F_n \in \mathcal{M} \). Since \( F_n \subset E \), we have \( E^C \subset F_n^C \) for every \( n \in \mathbb{N} \). Let \( A \subset S \). Then we have
\[
\mu^*(A) = \mu^*(A \cap F_n) + \mu^*(A \cap F_n^C) \geq \mu^*(A \cap F_n) + \mu^*(A \cap E^C) = \sum_{i=1}^n \mu^*(A \cap E_i) + \mu^*(A \cap E^C).
\]

The first equality follows since \( F_n \in \mathcal{M} \). The inequality follows from monotonicity because \( E^C \subset F_n^C \) (meaning that \( A \cap E^C \subset A \cap F_n^C \)). The final equality follows from part (c). The left side \( \mu^*(A) \)
does not depend on \( n \), so we may send \( n \to \infty \) to get
\[
\mu^*(A) \geq \sum_{i=1}^{\infty} m^*(A \cap E_i) + \mu^*(A \cap E^C) \geq \mu^* \left( \bigcup_{i=1}^{\infty} (A \cap E_i) \right) + \mu^*(A \cap E^C).
\]

The second inequality follows from countable sub-additivity. This is precisely the condition that \( E \in \mathcal{M} \), completing the proof of (d).

(2) We know \( \emptyset \in \mathcal{M} \), and \( \mu^*(\emptyset) = 0 \), so \( \mu(\emptyset) = 0 \). It remains to prove that if \( \{E_i\}_{i=1}^{\infty} \) is a disjoint family of measurable sets, then
\[
\mu \left( \bigcup_{i=1}^{\infty} E_i \right) = \sum_{i=1}^{\infty} \mu(E_i).
\]

We have already established \( \leq \) in (40) since \( \mu^* \) is countably sub-additive. Note that
\[
\mu^* \left( \bigcup_{i=1}^{\infty} E_i \right) \geq \mu^* \left( \bigcup_{i=1}^{n} E_i \right) = \sum_{i=1}^{n} \mu^*(E_i).
\]

Here, the inequality follows by monotonicity and the equality follows from (1)(c) by taking \( A = S \). The left-hand side of this inequality does not depend on \( n \), so we may send \( n \to \infty \) to get
\[
\mu^* \left( \bigcup_{i=1}^{\infty} E_i \right) \geq \sum_{i=1}^{\infty} \mu^*(E_i),
\]

which establishes the inequality \( \geq \) from (40). This proves (2).

(3) Let \( E \in \mathcal{M} \) with \( \mu(E) = 0 \). Let \( E' \subseteq E \). Then \( \mu^*(E) = 0 \), and so \( \mu^*(E') = 0 \) by monotonicity. We claim this means \( E' \in \mathcal{M} \). In particular, we must show
\[
\mu^*(A) \geq \mu^*(A \cap E') + \mu^*(A \cap (E')^C),
\]

provided \( \mu^*(A) < \infty \). But \( \mu^*(A \cap E') = 0 \) by monotonicity, and \( \mu^*(A \cap (E')^C) \leq m^*(A) \) by monotonicity, so the result follows and \( E' \in \mathcal{M} \). This completes the proof of (3). \( \square \)

17.3. Constructing Outer Measures. One major difference between Lebesgue measure on Euclidean space and general measure is the existence of “standard” sets; in \( \mathbb{R} \) the standard sets are open intervals, and in \( \mathbb{R}^n \) for \( n > 1 \), the standard sets are closed cubes. However, we can still provide an analogue to this framework for general measures by declaring in advance what our “standard” sets are. Consider a set function \( \mu : S \to [0, \infty] \), where \( S \) is a family of subsets of \( X \), and will play the role of our “standard” sets. Define a set function \( \mu^* : \mathcal{P}(X) \to [0, \infty] \) by
\[
\mu^*(E) = \inf \left\{ \sum_{k=1}^{\infty} \mu(E_k) : E \subseteq \bigcup_{k=1}^{\infty} E_k, \ \{E_k\}_{k=1}^{\infty} \subset S \right\},
\]

where \( E \neq \emptyset \), and \( \mu^*(\emptyset) = 0 \). We also follow the convention that \( \inf \emptyset = \infty \), a scenario that would arise if there were no admissible cover of \( E \) from sets in \( S \). Then \( \mu^* \) is an outer measure.

Thus, if we start with a set function \( \mu \) on \( S \subseteq \mathcal{P}(X) \), the collection of “standard” sets, we can form the outer measure \( \mu^* \) on \( \mathcal{P}(X) \), and from this we can construct the Carathéodory measure.
on the collection of measurable sets $\mathcal{M}$. In general, however, we cannot say anything about the relationship between $S$ and $\mathcal{M}$.

**Note 17.28.** Let $S \subset \mathcal{P}(X)$ be a collection of “standard” sets, with set function $\mu : S \to [0, \infty]$, and let $\mathcal{M}$ and $\varpi$ be the Carathéodory $\sigma$-algebra and measure, respectively, as discussed in the preceding paragraph. Denote by $\sigma(S)$ the smallest $\sigma$-algebra containing all the sets of $S$; that is, the $\sigma$-algebra generated by $S$. We ask the following questions.

1. When is $\sigma(S) \subset \mathcal{M}$?
2. If $\sigma(S) \subset \mathcal{M}$, when does $\varpi$ extend $\mu$ from $S$ to $\mathcal{M}$ (i.e. if $E \in S$, then $E \in \mathcal{M}$ and $\varpi(E) = \mu(E)$)?

The next important theorem gives an answer to these questions. We start with a definition.

**Definition 17.29.** A set function $\mu : S \to [0, \infty]$ is a **pre-measure on** $S$ if

1. If $\emptyset \in S$, then $\mu(\emptyset) = 0$,
2. if $E_1, E_2 \in S$ and $E_1 \subseteq E_2$, then $\mu(E_1) \leq \mu(E_2)$,
3. if $\{E_k\} \subset S$, and if $\bigcup_{k=1}^{\infty} E_k \in S$, then $\mu(\bigcup_{k=1}^{\infty} E_k) \leq \sum_{k=1}^{\infty} \mu(E_k)$, and
4. if $E_1, \ldots, E_n \in S$ is a disjoint collection in $S$, and if $\bigcup_{k=1}^{n} E_k \in S$, then
   $$\mu \left( \bigcup_{k=1}^{n} E_k \right) = \sum_{k=1}^{n} \mu(E_k).$$

**Theorem 17.30** (Carathéodory Extension). Suppose the following conditions are met.

1. $\mu : S \to [0, \infty]$ is a pre-measure on $S$, and
2. If $E_1, E_2 \in S$, then $E_1 \setminus E_2 \in S$. (We say $S$ is **closed under relative complements**.)

Then $S \subset \mathcal{M}$, and if $E \in S$, then $\varpi(E) = \mu(E)$.

**Note 17.31.** To get the Lebesgue measure, we apply the Carathéodory Extension Theorem to the collection

$$S = \{\text{all finite almost disjoint unions of closed rectangles}\}.$$ 

Then if $E \in S$, we define $\mu(E) = \sum_{k=1}^{n} |R_k|$. 


18. GENERAL MEASURABLE FUNCTIONS

Definition 18.1. Let \((S, \mathcal{M})\) be a measurable space. Let \(f : S \to \mathbb{R}\) be a real-valued function. Then \(f\) is measurable if for every \(c \in \mathbb{R}\), the set
\[
\{ f > c \} = \{ x \in S : f(x) > c \} \in \mathcal{M}.
\]

Though it is not standard notation, we denote the set of measurable functions by \(\mathcal{M}\).

Note 18.2. As in the Lebesgue case, this definition is equivalent if we replace \(>\) with the other inequalities \(\geq, \leq,\) and \(<\).

Note 18.3 (Measurability and Operations). The following are ways to combine measurable functions to produce another measurable function.

1. If \(f, g \in \mathcal{M}\), then \(f + g, f - g, fg, \) and \(f/g\) (where \(g \neq 0\)) are all measurable.
2. If \(\{f_k\}_{k=1}^{n} \subseteq \mathcal{M}\), then \(\max\{f_k\}\) and \(\min\{f_k\}\) are measurable.
3. If \(\{f_k\}_{k=1}^{\infty} \subseteq \mathcal{M}\), then \(\sup\{f_k\}, \inf\{f_k\}, \limsup f_k,\) and \(\liminf f_k\) are all measurable, and \(\lim f_k\) is measurable where it exists.

Note 18.4. Suppose \(f\) and \(g\) are functions on \((S, \mathcal{M}, \mu)\). Suppose further that \(f\) is measurable and \(f = g\) almost everywhere with respect to \(\mu\). Then \(g\) is measurable provided that \((S, \mathcal{M}, \mu)\) is a complete measure space.

Proof. Let \(A = \{ x \in S : f(x) = g(x) \} \in \mathcal{M}\) and \(B = \{ x \in S : f(x) \neq g(x) \} \in \mathcal{M}\). By assumption, \(\mu(B) = 0\), and since \((X, \mathcal{M}, \mu)\) is a complete measure space, any subset of \(B\) is also measurable.

Let \(c \in \mathbb{R}\). Then the set \(B \cap \{ g > c \} \subseteq B\), and so \(B \cap \{ g > c \} \in \mathcal{M}\). Thus the set
\[
\{ g > c \} = \left( \{ f = g \} \cap \{ f > c \} \right) \cup \left( \{ f \neq g \} \cap \{ g > c \} \right)
\]
is measurable, which completes the proof. \(\square\)

We now provide a counter-example in the case that \((S, \mathcal{M}, \mu)\) is not complete. Consider \((\mathbb{R}, \mathcal{B}, m)\), which as previously discussed is not complete. We recall that there is a set \(C' \subset \mathcal{C}\) which is not Borel, where \(\mathcal{C}\) is the Cantor set. Let
\[
f(x) = \begin{cases} 1 & \text{if } x \in [0, 1] \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad g(x) = \begin{cases} 1 & \text{if } x \in [0, 1] \setminus C' \\ 2 & \text{if } x \in C' \\ 0 & \text{otherwise.} \end{cases}
\]

Now \(f = g\) almost everywhere on \((\mathbb{R}, \mathcal{B}, m)\) since \(f = g\) on \(\mathbb{R} \setminus \mathcal{C}\) and \(m(C) = 0\) (we cannot use \(\mathbb{R} \setminus \mathcal{C}'\) because this is not a Borel set). Now \(\{ g > 3/2 \} = C' \notin \mathcal{B}\), so \(g\) is not a measurable function.

We now discuss simple approximation.

Definition 18.5. A simple function on \((S, \mathcal{M})\) has the form
\[
s(x) = \sum_{k=1}^{n} a_k \chi_{E_k}(x),
\]
where \(\{ E_k \}_{k=1}^{n} \subseteq \mathcal{M}\). 

100
Theorem 18.6 (Simple Approximation Theorem). The following hold for a measurable function \( f \) on a measurable space \((X, \mathcal{M}, \mu)\).

(1) Suppose \( f(x) \geq 0 \) for all \( x \in X \). Then there exists an increasing sequence of non-negative simple functions \( \{\varphi_n\}_{n=1}^\infty \) such that \( \lim_{n \to \infty} \varphi_n(x) = f(x) \).

(2) If \( f \) is not assumed to be non-negative, then there exists a sequence of simple functions \( \{\varphi_n\}_{n=1}^\infty \) such that \( |\varphi_n| \) is an increasing sequence and \( \lim_{n \to \infty} \varphi_n(x) = f(x) \).

Note 18.7. Suppose we are working in the measure space \((\mathbb{R}^n, \mathcal{M}_n, m_n)\), and let \( f \in \mathcal{M}_n \). By the Simple Approximation Theorem, we can find a sequence of simple function \( \{\tilde{\varphi}_n\}_{n=1}^\infty \) such that \( |\tilde{\varphi}_n| \) is an increasing sequence and \( \lim_{n \to \infty} \tilde{\varphi}_n(x) = f(x) \).

Example 18.8. Consider the measure space \(((0, 1], \mathcal{P}([0, 1]), c)\) where \( c \) is the counting measure given in Example 17.5(3). Let \( f(x) = 1 \) for all \( x \in [0, 1] \). The Simple Approximation Theorem works trivially since \( f \) itself is a simple function. However, it is impossible to approximate \( f \) by a sequence of simple functions supported on sets of finite measure. Suppose to the contrary that there exists an increasing sequence of simple functions \( \{s_k\}_{k=1}^\infty \) with \( \lim_{k \to \infty} s_k(x) = 1 \) for all \( x \in [0, 1] \), and where \( s_k \) is supported on a set of finite measure for all \( k \in \mathbb{N} \). Let

\[
A = \bigcup_{k=1}^\infty \text{supp}(s_k).
\]

Then \( s_k(x) = 0 \) for all \( x \in A^C \) and for all \( k \in \mathbb{N} \). This means that \( \lim_{k \to \infty} s_k(x) = 0 \) for all \( x \in [0, 1] \setminus A \). Since \( A \) is countable, being the countable union of finite sets, the set \( [0, 1] \setminus A \) is uncountable. This shows that there are uncountably many points for which \( \lim_{k \to \infty} s_k(x) \neq 1 \). This contradiction shows no such sequence of simple functions supported on sets of finite measure can exist.

Theorem 18.9 (Egorov’s Theorem). Let \((X, \mathcal{M}, \mu)\) be a measure space and let \( E \in \mathcal{M} \) with \( \mu(E) < \infty \). Suppose \( f_k \to f \) point wise, where \( \{f_k\}_{k=1}^\infty \subset \mathcal{M} \). Then for every \( \epsilon > 0 \), there exists a measurable function \( E_\epsilon \in \mathcal{M} \) such that \( E_\epsilon \subset E, \mu(E \setminus E_\epsilon) < \epsilon \), and \( f_k \to f \) uniformly on \( E_\epsilon \).

Note 18.10. Recall that Lusin’s Theorem involved continuous functions. In the setting of general measure spaces, however, we do not assume a topology on our set \( X \), and so Lusin’s Theorem does not exist in these general settings.
19. General Integration Theory

Suppose \((X, \mathcal{M}, \mu)\) is a measure space. We seek to define

\[
\int_X f \, d\mu
\]

for all non-negative functions and all integrable functions. In our construction for the Lebesgue case, we had four stages corresponding to four classes of functions \(K_1, K_2, K_3,\) and \(K_4\). However, we will skip the class \(K_2\) in this general case.

To see why we must skip \(K_2\), consider the “minimal” measurable space \((X, \mathcal{M})\), where \(\mathcal{M} = \{X, \emptyset\}\), and suppose \(\mu\) is a measure for this space given by \(\mu(\emptyset) = 0\) and \(\mu(X) = \infty\). Consider the non-negative function \(f = \chi_X \in K_3\), so that

\[
\int_X f \, d\mu = \infty.
\]

The only bounded measurable functions supported on a set of finite measure in this space are identically zero, so \(K_2 = \{0\}\), where 0 is the zero function. Since \(\int_X 0 \, d\mu = 0\), we cannot define integrals from \(K_3\) by using functions from \(K_2\). Hence, \(K_2\) should be removed from the construction.

Note 19.1 (Step 1). We begin with simple functions on a measurable space \((X, \mathcal{M}, \mu)\). Indeed, since we only use the class of simple functions to approximate non-negative measurable functions, we may assume all of our simple functions are non-negative. Recall by a simple function we mean a function \(s\) of the form

\[
s(x) = \sum_{i=1}^k a_i \chi_{E_i}(x),
\]

where in this case we assume \(a_i \geq 0\) for all \(i \in \{1, \ldots, k\}\). A simple function takes only a finite number of values \(c_1, \ldots, c_m\), where each \(c_j\) is some positive sum involving the terms \(a_i\). Hence we may form the standard representation, which is given by

\[
s(x) = \sum_{i=1}^m c_i \chi_{\mu^{-1}(c_i)},
\]

where clearly the sets \(E_i = \mu^{-1}(c_i)\) are disjoint and measurable. Then we define the integral of such a simple function as

\[
\int s \, d\mu = \sum_{i=1}^m c_i \mu(E_i).
\]

We will consider some properties of this integral later.

Note 19.2 (Step 3). Suppose \(f\) is a non-negative measurable function. Then we define

\[
\int f \, d\mu = \sup_{0 \leq s \leq f \text{ is simple}} \left\{ \int s \, d\mu \right\}.
\]

Note that we have skipped Step 2 in our construction per our notes at the beginning of the section.

Note 19.3 (Properties). The following are properties of the integral for non-negative measurable functions.

1. \(\int (f + g) \, d\mu = \int f \, d\mu + \int g \, d\mu\),
If \( \alpha \geq 0 \) and \( f \) is non-negative and measurable, then \( \int (\alpha f) \, d\mu = \alpha \int f \, d\mu \),

(3) If \( f_1 \) and \( f_2 \) are non-negative and measurable and \( f_1(x) \leq f_2(x) \) for all \( x \in X \), then \( \int f_1 \, d\mu \leq \int f_2 \, d\mu \), and

(4) If \( E_1, E_2 \in \mathcal{M} \) are disjoint, then

\[
\int_{E_1 \cup E_2} f \, d\mu = \int_{E_1} f \, d\mu + \int_{E_2} f \, d\mu.
\]

**Definition 19.4.** A measurable function \( f \) on the measure space \((X, \mathcal{M}, \mu)\) is called **integrable** with respect to \( \mu \) if

\[
\int |f| \, d\mu < \infty.
\]

The integral for such a function is defined to be

\[
\int f \, d\mu = \int f^+ \, d\mu - \int f^- \, d\mu.
\]

This definition is valid since \( \int f^+ \, d\mu \leq \int |f| \, d\mu \) and \( \int f^- \, d\mu \leq \int |f| \, d\mu \). We denote

\[
L^1(\mu) = \{ f \text{ is } \mu\text{-integrable} \}.
\]

All previous properties of integration hold.

**Note 19.5.** From this definition, we also get immediately that

\[
\left| \int f \, d\mu \right| \leq \left| \int f^+ \, d\mu \right| + \left| \int f^- \, d\mu \right| = \int f^+ \, d\mu - \int f^- \, d\mu = \int |f| \, d\mu,
\]

or \( |\int f \, d\mu| \leq \int |f| \, d\mu \) for short.

**Theorem 19.6** (Chebyshev’s Inequality). Let \( f \in L^1(\mu) \) and \( \lambda > 0 \). Then

\[
\mu \{ x \in X : |f(x)| > \lambda \} \leq \frac{1}{\lambda} \int |f| \, d\mu.
\]

**Lemma 19.7.** Let \( f \) be a measurable function on \((X, \mathcal{M}, \mu)\) and let \( g : \mathbb{R} \to \mathbb{R} \) be continuous. Then the function \( h = g \circ f : X \to \mathbb{R} \) is measurable.

**Proof.** By the definition of measurability, we must show that \( \{ h > c \} = h^{-1}((c, \infty)) \in \mathcal{M} \). Now \( h^{-1} = f^{-1} \circ g^{-1} \), and so

\[
h^{-1}((c, \infty)) = f^{-1} \circ g^{-1}((c, \infty)) = \text{Open}.
\]

Let \( O = g^{-1}((c, \infty)) \) be an open set since \( g \) is continuous. By the Structure Theorem for open sets, we have \( O = \bigcup_{k=1}^{\infty} I_k \) for disjoint open intervals \( I_k \). Now since \( f \) is measurable, it follows that \( f^{-1}(I_k) \) is a measurable set (i.e. if \( I_k = (a_k, b_k) \), then \( f^{-1}(I_k) = \{ x \in X : f(x) > a_k \} \cap \{ x \in X : f(x) < b_k \} \in \mathcal{M} \). It follows that

\[
h^{-1}((c, \infty)) = f^{-1} \left( \bigcup_{k=1}^{\infty} I_k \right) = \bigcup_{k=1}^{\infty} f^{-1}(I_k) \in \mathcal{M},
\]

as desired to finish the proof. \( \square \)

\(^{†}\)This definition will be modified slightly later.
Corollary 19.8. Let $f$ be a measurable function on $(X, \mathcal{M}, \mu)$ and let $p > 0$. Then the function $|f|^p$ is a measurable function.

Proof. Take $g : \mathbb{R} \to \mathbb{R}$ to be defined by $g(x) = |x|^p$, which is a continuous function. Then the function $h = g \circ f = |f|^p$ is a measurable function by Lemma 19.7. □

Definition 19.9. † Let $p > 0$. We say that a measurable function $f$ on the measure space $(X, \mathcal{M}, \mu)$ belongs to the class $L^p(\mu)$ if

$$\int |f|^p \, d\mu < \infty.$$  

†This definition will be modified slightly later.
20. The Radon-Nikodym Theorem and The Lebesgue Decomposition Theorem

We now emphasize that the property of a measure space being σ-finite is a particularly important property. Recall for a measure space \((X, \mathcal{M}, \mu)\) that we defined \(\int_E f \, d\mu = \int_X f \chi_E \, d\mu\).

**Lemma 20.1.** Let \((X, \mathcal{M}, \mu)\) be a measure space and let \(f\) be a non-negative measurable function. Then, if \(\int f \, d\mu < \infty\), then \(\mu(\{x \in X : f(x) = \infty\}) = 0\) (that is, \(f\) is finite almost everywhere) and \(\mu(\{x \in X : f(x) > 0\})\) is σ-finite.

**Proof.** Let \(A = \{x \in X : f(x) = \infty\}\). First, suppose \(\mu(A) > 0\). Recall that \(A \in \mathcal{M}\) since \(A = \bigcap_{n=1}^{\infty} \{x \in X : f(x) > n\}\). Then we have

\[
\int_X f \, d\mu \geq \int_A f \, d\mu = \int_X f \chi_A \, d\mu = \infty \cdot \mu(A) = \infty,
\]

contradicting our assumption. Hence \(\mu(A) = 0\), as desired.

Now let \(B = \{x \in X : f(x) > 0\}\) and let \(B_n = \{x \in X : f(x) \geq 1/n\}\). Then \(B = \bigcup_{n=1}^{\infty} B_n\), so to show \(B\) is σ-finite it suffices to show \(\mu(B_n) < \infty\) for every \(n \in \mathbb{N}\). By Chebyshev’s Inequality, since \(f \in L^1(\mu)\) by assumption we have

\[
\mu(B_n) = \mu\left(\left\{x \in X : f(x) \geq \frac{1}{n}\right\}\right) \leq n \int_{B_n} f \, d\mu < \infty,
\]

as desired to finish the proof. \(\square\)

**Definition 20.2.** Suppose \(\nu\) and \(\mu\) are measures on a measurable space \((X, \mu)\). The measure \(\nu\) is called **absolutely continuous with respect to** \(\mu\) if whenever \(E \in \mathcal{M}\) such that \(\mu(E) = 0\), we have \(\nu(E) = 0\). We denote this by \(\nu \ll \mu\).

**Lemma 20.3.** Let \((X, \mathcal{M}, \mu)\) be a finite measure space and let \(\nu\) be another finite measure on \((X, \mathcal{M})\). Suppose \(\nu \ll \mu\). Then if \(\nu \neq 0\), then there exists a non-negative measurable function \(f\) such that

\[
\int_X f \, d\mu > 0 \quad \text{and} \quad \int_E f \, d\mu \leq \nu(E)
\]

for all \(E \in \mathcal{M}\).

**Proof.** Let \(\lambda > 0\) and consider the signed measure \(\nu - \lambda \mu\). Since \(\nu\) and \(\mu\) are finite, this signed measure omits both \(\infty\) and \(-\infty\). Let us consider a Hahn decomposition \(\{P_\lambda, N_\lambda\}\) for this signed measure.

We claim that there exists a \(\lambda > 0\) such that \(\mu(P_\lambda) > 0\). Suppose to the contrary that no such \(\lambda\) exists. Then \(\mu(P_\lambda) = 0\) for all \(\lambda > 0\). Note, then, that for all \(E \in \mathcal{M}\) such that \(E \subseteq P_\lambda\), we have \(\mu(E) = 0\), and since \(\nu \ll \mu\), we also have \(\nu(E) = 0\). Also, for all \(F \in \mathcal{M}\) such that \(F \subseteq N_\lambda\), since \(N_\lambda\) is a negative set for \((\nu - \lambda \mu)\), we have \((\nu - \lambda \mu)(F) \leq 0\), which implies \(\nu(F) \leq \lambda \mu(F)\).

With these observations in mind, take \(A \in \mathcal{M}\). Then

\[
A = (A \cap P_\lambda) \cup (A \cap N_\lambda).
\]
Then $E, F \in \mathcal{M}$ and $E \subset P_\lambda$ and $F \subset N_\lambda$. Since $A = E \cup F$ is a disjoint union, by our preceding comments that $\nu(E) = 0$ and $\nu(F) \leq \lambda \mu(F)$, we have

$$\nu(A) = \nu(E) + \nu(F) = \nu(F) \leq \lambda \mu(F) \leq \lambda \mu(X).$$

This inequality holds for every $\lambda > 0$. Since $\mu$ is a finite measure, it follows that $\mu(X) < \infty$. Hence we may send $\lambda \to 0$ in this inequality to get $\nu(A) = 0$. Since $A \in \mathcal{M}$ was arbitrarily chosen, it follows that $\nu(A) = 0$ for all $A \in \mathcal{M}$, which is a contradiction since we are assuming $\nu \not\equiv 0$. We have established our claim that there is a $\lambda > 0$ such that $\mu(P_\lambda) > 0$.

Using this fact, fix $\lambda_0 > 0$ for which $\mu(P_{\lambda_0}) > 0$. Then define

$$f(x) = \lambda_0 \chi_{P_{\lambda_0}}(x).$$

It is clear that $f$ is non-negative and measurable. It is also clear that

$$\int_X f \, d\mu = \int_X \lambda_0 \chi_{P_{\lambda_0}} \, d\mu = \int_{P_{\lambda_0}} \lambda_0 \, d\mu = \lambda_0 \mu(P_{\lambda_0}) > 0$$

since both $\lambda_0 > 0$ and $\mu(P_{\lambda_0}) > 0$. It remains to show that $\int_E f \, d\mu \leq \nu(E)$ for every $E \in \mathcal{M}$. To this end, let $E \in \mathcal{M}$. For any two sets $A$ and $B$, note that

$$\chi_A \cdot \chi_B = \chi_{A \cap B}.$$

Using this identity, we have

$$\int_E f \, d\mu = \int_X \lambda_0 \chi_{P_{\lambda_0}} \cdot \chi_E \, d\mu = \int_X \lambda_0 \chi_{P_{\lambda_0} \cap E} \, d\mu = \int_{P_{\lambda_0} \cap E} \lambda_0 \, d\mu = \lambda_0 \mu(P_{\lambda_0} \cap E).$$

Recall that $P_{\lambda_0}$ is a positive set for $\nu - \lambda_0 \mu$, and so

$$\nu(P_{\lambda_0} \cap E) - \lambda_0 \mu(P_{\lambda_0} \cap E) \geq 0, \quad \text{or equivalently} \quad \lambda_0 \mu(P_{\lambda_0} \cap E) \leq \nu(P_{\lambda_0} \cap E).$$

Putting this all together gives us

$$\int_E f \, d\mu = \lambda_0 \mu(P_{\lambda_0} \cap E) \leq \nu(P_{\lambda_0} \cap E) \leq \nu(E),$$

as desired to finish the proof. \qed

**Lemma 20.4.** Let $(X, \mathcal{M}, \mu)$ be a finite measure space and let $\nu$ be another finite measure on $(X, \mathcal{M})$. Further, suppose that $\nu \ll \mu$ and $\nu \not\equiv 0$. Let

$$\mathcal{F} = \left\{ f \in K_3 : \int_X f \, d\mu > 0 \right\}.$$

(Recall that $f \in K_3$ means $f$ is non-negative and measurable.) Let $M = \sup_{\mathcal{F}} \left\{ \int_X f \, d\mu \right\}$. Then

1. $M < \infty$,
2. If $f_1, \ldots, f_n \in \mathcal{F}$, then $\max\{f_1, \ldots, f_n\} \in \mathcal{F}$.

**Proof.** (1) Let $f \in \mathcal{F}$ (which is non-empty by Lemma 20.3), and so by definition $\int_E f \, d\mu \leq \nu(E)$ for all $E \in \mathcal{M}$. Hence

$$\int_X f \, d\mu \leq \nu(X) < \infty,$$
and since the right-hand side does not depend on the choice of \(f \in \mathcal{F}\), we get

\[
M = \sup_{\mathcal{F}} \left\{ \int_X f \, d\mu \right\} \leq \nu(X) < \infty,
\]
as desired.

(2) By finite induction, it suffices to prove the result for two functions. Let \(f, g \in \mathcal{F}\). Let \(h = \max\{f, g\}\) for simplicity of notation. It is clear that \(h\) is non-negative and measurable. We also have by monotonicity of integration that

\[
\int_X h \, d\mu \geq \int_X f \, d\mu > 0.
\]

It remains to show \(\int_E h \, d\mu \leq \nu(E)\) for every \(E \in \mathcal{M}\). Since \(f\) and \(g\) are both measurable, it follows that \(f - g\) is measurable and so \(E_1 = \{x \in X : f - g \geq 0\} = \{x \in X : f \geq g\} \in \mathcal{M}\). Likewise, \(E_2 = \{x \in X : g > f\} \in \mathcal{M}\), and these two sets are disjoint. Thus we may write

\[
h(x) = f(x)\chi_{E_1}(x) + g(x)\chi_{E_2}(x).
\]

For \(E \in \mathcal{M}\), we can write \(E = (E \cap E_1) \cup (E \cap E_2)\), which is a disjoint union. Thus, by the additive property of the integral over disjoint sets we have

\[
\int_E h \, d\mu = \int_{E \cap E_1} h \, d\mu + \int_{E \cap E_2} h \, d\mu
\]
\[
= \int_{E \cap E_1} (f\chi_{E_1} + g\chi_{E_2}) \, d\mu + \int_{E \cap E_2} (f\chi_{E_1} + g\chi_{E_2}) \, d\mu
\]
\[
= \int_{E \cap E_1} f\chi_{E_1} \, d\mu + \int_{E \cap E_1} g\chi_{E_2} \, d\mu + \int_{E \cap E_2} f\chi_{E_1} \, d\mu + \int_{E \cap E_2} g\chi_{E_2} \, d\mu
\]
\[
= \int_X f\chi_{E_1}\chi_{E \cap E_1} \, d\mu + \int_X g\chi_{E_2}\chi_{E \cap E_2} \, d\mu
\]
\[
= \int_X f\chi_{E \cap E_1} \, d\mu + \int_X g\chi_{E \cap E_2} \, d\mu
\]
\[
= \int_{E \cap E_1} f \, d\mu + \int_{E \cap E_2} g \, d\mu
\]
\[
\leq \nu(E \cap E_1) + \nu(E \cap E_2)
\]
\[
= \nu(E).
\]

Here the inequality follows from the assumption that \(f, g \in \mathcal{F}\). This shows that \(h \in \mathcal{F}\), as desired to complete the proof. \(\Box\)

20.1. **Radon-Nikodym.**

**Theorem 20.5** (Radon-Nikodym). Let \((X, \mathcal{M}, \mu)\) be a \(\sigma\)-finite measure space and let \(\nu\) be another \(\sigma\)-finite measure on \((X, \mathcal{M})\). Then \(\nu \ll \mu\) if and only if there exists a non-negative measurable function \(f\) on \(X\) such that

\[
\nu(E) = \int_E f \, d\mu
\]

for all \(E \in \mathcal{M}\). This function \(f\) is unique in the sense that if \(g\) is any non-negative measurable function on \(X\) that also satisfies this property, then \(g = f\) almost everywhere with respect to \(\mu\).
Proof. First, suppose there exists a non-negative measurable function \( f \) for which \( \nu(E) = \int_E f \, d\mu \) for every \( E \in \mathcal{M} \). Suppose \( \mu(E) = 0 \) for some \( E \in \mathcal{M} \). Then \( \nu(E) = \int_E f \, d\mu = 0 \), since the integral over any set of measure zero is zero.

Now let \( \nu \ll \mu \). Suppose we can prove the theorem for any pair of finite measures on \((X, \mathcal{M}, \mu)\). Since we are assuming \( \nu \) and \( \mu \) are \( \sigma \)-finite measures, we can find disjoint families of sets \( \{A_k\}_{k=1}^\infty \) and \( \{B_k\}_{k=1}^\infty \) such that \( \nu(A_k) \ll \infty \) and \( \mu(B_k) \ll \infty \) for all \( k \in \mathbb{N} \). Then if we consider the disjoint family of sets \( \{E_k\}_{k=1}^\infty \), where the sets \( E_k \) are pairwise intersections of sets from the family \( \{A_k\} \) with the family \( \{B_k\} \), we have \( \bigcup_{k=1}^\infty E_k = X \) and \( \nu(E_k) \ll \infty \) and \( \mu(E_k) \ll \infty \) for each \( k \in \mathbb{N} \).

For each \( k \in \mathbb{N} \), let \( \nu_k \) and \( \mu_k \) be measures defined by \( \nu_k(E) = \nu(E \cap E_k) \) and \( \mu_k(E) = \mu(E \cap E_k) \), which are finite measures on \((X, \mathcal{M})\) for all \( k \in \mathbb{N} \). Also, whenever \( \mu_k(E) = 0 \), since \( \mu_k(E) = \mu(E \cap E_k) = 0 \) and \( \nu \ll \mu \), we have

\[
\nu_k(E) = \nu(E \cap E_k) = 0,
\]

and so \( \nu_k \ll \mu_k \) for all \( k \in \mathbb{N} \). Since we are assuming the finite measure version of the Radon-Nikodym Theorem, we know for every \( k \in \mathbb{N} \) that there exists a non-negative measurable function \( f_k \) defined on \( X \) such that

\[
\nu(E \cap E_k) = \nu_k(E) = \int_E f_k \, d\mu_k
\]

for every \( E \in \mathcal{M} \). Using this, we define a non-negative measurable function \( f \) by the following rule. For each \( x \in X \), we have \( x \in E_k \) for exactly one \( k \in \mathbb{N} \). Thus for this \( x \), we assign \( f(x) = f_k(x) \). This uniquely defines a function \( f \) on all of \( X \), and in fact, this gives us the useful characterization \( f_k = f \chi_{E_k} \). Now \( f \) is clearly non-negative since each function \( f_k \) is non-negative. To see that \( f \) is measurable, we note that

\[
\{ x \in E : f(x) > c \} = \bigcup_{k=1}^\infty \{ x \in E_k : f_k(x) > c \} \in \mathcal{M}.
\]

Also note that \( \nu(E) = \sum_{k=1}^\infty \nu(E \cap E_k) \) for any \( E \in \mathcal{M} \) since \( E = \bigcup_{k=1}^\infty (E \cap E_k) \) is a disjoint union. Thus we have

\[
\nu(E) = \sum_{k=1}^\infty \nu(E \cap E_k) = \sum_{k=1}^\infty \int_E f_k \, d\mu_k = \sum_{k=1}^\infty \int_E f \chi_{E_k} \, d\mu_k = \sum_{k=1}^\infty \int_E f \chi_{E_k} \, d\mu
\]

We may replace \( d\mu_k \) with \( d\mu \) in the final equality on the first line of this chain of equalities because \( E \cap E_k \) (and any measurable subset of \( E \cap E_k \)) has the same measure with respect to both \( \mu \) and \( \mu_k \). This shows how the result for the \( \sigma \)-finite case follows from the finite case. Hence, it is enough to assume from the beginning that \( \nu \) and \( \mu \) are finite measures. Furthermore, if \( \nu \equiv 0 \), then \( f \equiv 0 \) trivially and we are done, so we may assume \( \nu \not\equiv 0 \).

Let \( \mathcal{F} \) be as defined in Lemma 20.4. By Lemma 20.3 it follows that \( \mathcal{F} \neq \emptyset \). We also know from Lemma 20.4 that

\[
M = \sup_{\mathcal{F}} \left\{ \int_X f \, d\mu \right\} < \infty.
\]
We wish to find a function \( F \in \mathcal{F} \) such that \( \int_X F \, d\mu = M \). Using the definition of supremum, there is a sequence of (non-necessarily distinct) functions \( \{f_n\}_{n=1}^\infty \subset \mathcal{F} \) such that \( \int_X f_n \, d\mu \to M \). Without loss of generality, we may assume \( \{f_n\} \) is an increasing sequence. To see this, we remark that if \( \{f_n\} \) is not an increasing sequence, we form a new sequence \( \{g_n\} \subset \mathcal{F} \) where \( g_n = \max\{f_1, \ldots, f_n\} \), which is a member of \( \mathcal{F} \) by Lemma 20.4(2). Then \( \{g_n\} \) is an increasing sequence of functions, and since we have

\[
\int_X f_n \, d\mu \leq \int_X g_n \, d\mu \leq M
\]

for all \( n \in \mathbb{N} \), it follows that \( \int_X g_n \, d\mu \to M \). Thus it is safe to assume \( \{f_n\} \) is an increasing sequence of functions.

Now increasing sequences have limits (which may be infinite), so let \( F = \lim_{n \to \infty} f_n \). Now

\[
0 < \int_X f_n \, d\mu \quad \text{and} \quad \int_E f_n \, d\mu \leq \nu(E).
\]

for every \( n \in \mathbb{N} \). Notice that \( F \) must be finite almost everywhere. Otherwise, there would be a set of positive measure \( A \) for which \( f_n|_A \) would be unbounded as \( n \to \infty \), which in turn implies \( \int_A f_n \, d\mu \) would be unbounded as \( n \to \infty \). But since

\[
\int_A f_n \, d\mu \leq \int_X f_n \, d\mu \leq M < \infty,
\]

we see no such set \( A \) of positive measure can exist, so \( F \) is finite almost everywhere. Thus we may use the Monotone Convergence Theorem (which holds for general measure spaces) to conclude that

\[
\int_E F \, d\mu = \lim_{n \to \infty} \int_E f_n \, d\mu \leq \nu(E).
\]

Furthermore, since \( F \geq f_n \) for every \( n \in \mathbb{N} \), we have \( 0 < \int_X f_n \, d\mu \leq \int_X F \, d\mu \). It follows that \( F \in \mathcal{F} \) and \( \int_X F \, d\mu = M \).

Now set \( \eta(E) = \nu(E) - \int_E F \, d\mu \geq 0 \), which is a measure (seen by using both the fact that \( \nu \) is a measure and the additivity property of the integral over disjoint families of sets). Now

\[
\eta(X) = \nu(X) - \int_X F \, d\mu < \infty,
\]

and so \( \eta \) is a finite measure. We claim \( \eta \ll \mu \). Let \( E \in \mathcal{M} \) with \( \mu(E) = 0 \). Using that fact that \( \nu \ll \mu \), we have

\[
\eta(E) = \nu(E) - \int_E F \, d\mu = 0 - 0 = 0,
\]

where \( \int_E F \, d\mu = 0 \) since \( \mu(E) = 0 \). Hence \( \eta \ll \mu \), as claimed. We consider cases.

First, suppose \( \eta \equiv 0 \). Then we are done and \( F = \frac{d\nu}{d\mu} \). (See Note 20.6 for this notation.)

For the second case, suppose \( \eta \neq 0 \). Then we apply the exact same procedures we have already used to the pair of finite measures \( \mu \) and \( \eta \). Thus, there exists a non-negative measurable function \( \tilde{f} \) such that

\[
\int_X \tilde{f} \, d\mu > 0 \quad \text{and} \quad \int_E \tilde{f} \, d\mu \leq \eta(E) = \nu(E) - \int_E F \, d\mu.
\]
Using the latter inequality, this implies $\int_E (\tilde{f} + F) \leq \nu(E)$, and clearly $\int_X (\tilde{f} + F) \, d\mu > 0$. Thus we have $\tilde{f} + F \in \mathcal{F}$. But

$$\int_X (\tilde{f} + F) \, d\mu = \int_X \tilde{f} \, d\mu + \int_X F \, d\mu > M$$

since $\int_X F \, d\mu = M$ and $\int_X \tilde{f} \, d\mu > 0$ by assumption. This contradicts $M = \sup_{\mathcal{F}} \{ \int_X f \, d\mu \}$. Hence this case does not occur. We have now shown the existence of the non-negative function $f$ on $X$ such that

$$\nu(E) = \int_E f \, d\mu.$$  

The proof will be complete upon showing $f$ is unique. Suppose $g$ has the properties mentioned in the Radon-Nikodym Theorem. Let

$$E = \{ x \in X : g(x) = f(x) \} \quad \text{and} \quad G = \{ x \in X : f(x) \neq g(x) \}.$$  

Note that $G = \{ x \in X : f(x) - g(x) > 0 \} \cup \{ x \in X : f(x) - g(x) < 0 \}$, which is a union of measurable sets since $f - g$ and $g - f$ are measurable functions. Thus $G \in \mathcal{M}$, and by consequence, $E = G^C \in \mathcal{M}$.

Our goal is to show $f = g$ almost everywhere with respect to $\mu$, or more precisely, that $\mu(G) = 0$. Now we have by assumption that

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

for every $A \in \mathcal{M}$, from which it follows that

$$\int_A (f - g) \, d\mu = 0$$

for all $A \in \mathcal{M}$. Now $A = (E \cap A) \cup (G \cap A)$ is a disjoint union, and so

$$\int_{E \cap A} (f - g) \, d\mu + \int_{G \cap A} (f - g) \, d\mu = 0, \text{ or simply } \int_{G \cap A} (f - g) \, d\mu = 0 \tag{41}$$

for any $A \in \mathcal{M}$. Suppose now to the contrary that $\mu(G) > 0$. Now let

$$P = \{ x \in X : f(x) > g(x) \} \quad \text{and} \quad Q = \{ x \in X : f(x) < g(x) \},$$

and so $G = P \cup Q$ is a disjoint union of measurable sets; hence, $\mu(G) = \mu(P) + \mu(Q)$, and so either $\mu(P) > 0$ or $\mu(Q) > 0$. Without loss of generality, suppose $\mu(P) > 0$. Then from (41) and using the fact that $G \cap P = P$ we get

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

But $f - g > 0$ on $P$ and $\mu(P) > 0$, so this integral must be greater than zero as well, a contradiction. Therefore, we must have $\mu(G) = 0$, completing the proof. □

**Note 20.6.** The function $f = \frac{d\nu}{d\mu}$ of the Radon-Nikodym Theorem is unique and is called the **Radon-Nikodym derivative** of $\nu$ with respect to $\mu$.

**Example 20.7.** The following example illustrates the necessity of the $\sigma$-finite condition of the measures. Consider the measurable space $([0, 1], \mathcal{B})$, where $\mathcal{B}$ is the Borel sets restricted to $[0, 1]$, and let $\mu = c$ (the counting measure) and $\nu = m$ (the Lebesgue measure). We note that
(1) $\nu$ is a finite measure,
(2) $\mu$ is not a $\sigma$-finite measure, and
(3) $\nu \ll \mu$ trivially, since the only set of zero measure for $\mu$ is $\emptyset$.

Suppose the Radon-Nikodym Theorem holds for $\nu$ and $\mu$. Then there is a non-negative measurable function $f$ such that

$$m(E) = \int_E f \, dc.$$  

Note that the function $f \neq 0$. For if $f \equiv 0$, then $0 = \int_E f \, dc = m(E)$ for all $E \in \mathcal{B}$, which is clearly not true. Hence, take $x \in [0, 1]$ such that $f(x) > 0$. Let $E = \{x\}$. Then $m(E) = 0$ but $\int_E f \, dc = f(x) > 0$. Hence, we see that $\sigma$-finiteness is essential for the Radon-Nikodym Theorem.

20.2. Lebesgue Decomposition.

**Lemma 20.8.** Let $\nu$ and $\mu$ be measures on a measurable space $(X, \mathcal{M})$. Let $f$ be a non-negative measurable function. Then $\nu + \mu$ is a measure on $(X, \mathcal{M})$ and

$$\int_E f \, d(\nu + \mu) = \int_E f \, d\nu + \int_E f \, d\mu.$$  

**Proof.** It is clear that $\nu + \mu$ is a measure on $(X, \mathcal{M})$. We start by showing the equality for simple functions. If $s(x) = \sum_{k=1}^n c_k \chi_{E_k}$ is a simple function, then

$$\int_E s \, d(\nu + \mu) = \sum_{k=1}^n c_k (\nu + \mu)(E_k) = \sum_{k=1}^n [c_k \nu(E_k) + c_k \mu(E_k)]$$

$$= \sum_{k=1}^n c_k \nu(E_k) + \sum_{k=1}^n c_k \mu(E_k) = \int_E s \, d\nu + \int_E s \, d\mu.$$  

This establishes the result for simple functions. Now for the general case where $f$ is a non-negative measurable function, we have

$$\int_E f \, d(\nu + \mu) = \sup_{0 \leq s \leq f, \ s \text{ is simple}} \left\{ \int_E s \, d(\nu + \mu) \right\}$$

$$= \sup_{0 \leq s \leq f, \ s \text{ is simple}} \left\{ \int_E s \, d\nu + \int_E s \, d\mu \right\}$$

$$\leq \sup_{0 \leq s \leq f, \ s \text{ is simple}} \left\{ \int_E s \, d\nu \right\} + \sup_{0 \leq t \leq f, \ t \text{ is simple}} \left\{ \int_E t \, d\mu \right\}$$

$$= \int_E f \, d\nu + \int_E f \, d\mu.$$  

On the other hand, whenever $s, t$ are simple functions, we know $u = \max\{s, t\}$ is a simple function, and so we have

$$\int_E f \, d\nu + \int_E f \, d\mu = \sup_{0 \leq s \leq f, \ s \text{ is simple}} \left\{ \int_E s \, d\nu \right\} + \sup_{0 \leq t \leq f, \ t \text{ is simple}} \left\{ \int_E t \, d\mu \right\}$$

$$= \sup_{0 \leq s, t \leq f, \ s, t \text{ are simple}} \left\{ \int_E s \, d\nu + \int_E t \, d\mu \right\}.$$
\[
\begin{align*}
\leq & \sup_{0 \leq s, t \leq f, s, t \text{ are simple}} \left\{ \int_E u \, d\nu + \int_E u \, d\mu \right\} \\
= & \sup_{0 \leq g \leq f, g \text{ is simple}} \left\{ \int_E g \, d(\nu + \mu) \right\} \\
\leq & \sup_{0 \leq g \leq f, g \text{ is simple}} \left\{ \int_E g \, d(\nu + \mu) \right\} \\
= & \int_E f \, d(\nu + \mu).
\end{align*}
\]

Here, the second inequality follows since the set
\[
\{ u : u = \max\{s, t\} \text{ where } 0 \leq s, t \leq f \text{ and } s \text{ and } t \text{ are simple functions} \}
\]
is a subset of all simple functions bounded below by 0 and above by \( f \). Thus, since we are taking the supremum of a larger set, the inequality is justified. This completes the proof. \( \square \)

**Theorem 20.9** (Lebesgue Decomposition). Let \((X, \mathcal{M})\) be a measurable space and \(\mu, \nu\) be \(\sigma\)-finite measures on \((X, \mathcal{M})\). Then there exists measures \(\nu_0\) and \(\nu_1\) such that

1. \(\nu = \nu_0 + \nu_1\),
2. \(\nu_0 \perp \mu\), and
3. \(\nu_1 \ll \mu\).

The pair \((\nu_0, \nu_1)\) is unique.

**Proof.** Set \(\lambda = \mu + \nu\), which is clearly a \(\sigma\)-finite measure. If \(\lambda(E) = 0\), then we must have \(\mu(E) = 0\), and so \(\mu \ll \lambda\). By the Radon-Nikodym Theorem, there exists a non-negative measurable function \(f\) such that
\[
\mu(E) = \int_E f \, d\lambda
\]
for all \(E \in \mathcal{M}\). Define
\[
X_+ = \{ x \in X : f(x) > 0 \} \quad \text{and} \quad X_0 = \{ x \in X : f(x) = 0 \},
\]
and note that \(X = X_+ \cup X_0\). Set \(\nu_0(E) = \nu(E \cap X_0)\) and \(\nu_1(E) = \nu(E \cap X_+),\) which are measures. We now check the properties we require for the conclusion of the proof.

1. We have
\[
\nu_0(E) + \nu_1(E) = \nu(E \cap X_0) + \nu(E \cap X_+) = \nu(E)
\]
since \(X = X_0 \cup X_+\) is a disjoint union. Thus \(\nu = \nu_0 + \nu_1\).

2. We have
\[
\mu(X_0) = \int_{X_0} f \, d\lambda = 0
\]
since \(f \equiv 0\) on \(X_0\). We also have \(\nu_0(X_+) = \nu(X_+ \cap X_0) = \nu(\emptyset) = 0.\) Hence \(\nu_0 \perp \mu\).

3. Let \(\mu(E) = 0.\) Then \(\int_E f \, d\mu = 0\), and so, using Lemma 20.8, we have
\[
0 = \mu(E) = \int_E f \, d\lambda = \int_E f \, d\mu + \int_E f \, d\nu = \int_E f \, d\nu.
\]
From this it follows that
\[ \int_{E \cap X_+} f \, d\nu = 0, \]
and since \( f > 0 \) on \( X_+ \), it consequently must be that \( \nu(E \cap X_+) = 0 \). Hence \( \nu_1(E) = \nu(E \cap X_+) = 0 \), and so we see \( \nu_1 \ll \mu \), as desired.

The uniqueness is left to the reader. \( \square \)

**Note 20.10.** Since we have shown \( \nu_1 \ll \mu \) in the Lebesgue Decomposition Theorem, by the Radon-Nikodym Theorem there is a non-negative measurable function \( h \) such that \( \nu_1(E) = \int_E h \, d\mu \) for every \( E \in \mathcal{M} \). Hence we have
\[ \nu(E) = \int_E h \, d\mu + \nu_0(E), \]
where \( \nu_0 \perp \mu \). This is the **Lebesgue decomposition** of \( \nu \) with respect to \( \mu \). We call \( \int_E h \, d\mu \) the **absolutely continuous part** and \( \nu_0(E) \) the **singular part**.
21. $L^p$ Spaces

Let $(X, \mathcal{M}, \mu)$ be a measure space. (In fact, we might as well fix a general measure space now. Everything in this section will be assumed to hold on this general measure space unless otherwise specified.) Let $0 < p < \infty$ (in fact, most of the time we require $p \geq 1$). The **conjugate number** for $p$ is the number $q$ for which

$$ \frac{1}{p} + \frac{1}{q} = 1. $$

That is, $q = \frac{p}{p-1}$. If $p = 1$, then we define $q = \infty$.

**Definition 21.1.** For a measurable function $f$, we say that $f \in L^p$ if $\int_X |f|^p d\mu < \infty$. This makes the $L^p$ space a function class on $X$.

**Lemma 21.2.** Suppose $p > 0$ and $a, b \in \mathbb{R}$. Then $|a + b|^p \leq 2^p (|a|^p + |b|^p)$.

**Proof.** We use the triangle inequality to conclude

$$ |a + b| \leq |a| + |b| \leq 2 \max\{|a|, |b|\}. $$

From this we have

$$ |a + b|^p \leq 2^p (\max\{|a|, |b|\})^p \leq 2^p (|a|^p + |b|^p), $$

which completes the proof. \qed

**Lemma 21.3.** $L^p$ is a linear class.

**Proof.** First, let $f \in L^p$ and let $\alpha \in \mathbb{R}$. Then

$$ \int_X |\alpha f|^p d\mu = |\alpha|^p \int_X |f|^p d\mu < \infty, $$

and so $\alpha f \in L^p$.

It immediately follows from Lemma 21.2 that if $f, g \in L^p$, then $f + g \in L^p$. This shows that $L^p$ is a linear class. \qed

One goal we have is to introduce a **norm** on $L^p$. If $L$ is a linear class, then a norm is a function $\| \cdot \| : L \to [0, \infty)$ that satisfies the following axioms.

1. $\|0\| = 0$, where $0$ is the zero element of $L$.
2. If $\|x\| = 0$, then $x = 0$.
3. If $\alpha \in \mathbb{R}$ and $x \in L$, then $\|\alpha x\| = |\alpha|\|x\|$.
4. $\|x + y\| \leq \|x\| + \|y\|$. (Triangle inequality.)

Satisfying these axioms, we have a **normed linear space**, or simply a **normed space**. In addition, $d(x, y) = \|x - y\|$ is a translation invariant distance function, giving a **metric space**.

**Note 21.4.** We would like to define a norm on $L^p$ if $p \geq 1$. A natural choice is

$$ \|f\|_p = \left( \int_X |f|^p d\mu \right)^{1/p}. $$

However, a problem arises with the second axiom in the definition of a normed space. In particular, if $f = 0$ almost everywhere with respect to $\mu$, but $f \neq 0$, then we will have

$$ \|f\|_p = \left( \int_X |f|^p d\mu \right)^{1/p} = 0, $$
contradicting the second axiom. The remedy this apparent flaw, we do not distinguish between functions which are equal almost everywhere with respect to $\mu$.

**Note 21.5** (Equivalence Relation on $\mathcal{M}$ ($\mu$-measurable functions)). We say $f \sim g$ if and only if $f = g$ almost everywhere with respect to $\mu$. This is clearly an equivalence relation on $\mathcal{M}$, and so $\mathcal{M}$ can be represented as a disjoint union of the $\sim$ equivalence classes.

**Definition 21.6** (Modified Definition of $L^p$). We now think of the elements of $L^p$ as equivalence classes of $\sim$; that is, $[f] \in L^p$ if and only if $\int_X |f|^p \, d\mu < \infty$. This takes care of the problem with the second axiom and makes $L^p$ ($p \geq 1$) a normed linear space.

**Note 21.7.** So far we have restricted our attention to $p \geq 1$, but what if $0 < p < 1$? We may try to define a norm in the same way:

$$\|f\|_p = \left\{ \int_X |f|^p \, d\mu \right\}^{1/p}.$$  

However, in this case the triangle inequality fails. We could also try

$$\|f\|_p^* = \int_X |f|^p \, d\mu.$$  

However, in this case $\|\alpha f\|_p^* \neq |\alpha| \|f\|_p^*$, so in either case, this is not a normed space. However, if we define

$$d(f, g) = \int_X |f - g|^p \, d\mu$$

for $p < 1$, then we have a metric space structure on our normed linear space $L^p$.

**Note 21.8.** Now consider $\|f\|_{\infty} = \sup_{x \in X} |f(x)|$ for $f \in \mathcal{M}$. This definition poses a problem in extending the functions that are equal to $f$ almost everywhere, since this supremum may not be unique. However, for the class $[f]$ we write

$$\|[f]\|_{\infty} = \inf_{g \sim f} \|g\|_{\infty} = \inf_{g \sim f} \left( \sup_{x \in X} |g(x)| \right),$$

which is called the *essential supremum*. We denote this as

$$\|[f]\|_{\infty} = \text{ess sup}_{x \in X} |f(x)|.$$  

This makes $L^\infty$ a normed linear space.

**Note 21.9.** Even though we now talk about elements of $L^p$ being equivalence classes, we will often revert to the notation $f$ instead of $[f]$ (and $\|f\|_{\infty}$ instead of $\|[f]\|_{\infty}$), but we must always keep in mind we mean the equivalence class of $f$, not the function $f$ itself. As one example, we will write

$$\|f\|_{\infty} = \text{ess sup}_{x \in X} |f(x)|,$$

rather than writing the brackets.

**Note 21.10.** Let $f \in L^\infty$. Then for each $n \in \mathbb{N}$ there is a set $E_n \in \mathcal{M}$ such that

$$|f| \leq \|f\|_{\infty} + \frac{1}{n} \quad \text{on} \quad X \setminus E_n, \quad \text{where} \quad \mu(E_n) = 0.$$
(To see this, note that by the definition of the infimum there is a function \( g_n \sim f \) such that 
\[ |g_n| \leq \|f\|_\infty + 1/n \] on \( X \). Since \( g_n \sim f \), we choose \( E_n \) such that \( \{x \in X : f \neq g_n\} \subset E_n \) with 
\( \mu(E_n) = 0 \). Let \( E_\infty = \bigcup_{n=1}^{\infty} E_n \). Then it follows that
\[ |f| \leq \|f\|_\infty \] on \( X \setminus E_\infty \), and \( \mu(E_\infty) = 0 \).
That is, \( |f| \leq \|f\|_\infty \) almost everywhere with respect to \( \mu \).

We now discuss some very important inequalities dealing with \( L^p \) spaces that are also covered in functional analysis.

**Lemma 21.11 (Young’s Inequality).** Suppose \( 1 < p < \infty \). If \( a, b > 0 \), then
\[ ab \leq \frac{a^p}{p} + \frac{b^q}{q} \]

**Proof.** Consider the function \( g(x) = \frac{1}{p} x^p + \frac{1}{q} - x \). Then \( g'(x) = x^{p-1} - 1 \). Since \( p > 1 \), we have
\[ g'(x) = \begin{cases} 
> 0 & \text{if } x > 1 \\
0 & \text{if } x = 1 \\
< 0 & \text{if } x < 1 
\end{cases} \]
It follows that we have an absolute minimum at \( x = 1 \), and the absolute minimum value is
\[ g(1) = \frac{1}{p} + \frac{1}{q} - 1 = 0. \]
Thus \( g(x) \geq 0 \) for all \( x \), and consequently
\[ x \leq \frac{1}{p} x^p + \frac{1}{q} \]
for all \( x \). In particular, let \( x = \frac{a^p}{b^q-1} \). Then
\[ \frac{a}{b^{q-1}} \leq \frac{1}{p} \left( \frac{1}{b^{q-1}} \right)^p + \frac{1}{q}. \]
Simplifying this expression, we have
\[ \frac{a}{b^{q-1}} \leq \frac{1}{p} \cdot \frac{a^p}{b^p(q-1)} + \frac{1}{q}. \]
Note that \( p(q-1) = q \), and so
\[ \frac{a}{b^{q-1}} \leq \frac{1}{p} \cdot \frac{a^p}{b^q} + \frac{1}{q}. \]
Multiplying both sides by \( b^q \) completes the proof. \( \square \)

**Theorem 21.12 (Hölder Inequality).** Assume \( 1 \leq p < \infty \) and \( f \in L^p \), \( g \in L^q \), where \( q \) is the conjugate to \( p \). Recall that if \( p = 1 \), then we define \( q = \infty \). Then \( \|fg\|_1 \leq \|f\|_p \|g\|_q \). That is,
\[ \int_X |fg| \, d\mu \leq \left( \int_X |f|^p \, d\mu \right)^{1/p} \left( \int_X |g|^q \, d\mu \right)^{1/q}. \]

**Proof.** First, consider \( p = 1 \). Then we must show
\[ \int_X |fg| \, d\mu \leq \|g\|_\infty \int_X |f| \, d\mu = \int_X |f| \|g\|_\infty \, d\mu. \]  \( (42) \)
Note that by Note 21.10 we have $|g| \leq \|g\|_{\infty}$ almost everywhere with respect to $\mu$. Hence $|fg| \leq \|f\|_{\infty} \|g\|_{\infty}$ almost everywhere with respect to $\mu$, from which (42) follows by monotonicity of integration. Hence we have shown the result for the case $p = 1$.

Let $1 < p < \infty$. Let $a = \frac{|f(x)|}{\|f\|_p}$ and $b = \frac{|g(x)|}{\|g\|_q}$ for some fixed $x \in X$. By Young’s Inequality, we have

$$\frac{|f(x)g(x)|}{\|f\|_p \|g\|_q} \leq \frac{|f(x)|^p}{(\|f\|_p)^p \cdot p} + \frac{|g(x)|^q}{(\|g\|_q)^q \cdot q}.$$

Now our $x$ was chosen arbitrarily, so this holds for every $x \in X$. Thus we can integrate over $X$ using the monotonicity and linearity of integration to get

$$\frac{1}{\|f\|_p \|g\|_q} \int_X |fg| \, d\mu \leq \frac{1}{(\|f\|_p)^p \cdot p} \int_X |f|^p \, d\mu + \frac{1}{(\|g\|_q)^q \cdot q} \int_X |g|^q \, d\mu$$

$$= \frac{(\|f\|_p)^p}{(\|f\|_p)^p \cdot p} + \frac{(\|g\|_q)^q}{(\|g\|_q)^q \cdot q}$$

$$= \frac{1}{p} + \frac{1}{q} = 1,$$

where the last equality follows since $q$ is the conjugate of $p$. Multiplying both sides by $\|f\|_p \|g\|_q$ finishes the proof since $\int_X |fg| \, d\mu = \|fg\|_1$. \qed

Note 21.13. By Hölder’s Inequality, for $p \geq 1$, $h \in L^p$ and $g \in L^q$, we have

$$\int_X |hg| \, d\mu \leq \|h\|_p \|g\|_q.$$

(43)

Fix a function $h = f$. We want to consider functions $g \in L^q$ with $\|g\|_q = 1$, making (43) become

$$\int_X |fg| \, d\mu \leq \|f\|_p.$$

Now we would like to find such a function $g$ such that

$$\int_X |fg| \, d\mu = \int_X f g \, d\mu \quad \text{and} \quad \int_X fg \, d\mu = \|f\|_p.$$

That is, we want to find a function in $L^q$ to make Hölder’s inequality an equality. We do this now, but first we define a new type of function.

Definition 21.14. The sign function $\text{sgn}(u)$ is defined by

$$\text{sgn}(u) = \begin{cases} 
1 & \text{if } u > 0 \\
0 & \text{if } u = 0 \\
-1 & \text{if } u < 0
\end{cases}$$

Theorem 21.15 (Addition to Hölder’s Inequality). Consider a function $f \in L^p$ on $(X, \mathcal{M}, \mu)$, where $f \neq 0$ and $p \geq 1$. Set

$$f^*(x) = (\|f\|_p)^{1-p} \cdot \text{sgn}[f(x)] \cdot |f(x)|^{p-1}.$$

Then
(1) \( f^* \in L^q \), where \( q \) is the conjugate number to \( p \),
(2) \( \|f^*\|_q = 1 \), and
(3) \( \int_X f \cdot f^* \, d\mu = \|f\|_p \).

Proof. Recall that \( f^* \in L^q \) if and only if \( \int_X |f^*|^q \, d\mu < \infty \). Hence, establishing (2) will simultaneously establish (1). With this in mind, we compute

\[
\|f^*\|_q = \left\{ \int_X |(\|f\|_p)^{1-p} \cdot \text{sgn}(f(x)) \cdot |f(x)|^{p-1}| \cdot \|f\|_p^q \, d\mu \right\}^{1/q}
\]

\[
= (\|f\|_p)^{1-p} \left\{ \int_X |f(x)|^{p-1} \, d\mu \right\}^{1/q}
\]

\[
= \left\{ \int_X |f| \, d\mu \right\}^{\frac{1}{p} (1-p)} \left\{ \int_X |f| \, d\mu \right\}^{1/q}
\]

\[
= \left\{ \int_X |f| \, d\mu \right\}^0
\]

\[
= 1.
\]

The equalities \( q(p-1) = p \) and \( \frac{1}{p}(1-p) + \frac{1}{q} = \frac{1}{p} - 1 + \frac{1}{q} = 0 \) were used in this calculation. Thus we have established both (2) and (1).

For (3), we compute

\[
\int_X f \cdot f^* \, d\mu = (\|f\|_p)^{1-p} \int_X \underbrace{f(x) \cdot \text{sgn}(f(x)) \cdot |f(x)|^{p-1}}_{=|f(x)|} \, d\mu
\]

\[
= (\|f\|_p)^{1-p} \int_X |f(x)| \cdot |f(x)|^{p-1} \, d\mu
\]

\[
= (\|f\|_p)^{1-p} \int_X |f| \, d\mu
\]

\[
= (\|f\|_p)^{1-p} \cdot (\|f\|_p)^p
\]

\[
= \|f\|_p.
\]

This completes the proof. □

Using these results, we state and prove Minkowski's Inequality, which is essential, as it supplies the triangle inequality in showing that \( L^p \) is a normed linear space.

**Theorem 21.16** (Minkowski’s Inequality). If \( 1 \leq p \leq \infty \), and \( f, g \in L^p \), then

\[
\|f + g\|_p \leq \|f\|_p + \|g\|_p.
\]

**Proof.** For \( p = 1 \), the result is immediate by the ordinary triangle inequality. Next, consider \( p = \infty \).

By Note 21.10, we have \( |f| \leq \|f\|_\infty \) and \( |g| \leq \|g\|_\infty \) almost everywhere with respect to \( \mu \). Hence

\[
|f + g| \leq |f| + |g| \leq \|f\|_\infty + \|g\|_\infty
\]

almost everywhere with respect to \( \mu \). It follows that

\[
\|f + g\|_p = \inf_{h \sim f + g} \left( \sup_{x \in X} |h(x)| \right) \leq \|f\|_\infty + \|g\|_\infty.
\]
Hence, let $1 < p < \infty$. From Theorem 21.15(3), we get
\[
\|f + g\|_p = \int_X (f + g)(f + g)^* \, d\mu
\]
\[
= \int_X [f(f + g)^* + g(f + g)^*] \, d\mu
\]
\[
= \int_X f(f + g)^* \, d\mu + \int_X g(f + g)^* \, d\mu
\]
\[
\leq \|f\|_p \|f + g\|^* + q + \|g\|_p \|f + g\|^* q
\]
\[
= \|f\|_p + \|g\|_p.
\]
The inequality follows from Hölder’s Inequality. The final equality follows from Theorem 21.15(2). Thus the proof is complete. \[\square\]

Minkowski’s Inequality was the final piece in proving:

**Theorem 21.17.** If $1 \leq p \leq \infty$, then $L^p$ is a normed linear space of equivalence classes.

In fact, for $1 \leq p \leq \infty$, $L^p$ is a special type of normed linear space.

**Definition 21.18.** Recall that $d(x, y) = \|x - y\|$ defines a metric on a normed linear space $(X, \| \cdot \|)$. A sequence $\{x_n\}_{n=1}^{\infty}$ is called a **Cauchy sequence** if $\|x_n - x_m\| \to 0$ as $n \to \infty$ and $m \to \infty$. A normed linear space $(X, \| \cdot \|)$ is called **complete** if for every Cauchy sequence $\{x_n\}_{n=1}^{\infty}$, there is an element $x \in X$ for which $\|x_n - x\| \to 0$ as $n \to \infty$.

**Definition 21.19.** A normed linear space $X$ is called a **Banach space** if $(X, \| \cdot \|)$ is complete.

**Theorem 21.20.** For $1 \leq p \leq \infty$, the space $L^p$ is a Banach space.

*Proof.* Omitted, as this is a result of functional analysis. \[\square\]

**Example 21.21.** We have two important examples of $L^p$ spaces.

1. The Lebesgue $L^p$ space on $(\mathbb{R}^n, \mathcal{M}_n, m_n)$. We have
   \[
   \|f\|_p = \left\{ \int_{\mathbb{R}^n} |f|^p \, dm_n \right\}^{1/p}.
   \]
2. Take $\omega = \{0, 1, 2, \ldots \}$ and consider $(\omega, \mathcal{P}(\omega), c)$, where $c$ is the counting measure. Functions on this space are simply sequences. Because our $\sigma$-algebra is $\mathcal{P}(\omega)$, all sequences (i.e. functions) are measurable. Also, to be equal almost everywhere is to be equal everywhere since we are working with the counting measure $c$.

   Note that the space $L^\infty$ consists of all bounded sequences, since for $\alpha = \{\alpha_n\}_{n=0}^{\infty} \in \mathcal{M}$, we have
   \[
   \|\alpha\|_\infty = \sup_{n \geq 0} \{ |\alpha_n| \},
   \]
   which is only finite if and only if $\alpha$ is a bounded sequence. (Here we referred only to the supremum and not the essential supremum because the two are the same in this space. This is because almost everywhere is the same as everywhere.) We denote this space by $\ell^\infty$, and in general, for $1 \leq p \leq \infty$, we denote the space by $\ell^p$. 

119
Note that $\ell^1$ is the space of absolutely continuous sequences because

$$\|\alpha\|_1 = \int \omega |\alpha| \, dc = \sum_{n=0}^{\infty} |\alpha_n| < \infty.$$ 

For $1 < p < \infty$, we define the norm on $\ell^p$ by

$$\|\alpha\|_p = \left\{ \int_X |\alpha|^p \, dc \right\}^{1/p} = \left\{ \sum_{n=0}^{\infty} |\alpha_n|^p \right\}^{1/p}.$$ 

**Lemma 21.22.** Suppose $\mu(X) < \infty$. Then if $1 \leq p_1 \leq p_2 \leq \infty$, then $L^{p_2} \subset L^{p_1}$.

**Proof.** Clearly the result holds if $p_2 = \infty$ since if $f \in L^{\infty}$, then $f$ is bounded almost everywhere, and since $\mu(X) < \infty$, we have $\int_X |f|^p \, d\mu < \infty$ for any $p < \infty$. Now let $p_2 < \infty$. Let $f \in L^{p_2}$. We want to show $\int_X |f|^{p_1} \, d\mu < \infty$. Let $h = |f|^{p_1}$, $j \equiv 1$, $p = \frac{p_2}{p_1} \geq 1$, and $q$ the conjugate number to $p$. Then by the Hölder Inequality, we have

$$\int_X |f|^{p_1} \, d\mu = \int_X |h|^{p} \, d\mu \leq \|h\|_p \|j\|_q = \left\{ \int_X |h|^p \, d\mu \right\}^{1/p} \left\{ \int_X 1 \, d\mu \right\}^{1/q}$$

$$= \left\{ \int_X |f|^{p_2} \, d\mu \right\}^{1/p} \left\{ \mu(X) \right\}^{1/q} < \infty,$$

and so we see $f \in L^{p_1}$, as desired to finish the proof. \(\square\)

**Note 21.23.** The previous result is known as an embedding theorem, since we are showing that one space is embedded in another.

**Example 21.24.** The assumption that $\mu(X) < \infty$ was a necessary assumption. For example, consider the measure space $(\mathbb{R}, \mathcal{M}, m_1)$, the one-dimensional Lebesgue measure space. Consider the function $f \equiv 1$ on $\mathbb{R}$. Then $f \in L^\infty$ since $\|f\|_\infty = \text{ess sup}_{x \in \mathbb{R}} |f(x)| = 1 < \infty$, but clearly $f \notin L^p$ for any $1 \leq p < \infty$.

**Lemma 21.25.** Let $\mu(X) \leq \infty$. If $1 \leq p_1 \leq p_2 \leq \infty$, then $L^\infty \cap L^{p_1} \subset L^{p_2}$.

**Proof.** If $\mu(X) < \infty$, then the result follows from Lemma 21.22 since $L^\infty \subset L^{p_2}$. Now suppose $\mu(X) = \infty$. If $p_2 = \infty$, then the result is immediate, so let $p_2 < \infty$. Let $f \in L^\infty \cap L^{p_1}$. Let

$$f_1 = f \cdot \chi_{\{|f| > 1\}} \quad \text{and} \quad f_2 = f \cdot \chi_{\{|f| \leq 1\}},$$

so that $f = f_1 + f_2$. If $f(x) = \infty$, then $f_2(x) = \infty \cdot 0$; however, this will happen at most on a set of measure zero since $f \in L^\infty$, so without loss of generality we can assume $\text{supp}(f) \subset [-1, 1]$. Since $p_1 \leq p_2$, then it follows that $|f_2(x)|^{p_2} \leq |f_2(x)|^{p_1}$. Also, note that $|f_2(x)| \leq |f(x)|$. Since $f \in L^{p_1}$, we have

$$\int_X |f_2|^{p_2} \, d\mu \leq \int_X |f|^{p_1} \, d\mu \leq \int_X |f|^{p_1} \, d\mu < \infty.$$ 

It follows that $f_2 \in L^{p_2}$.

Now we also have $|f_1| \leq |f|$. Since $f \in L^\infty$, there is an $M > 0$ such that $|f| \leq M$ almost everywhere. Hence $|f|^{p_2 - p_1} \leq M^{p_2 - p_1}$ since $p_2 \geq p_1$ and so we have

$$\int_X |f_1|^{p_2} \, d\mu = \int_X |f_1|^{p_1} |f_1|^{p_2 - p_1} \, d\mu \leq \int_X |f|^{p_1} |f|^{p_2 - p_1} \, d\mu \leq M^{p_2 - p_1} \int_X |f|^{p_1} \, d\mu < \infty,$$

for $1 < p < \infty$.

$$\int_X |f_1|^{p_2} \, d\mu = \int_X |f_1|^{p_1} |f_1|^{p_2 - p_1} \, d\mu \leq \int_X |f|^{p_1} |f|^{p_2 - p_1} \, d\mu \leq M^{p_2 - p_1} \int_X |f|^{p_1} \, d\mu < \infty.$$
where the last inequality follows since \( f \in L^p \). Hence, \( f_1 \in L^p \). Since \( L^p \) is a linear class, we have \( f_1 + f_2 = f \in L^p \), which finishes the proof.

21.1. **Overview of Functional Analysis.** Consider a function \( \gamma : L^p \to \mathbb{R} \). We first consider linear functionals on \( L^p \). A linear functional on \( L^p \) is a function \( \gamma : L^p \to \mathbb{R} \) such that

1. \( \gamma(f + g) = \gamma(f) + \gamma(g) \), and
2. \( \gamma(\alpha f) = \alpha \gamma(f) \) for \( \alpha \in \mathbb{R} \).

Clearly the zero functional \( \theta(f) \equiv 0 \) for all \( f \in L^p \) is a linear functional. However, the natural question arises: are there any non-trivial linear functionals? We address this question soon.

**Definition 21.26.** A linear functional \( \gamma \) is called **bounded** if there exists a number \( c > 0 \) such that \( |\gamma(f)| \leq c \|f\|_p \) for all \( f \in L^p \). This is the **boundedness condition**. We denote

\[
\|\gamma\|_{(L^p)^*} = \inf \{c \mid \gamma \text{ satisfies boundedness condition}\},
\]

where this infimum is taken over all numbers \( c \in \mathbb{R} \) that satisfy the boundedness condition. This number is the **norm** of \( \gamma \).

**Definition 21.27.** The space of all bounded linear functionals on \( L^p \) is called the **dual space** of \( L^p \), and is denoted \((L^p)^*\). We can define operations of addition and scalar multiplication on \((L^p)^*\). We have

1. For \( \gamma_1, \gamma_2 \in (L^p)^* \), we have \( \gamma_1 + \gamma_2 \in (L^p)^* \), and
2. For \( \alpha \in \mathbb{R} \) and \( \gamma \in (L^p)^* \), we have \( \alpha \gamma \in (L^p)^* \).

This makes \((L^p)^*\) a linear space. As alluded to earlier, this is a normed linear space under the norm \( \| \cdot \|_{(L^p)^*} \), given in Definition 21.26.

**Note 21.28.** Bounded linear functionals take bounded sets to bounded sets; that is, if \( \gamma \in (L^p)^* \), then \( \gamma(B(0, 1)) \) is bounded in \( \mathbb{R} \), where \( B(0, 1) = \{ f \in L^p : \|f\|_p < 1 \} \).

**Proof.** Suppose \( \gamma(B(0, 1)) \) is not bounded in \( \mathbb{R} \). Then there exists a sequence \( \{f_n\}_{n=1}^{\infty} \subset B(0, 1) \) such that \( |\gamma(f_n)| > n \) for every \( n \in \mathbb{N} \). But since \( \|f_n\|_p < 1 \) for all \( n \in \mathbb{N} \), we have

\[
|\gamma(f_n)| > n > n \cdot \|f_n\|_p.
\]

This contradicts \( \gamma \) being bounded since it shows there is no such \( c > 0 \) with \( |\gamma(f)| \leq c \|f\|_p \) for all \( f \in L^p \). Hence, we conclude \( \gamma(B(0, 1)) \) is bounded in \( \mathbb{R} \).

**Note 21.29.** Let \( c > 0 \) be the constant for which \( c = \|\gamma\|_{(L^p)^*} \). Then for every \( f \in L^p \), we have \( |\gamma(f)| \leq c \|f\|_p \), and there is no \( k < c \) such that this inequality holds for every \( f \in L^p \). Note, then, that this implies

\[
|\gamma \left( \frac{f}{\|f\|_p} \right) | \leq c, \quad \text{with} \quad \left\| \frac{f}{\|f\|_p} \right\|_p = 1.
\]

This implies that \( c = \sup_{\|h\|_p = 1} |\gamma(h)| \). Now if \( \|f\|_p < 1 \), then we have \( |\gamma(f)| \leq c \|f\|_p < c \), and so \( \sup_{\|h\|_p < 1} |\gamma(h)| \leq c \). Thus we have shown

\[
\|\gamma\|_{(L^p)^*} = \sup_{\|h\|_p = 1} |\gamma(h)| = \sup_{\|h\|_p \leq 1} |\gamma(h)|.
\]

This gives a nice characterization of the norm of \((L^p)^*\).
Example 21.30. As promised, let us consider some types of bounded non-trivial linear functionals.

(1) Let \( \gamma : L^1 \to \mathbb{R} \) be defined by \( \gamma(f) = \int_X f \, d\mu \). Clearly \( \gamma \) is a linear functional. We have

\[
|\gamma(f)| = \left| \int_X f \, d\mu \right| \leq \int_X |f| \, d\mu = \|f\|_1.
\]

Thus \( \gamma \in (L^1)^* \) since \( \gamma \) is a bounded linear functional.

Consider the case when \( 1 < p \leq \infty \). Let \( q \) be the conjugate number to \( p \) satisfying \( 1/p + 1/q = 1 \). First, suppose \( \mu(X) < \infty \). We have shown in the preceding paragraph that \( \gamma \in (L^1)^* \). By the Riesz Representation Theorem (see below), we have \((L^p)^* \simeq L^q\) and \((L^1)^* \simeq L^\infty\), and so since \( L^\infty \subset L^q \) by Lemma 21.22, we have \((L^1)^* \subset (L^p)^*\). Hence, \( \gamma \in (L^p)^* \), and so the result still holds.

A problem arises when we consider this definition of \( \gamma \) when \( \mu(X) = \infty \). Consider the function \( f \) defined by

\[
f = \begin{cases} 
\frac{1}{n} & x \in [n-1, n) \text{ for every } n \in \mathbb{N} \\
0 & x < 0.
\end{cases}
\]

on the measure space \((\mathbb{R}, \mathcal{M}, m_1)\), the Lebesgue measure space. Now clearly \( m_1(\mathbb{R}) = \infty \), and in this case, it leads to a problem. If \( p > 1 \), then

\[
\int_{\mathbb{R}} |f|^p \, dm_1 = \sum_{n=1}^{\infty} \left| \frac{1}{n} \right|^p = \sum_{n=1}^{\infty} \frac{1}{n^p} < \infty,
\]

and so \( f \in L^p \). However,

\[
|\gamma(f)| = \left| \int_X f \, d m_1 \right| = \sum_{n=1}^{\infty} \frac{1}{n} = \infty,
\]

and so \( \gamma \not\in (L^p)^* \) since \( \gamma \) is not a bounded linear functional. Hence, we may want to consider another type of linear functional that is bounded in more generality than \( \gamma \) from this example.

(2) Fix a function \( g \in L^\infty \). Then we define \( \gamma_g \) by

\[
\gamma_g(f) = \int_X f g \, d\mu
\]

for every \( f \in L^1 \). Then by Hölder’s Inequality we have

\[
|\gamma_g(f)| = \left| \int_X f g \, d\mu \right| \leq \int_X |f g| \, d\mu \leq \|f\|_1 \|g\|_\infty
\]

for every \( f \in L^1 \), and so \( \gamma_g \in (L^1)^* \) with \( \|\gamma_g\|_{(L^1)^*} \leq \|g\|_\infty \).

(3) Now fix \( g \in L^q \), where \( q \) is the conjugate number to \( p \) for \( 1 < p < \infty \). We define \( \gamma_g \) by

\[
\gamma_g(f) = \int_X f g \, d\mu
\]

for every \( f \in L^p \). Just as in the last case, we have

\[
|\gamma_g(f)| = \left| \int_X f g \, d\mu \right| \leq \int_X |f g| \, d\mu \leq \|f\|_p \|g\|_q
\]

by Hölder’s Inequality. Hence \( \gamma_g \in (L^p)^* \) and \( \|\gamma_g\|_{(L^p)^*} \leq \|g\|_q \).
Both of the previous two examples are part of the following fundamental result.

**Theorem 21.31 (Riesz Representation Theorem).** Let \((X, \mathcal{M}, \mu)\) be a measure space. Let \(1 \leq p < \infty\) and let \(q\) be the conjugate number to \(p\). For every \(g \in L^q\), the function \(\gamma_g : L^p \to \mathbb{R}\) given by
\[
\gamma_g(f) = \int_X fg \, d\mu
\]
defines a bounded linear functional in \((L^p)^*\). Furthermore,

1. \(\|\gamma_g\|_{(L^p)^*} = \|g\|_q\), and
2. For every \(\gamma \in (L^p)^*\), there exists a function \(g \in L^q\) such that \(\gamma = \gamma_g\) as defined above.

In particular, we have \((L^p)^* \simeq L^q\).

**Proof.** Omitted, as this is a result of functional analysis. Note, however, that we have already shown some of this result. \(\square\)

**Note 21.32.** A particularly important case arises when \(p = q = 2\). In this case, we have \((L^2)^* \simeq L^2\), and so \(L^2\) is a **self-dual space**. A self-dual Banach space is called a **Hilbert Space**. Hilbert spaces are also characterized by having an inner product. For \(L^2\), we define the inner product as
\[
\langle f, g \rangle = \int_X fg \, d\mu.
\]
In addition, the Hölder Inequality has a special name for the case \(p = q = 2\); in this case it is called the **Cauchy-Schwartz Inequality**. Note the property
\[
\langle f, f \rangle = \int_X f^2 \, d\mu = \int_X f^2 \, d\mu = (\|f\|_2^2).
\]

**Definition 21.33.** Let \(1 \leq p \leq \infty\) and suppose \(\{f_n\}_{n=1}^\infty \subset L^p\) and \(f \in L^p\). We say \(f_n \to f\) in \(L^p\) if
\[
\|f_n - f\|_p = \left\{ \int_X |f_n - f|^p \, d\mu \right\}^{1/p} \to 0
\]
as \(n \to \infty\). This is called **convergence in** \(L^p\), or **strong convergence**.

**Definition 21.34.** Let \(1 \leq p \leq \infty\) and let \(q\) be the conjugate number to \(p\). Suppose \(\{f_n\}_{n=1}^\infty \subset L^p\) and \(f \in L^p\). We say \(f_n \to f\) **weakly in** \(L^p\) if one (and hence both) of the following equivalent definitions hold.

1. For every \(\gamma \in (L^p)^*\), we have \(\gamma(f_n) \to \gamma(f)\).
2. For every \(g \in L^q\), we have \(\int_X f_ng \, d\mu \to \int_X fg \, d\mu\).

**Lemma 21.35.** Strong convergence implies weak convergence in \(L^p\).

**Proof.** Suppose \(f_n \to f\) strongly in \(L^p\). Take \(g \in L^q\), where \(q\) is the conjugate number to \(p\). Now we have
\[
\left| \int_X f_n g \, d\mu - \int_X fg \, d\mu \right| = \left| \int_X (f_n - f)g \, d\mu \right| \\
\leq \int_X |f_n - f|g| \, d\mu \\
\leq \|f_n - f\|_p \|g\|_q \to 0,
\]
where the last inequality follows by the H"older Inequality. This sequence tends to zero as \( n \to \infty \) because \( \|g\|_q \) is a finite number and \( \|f_n - f\| \to 0 \) as \( n \to \infty \) by assumption.

Alternate Proof. Using Definition 21.34(1) along with the linearity of \( \gamma \in (L^p)^* \), we have
\[
|\gamma(f_n) - \gamma(f)| = |\gamma(f_n - f)| \leq \|\gamma\|_{(L^p)^*} \|f_n - f\|_p \to 0,
\]

since \( \|\gamma\|_{(L^p)^*} < \infty \) and \( \|f_n - f\| \to 0 \) by assumption.

\[\square\]

**Theorem 21.36** (Chebyshev’s Inequality in \( L^p \)). Let \( 1 \leq p < \infty \) and let \( f \in L^p \). For \( \lambda > 0 \), it follows that
\[
\mu\{x \in X : |f(x)| > \lambda\} \leq \frac{1}{\lambda^p} \int_X |f|^p \, d\mu.
\]

**Proof.** Let \( E = \{x \in X : |f(x)| > \lambda\} \) for some fixed \( \lambda > 0 \). Note this implies \( |f|^p > \lambda^p \) on \( E \). Thus
\[
\int_X |f|^p \, d\mu \geq \int_E |f|^p \, d\mu \geq \int_E \lambda^p \, d\mu = \lambda^p \cdot \mu(E).
\]
Dividing both sides by \( \lambda^p \) finishes the proof. \[\square\]

Chebyshev’s Inequality can be used to show the following.

**Lemma 21.37.** Strong convergence in \( L^p \) implies convergence in measure.

**Proof.** Let \( \{f_n\}_{n=1}^\infty \subset L^p \) and \( f \in L^p \), and let \( f_n \to f \) strongly in \( L^p \). Recall that convergence in measure is defined such that for every \( \lambda > 0 \), we have
\[
\mu\{x \in X : |f_n(x) - f(x)| > \lambda\} \to 0 \quad \text{as} \quad n \to \infty.
\]

By Chebyshev’s Inequality, we have
\[
\mu\{x \in X : |f_n(x) - f(x)| > \lambda\} \leq \frac{1}{\lambda^p} \int_X |f - f_n|^p \, d\mu = \frac{1}{\lambda^p} (\|f - f_n\|_p)^p \to 0
\]
as \( n \to \infty \), since \( \frac{1}{\lambda^p} < \infty \) and \( \|f - f_n\|_p \to 0 \) as \( n \to \infty \). Thus we have established convergence in measure, completing the proof. \[\square\]
22. Product Measures; Fubini and Tonelli

**Note 22.1.** The following section was covered, more or less, in the last lecture of the year. Therefore, many details were omitted. I will not try to reproduce these omitted details in this section, so this section may leave more to the reader than in previous sections. However, Dr. Gulisashvili cautioned this material would appear on comprehensive exams.

Let \((X_1, \mathcal{M}_1, \mu_1)\) and \((X_2, \mathcal{M}_2, \mu_2)\) be \(\sigma\)-finite measure spaces. There is a natural construction, namely the Cartesian Product

\[
X_1 \times X_2 = \{(x_1, x_2) : x_1 \in X_2, x_2 \in X_2\},
\]

which is a product of sets. Can we likewise define a natural \(\sigma\)-algebra from \(\mathcal{M}_1\) and \(\mathcal{M}_2\) and measure from \(\mu_1\) and \(\mu_2\)? The answer to this question is yes. In particular, we will construct \(\mathcal{M}_1 \otimes \mathcal{M}_2\), called the tensor product of \(\mathcal{M}_1\) and \(\mathcal{M}_2\), and \(\mu_1 \times \mu_2\), called the product measure of \(\mu_1\) and \(\mu_2\), such that \((X_1 \times X_2, \mathcal{M}_1 \otimes \mathcal{M}_2, \mu_1 \times \mu_2)\) is a measure space. This construction uses the Carathéodory Extension Theorem. We now describe the construction.

**Construction 22.2.** There are generalized rectangles in \(X_1 \times X_2\):

\[
R = E_1 \times E_2, \text{ where } E_1 \in \mathcal{M}_1, \ E_2 \in \mathcal{M}_2.
\]

If we let \(\mathcal{R}\) denote the collection of all generalized rectangles, then we can form a set function \(\mu : \mathcal{R} \to [0, \infty]\) by

\[
\mu(R) = \mu_1(E_1) \times \mu_2(E_2).
\]

Now define a new collection \(\tilde{\mathcal{M}}\) by the rule

\[
E \in \tilde{\mathcal{M}} \iff E = \bigcup_{k=1}^{m} R_k = \bigcup_{k=1}^{m} (E_{1,k} \times E_{2,k})
\]

for some disjoint collection \(\{R_k\}_{k=1}^{m} \subset \mathcal{R}\). Now we can extend \(\mu\) to \(\tilde{\mathcal{M}}\) by

\[
\mu(E) = \sum_{k=1}^{m} \mu_1(E_{1,k}) \cdot \mu_2(E_{2,k}).
\]

The representation of \(E\) may not be unique, but the number \(\mu(E)\) does not depend on the choice of representation of \(E\). Now we define the outer measure \(\mu^* : \mathcal{P}(X_1 \times X_2) \to [0, \infty]\) in the usual way by

\[
\mu^*(E) = \inf \left\{ \sum_{k=1}^{\infty} \mu(E_k) : E \subseteq \bigcup_{k=1}^{\infty} E_k, \ E_k \in \tilde{\mathcal{M}} \right\}.
\]

Now define \(\mathcal{M}\) to be the Carathéodory \(\sigma\)-algebra, and define \(\overline{\mu} = \mu^*|_\mathcal{M}\). We need the Carathéodory Extension Theorem to show \(\overline{\mu}\) extends \(\mu\); in particular, we must show \(\tilde{\mathcal{M}}\) is closed under relative complements and that \(\mu\) is a pre-measure. These conditions do hold for our construction. We denote \(\mathcal{M} = \mathcal{M}_1 \otimes \mathcal{M}_2\) and \(\overline{\mu} = \mu_1 \times \mu_2\). The Carathéodory construction produces a complete measure space, so the space \((X_1 \times X_2, \mathcal{M}_1 \otimes \mathcal{M}_2, \mu_1 \times \mu_2)\) is complete.

22.1. **Fubini and Tonelli.** To motivate the idea behind Fubini’s Theorem, consider the Lebesgue measure space \((\mathbb{R}, \mathcal{M}, m_1)\) under the setup given in Figure 6. In the figure, the bold stripe through \(R\) represents an \(x_1\)-slice, which we denote \(J_{x_1}\). We note that \(m_1(J_{x_1}) = m_1(I_2)\). Since \(m_2 = m_1 \times m_1\),
we have $m_2(R) = m_1(I_1) \cdot m_1(I_2)$. Thus we think of $m_1(J_{x_1})$ as a function on $I_1$ for different values of $x_1$, and so

$$\int_{I_1} m_1(J_{x_1}) \, dx_1 = m_1(I_1) \cdot m_1(I_2).$$

We are effectively “slicing” along $I_1$ and integrating along $I_1$, effectively “adding up the slices.” This is a familiar concept from calculus.

**Theorem 22.3** (Fubini’s Theorem). Let $(X_1 \times X_2, \mathcal{M}_1 \otimes \mathcal{M}_2, \mu_1 \times \mu_2)$ be a product measure space for measure spaces $(X_1, \mathcal{M}_1, \mu_1)$ and $(X_2, \mathcal{M}_2, \mu_2)$. Let $f \in L^1(X_1 \times X_2)$. We can write $f(x_1, x_2)$ for $x_1 \in X_1$ and $x_2 \in X_2$. Consider an $x_1$-slice of $f$: Fix $x_1 \in X_1$ and let $f_{x_1}(x_2) = f(x_1, x_2)$, which is a function on $X_2$. Then

1. The function $f_{x_1}(x_2)$ is a $\mathcal{M}_2$-measurable and integrable function for almost all $x_1 \in X_1$,
2. The function $x_1 \mapsto \int_{X_2} f_{x_1} \, d\mu_2$ is $\mathcal{M}_1$-measurable and integrable, and
3. (Fubini) We have

$$\int_{X_1 \times X_2} f \, d(\mu_1 \times \mu_2) = \int_{X_1} \, d\mu_1 \int_{X_2} f_{x_1} \, d\mu_2 = \int_{X_2} \, d\mu_2 \int_{X_1} f_{x_1} \, d\mu_1.$$

**Note 22.4.** As a matter of notation, we have

$$\int_{X_1} \, d\mu_1 \int_{X_2} f_{x_1} \, d\mu_2 = \int_{X_1} \int_{X_2} f_{x_1} \, d\mu_2 \, d\mu_1.$$

**Theorem 22.5** (Tonelli’s Theorem). Let $(X_1 \times X_2, \mathcal{M}_1 \otimes \mathcal{M}_2, \mu_1 \times \mu_2)$ be a product measure space for measure spaces $(X_1, \mathcal{M}_1, \mu_1)$ and $(X_2, \mathcal{M}_2, \mu_2)$. Let $f$ be non-negative and $\mathcal{M}_1 \otimes \mathcal{M}_2$-measurable. Then all of the conclusions of Fubini’s Theorem hold, and in particular,

$$\int_{X_1 \times X_2} f \, d(\mu_1 \times \mu_2) = \int_{X_1} \, d\mu_1 \int_{X_2} f_{x_1} \, d\mu_2 = \int_{X_2} \, d\mu_2 \int_{X_1} f_{x_1} \, d\mu_1.$$

**Note 22.6.** Note that there is no assumption of integrability for Tonelli’s Theorem.
**Note 22.7.** Why is Tonelli’s Theorem so important? Suppose we know that $f$ is $\mathcal{M}_1 \otimes \mathcal{M}_2$-measurable. Then $|f|$ satisfies the hypotheses of Tonelli’s Theorem, and so

$$
\int_{X_1 \times X_2} |f| \, d(\mu_1 \times \mu_2) = \int_{X_1} d\mu_1 \int_{X_2} |f| \, d\mu_2 = \int_{X_2} d\mu_2 \int_{X_1} |f| \, d\mu_1,
$$

and so if we can show either of these repeated integrals on the right is finite, then we have shown $f \in L^1(X_1 \times X_2)$, from which we can apply Fubini’s Theorem for the general case.

**Example 22.8.** We give an example of how to use Tonelli’s Theorem. Let us consider the measure space $(\mathbb{R}, \mathcal{B}, \mu)$ for some measure $\mu$, where $\mathcal{B}$ is the Borel $\sigma$-algebra on $\mathbb{R}$. Let $f$ be a non-negative measurable function on $\mathbb{R}$. Consider level sets given by

$$
E_\lambda = \{ x \in \mathbb{R} : f(x) > \lambda \} \text{ for } \lambda \geq 0.
$$

We wish to prove that

$$
\int_{\mathbb{R}} f \, d\mu = \int_0^\infty \mu(E_\lambda) \, d\lambda.
$$

(The integral on the right is integrating with respect to 1-dimensional Lebesgue measure.) We have

$$
\int_0^\infty \mu(E_\lambda) \, d\lambda = \int_0^\infty \left( \int_{E_\lambda} \mu \right) \, d\lambda = \int_0^\infty \left( \int_\mathbb{R} \chi_{E_\lambda} \mu \right) \, d\lambda
$$

$$
= \int_\mathbb{R} \left( \int_0^\infty \chi_{E_\lambda} \, d\lambda \right) \mu = \int_\mathbb{R} \left( \int_0^{f(x)} 1 \, d\lambda \right) \mu = \int_\mathbb{R} f \, d\mu.
$$

Here, the third equality follows from Tonelli’s Theorem. The fourth equality follows since $\chi_{E_\lambda} = 1$ whenever $0 \leq \lambda < f(x)$ by how $E_\lambda$ is defined, and 0 otherwise.

Dr. Gulisashvili also gave two problems to practice for the comprehensive exam. These are #5 and #6 on page 423 of the textbook.

**Exercise 22.9.** Let $(X, A, \mu) = (Y, B, \nu) = (\mathbb{N}, \mathcal{M}, c)$, where $\mathcal{M} = \mathcal{P}(\mathbb{N})$ and $c$ is the counting measure. Define $f : \mathbb{N} \times \mathbb{N} \to \mathbb{R}$ by setting

$$
f(x, y) = \begin{cases} 
2 - 2^{-x} & \text{if } x = y \\
-2 + 2^{-x} & \text{if } x = y + 1 \\
0 & \text{otherwise}.
\end{cases}
$$

Show that $f$ is measurable with respect to the product measure $c \times c$. Also show that

$$
\int_{\mathbb{N}} \left[ \int_{\mathbb{N}} f(m, n) \, dc(m) \right] \, dc(n) \neq \int_{\mathbb{N}} \left[ \int_{\mathbb{N}} f(m, n) \, dc(n) \right] \, dc(m).
$$

Is this a contradiction either of Fubini’s Theorem or Tonelli’s Theorem?

**Exercise 22.10.** Let $X = Y$ be the interval $[0, 1]$, with $\mathcal{A} = \mathcal{B}$ the class of Borel sets. Let $\mu$ be Lebesgue measure and $\nu = c$ the counting measure. Show that the diagonal $\Delta = \{(x, y) : x = y\}$ is measurable with respect to the product measure $\mu \times c$ (is an $\mathcal{R}_{\sigma \delta}$, in fact). Show that if $f$ is the characteristic function of $D$, then

$$
\int_{[0,1] \times [0,1]} f \, d(\mu \times c) \neq \int_{[0,1]} \left[ \int_{[0,1]} f(x, y) \, dc(y) \right] \, d\mu(x).
$$

Is this a contradiction either of Fubini’s Theorem or Tonelli’s Theorem?
$L^1$-Convergence, 57, 70
$L^p$ space, 114
$L^p(\mu)$, 104
$\ell^p$ space, 119
$\sigma$-algebra, 7, 88
Borel, 7, 63
Carathéodory, 13
Generated, 7
Smallest, 7
Tensor Product, 125
“Sliding Hump”, 35

Absolute Continuity of the Lebesgue Integral, 75
Absolute Value, 2
Absolutely Continuous with Respect to a Measure, 105
Almost Disjoint, 60
Almost Everywhere, 18
Almost Uniformly Convergent, 32
Approximation Lemma, 38
Archimedean Property of Real Numbers, 2

Banach Space, 119
Bolzano-Weierstrass Theorem, 8
Borel-Cantelli Lemma, 19
Bounded, 1
Bounded Convergence Theorem, 49, 53, 69
Boundedness Condition, 121

Cantor Set, 23
Cantor-Lebesgue Function, 24, 77
Carathéodory Extension Theorem, 99
Cauchy Sequence, 119
Cauchy-Schwartz Inequality, 123
Chebyshev’s Inequality, 51, 70
For General Measures, 103
In $L^p$, 124
Complete, 119
Completeness Axiom, 1
Conjugate Number, 114
Convergence in $L^1$, 70
Convergence in $L^p$, 123
Convergent Almost Everywhere, 32
Convergent in Measure, 32, 70
Countable, 4
Countably Infinite, 4
Counting Measure, 89
Derivative, 80
Distance Function, 58

Dominated Convergence Theorem, 55, 70
Dual Space, 121
Dyadic Rational, 23

Egorov’s Theorem, 36, 37, 67, 101
Embedding Theorem, 120
Essential Supremum, 115

Fatou’s Lemma, 54, 70
Field, 1
Finite, 4
Fubini’s Theorem, 126

Function

Absolutely Continuous, 74
Characteristic, 27
Indicator, 27
Lebesgue Measurable, 27, 66
Lipschitz, 74
Measurability and Operations, 29, 66, 100
Negative Part, 31
of Bounded Variation, 72
Positive Part, 31
Simple, 38, 66, 100
Standard Representation, 40, 102
Support, 31, 45
Total Variation Function, 78

Fundamental Theorem of Calculus, 81
Part One, 81
Part Two, 86

Hölder’s Inequality, 116
Addition to, 117
Hahn’s Decomposition Theorem, 95
Hahn’s Lemma, 93
Heine-Borel Theorem, 6, 58
Hilbert Space, 123

Inner Regularity, 16
Integrable, 46, 51
With Respect to General Measure, 103

Integral

Bounded Measurable Functions Supported on a Set
of Finite Measure, 45, 68
Integrable Functions $L^1$, 51, 53, 69
Lower Lebesgue, 46, 68
Non-Negative Measurable Functions, 50, 68
Properties, 43, 56, 69
Simple Functions, 43, 67
Upper Lebesgue, 46, 68
Uncountable, 4
Uniformly Convergent, 32

Variation, 72
  Total, 72
    With Respect to a Partition, 72
Vitali Covering, 82
Vitali Covering Argument, 82
Vitali Covering Lemma, 83

Weak Convergence, 123