Contents

Chapter 1

Abstract Measure Theory

1.1 Outer Measures and Carathéodory Measurable Sets

In this section, we discuss outer measures which allow us to define "size" or measure for every subset of a given set X. This has certain advantages and disadvantages compared with the approach of defining a measure (as we do in the next section).

    The advantage is that we have a notion of measure for all subsets, but the disadvantage is that it turns out there are some extremely pathological sets which are not "measurable" in a reasonable sense. Another advantage is that the two measures we are most interested in — the Lebesgue measure and the Hausdorff measure — can both be easily constructed as outer measures.

1.1.1 Outer Measures and Measurable Sets

Definition 1.1.1 (Outer Measure). Let X be a set. A function $\mu : 2^X \to [0, \infty]$ is called an *outer measure* on X if it satisfies the following properties:

(i) $\mu(\varnothing) = 0$,

(ii) (monotonicity) if $A \subseteq B \subseteq X$, then $\mu(A) \leqslant \mu(B)$,

(iii) (countable subadditivity) if $\{A_j\}_{j=1}^{\infty}$ is a countable collection of subsets of X, then

$$\mu \left(\bigcup_{j=1}^{\infty} A_j \right) \leqslant \sum_{j=1}^{\infty} \mu(A_j).$$

Definition 1.1.2 (Restriction of an Outer Measure). Let μ be an outer measure on a set X, and let $A \subseteq X$ be an arbitrary subset. The *restriction* of μ to A is the outer measure

$\mu \downharpoonright A$ on X defined by

$$\mu \downharpoonright A(B) := \mu(A \cap B)$$

for every subset $B \subseteq X$.

Notice that the restriction is still defined on all subsets of X, not just those contained in A; basically, the restriction just ignores the part of each set that lies outside of A. It is clear that for each subset $A \subseteq X$, the restriction $\mu \downharpoonright A$ is indeed an outer measure on X.

Definition 1.1.3 (Carathéodory Measurable)**.** Let μ be an outer measure on a set X. A set $A \subseteq X$ is called μ-*measurable* (or *Carathéodory measurable*) if for every subset $B \subseteq X$, we have

$$\mu(B) = \mu(B \cap A) + \mu(B \backslash A).$$

Remark 1.1.4. In words, this says that A is μ-measurable if it splits every set B into two pieces without adding any extra measure. We note that if $A, B \subseteq X$ are arbitrary subsets, then $B = (B \cap A) \cup (B \backslash A)$ so by countable subadditivity of μ, we always have

$$\mu(B) \leqslant \mu(B \cap A) + \mu(B \backslash A).$$

Thus the nontrivial part of the definition is the reverse inequality. Furthermore, if $\mu(B) = \infty$, then the reverse inequality is automatically satisfied, so in order to check that A is μ-measurable, it suffices to check the definition for those sets $B \subseteq X$ satisfying $\mu(B) < \infty$.

Proposition 1.1.5 (Properties of μ-Measurable Sets)**.** *Let μ be an outer measure on a set X. Then the following properties hold:*

(i) *A subset $A \subseteq X$ is μ-measurable if and only if its complement A^c is μ-measurable.*

(ii) *The empty set $\varnothing$ and the whole set X are μ-measurable.*

(iii) *If $A \subseteq X$ satisfies $\mu(A) = 0$, then A is μ-measurable.*

(iv) *For each subset $A \subseteq X$, each μ-measurable set is also $\mu \downharpoonright A$-measurable.*

Proof.   1. Let $A \subseteq X$ be an arbitrary subset.

($\implies$) Suppose A is μ-measurable. Then for each subset $B \subseteq X$, we have

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

Since $B \backslash A^c = B \cap A$, this shows that A^c is μ-measurable.

($\impliedby$) Suppose A^c is μ-measurable. Then by repeating the above argument, we see that $(A^c)^c = A$ is μ-measurable.

2. To see that $\varnothing$ is μ-measurable, let $B \subseteq X$ be an arbitrary subset. Then

$$\mu(B \cap \varnothing) + \mu(B \backslash \varnothing) = 0 + \mu(B) = \mu(B).$$

Thus $\varnothing$ is μ-measurable.

To see that X is μ-measurable, let $B \subseteq X$ be an arbitrary subset. Then

$$\mu(B \cap X) + \mu(B \backslash X) = \mu(B) + 0 = \mu(B).$$

Thus X is μ-measurable.

3. More generally, let $A \subseteq X$ be an arbitrary subset satisfying $\mu(A) = 0$. Then for each subset $B \subseteq X$, we have

$$\mu(B \cap A) + \mu(B \backslash A) \leqslant \mu(A) + \mu(B) = 0 + \mu(B) = \mu(B).$$

By Remark 1.1.4, this shows that A is μ-measurable.

4. Let $A \subseteq X$ be an arbitrary subset, and let $E \subseteq X$ be a μ-measurable set. Then for each subset $B \subseteq X$, we have

$$
\begin{aligned}
\mu \restriction A(B) &= \mu(A \cap B) \\
&= \mu((A \cap B) \cap E) + \mu((A \cap B) \backslash E) \quad \text{(since E is μ-measurable)} \\
&= \mu((B \cap E) \cap A) + \mu((B \backslash E) \cap A) \\
&= \mu \restriction A(B \cap E) + \mu \restriction A(B \backslash E).
\end{aligned}
$$

Thus E is $\mu \restriction A$-measurable.

$\square$

Proposition 1.1.6 (Sequences of Measurable Sets). *Let μ be an outer measure on a set X, and let $\{A_j\}_{j=1}^{\infty}$ be a sequence of μ-measurable sets.*

(i) *The sets $\bigcup_{j=1}^{\infty} A_j$ and $\bigcap_{j=1}^{\infty} A_j$ are μ-measurable.*

(ii) *If the sets A_j are pairwise disjoint, then*

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

(iii) *If $A_1 \subseteq A_2 \subseteq \cdots \subseteq A_j \subseteq \cdots$ is an increasing sequence, then*

$$\mu \left(\bigcup_{j=1}^{\infty} A_j \right) = \lim_{j \to \infty} \mu(A_j).$$

(iv) If $A_1 \supseteq A_2 \supseteq \cdots \supseteq A_j \supseteq \cdots$ is a decreasing sequence and $\mu(A_1) < \infty$, then

$$\mu\left(\bigcap_{j=1}^{\infty} A_j\right) = \lim_{j\to\infty} \mu(A_j).$$

Proof. Step 1: We claim that finite unions and intersections of μ-measurable sets are also μ-measurable.

If $A_1, A_2 \subseteq X$ are μ-measurable, then for each subset $B \subseteq X$, we have

$$
\begin{aligned}
\mu(B) &= \mu(B \cap A_1) + \mu(B \backslash A_1) \quad \text{(since A_1 is μ-measurable)} \\
&= \mu(B \cap A_1) + \mu((B \backslash A_1) \cap A_2) + \mu((B \backslash A_1) \backslash A_2) \quad \text{(since A_2 is μ-measurable)} \\
&\geqslant \mu((B \cap A_1) \cup ((B \backslash A_1) \cap A_2)) + \mu(B \backslash (A_1 \cup A_2)) \quad \text{(by countable subadditivity)} \\
&= \mu(B \cap A_1 \cap A_2) + \mu(B \backslash (A_1 \cup A_2)).
\end{aligned}
$$

By Remark 1.1.4, this shows that $A_1 \cup A_2$ is μ-measurable.

Similarly, if $A_1, A_2 \subseteq X$ are μ-measurable, then for each subset $B \subseteq X$, we know that $A_1^c, A_2^c \subseteq X$ are also μ-measurable by Proposition 1.1.5 (i), and we see that

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

by De Morgan's laws; thus $(A_1 \cap A_2)$ is μ-measurable by the above argument, and hence $A_1 \cap A_2$ is μ-measurable by Proposition 1.1.5 (i).

By induction, we conclude that finite unions and intersections of μ-measurable sets are also μ-measurable. This proves our claim.

We will actually prove part (i) in the last step of this proof, after we have established the other parts. *Step 2*: We prove part (ii). Suppose that $\{A_j\}_{j=1}^{\infty}$ is a sequence of pairwise disjoint μ-measurable sets. Then for each $N \in \mathbb{N}$, the finite union $\bigcup_{j=1}^{N} A_j$ is μ-measurable by Step 1, so

$$
\begin{aligned}
\mu\left(\bigcup_{j=1}^{N+1} A_j\right) &= \mu\left(\left(\bigcup_{j=1}^{N} A_j\right) \cap A_{N+1}\right) + \mu\left(\left(\bigcup_{j=1}^{N} A_j\right) \backslash A_{N+1}\right) \\
&= \mu(A_{N+1}) + \mu\left(\bigcup_{j=1}^{N} A_j\right).
\end{aligned}
$$

By induction, we conclude that

$$\mu\left(\bigcup_{j=1}^{N} A_j\right) = \sum_{j=1}^{N} \mu(A_j)$$

for each $N \in \mathbb{N}$. Using monotonicity of μ, we have

$$\sum_{j=1}^{N} \mu(A_j) \leqslant \mu\left(\bigcup_{j=1}^{\infty} A_j\right)$$

for each $N \in \mathbb{N}$, so by letting $N \to \infty$, we obtain

$$\sum_{j=1}^{\infty} \mu(A_j) \leqslant \mu\left(\bigcup_{j=1}^{\infty} A_j\right).$$

The reverse inequality follows from countable subadditivity of μ.

This proves part (ii). *Step 3*: We prove part (iii) and (iv)

Suppose that $A_1 \subseteq A_2 \subseteq \cdots \subseteq A_j \subseteq \cdots$ is an increasing sequence of μ-measurable subsets of X. Then we consider the pairwise disjoint μ-measurable sets

$$B_1 := A_1, \quad B_j := A_j \backslash A_{j-1} \quad \text{for } j \geqslant 2.$$

By part (ii), we have

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

That is,

$$\mu\left(\bigcup_{j=1}^{\infty} A_j\right) = \mu(A_1) + \sum_{j=2}^{\infty}[\mu(A_j) - \mu(A_{j-1})] = \lim_{N \to \infty} \mu(A_N).$$

This proves part (iii).

Now suppose that $A_1 \supseteq \cdots \supseteq A_j \supseteq \cdots$ is a decreasing sequence of μ-measurable subsets of X satisfying $\mu(A_1) < \infty$. Then we see that

$$\mu(A_1) - \lim_{j \to \infty} \mu(A_j) = \lim_{j \to \infty}[\mu(A_1) - \mu(A_j)]$$

$$= \lim_{j \to \infty} \mu\left(\bigcup_{k=j}^{\infty}(A_1 \backslash A_k)\right)$$

$$\geqslant \mu(A_1) - \mu\left(\bigcap_{j=1}^{\infty} A_j\right)$$

by part (iii) and monotonicity of μ; it follows that

$$\mu\left(\bigcap_{j=1}^{\infty} A_j\right) \geqslant \lim_{j \to \infty} \mu(A_j).$$

The reverse inequality follows from monotonicity of μ. This proves part (iv). *Step 4*: We prove part (i).

Suppose that $\{A_j\}_{j=1}^\infty$ is a sequence of μ-measurable sets. We use step 1 to see that the finite unions $\bigcup_{j=1}^N A_j$ are μ-measurable for each $N \in \mathbb{N}$, and form an increasing sequence:

$$A_1 \subseteq A_1 \cup A_2 \subseteq \cdots \subseteq \bigcup_{j=1}^N A_j \subseteq \cdots .$$

Hence for each set $B \subseteq X$ with $\mu(B) < \infty$, we have

$$\mu\left(B \cap \bigcup_{j=1}^\infty A_j\right) + \mu\left(B \backslash \bigcup_{j=1}^\infty A_j\right) = \mu\left(B \cap \bigcup_{N=1}^\infty \bigcup_{j=1}^N A_j\right) + \mu\left(B \cap \bigcap_{N=1}^\infty \left(\bigcup_{j=1}^N A_j\right)^c\right)$$

$$= \mu \!\restriction\! B\left(\bigcup_{N=1}^\infty \bigcup_{j=1}^N A_j\right) + \mu \!\restriction\! B\left(\bigcap_{N=1}^\infty \bigcap_{j=1}^N A_j^c\right)$$

$$= \lim_{N\to\infty} \mu \!\restriction\! B\left(\bigcup_{j=1}^N A_j\right) + \lim_{N\to\infty} \mu \!\restriction\! B\left(\bigcap_{j=1}^N A_j^c\right) \quad \text{(by parts (iii) and (iv))}$$

$$= \lim_{N\to\infty}\left[\mu \!\restriction\! B\left(X \cap \bigcup_{j=1}^N A_j\right) + \mu \!\restriction\! B\left(X \cap \bigcap_{j=1}^N A_j^c\right)\right]$$

$$= \lim_{N\to\infty}\left[\mu \!\restriction\! B\left(X \cap \bigcup_{j=1}^N A_j\right) + \mu \!\restriction\! B\left(X \backslash \bigcup_{j=1}^N A_j\right)\right]$$

$$= \mu \!\restriction\! B(X) = \mu(B).$$

Since this holds for each set $B \subseteq X$ with $\mu(B) < \infty$, we conclude that $\bigcup_{j=1}^\infty A_j$ is μ-measurable by Remark 1.1.4.

Finally, since $\bigcap_{j=1}^\infty A_j = \left(\bigcup_{j=1}^\infty A_j^c\right)^c$ by De Morgan's laws, we see that $\bigcap_{j=1}^\infty A_j$ is also μ-measurable by the above argument and Proposition 1.1.5. This proves part (i). $\square$

1.1.2 σ-Algebras and the Carathéodory Criterion

Definition 1.1.7 (σ-Algebra)**.** Let X be a set. A collection $\mathcal{A} \subseteq 2^X$ of subsets of X is called a *σ-algebra* on X if it satisfies the following properties:

(i) $\varnothing, X \in \mathcal{A}$,

(ii) (closed under complements) if $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$,

(iii) (closed under countable unions) if $\{A_j\}_{j=1}^\infty \subseteq \mathcal{A}$ is a countable collection of sets, then $\bigcup_{j=1}^\infty A_j \in \mathcal{A}$,

(iv) (closed under countable intersections) if $\{A_j\}_{j=1}^\infty \subseteq \mathcal{A}$ is a countable collection of sets, then $\bigcap_{j=1}^\infty A_j \in \mathcal{A}$.

Note that properties (ii) and (iii) imply property (iv) by De Morgan's laws. We list this property explicitly for emphasis, and because it is very convenient to use in proofs (once you have a σ-algebra) and as a stepping stone to proving the other properties. In particular, it is sometimes easier to show that (iii) and (iv) hold before showing that (ii) holds.

Lemma 1.1.8 (Measurable Sets Form a σ-Algebra)**.** *Let μ be an outer measure on a set X, and let*

$$\mathcal{M}_\mu := \{A \subseteq X : A \text{ is } \mu\text{-measurable}\}.$$

Then $\mathcal{M}_\mu$ is a σ-algebra on X which contains all subsets of X with μ-measure zero.

Proof. This follows from Proposition 1.1.5 and Proposition 1.1.6. $\qquad\square$

Exercise 1.1.1. Let X be a set, and let $\{\mathcal{A}_j\}_{j \in J}$ be an arbitrary collection of σ-algebras on X (where J is an arbitrary index set). Then the intersection $\bigcap_{j \in J} \mathcal{A}_j$ is also a σ-algebra on X.

Proof. Since $\varnothing, X \in \mathcal{A}_j$ for each $j \in J$, we have $\varnothing, X \in \bigcap_{j \in J} \mathcal{A}_j$.

Now let $A \in \bigcap_{j \in J} \mathcal{A}_j$ be an arbitrary set. Then $A \in \mathcal{A}_j$ for each $j \in J$, so $A^c \in \mathcal{A}_j$ for each $j \in J$ by property (ii) of Definition 1.1.7. Thus $A^c \in \bigcap_{j \in J} \mathcal{A}_j$. This shows that $\bigcap_{j \in J} \mathcal{A}_j$ is closed under complements.

Finally, let $\{A_k\}_{k=1}^\infty \subseteq \bigcap_{j \in J} \mathcal{A}_j$ be an arbitrary countable collection of sets. Then for each $k \in \mathbb{N}$, we have $A_k \in \mathcal{A}_j$ for each $j \in J$, so

$$\bigcup_{k=1}^\infty A_k \in \mathcal{A}_j$$

for each $j \in J$ by property (iii) of Definition 1.1.7. Thus $\bigcup_{k=1}^\infty A_k \in \bigcap_{j \in J} \mathcal{A}_j$. This shows that $\bigcap_{j \in J} \mathcal{A}_j$ is closed under countable unions, and hence also closed under countable intersections by De Morgan's laws.

We conclude that $\bigcap_{j \in J} \mathcal{A}_j$ is a σ-algebra on X. $\qquad\square$

Definition 1.1.9 (Smallest σ-Algebra Containing a Collection of Sets)**.** Let X be a set, and let $\mathcal{C} \subseteq 2^X$ be an arbitrary collection of subsets of X. The *smallest σ-algebra containing $\mathcal{C}$* is the intersection of all σ-algebras on X which contain $\mathcal{C}$, i.e.

$$\sigma(\mathcal{C}) := \bigcap_{\substack{\mathcal{A} \text{ is a } \sigma\text{-algebra on } X \\ \mathcal{C} \subseteq \mathcal{A}}} \mathcal{A}.$$

This is a well-defined σ-algebra on X by the previous exercise 1.1.1.

Definition 1.1.10 (Borel Sets, Borel Measure)**.** Let X be a topological space. The *Borel σ-algebra* on X is the smallest σ-algebra containing all open sets in X. The elements of the Borel σ-algebra are called *Borel sets*.
  We say an outer measure μ on X is *Borel* if every Borel set is μ-measurable.

Since the Borel σ-algebra is generated by the open sets and is closed under complements, it also contains all closed sets, and can be equivalently be defined as the smallest σ-algebra containing all closed sets.

Theorem 1.1.11 (Carathéodory Criterion)**.** *Let (X, d) be a metric space, and let μ be an outer measure on X with the property that*

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

for all sets $A, B \subseteq X$ satisfying $d(A, B) > 0$. Then every Borel set is μ-measurable, i.e. μ is a Borel outer measure on X.

Such outer measures are sometimes called *metric outer measures.*

Proof. Since the Borel σ-algebra is the smallest σ-algebra containing all closed sets, it is enough to show that every closed set is μ-measurable.
  Let $C \subseteq X$ be a closed set, and let $B \subseteq X$ be an arbitrary subset satisfying $\mu(B) < \infty$. For each $j \in \mathbb{N}$, we define

$$C_j := \left\{ x \in X : \operatorname{dist}(x, C) \leqslant \frac{1}{j} \right\}$$

which is a closed set; this implies

$$\operatorname{dist}(B \backslash C_j, B \cap C_j) > \frac{1}{j} > 0.$$

Hence

$$\mu(B) \geqslant \mu((B \backslash C_j) \cup (B \cap C_j)) = \mu(B \backslash C_j) + \mu(B \cap C_j)$$

by monotonicity and the assumption on μ.

We claim that

$$\lim_{j\to\infty} \mu(B\backslash C_j) = \mu(B\backslash C). \tag{$\star$}$$

To see this, note that since C is closed we can write

$$B\backslash C = \{x \in X : \mathrm{dist}(x, C) > 0\} = (B\backslash C_j) \cup \bigcup_{k=j}^{\infty} R_k$$

where

$$R_k := \left\{ x \in B : \frac{1}{k+1} < \mathrm{dist}(x, C) \leqslant \frac{1}{k} \right\}$$

for each $k \in \mathbb{Z}^+$.

Using subadditivity of μ, we have

$$\mu(B\backslash C_j) \leqslant \mu(B\backslash C) \leqslant \mu(B\backslash C_j) + \sum_{k=j}^{\infty} \mu(R_k) \tag{$\star\star$}$$

for each $j \in \mathbb{Z}^+$. Since $\mu(B) < \infty$, for each $k \in \mathbb{Z}^+$ we have $\mu(R_k) \leqslant \mu(B) < \infty$, and we also have that

$$\mathrm{dist}(R_j, R_k) > 0 \qquad \forall \ k > j + 1$$

by definition of these sets. By assumption on μ, this implies

$$\sum_{k=1}^{N} \mu(R_{2k}) = \mu \left(\bigcup_{k=1}^{N} R_{2k} \right) \leqslant \mu(B) < \infty,$$

and

$$\sum_{k=1}^{N} \mu(R_{2k-1}) = \mu \left(\bigcup_{k=1}^{N} R_{2k-1} \right) \leqslant \mu(B) < \infty$$

for all $N \in \mathbb{Z}^+$. In particular, we conclude that $\sum_{k=1}^{\infty} \mu(R_k) < \infty$. Hence for each $\epsilon > 0$, there exists $J \in \mathbb{Z}^+$ such that

$$\sum_{k=j}^{\infty} \mu(R_k) < \epsilon$$

for all $j \geqslant J$. Using this in ($\star\star$), we obtain

$$0 \leqslant \mu(B\backslash C) - \mu(B\backslash C_j) \leqslant \sum_{k=j}^{\infty} \mu(R_k) < \epsilon$$

for all $j \geqslant J$, which proves ($\star$).

Using this in the inequality before ($\star$), we obtain

$$\mu(B) \geqslant \lim_{j\to\infty} [\mu(B\backslash C_j) + \mu(B \cap C_j)] = \mu(B\backslash C) + \lim_{j\to\infty} \mu(B \cap C_j)$$

by ($\star$). Since $B \subseteq X$ is an arbitrary subset satisfying $\mu(B) < \infty$, this shows that C is μ-measurable by Remark 1.1.4. Since $C \subseteq X$ is an arbitrary closed set, we conclude that every closed set is μ-measurable. $\qquad\square$

1.1.3   Regular and Borel Regular Outer Measures

Definition 1.1.12 (Regular, Borel Regular, Open σ-Finite). Let X be a set, and let μ be an outer measure on X. We say that μ is *regular* if for every subset $A \subseteq X$ there exists a μ-measurable set $B \subseteq X$ such that $A \subseteq B$ and $\mu(A) = \mu(B)$.

Now let X be a topological space, and let μ be an outer measure on X. We say that μ is *Borel regular* if μ is a Borel measure and for every subset $A \subseteq X$ there exists a Borel set $B \subseteq X$ such that $A \subseteq B$ and $\mu(A) = \mu(B)$.

If μ is a Borel regular measure on X, then we say that the pair (X, μ) is *open σ-finite* if X can be covered by a countable collection of open sets with finite μ-measure.

Observe that if (X, d) is a seperable metric space and μ is locally finite (i.e. for each $x \in X$ there exists $r > 0$ such that $\mu(B(x, r)) < \infty$), then (X, μ) is open σ-finite. Why? — seperability gives a countable dense set $\{x_j\}_{j=1}^{\infty} \subseteq X$, so let $r_j > 0$ be such that $\mu(B(x_j, r_j)) < \infty$ for each $j \in \mathbb{Z}^+$. Then $\{B(x_j, r_j)\}_{j=1}^{\infty}$ is a countable cover of X by open sets with finite μ-measure.

Lemma 1.1.13 (Restriction of Borel Regular Outer Measure). *Let X be a topological space, and let μ be a Borel regular outer measure on X. Let A be an arbitrary subset of X satisfying $\mu(A) < \infty$. Then the restriction $\mu \restriction A$ is also a Borel regular outer measure on X.*

Proof. Since μ is a Borel regular outer measure on X, we can choose a Borel set B_1 with $A \subseteq B_1$ such that $\mu(B_1 \backslash A) = 0$. Then choose another Borel set B_2 with $B_1 \backslash A \subseteq B_2$ such that $\mu(B_2) = 0$. Fix a subset $E \subseteq X$. Then we have

$$
\begin{aligned}
E = (E \cap A) \cup (E \backslash A) &\subseteq (E \cap A) \cup (X \backslash A) \\
&= (E \cap A) \cup (X \backslash B_1) \cup (B_1 \backslash A) \\
&\subseteq (E \cap A) \cup (X \backslash B_1) \cup B_2.
\end{aligned}
$$

Finally choose a Borel set B_3 with $E \cap A \subseteq B_3$ such that $\mu(B_3) = \mu(E \cap A)$. Then we have

$$
E \subseteq (X \backslash B_1) \cup B_2 \cup B_3
$$

which is a Borel set containing E. Thus

$$
E \cap A \subseteq \big((X \backslash B_1) \cup B_2 \cup B_3 \big) \cap A
$$

which implies that

$$
(\mu \restriction A)(E) = \mu(E \cap A) \leqslant \mu \Big(\big((X \backslash B_1) \cup B_2 \cup B_3 \big) \cap A \Big) = (\mu \restriction A) \big((X \backslash B_1) \cup B_2 \cup B_3 \big).
$$

Also, we have

$$(\mu \downharpoonright A)((X\backslash B_1) \cup B_2 \cup B_3) = (\mu \downharpoonright A)((X\backslash B_1) \cup B_3) \qquad \text{since } \mu(B_2) = 0$$
$$= \mu((A \cap (X\backslash B_1)) \cup (A \cap B_3))$$
$$= \mu(A \cap B_3) \qquad \text{since } A \subseteq B_1 \implies A \cap (X\backslash B_1) = \varnothing$$
$$\leqslant \mu(B_3)$$
$$= \mu(E \cap A)$$
$$= (\mu \downharpoonright A)(E).$$

This shows that

$$(\mu \downharpoonright A)(E) = (\mu \downharpoonright A)((X\backslash B_1) \cup B_2 \cup B_3).$$

Since $(X\backslash B_1) \cup B_2 \cup B_3$ is a Borel set containing E, and $E \subseteq X$ was arbitrary, we see that $\mu \downharpoonright A$ is a Borel regular outer measure on A. $\qquad \square$

We come to the last result of this section. The first states that on a nice topological space, every Borel regular measure is "inner regular" and "outer regular".

Theorem 1.1.14 (Borel Regular Measures are Inner and Outer Regular). *Let X be a topological space which has the property that every closed subset of X is a countable intersection of open sets, and let μ be a Borel regular measure on X such that (X, μ) is open σ-finite. Then the following two properties hold:*

(i) for each subset $A \subseteq X$, we have

$$\mu(A) = \inf\{\mu(U) : A \subseteq U, U \ open\}$$

and

(ii) for each μ-measurable set $A \subseteq X$, we have

$$\mu(A) = \sup\{\mu(K) : K \subseteq A, K \ closed\}.$$

Remark 1.1.15. Before the proof, we make two remarks. First is that the condition that every closed set is a countable intersection of open sets is satisfied by all metric spaces.

Second, if X is a Hausdorff space which is a union of countably many compact sets, then the conclusion (ii) above guaruntees that

$$\mu(A) = \sup\{\mu(K) : K \subseteq A, K \text{ compact}\}$$

for each μ-measurable set $A \subseteq X$ with $\mu(A) < \infty$.

We check this — write $X = \bigcup_{j=1}^{\infty} K_j$ where $K_j \subseteq X$ is compact for each $j \in \mathbb{Z}^+$. Then we can write X as the increasing union of compact sets

$$X = \bigcup_{m=1}^{\infty} \bigcup_{j=1}^{m} K_j$$

and each closed set C can be written as the increasing union of compact sets

$$C = \bigcup_{m=1}^{\infty} \left(\bigcup_{j=1}^{m} K_j \cap C \right).$$

Here we have used that $K_j \cap C$ is compact, since X is Hausdorff.

Therefore, if $A \subseteq X$ is a μ-measurable set with $\mu(A) < \infty$, then by (ii) we have

$$\sup\{\mu(K) : K \subseteq A, K \text{ compact}\} \leqslant \mu(A)$$

since every compact set is closed, and for each closed set $C \subseteq A$, we have

$$\mu(C) = \lim_{m \to \infty} \mu \left(\bigcup_{j=1}^{m} K_j \cap C \right) \leqslant \sup\{\mu(K) : K \subseteq A, K \text{ compact}\}$$

by monotonicity and the above argument. Thus

$$\mu(A) = \sup\{\mu(K) : K \subseteq A, K \text{ compact}\}.$$

Proof. We begin by assuming we have $\mu(X) < \infty$. We will remove this assumption at the end of the proof. First note that in this case, conclusion (ii) follows by applying conclusion (i) to the complement A^c of a μ-measurable set $A \subseteq X$ and using that $\mu(A^c) = \mu(X) - \mu(A)$.

Also since μ is Borel regular, it suffices to prove conclusion (i) in the case that $A \subseteq X$ is a Borel set. Let

$$\mathcal{A} := \{A \subseteq X : A \text{ is a Borel set such that (i) holds}\}.$$

Trivially we see that $\mathcal{A}$ contains all open sets; we claim that $\mathcal{A}$ is closed under countable unions and countable intersections.

Let $\{A_j\}_{j=1}^{\infty} \subseteq \mathcal{A}$ be an arbitrary countable collection of Borel sets satisfying (i). Let $\varepsilon > 0$ be arbitrary. For each $j \in \mathbb{N}$, we can find an open set U_j such that $A_j \subseteq U_j$ and $\mu(U_j) \leqslant \mu(A_j) + \varepsilon/2^j$. Then

$$\left(\bigcup_{j=1}^{\infty} U_j \right) \setminus \left(\bigcup_{k=1}^{\infty} A_k \right) = \bigcup_{j=1}^{\infty} \left(U_j \setminus \bigcup_{k=1}^{\infty} A_k \right) \subseteq \bigcup_{j=1}^{\infty} U_j \setminus A_j$$

and

$$\left(\bigcap_{j=1}^{\infty} U_j\right) \setminus \left(\bigcap_{k=1}^{\infty} A_k\right) = \bigcap_{j=1}^{\infty} \left(U_j \setminus \bigcap_{k=1}^{\infty} A_k\right) = \bigcap_{j=1}^{\infty} \left(\bigcup_{k=1}^{\infty} U_j \setminus A_k\right) \subseteq \bigcup_{j=1}^{\infty} U_j \setminus A_j$$

so by subadditivity we have

$$\mu\left(\left(\bigcup_{j=1}^{\infty} U_j\right) \setminus \left(\bigcup_{k=1}^{\infty} A_k\right)\right) \leqslant \sum_{j=1}^{\infty} \mu(U_j \setminus A_j) \leqslant \sum_{j=1}^{\infty} [\mu(U_j) - \mu(A_j)] \leqslant \varepsilon,$$

and

$$\lim_{N \to \infty} \mu\left(\left(\bigcap_{j=1}^{N} U_j\right) \setminus \left(\bigcap_{k=1}^{\infty} A_k\right)\right) = \mu\left(\left(\bigcap_{j=1}^{\infty} U_j\right) \setminus \left(\bigcap_{k=1}^{\infty} A_k\right)\right)$$
$$\leqslant \sum_{j=1}^{\infty} \mu(U_j \setminus A_j)$$
$$\leqslant \sum_{j=1}^{\infty} [\mu(U_j) - \mu(A_j)] \leqslant \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary we conclude that $\bigcup_{j=1}^{\infty} A_j$ and $\bigcap_{j=1}^{\infty} A_j$ satisfy (i).

(The reason we needed to take the limit in the second inequality is that $\bigcap_{j=1}^{N} U_j$ is a decreasing sequence of open sets, but the infinite intersection need not be open; taking limits allows us to show the infimum condition in (i) holds.)

This proves our claim that $\mathcal{A}$ is closed under countable unions and countable intersections. Since $\mathcal{A}$ contains all open sets and is closed under countable intersections, it contains all closed sets by the assumption on X. Notice that we do not yet know that $\mathcal{A}$ is a σ-algebra, since we do not know that it is closed under complements.

For this reason, we let $\mathcal{A}^*$ be defined by

$$\mathcal{A}^* := \{A \subseteq X : A^c \in \mathcal{A}\}$$

and we claim that $\mathcal{A}^*$ is a σ-algebra. Clearly $\varnothing, X \in \mathcal{A}^*$ since $\varnothing^c = X$ and $X^c = \varnothing$ are both open sets, so they belong to $\mathcal{A}$. Also $\mathcal{A}^*$ is closed under complements since if $A \in \mathcal{A}^*$, then $A^c \in \mathcal{A}$ by definition, so $(A^c)^c = A \in \mathcal{A}^*$.

Finally, if $\{A_j\}_{j=1}^{\infty} \subseteq \mathcal{A}^*$ is an arbitrary countable collection of sets, then $A_j^c \in \mathcal{A}$ for each $j \in \mathbb{N}$, so that

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

becuase $\mathcal{A}$ is closed under countable intersections. Thus $\bigcup_{j=1}^{\infty} A_j \in \mathcal{A}^*$.

This proves our claim that $\mathcal{A}^*$ is a σ-algebra. Since $\mathcal{A}^*$ contains all closed sets and hence all open sets, we conclude that $\mathcal{A}^*$ contains all Borel sets. Thus $\mathcal{A}$ also contains all Borel sets (and is actually equal to the Borel σ-algebra).

Thus every Borel set satisfies (i), which completes the proof in the case that $\mu(X) < \infty$.

Now we remove the assumption that $\mu(X) < \infty$. Since (X, μ) is open σ-finite, we can write $X = \bigcup_{j=1}^{\infty} V_j$ where $V_j \subseteq X$ is open and $\mu(V_j) < \infty$ for each $j \in \mathbb{Z}^+$. Since μ is Borel regular, it suffices to prove conclusion (i) in the case that $A \subseteq X$ is a Borel set.

Let $A \subseteq X$ be an arbitrary Borel set, and let $\varepsilon > 0$ be arbitrary. For each $j \in \mathbb{N}$, we can apply the first part of the proof to the measure $\mu \restriction V_j$ which satisfies $\mu \restriction V_j(X) = \mu(V_j) < \infty$ to find an open set $U_j \subseteq X$ such that $A \subseteq U_j$ and

$$\mu((V_j \cap U_j)\backslash A) = \mu \restriction V_j(U_j\backslash A) < \frac{\varepsilon}{2^j}.$$

By subadditivity of μ and summing over $j \in \mathbb{Z}^+$ we get

$$\mu\left(\bigcup_{j=1}^{\infty}(V_j \cap U_j)\backslash A\right) < \sum_{j=1}^{\infty} \frac{\varepsilon}{2^j} = \varepsilon.$$

Since $\bigcup_{j=1}^{\infty}(V_j \cap U_j)$ is open and contains A, this shows that A satisfies (i).

Now to complete the proof of conclusion (ii), let $A \subseteq X$ be a Borel set and let $\varepsilon > 0$ be arbitrary. For each $j \in \mathbb{N}$, the fact that $\mu \restriction V_j(X) < \infty$ implies we can apply conclusion (ii) to the measure $\mu \restriction V_j$ to find a closed set $C_j \subseteq X$ such that $C_j \subseteq A$ and

$$\mu((A \cap V_j)\backslash C_j) = \mu \restriction V_j(A\backslash C_j) < \frac{\varepsilon}{2^j}.$$

Since

$$\bigcup_{j=1}^{\infty} V_j \backslash \bigcup_{k=1}^{\infty} C_k = \bigcup_{j=1}^{\infty}\left(V_j\backslash \bigcup_{k=1}^{\infty} C_k\right) \subseteq \bigcup_{j=1}^{\infty} V_j\backslash C_j$$

the countable subadditivity of μ implies

$$\mu\left(A\backslash \bigcup_{k=1}^{\infty} C_k\right) = \mu\left(\bigcup_{j=1}^{\infty}(A \cap V_j)\backslash \bigcup_{k=1}^{\infty} C_k\right)$$

$$\leqslant \mu\left(\bigcup_{j=1}^{\infty}(A \cap V_j)\backslash C_j\right)$$

$$\leqslant \sum_{j=1}^{\infty} \mu((A \cap V_j)\backslash C_j) < \varepsilon.$$

Thus either $\mu(A) = \infty$ and $\mu\left(A\backslash \bigcup_{k=1}^{N} C_k\right) \to \infty$ as $N \to \infty$ or $\mu(A) < \infty$ and $\mu\left(A\backslash \bigcup_{k=1}^{N} C_k\right) < 2\epsilon$ for all sufficiently large N. In either case, since $\varepsilon > 0$ was arbitrary we conclude that A satisfies (ii). This completes the proof. $\qquad\square$

1.1.4 π-λ Theorem

It will be useful later to have a way of showing that two measures are equal. The π-λ theorem is a tool for doing this.

Definition 1.1.16 (π-system, λ-system)**.** Let X be a set. A collection $\mathcal{P}$ of subsets of X is called a π-*system* if it is closed under finite intersections, i.e. if $A, B \in \mathcal{P}$, then $A \cap B \in \mathcal{P}$ as well.

A collection $\mathcal{L}$ of subsets of X is called a λ-*system* if it satisfies the following properties:

- $X \in \mathcal{L}$.

- If $A, B \in \mathcal{L}$ and $A \subseteq B$, then $B \backslash A \in \mathcal{L}$ as well.

- If $\{A_j\}_{j=1}^{\infty}$ is an increasing sequence of sets in $\mathcal{L}$, then $\bigcup_{j=1}^{\infty} A_j \in \mathcal{L}$ as well.

Since π-systems and λ-systems are less restrictive than σ-algebras, it is easier in practice to check that a collection of sets is a π-system or a λ-system than to check that it is a σ-algebra. Of course, σ-algebras are both π-systems and λ-systems. The converse is contained in Step 3 of the following theorem.

Theorem 1.1.17 (Dynkin's π-λ Theorem)**.** *Let X be a set. If $\mathcal{P}$ is a π-system and $\mathcal{L}$ is a λ-system such that $\mathcal{P} \subseteq \mathcal{L}$, then $\sigma(\mathcal{P}) \subseteq \mathcal{L}$.*

Proof. Step 1. We define

$$\mathcal{D} := \bigcap \{\mathcal{L} : \text{ is a } \lambda\text{-system containing } \mathcal{P}\}.$$

Then $\mathcal{D}$ is a λ-system and we have $\mathcal{P} \subseteq \mathcal{D} \subseteq \mathcal{L}$.

We check this. First, let $P \in \mathcal{P}$. Then for each λ-system $\mathcal{L}'$ containing $\mathcal{P}$, we have $P \in \mathcal{L}'$, and hence $P \in \mathcal{D}$. This shows that $\mathcal{P} \subseteq \mathcal{D}$. Also, since $\mathcal{L}$ is a λ-system containing $\mathcal{P}$, we have $\mathcal{D} \subseteq \mathcal{L}$ by definition.

Finally we see that $\mathcal{D}$ is a λ-system.

- Since $X \in \mathcal{L}'$ for each λ-system $\mathcal{L}'$ containing $\mathcal{P}$, we have $X \in \mathcal{D}$ as well.

- If $A, B \in \mathcal{D}$ and $A \subseteq B$, then $A, B \in \mathcal{L}'$ for each λ-system $\mathcal{L}'$ containing $\mathcal{P}$, and hence $B \backslash A \in \mathcal{L}'$. This shows that $B \backslash A \in \mathcal{D}$ as well.

- If $\{A_j\}_{j=1}^{\infty}$ is an increasing sequence of sets in $\mathcal{D}$, then $A_j \in \mathcal{L}'$ for each $j \in \mathbb{Z}^+$ and each λ-system $\mathcal{L}'$ containing $\mathcal{P}$, and hence $\bigcup_{j=1}^{\infty} A_j \in \mathcal{L}'$. This shows that $\bigcup_{j=1}^{\infty} A_j \in \mathcal{D}$.

Step 2. We claim that $\mathcal{D}$ is also a π-system.

Let $A, B \in \mathcal{D}$. We want to show that $A \cap B \in \mathcal{D}$. Define

$$\mathcal{D}_A := \{C \subseteq X : A \cap C \in \mathcal{D}\}.$$

We claim that $\mathcal{D}_A$ is a λ-system containing $\mathcal{P}$; this follows from the fact that $\mathcal{D}$ is a λ-system.

- Since $A \cap X = A \in \mathcal{D}$, we have $X \in \mathcal{D}_A$.

- If $C_1, C_2 \in \mathcal{D}_A$ and $C_1 \subseteq C_2$, then $A \cap C_1, A \cap C_2 \in \mathcal{D}$ and $A \cap C_1 \subseteq A \cap C_2$, so that $(A \cap C_2)\backslash(A \cap C_1) = A \cap (C_2\backslash C_1) \in \mathcal{D}$. This shows that $C_2\backslash C_1 \in \mathcal{D}_A$.

- If $\{C_j\}_{j=1}^{\infty}$ is an increasing sequence of sets in $\mathcal{D}_A$, then $A \cap C_j \in \mathcal{D}$ for each $j \in \mathbb{Z}^+$, and hence

$$A \cap \left(\bigcup_{j=1}^{\infty} C_j \right) = \bigcup_{j=1}^{\infty} (A \cap C_j) \in \mathcal{D}$$

  since $\mathcal{D}$ is a λ-system. This shows that $\bigcup_{j=1}^{\infty} C_j \in \mathcal{D}_A$.

Therefore, $\mathcal{D}_A$ is a λ-system. Now, if $P \in \mathcal{P}$, then $A \cap P \in \mathcal{D}$ since $\mathcal{P}$ is a π-system and $A \in \mathcal{D}$; hence $P \in \mathcal{D}_A$. This shows that $\mathcal{P} \subseteq \mathcal{D}_A$. By the definition of $\mathcal{D}$, we have $\mathcal{D}_A \subseteq \mathcal{D}$, and hence $B \in \mathcal{D}_A$ since $B \in \mathcal{D}$ by assumption. This shows that $A \cap B \in \mathcal{D}$, as desired.

Since $A, B \in \mathcal{D}$ were arbitrary, this shows that $\mathcal{D}$ is a π-system.

Step 3. Now we claim that $\mathcal{D}$ is a σ-algebra. This follows from the fact that $\mathcal{D}$ is both a π-system and a λ-system.

Since $\mathcal{D}$ is a λ-system, we have $X \in \mathcal{D}$, and the closure of $\mathcal{D}$ under complements implies that $X\backslash A = \varnothing \in \mathcal{D}$. Also it is clear that $\mathcal{D}$ is closed under finite unions since it is closed under finite intersections and complements. Finally, let $\{A_j\}_{j=1}^{\infty}$ be an arbitrary countable collection of sets in $\mathcal{D}$. Then for each $n \in \mathbb{Z}^+$ define $B_n := \bigcup_{j=1}^{n} A_j$ which is in $\mathcal{D}$ since $\mathcal{D}$ is closed under finite unions; also note that $\{B_n\}_{n=1}^{\infty}$ is an increasing sequence of sets in $\mathcal{D}$. Since $\mathcal{D}$ is a λ-system, we have

$$\bigcup_{j=1}^{\infty} A_j = \bigcup_{n=1}^{\infty} \bigcup_{j=1}^{n} A_j = \bigcup_{n=1}^{\infty} B_n \in \mathcal{D}.$$

This shows that $\mathcal{D}$ is closed under countable unions, and hence is a σ-algebra.

Since $\mathcal{D}$ is a σ-algebra containing $\mathcal{P}$, we have $\sigma(\mathcal{P}) \subseteq \mathcal{D} \subseteq \mathcal{L}$ as desired. $\square$

Corollary 1.1.18 (Borel Outer Measures on $\mathbb{R}^n$)**.** *If μ and ν are two Borel outer measures on $\mathbb{R}^n$ such that*

$$\mu(R) = \nu(R)$$

for each closed rectangle $R = [a_1, b_1] \times \cdots \times [a_n, b_n]$, then $\mu(E) = \nu(E)$ for every Borel set $E \subseteq \mathbb{R}^n$. The same is true if we replace closed rectangles with open rectangles or half-open rectangles.

Proof. Let $\mathcal{P}$ be the collection of all closed (open, half-open) rectangles in $\mathbb{R}^n$. Then $\mathcal{P}$ is a π-system which generates the Borel σ-algebra on $\mathbb{R}^n$. Let $\mathcal{L} = \{E \subseteq \mathbb{R}^n : \mu(E) = \nu(E)\}$. Then $\mathcal{L}$ is a λ-system containing $\mathcal{P}$, since $\mu(R) = \nu(R)$ for every $R \in \mathcal{P}$ by assumption. Thus by the π-λ theorem, we have $\sigma(\mathcal{P}) \subseteq \mathcal{L}$.

We know that $\sigma(\mathcal{P})$ is the Borel σ-algebra on $\mathbb{R}^n$ since each open set in $\mathbb{R}^n$ can be written as a countable union of closed (open, half-open) rectangles. Thus $\mu(E) = \nu(E)$ for every Borel set $E \subseteq \mathbb{R}^n$ as desired. $\qquad\square$

Hopefully this proof shows the utility of the π-λ theorem, as it is not obvious that the collection $\mathcal{L}$ is a σ-algebra (closed under intersections is not obvious), but it is easy to check that it is a λ-system.

1.2   Lebesgue Measure

Up to now, we have not really had any interesting examples. Sorry. Of course, it is examples which motivated all the abstract theory we have developed so far. The most important and motivating example is Lebesgue measure on the line $\mathbb{R}$ and in Euclidean space $\mathbb{R}^n$.

We will look at another example, the Hausdorff measure(s), in a later section.

1.2.1   Boxes and Cubes

Definition 1.2.1 (Boxes and Cubes). A *closed box* in $\mathbb{R}^n$ is a set of the form

$$R = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$$

for some real numbers $a_j \leqslant b_j$ for $j = 1, \ldots, n$. The *volume* of such a box is defined to be

$$\text{vol}(R) := \prod_{j=1}^{n} (b_j - a_j).$$

An *open box* is a set of the form

$$(a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_n, b_n)$$

for some real numbers $a_j < b_j$ for $j = 1, \ldots, n$. A *half-open* box is a set of the form

$$[a_1, b_1) \times [a_2, b_2) \times \cdots \times [a_n, b_n)$$

for some real numbers $a_j \leqslant b_j$ for $j = 1, \ldots, n$. The volume of an open or half-open box is defined in the same way as for closed boxes.

A *cube* is a special case of a box where all the side lengths are equal, i.e. $b_j - a_j = b_i - a_i$ for all $1 \leqslant i, j \leqslant n$. If $Q \subset \mathbb{R}^n$ is a cube with side length $\ell > 0$, then $\mathrm{vol}(Q) = \ell^n$.

We remark that we allow degenerate closed and half-open boxes, where $a_j = b_j$ for some j. Degenerate open boxes are empty, so we do not allow them. Also, in our definition, boxes and cubes have sides parallel to the coordinate axes.

Definition 1.2.2 (Almost Disjoint). Say that a collection of boxes $\{R_j\}_{j \in J}$ in $\mathbb{R}^n$ is *almost disjoint* if the interiors of the boxes are pairwise disjoint.

Intuitively, a collection of boxes is almost disjoint if they only overlap on their boundaries.

Lemma 1.2.3. *Let $R \subseteq \mathbb{R}^n$ be a closed box, and let $\{R_j\}_{j=1}^N$ be a finite collection of almost disjoint closed boxes in $\mathbb{R}^n$ such that*

$$R = \bigcup_{j=1}^N R_j.$$

Then

$$\mathrm{vol}\left(\bigcup_{j=1}^N R_j\right) = \sum_{j=1}^N \mathrm{vol}(R_j).$$

Proof. For each $1 \leqslant j \leqslant N$, let the box R_j be given by

$$R_j = [a_{j,1}, b_{j,1}] \times [a_{j,2}, b_{j,2}] \times \cdots \times [a_{j,n}, b_{j,n}]$$

where $a_{j,i} \leqslant b_{j,i}$ for $i = 1, \ldots, n$. Then for each index $1 \leqslant i \leqslant n$, we collect the endpoints of the i-th intervals of all the boxes into a set,

$$S_i = \{a_{j,i}, b_{j,i} : 1 \leqslant j \leqslant N\}$$

which is a finite set of real numbers, and therefore can be written as

$$S_i = \{c_{i,1}, c_{i,2}, \ldots, c_{i,m_i}\}$$

where $c_{i,1} < c_{i,2} < \cdots < c_{i,m_i}$.

For each $1 \leqslant i \leqslant n$ and each $1 \leqslant k \leqslant m_i$ define the interval

$$I_{i,k} := [c_{i,k-1}, c_{i,k}].$$

This gives a collection of closed boxes $\{\tilde{R}_{k_1,k_2,\ldots,k_n}\}_{1 \leqslant k_i \leqslant m_i, 1 \leqslant i \leqslant n}$ where

$$\tilde{R}_{k_1,k_2,\ldots,k_n} := I_{1,k_1} \times I_{2,k_2} \times \cdots \times I_{n,k_n}.$$

We note that there are at most $m_1 m_2 \cdots m_n$ such boxes, and that these boxes are almost disjoint because the interior of a box $\tilde{R}_{k_1,k_2,\ldots,k_n}$ is a product of open intervals $I^\circ_{1,k_1} \times I^\circ_{2,k_2} \times \cdots \times I^\circ_{n,k_n}$, and two such products of open intervals are either equal or disjoint by definition of endpoints $c_{i,j}$.

Define index sets

$$\mathcal{J} := \left\{ (k_1, k_2, \ldots, k_n) : \tilde{R}_{k_1,k_2,\ldots,k_n} \subseteq R \right\} \quad \text{and} \quad \mathcal{J}_j := \left\{ (k_1, k_2, \ldots, k_n) : \tilde{R}_{k_1,k_2,\ldots,k_n} \subseteq R_j \right\}$$

for $1 \leqslant j \leqslant N$. Then for each $1 \leqslant j \leqslant N$ we claim that we have

$$R_j = \bigcup_{(k_1,k_2,\ldots,k_n) \in \mathcal{J}_j} \tilde{R}_{k_1,k_2,\ldots,k_n}.$$

Fix $1 \leqslant j \leqslant N$. The inclusion $\supseteq$ holds by definition of $\mathcal{J}_j$. To see that the inclusion $\subseteq$ holds, let $x \in R_j$; then

$$x = (x_1, x_2, \ldots, x_n) \in [a_{j,1}, b_{j,1}] \times [a_{j,2}, b_{j,2}] \times \cdots \times [a_{j,n}, b_{j,n}].$$

Since $a_{j,i}, b_{j,i} \in S_i$ for each $1 \leqslant i \leqslant n$, there exist indices $1 \leqslant k_i \leqslant m_i$ such that $x_i \in I_{i,k_i}$ for each $1 \leqslant i \leqslant n$. Thus

$$x \in \tilde{R}_{k_1,k_2,\ldots,k_n},$$

and by definition of $\mathcal{J}_j$ we have $(k_1, k_2, \ldots, k_n) \in \mathcal{J}_j$. Therefore $x \in \bigcup_{(k_1,k_2,\ldots,k_n) \in \mathcal{J}_j} \tilde{R}_{k_1,k_2,\ldots,k_n}$. Since $x \in R_j$ was arbitrary, we have proven the reverse inclusion $\supseteq$ and shown the claim. As a consequence of this claim and the fact that $R = \bigcup_{j=1}^N R_j$, we also have

$$\mathcal{J} = \bigcup_{j=1}^N \mathcal{J}_j. \tag{$*$}$$

The fact that the boxes $\tilde{R}_{k_1,k_2,\ldots,k_n}$ are almost disjoint also implies that the sets $\{\mathcal{J}_j\}_{j=1}^N$ are pairwise disjoint.

We also claim that

$$\mathrm{vol}(R_j) = \sum_{(k_1,k_2,\ldots,k_n) \in \mathcal{J}_j} \mathrm{vol}(\tilde{R}_{k_1,k_2,\ldots,k_n}) \tag{$**$}$$

for each $1 \leqslant j \leqslant N$.

To see this, fix $1 \leqslant j \leqslant N$. Then for each $1 \leqslant i \leqslant n$ we have

$$a_{j,i} = c_{i,l_i} < c_{i,l_i+r_i} = b_{j,i}$$

for some $1 \leqslant l_i < l_i + r_i \leqslant m_i$ and some $r_i \geqslant 1$; in particular, the interval $[a_{j,i}, b_{j,i}]$ is partitioned by the intervals $I_{i,l_i}, I_{i,l_i+1}, \ldots, I_{i,l_i+r_i}$. As a telescoping sum, we have

$$b_{j,i} - a_{j,i} = c_{i,l_i+r_i} - c_{i,l_i} = \sum_{s=l_i}^{l_i+r_i-1} (c_{i,s+1} - c_{i,s})$$

for each $1 \leqslant i \leqslant n$. Thus, we obtain

$$
\begin{aligned}
\mathrm{vol}(R_j) &= \prod_{i=1}^{n}(b_{j,i} - a_{j,i}) = \prod_{i=1}^{n} \sum_{s=l_i}^{l_i+r_i-1} (c_{i,s+1} - c_{i,s}) \\
&= \sum_{s_1=l_1}^{l_1+r_1-1} \sum_{s_2=l_2}^{l_2+r_2-1} \cdots \sum_{s_n=l_n}^{l_n+r_n-1} \prod_{i=1}^{n}(c_{i,s_i+1} - c_{i,s_i}) \\
&= \sum_{s_1=l_1}^{l_1+r_1-1} \sum_{s_2=l_2}^{l_2+r_2-1} \cdots \sum_{s_n=l_n}^{l_n+r_n-1} \mathrm{vol}(\tilde{R}_{s_1,s_2,\dots,s_n}) \\
&= \sum_{(k_1,k_2,\dots,k_n) \in \mathcal{J}_j} \mathrm{vol}(\tilde{R}_{k_1,k_2,\dots,k_n}).
\end{aligned}
$$

where in the final equality we have used the fact that $\tilde{R}_{s_1,s_2,\dots,s_n} \subseteq R_j$ if and only if $l_i \leqslant s_i \leqslant l_i + r_i - 1$ for each $1 \leqslant i \leqslant n$, which is equivalent to $(s_1, s_2, \dots, s_n) \in \mathcal{J}_j$. We have thus shown the claim.

We make a third and final claim that

$$
\mathrm{vol}\,(R) = \sum_{(k_1,k_2,\dots,k_n) \in \mathcal{J}} \mathrm{vol}(\tilde{R}_{k_1,k_2,\dots,k_n}). \tag{\dagger}
$$

To see this, we write R as a box of

$$
R = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n].
$$

Then for each $1 \leqslant i \leqslant n$ we have $a_i, b_i \in S_i$, and in fact it must be the case that $a_i = c_{i,1}$ and $b_i = c_{i,m_i}$. Thus, for each $1 \leqslant i \leqslant n$ we have

$$
b_i - a_i = c_{i,m_i} - c_{i,1} = \sum_{s=1}^{m_i-1} (c_{i,s+1} - c_{i,s})
$$

and therefore

$$
\begin{aligned}
\mathrm{vol}(R) &= \prod_{i=1}^{n}(b_i - a_i) = \prod_{i=1}^{n} \sum_{s=1}^{m_i-1} (c_{i,s+1} - c_{i,s}) \\
&= \sum_{s_1=1}^{m_1-1} \sum_{s_2=1}^{m_2-1} \cdots \sum_{s_n=1}^{m_n-1} \prod_{i=1}^{n}(c_{i,s_i+1} - c_{i,s_i}) \\
&= \sum_{s_1=1}^{m_1-1} \sum_{s_2=1}^{m_2-1} \cdots \sum_{s_n=1}^{m_n-1} \mathrm{vol}(\tilde{R}_{s_1,s_2,\dots,s_n}) \\
&= \sum_{(k_1,k_2,\dots,k_n) \in \mathcal{J}} \mathrm{vol}(\tilde{R}_{k_1,k_2,\dots,k_n}).
\end{aligned}
$$

This proves the claim.

Now we see that putting (∗), (∗∗), and (†) together yields

$$\text{vol}(R) = \sum_{(k_1, k_2, \ldots, k_n) \in \mathcal{J}} \text{vol}(\tilde{R}_{k_1, k_2, \ldots, k_n})$$

$$= \sum_{j=1}^{N} \sum_{(k_1, k_2, \ldots, k_n) \in \mathcal{J}_j} \text{vol}(\tilde{R}_{k_1, k_2, \ldots, k_n})$$

$$= \sum_{j=1}^{N} \text{vol}(R_j).$$

This completes the proof. □

Lemma 1.2.4. *Let $R \subseteq \mathbb{R}^n$ be a closed box, and let $\{R_j\}_{j=1}^{N}$ be a finite collection of closed boxes such that*

$$R \subseteq \bigcup_{j=1}^{N} R_j.$$

Then

$$\text{vol}(R) \leqslant \sum_{j=1}^{N} \text{vol}(R_j).$$

The idea is similar to the last proof.

Proof. For each $1 \leqslant j \leqslant N$, let the box R_j be given by

$$R_j = [a_{j,1}, b_{j,1}] \times [a_{j,2}, b_{j,2}] \times \cdots \times [a_{j,n}, b_{j,n}]$$

where $a_{j,i} \leqslant b_{j,i}$ for $i = 1, \ldots, n$. Then for each index $1 \leqslant i \leqslant n$, we collect the endpoints of the i-th intervals of all the boxes into a set,

$$S_i = \{a_{j,i}, b_{j,i} : 1 \leqslant j \leqslant N\}$$

which is a finite set of real numbers, and therefore can be written as

$$S_i = \{c_{i,1}, c_{i,2}, \ldots, c_{i,m_i}\}$$

where $c_{i,1} < c_{i,2} < \cdots < c_{i,m_i}$.

For each $1 \leqslant i \leqslant n$ and each $1 \leqslant k \leqslant m_i$ define the interval

$$I_{i,k} := [c_{i,k-1}, c_{i,k}].$$

This gives a collection of closed boxes $\{\tilde{R}_{k_1, k_2, \ldots, k_n}\}_{1 \leqslant k_i \leqslant m_i, 1 \leqslant i \leqslant n}$ where

$$\tilde{R}_{k_1, k_2, \ldots, k_n} := I_{1,k_1} \times I_{2,k_2} \times \cdots \times I_{n,k_n}.$$

We note that there are at most $m_1 m_2 \cdots m_n$ such boxes, and that these boxes are almost disjoint because the interior of a box $\tilde{R}_{k_1,k_2,\ldots,k_n}$ is a product of open intervals $I_{1,k_1}^\circ \times I_{2,k_2}^\circ \times \cdots \times I_{n,k_n}^\circ$, and two such products of open intervals are either equal or disjoint by definition of endpoints $c_{i,j}$.

Define index sets

$$\mathcal{J} := \left\{ (k_1, k_2, \ldots, k_n) : \tilde{R}_{k_1,k_2,\ldots,k_n} \subseteq R \right\} \quad \text{and} \quad \mathcal{J}_j := \left\{ (k_1, k_2, \ldots, k_n) : \tilde{R}_{k_1,k_2,\ldots,k_n} \subseteq R_j \right\}$$

for $1 \leqslant j \leqslant N$. Then for each $1 \leqslant j \leqslant N$ we claim that we have

$$R_j = \bigcup_{(k_1,k_2,\ldots,k_n) \in \mathcal{J}_j} \tilde{R}_{k_1,k_2,\ldots,k_n}$$

and similarly

$$R = \bigcup_{(k_1,k_2,\ldots,k_n) \in \mathcal{J}} \tilde{R}_{k_1,k_2,\ldots,k_n}.$$

The proof of this claim is identical to the proof of the same claim in the previous lemma. As a consequence of this claim and the fact that $R \subseteq \bigcup_{j=1}^{N} R_j$, we also have

$$\mathcal{J} \subseteq \bigcup_{j=1}^{N} \mathcal{J}_j. \tag{$*$}$$

The main difference from the previous lemma is that the sets $\{\mathcal{J}_j\}_{j=1}^{N}$ are not necessarily pairwise disjoint. Because the boxes $\{R_j\}_{j=1}^{N}$ are not necessarily almost disjoint, the a single box $\tilde{R}_{k_1,k_2,\ldots,k_n}$ may be contained in more than one of the boxes R_j.

Now we see that

$$\begin{aligned}
\mathrm{vol}(R) &= \sum_{(k_1,k_2,\ldots,k_n) \in \mathcal{J}} \mathrm{vol}(\tilde{R}_{k_1,k_2,\ldots,k_n}) \\
&\leqslant \sum_{j=1}^{N} \sum_{(k_1,k_2,\ldots,k_n) \in \mathcal{J}_j} \mathrm{vol}(\tilde{R}_{k_1,k_2,\ldots,k_n}) \\
&= \sum_{j=1}^{N} \mathrm{vol}(R_j)
\end{aligned}$$

where we have used the previous lemma and our claim in the first equality, the inclusion $(*)$ in the inequality, and the previous lemma and our claim again in the final equality. $\qquad \square$

Exercise 1.2.1. Let $U \subseteq \mathbb{R}^n$ be a nonempty open set. Show that there exists a countable collection of almost disjoint closed cubes $\{Q_j\}_{j=1}^{\infty}$ such that

$$U = \bigcup_{j=1}^{\infty} Q_j.$$

Proof. For this proof, we will use the dyadic cubes

$$Q_{m,k} := \left\{ x = (x_1, \ldots, x_n) \in \mathbb{R}^n : 2^k m_j \leqslant x_j \leqslant 2^k (m_j + 1) \text{ for } j = 1, 2, \ldots, n \right\}, \quad m \in \mathbb{Z}^n, k \in \mathbb{Z}.$$

We say that $Q_{m,k}$ is a dyadic cube of generation k. It has edge length 2^k and volume 2^{kn}, and corner point $2^k m \in 2^k \mathbb{Z}^n$. Notice that $Q_{m,0}$ is a half-open cube of edge length 1 with corner point $m \in \mathbb{Z}^n$. Then $Q_{m,1}$ is a half-open cube of edge length 2 with corner point $2m \in 2\mathbb{Z}^n$, and so on. Similarly, $Q_{m,-1}$ is a half-open cube of edge length $\frac{1}{2}$ with corner point $\frac{m}{2} \in \frac{1}{2}\mathbb{Z}^n$, and so on. The dyadic cubes of generation k form are almost disjoint and their union is $\mathbb{R}^n$.

We define the desired collection of cubes $\{Q_j\}_{j=1}^{\infty}$ as follows.

At the first step, let $\mathcal{Q}_{-1}$ be the set of all dyadic cubes of generation -1, which have edge length $\frac{1}{2}$. There are countably many such cubes, and we sort them into three sets based on their relationship to U — we define

$$\text{green}_1 = \{Q_{m,-1} : m \in \mathbb{Z}^n \text{ and } Q_{m,-1} \subset U\}$$
$$\text{red}_1 = \{Q_{m,-1} : m \in \mathbb{Z}^n \text{ and } Q_{m,-1} \cap U = \varnothing\}$$
$$\text{yellow}_1 = \{Q_{m,-1} : m \in \mathbb{Z}^n \text{ and } Q_{m,-1} \cap U \neq \varnothing, Q_{m,-1} \not\subset U\}.$$

The green cubes are those which are completely contained in U, the red cubes are those which do not intersect U at all and are contained in the complement of U, and the yellow cubes are those which intersect U but are not completely contained in U. The set of green cubes green_1 and the set of yellow cubes yellow_1 cannot both be empty, since U is nonempty and open. Let $\mathcal{G}_1 := \text{green}_1$ be the set of green cubes at step 1.

At the second step, we look at the yellow cubes from the first step. Each yellow cube $Q_{m,-1}$ evolves into 2^n dyadic cubes of generation -2, each with edge length $\frac{1}{4}$. We sort these new cubes into three sets based on their relationship to U — we define

$$\text{green}_2 = \{Q_{m,-2} : Q_{m,-2} \subset U \text{ and } Q_{m,-1} \subset \text{yellow}_1\}$$
$$\text{red}_2 = \{Q_{m,-2} : Q_{m,-2} \cap U = \varnothing \text{ and } Q_{m,-1} \subset \text{yellow}_1\}$$
$$\text{yellow}_2 = \{Q_{m,-2} : Q_{m,-2} \cap U \neq \varnothing, Q_{m,-2} \not\subset U \text{ and } Q_{m,-1} \subset \text{yellow}_1\}.$$

The set of green cubes green_2 and the set of yellow cubes yellow_2 cannot both be empty, since each yellow cube from the first step intersects U and U is open. Let $\mathcal{G}_2 := \mathcal{G}_1 \cup \text{green}_2$ be the set of green cubes at step 2.

We continue in this manner, where at step k we look at the yellow cubes from step $k-1$, subdivide each of them into 2^n dyadic cubes of generation $-k$, and sort these new cubes into three sets based on their relationship to U. We define the sets green_k, red_k, and yellow_k in the obvious way, and let $\mathcal{G}_k := \mathcal{G}_{k-1} \cup \text{green}_k$ be the set of green cubes at step k. This

process results in a countable collection of green cubes $\mathcal{G} := \bigcup_{k=1}^{\infty} \mathcal{G}_k$. We claim that $\mathcal{G}$ is the desired collection of cubes, i.e

$$U = \bigcup_{Q \in \mathcal{G}} Q.$$

To see this, first note that every cube in $\mathcal{G}$ is contained in U by construction, so we have the inclusion $\supseteq$. To see the inclusion $\subseteq$, let $x \in U$. Then there exists an $N > 0$ and a cube $Q_{m,-N}$ of generation $-N$ such that $x \in Q_{m,-N} \subset U$. Either this cube $Q_{m,-N}$ is in $\mathcal{G}$, or it is contained in a cube which is in $\mathcal{G}$. Therefore $x \in \bigcup_{Q \in \mathcal{G}} Q$, and since $x \in U$ was arbitrary, we have shown the inclusion $\subseteq$.

Finally, we note that the cubes in $\mathcal{G}$ are almost disjoint, since they are dyadic cubes of various generations. Since U was an arbitrary nonempty open set, this completes the proof. $\qquad\square$

Corollary 1.2.5. *The Borel σ-algebra on $\mathbb{R}^n$ is generated by the collection of closed boxes in $\mathbb{R}^n$. Similarly, it is also generated by the collection of open boxes in $\mathbb{R}^n$, or by the collection of half-open boxes in $\mathbb{R}^n$.*

Proof. Let $\mathcal{A}$ be the smallest σ-algebra containing all closed (open, half-open) boxes in $\mathbb{R}^n$. By the previous exercise, every open set in $\mathbb{R}^n$ is a countable union of closed (open, half-open) boxes, so every open set is in $\mathcal{A}$. Since the Borel σ-algebra is the smallest σ-algebra containing all open sets, we have that the Borel σ-algebra is contained in $\mathcal{A}$.

Conversely, the collection of closed (open, half-open) boxes is a subset of the Borel σ-algebra and is a π-system which generates $\mathcal{A}$, so we know that $\mathcal{A}$ is contained in the Borel σ-algebra. $\qquad\square$

1.2.2 Lebesgue Measure on $\mathbb{R}^n$

With the theory we have put in place, this should be pretty straightforward.

We want to define an outer measure μ on $\mathbb{R}^n$ which assigns to each box its volume. This measure should also satisfy the following properties, to capture our intuition of volume:

- (Translation Invariance) If $A \subseteq \mathbb{R}^n$ and $x \in \mathbb{R}^n$, then $\mu(A + x) = \mu(A)$, where $A + x = \{a + x : a \in A\}$.

- (Homogeneity) If $A \subseteq \mathbb{R}^n$ and $t > 0$, then $\mu(tA) = t^n \mu(A)$, where $tA = \{ta : a \in A\}$.

- (Disjoint Additivity) If $A \subseteq \mathbb{R}^n$ and $B \subseteq \mathbb{R}^n$ are disjoint, then $\mu(A \cup B) = \mu(A) + \mu(B)$.

We will see that these properties are all satisfied by μ-measurable sets with respect to the Lebesgue measure μ we construct.

Definition 1.2.6 (Lebesgue Outer Measure)**.** The *Lebesgue outer measure* μ on $\mathbb{R}^n$ is defined for each $E \subseteq \mathbb{R}^n$ by

$$\mu(E) := \inf \left\{ \sum_{j=1}^{\infty} \mathrm{vol}(Q_j) : E \subseteq \bigcup_{j=1}^{\infty} Q_j, \text{ each } Q_j \text{ is a closed cube} \right\}$$

where $\mathrm{vol}(Q_j)$ is the volume of the cube Q_j.

That is, to define the Lebesgue outer measure of a set $E \subseteq \mathbb{R}^n$, we cover E by a countable collection of closed cubes $\{Q_j\}_{j=1}^{\infty}$ and take the infimum of the sums of the volumes of these cubes over all such covers. We see that for each $E \subseteq \mathbb{R}^n$, the set over which we are taking the infimum in the definition of $\mu(E)$ is nonempty, so $\mu(E)$ is well-defined and takes values in $[0, \infty]$.

One can instead use open, half-open, or closed boxes instead of closed cubes in the definition, and the resulting outer measure will be the same. One can even use dyadic cubes instead of arbitrary cubes or boxes. The theory is the same in all these cases. One can also cover by countably many open or closed balls instead of cubes or boxes, but it is nontrivial to show this gives the same theory (but it does).

What is crucial though is that we allow countably many cubes or boxes in the cover. Allowing only finitely many cubes or boxes would give the theory of *Jordan content*, instead of Lebesgue measure.

Let us begin by showing that the Lebesgue outer measure gives the correct value on boxes.

Lemma 1.2.7. *Let $R \subseteq \mathbb{R}^n$ be a box which is either closed, open, or half-open. Then $\mu(R) = \mathrm{vol}(R)$.*

Proof. Step 1: We will first show that $\mu(Q) = \text{vol}(Q)$ for every closed cube $Q \subseteq \mathbb{R}^n$.

Let $Q \subseteq \mathbb{R}^n$ be a cube of the form

$$Q = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n].$$

To see that $\mu(Q) = \text{vol}(Q)$, first note that $\mu(Q) \leqslant \text{vol}(Q)$ since Q can be covered by itself. This inequality shows that equality holds if $\text{vol}(Q) = 0$, so we may assume that $\text{vol}(Q) > 0$. Let $\{Q_j\}_{j=1}^{\infty}$ be a cover of Q by closed cubes; let $\epsilon > 0$ be given. For each $j \geqslant 1$ we choose an open cube $S_{j,\epsilon}$ containing Q_j such that $\text{vol}(S_{j,\epsilon}) \leqslant (1 + \epsilon) \text{vol}(Q_j)$. Then $\{S_{j,\epsilon}\}_{j=1}^{\infty}$ is an open cover of the compact set Q, so we can choose a finite subcover $\{S_{j,\epsilon}\}_{j=1}^{N}$ of Q. Taking the closure of these cubes, we obtain a finite cover of Q by closed cubes $\{\overline{S_{j,\epsilon}}\}_{j=1}^{N}$. Then by 1.2.4 we have

$$\text{vol}(Q) \leqslant \sum_{j=1}^{N} \text{vol}(\overline{S_{j,\epsilon}}) = \sum_{j=1}^{N} \text{vol}(S_{j,\epsilon}) \leqslant (1 + \epsilon) \sum_{j=1}^{N} \text{vol}(Q_j).$$

Since $\epsilon > 0$ was arbitrary, we have

$$\text{vol}(Q) \leqslant \sum_{j=1}^{\infty} \text{vol}(Q_j).$$

Taking the infimum over all such covers $\{Q_j\}_{j=1}^{\infty}$ of Q, we obtain $\text{vol}(Q) \leqslant \mu(Q)$. Thus we have shown $\mu(Q) = \text{vol}(Q)$ for every closed cube $Q \subseteq \mathbb{R}^n$.

Step 2: We will now extend this to all boxes.

Now let $R \subseteq \mathbb{R}^n$ be a box which is either closed, open, or half-open. If $\text{vol}(R) = 0$, then $\mu(R) = 0$ as R can be covered by a degenerate cube of volume 0. Thus we may assume that $\text{vol}(R) > 0$. For convenience, we let

$$\overline{R} = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$$

where $a_j < b_j$ for $j = 1, \ldots, n$.

Suppose that $\{Q_j\}_{j=1}^{\infty}$ is a cover of R by closed cubes. Then for each $\varepsilon > 0$ and each $j \geqslant 1$ we can choose an open cube $S_{j,\epsilon}$ containing Q_j such that $\text{vol}(S_{j,\epsilon}) \leqslant (1 + \epsilon) \text{vol}(Q_j)$. Then $\{S_{j,\epsilon}\}_{j=1}^{\infty}$ is an open cover of the compact set $\overline{R}$, so we can choose a finite subcover $\{S_{j,\epsilon}\}_{j=1}^{N}$ of $\overline{R}$. Taking the closure of these cubes, we obtain a finite cover of $\overline{R}$ by closed cubes $\{\overline{S_{j,\epsilon}}\}_{j=1}^{N}$. Then by 1.2.4 we have

$$\text{vol}(\overline{R}) \leqslant \sum_{j=1}^{N} \text{vol}(\overline{S_{j,\epsilon}}) = \sum_{j=1}^{N} \text{vol}(S_{j,\epsilon}) \leqslant (1 + \epsilon) \sum_{j=1}^{N} \text{vol}(Q_j) = (1 + \epsilon) \sum_{j=1}^{\infty} \text{vol}(Q_j).$$

Since $\epsilon > 0$ was arbitrary, we have

$$\mathrm{vol}(\overline{R}) \leqslant \sum_{j=1}^{\infty} \mathrm{vol}(Q_j).$$

Taking the infimum over all such covers $\{Q_j\}_{j=1}^{\infty}$ of R, we obtain $\mathrm{vol}(\overline{R}) \leqslant \mu(R)$. Since $\mathrm{vol}(\overline{R}) = \mathrm{vol}(R)$, we have shown that $\mathrm{vol}(R) \leqslant \mu(R)$.

To see the reverse inequality, fix $k \geqslant 1$; we let $\mathcal{Q}_k$ be given by

$$\mathcal{Q}_k = \big\{[m_1/k, (m_1+1)/k] \times [m_2/k, (m_2+1)/k] \times \cdots \times [m_n/k, (m_n+1)/k] : m_i \in \mathbb{Z}\big\}$$

which is a collection of almost disjoint closed cubes of side length $1/k$ that cover $\mathbb{R}^n$. We let $\mathcal{Q}_k^{(1)}$ be the subcollection of cubes in $\mathcal{Q}_k$ which are entirely contained in R, and we let $\mathcal{Q}_k^{(2)}$ be the subcollection of cubes in $\mathcal{Q}_k$ which intersect R but are not entirely contained in R. Then both $\mathcal{Q}_k^{(1)}$ and $\mathcal{Q}_k^{(2)}$ are finite collections of cubes since R is bounded, and we have

$$R \subseteq \bigcup_{Q \in \mathcal{Q}_k^{(1)} \cup \mathcal{Q}_k^{(2)}} Q.$$

See that for $1 \leqslant i \leqslant n$, the number of intervals of length $1/k$ which fit inside $[a_i, b_i]$ is at most $N_i = \lfloor k(b_i - a_i) \rfloor$; hence the number of cubes in $\mathcal{Q}_k^{(1)}$ is at most $N_1 N_2 \cdots N_n$. Thus we have

$$\sum_{Q \in \mathcal{Q}} \mathrm{vol}(Q) \leqslant \left(\prod_{i=1}^{n} N_i\right) \frac{1}{k^n} \leqslant \prod_{i=1}^{n} \frac{\lfloor k(b_i - a_i) \rfloor}{k^n} \leqslant \prod_{i=1}^{n} (b_i - a_i) = \mathrm{vol}(R). \tag{$\star$}$$

We claim that the number of cubes in $\mathcal{Q}_k^{(2)}$ is bounded above by Ck^{n-1} for some constant C depending only on R and n.

For a cube $Q \in \mathcal{Q}_k^{(2)}$, see that Q intersects the boundary of R. Since each cube $Q \in \mathcal{Q}_k^{(2)}$ has side length $1/k$, we see that there exists $1 \leqslant i \leqslant n$ such that the i-th closed interval $[m_i/k, (m_i+1)/k]$ defining Q intersects $[a_i, b_i]$ but is not contained in $[a_i, b_i]$. Thus either $m_i/k < a_i < (m_i+1)/k$ or $m_i/k < b_i < (m_i+1)/k$. In either case, there are at most 2 choices for m_i.

Say $Q \in \mathcal{Q}_k^{(2)}$ is defined by the intervals $[m_j/k, (m_j+1)/k]$ for $j = 1, \ldots, n$, and that there is an i such that $[m_i/k, (m_i+1)/k]$ intersects $[a_i, b_i]$ but is not contained in $[a_i, b_i]$. Then for each $j \neq i$, we have $a_j \leqslant (m_j+1)/k$ and $b_j \geqslant m_j/k$, so

$$m_j \in \{\lceil ka_j \rceil - 1, \lceil ka_j \rceil, \ldots, \lfloor kb_j \rfloor\}.$$

Thus for each $1 \leqslant i \leqslant n$, there are at most 2 choices for the integer m_i, and for each $j \neq i$ there are at most $\lfloor kb_j \rfloor - \lceil ka_j \rceil + 2$ choices for the integer m_j.

Thus the number of cubes in $\mathcal{Q}_k^{(2)}$ is at most

$$\sum_{i=1}^{n}\left(2\prod_{j\neq i}\left(\lfloor kb_j\rfloor - \lceil ka_j\rceil + 2\right)\right) \leqslant 2\sum_{i=1}^{n}\prod_{j\neq i}(k(b_j - a_j) + 2)$$

$$\leqslant 2k^{n-1}\sum_{i=1}^{n}\prod_{j\neq i}\left((b_j - a_j) + \frac{2}{k}\right)$$

$$\leqslant 2k^{n-1}\sum_{i=1}^{n}\prod_{j\neq i}(b_j - a_j + 2)$$

$$=: Ck^{n-1}$$

which proves the claim.

As a result, it follows that

$$\sum_{Q\in\mathcal{Q}_k^{(2)}} \mathrm{vol}(Q) \leqslant Ck^{n-1}\cdot\frac{1}{k^n} = \frac{C}{k}. \tag{$\star\star$}$$

Using $(\star)$ and $(\star\star)$, we have

$$\sum_{Q\in\mathcal{Q}_k^{(1)}\cup\mathcal{Q}_k^{(2)}} \mathrm{vol}(Q) \leqslant \mathrm{vol}(R) + \frac{C}{k}.$$

Since $k \geqslant 1$ was arbitrary, we have shown that for every error $\epsilon > 0$, there exists a cover of R by closed cubes whose total volume is at most $\mathrm{vol}(R) + \epsilon$. Taking the infimum over all such covers of R, we obtain $\mu(R) \leqslant \mathrm{vol}(R)$, finally proving that $\mu(R) = \mathrm{vol}(R)$. $\square$

Theorem 1.2.8. *The Lebesgue outer measure μ is a metric outer measure on $\mathbb{R}^n$ which is finite on compact sets.*

Proof. Step 1: We will verify the three properties of an outer measure.

First see that $\mu(\varnothing) = 0$ because the empty set can be covered by a degenerate cube of volume 0. Next, if $E \subseteq F \subseteq \mathbb{R}^n$, then any cover of F by closed cubes is also a cover of E by closed cubes, so

$$\mu(E) \leqslant \mu(F).$$

Finally, if $E = \bigcup_{j=1}^{\infty} E_j$, then for $\epsilon > 0$ and for each j we can choose a cover of E_j by closed cubes $\{Q_{j,k}\}_{k=1}^{\infty}$ such that

$$\sum_{k=1}^{\infty} \mathrm{vol}(Q_{j,k}) \leqslant \mu(E_j) + \frac{\epsilon}{2^j}$$

Then $\{Q_{j,k} : j, k \in \mathbb{Z}^+\}$ is a cover of E by closed cubes, so

$$\mu(E) \leqslant \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \text{vol}(Q_{j,k})$$

$$\leqslant \sum_{j=1}^{\infty} \left(\mu(E_j) + \frac{\epsilon}{2^j} \right)$$

$$= \sum_{j=1}^{\infty} \mu(E_j) + \epsilon.$$

Since $\epsilon > 0$ was arbitrary, we have

$$\mu(E) \leqslant \sum_{j=1}^{\infty} \mu(E_j).$$

Thus μ is an outer measure on $\mathbb{R}^n$.

Step 2: We will show that μ is finite on compact sets.

Now let $K \subseteq \mathbb{R}^n$ be compact. Then K is closed and bounded by the Heine-Borel theorem, so K is contained in some closed cube Q. Since $\mu(Q) \leqslant \text{vol}(Q) < \infty$, we have $\mu(K) \leqslant \mu(Q) < \infty$. Thus μ is finite on compact sets.

Step 3: We will show that μ is a metric outer measure. That is, if $E_1, E_2 \subseteq \mathbb{R}^n$ are such that $\text{dist}(E_1, E_2) > 0$, then

$$\mu(E_1 \cup E_2) = \mu(E_1) + \mu(E_2).$$

Let $E_1, E_2 \subseteq \mathbb{R}^n$ be such that $\text{dist}(E_1, E_2) > 0$. Then by countable subadditivity of μ, we have

$$\mu(E_1 \cup E_2) \leqslant \mu(E_1) + \mu(E_2).$$

To prove the reverse inequality, let δ be such that $0 < \delta < \text{dist}(E_1, E_2)$. Let $\epsilon > 0$ be arbitrary, and let $\{Q_j\}_{j=1}^{\infty}$ be a cover of $E_1 \cup E_2$ by closed cubes such that

$$\sum_{j=1}^{\infty} \text{vol}(Q_j) < \mu(E_1 \cup E_2) + \epsilon.$$

After subdividing the cubes if necessary, we may assume that $\text{diam}(Q_j) < \delta$ for each j. Then each cube Q_j intersects at most one of the sets E_1 or E_2, since $\text{dist}(E_1, E_2) > \delta$. Let

$$J_i := \{j \in \mathbb{N} : Q_j \cap E_i \neq \varnothing\}$$

for $i = 1, 2$. Then $J_1 \cap J_2 = \varnothing$ and we have

$$E_i \subseteq \bigcup_{j \in J_i} Q_j$$

for $i = 1, 2$. By definition of the Lebesgue outer measure μ, we have

$$
\begin{aligned}
\mu(E_1) + \mu(E_2) &\leqslant \sum_{j \in J_1} \operatorname{vol}(Q_j) + \sum_{j \in J_2} \operatorname{vol}(Q_j) \\
&= \sum_{j \in J_1 \cup J_2} \operatorname{vol}(Q_j) \\
&\leqslant \sum_{j=1}^{\infty} \operatorname{vol}(Q_j) \\
&< \mu(E_1 \cup E_2) + \epsilon.
\end{aligned}
$$

Since $\epsilon > 0$ was arbitrary, we have

$$\mu(E_1) + \mu(E_2) \leqslant \mu(E_1 \cup E_2)$$

which proves that $\mu(E_1 \cup E_2) = \mu(E_1) + \mu(E_2)$ as desired.

Since $E_1, E_2 \subseteq \mathbb{R}^n$ with $\operatorname{dist}(E_1, E_2) > 0$ were arbitrary, we have shown that μ is a metric outer measure on $\mathbb{R}^n$. $\square$

Corollary 1.2.9 (Unions of Almost Disjoint Boxes). *If $\{Q_j\}_{j=1}^{\infty}$ is a countable collection of almost disjoint boxes in $\mathbb{R}^n$, then*

$$\mu\left(\bigcup_{j=1}^{\infty} Q_j\right) = \sum_{j=1}^{\infty} \operatorname{vol}(Q_j).$$

Proof. Let $\epsilon > 0$ be arbitrary. For each $j \geqslant 1$ we let $S_{j,\epsilon}$ be a closed box contained in Q_j such that $\operatorname{vol}(Q_j) \leqslant \operatorname{vol}(S_{j,\epsilon}) + \epsilon/2^j$. Then for each $N \geqslant 1$ the collection of boxes $\{S_{j,\epsilon}\}_{j=1}^{N}$ is disjoint, so that $\operatorname{dist}(S_{i,\epsilon}, S_{j,\epsilon}) > 0$ for $i \neq j$. Since μ is a metric outer measure by 1.2.8, for each $N \geqslant 1$ we have

$$\mu\left(\bigcup_{j=1}^{N} S_{j,\epsilon}\right) = \sum_{j=1}^{N} \mu(S_{j,\epsilon}) = \sum_{j=1}^{N} \operatorname{vol}(S_{j,\epsilon}) \geqslant \sum_{j=1}^{N} \left(\operatorname{vol}(Q_j) - \frac{\epsilon}{2^j}\right) = \sum_{j=1}^{N} \operatorname{vol}(Q_j) - \epsilon$$

where the second equality follows from Lemma 1.2.7. Since $\bigcup_{j=1}^{N} S_{j,\epsilon} \subseteq \bigcup_{j=1}^{\infty} Q_j$ for each $N \geqslant 1$, monotonicity of μ gives

$$\mu\left(\bigcup_{j=1}^{\infty} Q_j\right) \geqslant \mu\left(\bigcup_{j=1}^{N} S_{j,\epsilon}\right) \geqslant \sum_{j=1}^{N} \operatorname{vol}(Q_j) - \epsilon,$$

and taking the limit as $N \to \infty$ gives

$$\mu\left(\bigcup_{j=1}^{\infty} Q_j\right) \geq \sum_{j=1}^{\infty} \text{vol}(Q_j) - \epsilon.$$

Since $\epsilon > 0$ was arbitrary, we have shown that

$$\mu\left(\bigcup_{j=1}^{\infty} Q_j\right) \geq \sum_{j=1}^{\infty} \text{vol}(Q_j).$$

The reverse inequality follows from the fact that $\{\overline{Q_j}\}_{j=1}^{\infty}$ is a cover of $\bigcup_{j=1}^{\infty} Q_j$ by closed boxes, so we have

$$\mu\left(\bigcup_{j=1}^{\infty} Q_j\right) \leq \sum_{j=1}^{\infty} \text{vol}(\overline{Q_j}) = \sum_{j=1}^{\infty} \text{vol}(Q_j).$$

$\square$

As another corollary of the fact that Lebesgue outer measure is a metric outer measure, we have the following.

Corollary 1.2.10 (Borel Regularity of Lebesgue Measure). *The Lebesgue outer measure μ is a Borel regular outer measure on $\mathbb{R}^n$, and the Borel σ-algebra is contained in the σ-algebra of μ-measurable sets. That is, every Borel set in $\mathbb{R}^n$ is Lebesgue measurable.*

Proof. Since μ is a metric outer measure by 1.2.8, it is Borel regular by 1.1.11. $\square$

Remark 1.2.11. Recall that the σ-algebra of μ-measurable sets is denoted by $\mathcal{M}_\mu$, and sets in $\mathcal{M}_\mu$ are called μ-measurable. In this case of Lebesgue outer measure μ, sets in $\mathcal{M}_\mu$ are called *Lebesgue measurable.*

Corollary 1.2.12 (Countable Disjoint Additivity). *Let $\{E_j\}_{j=1}^{\infty}$ be a countable collection of disjoint Lebesgue measurable sets in $\mathbb{R}^n$. Then*

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

Proof. This is just a special case of 1.1.6. $\square$

Instead of just boxes, we can compute the Lebesgue outer measure of more interesting sets.

Example 1.2.13 (The Cantor Set)**.** Let's look at a classical example. The *Cantor set* $C \subseteq [0, 1]$ is constructed by starting with the interval $[0, 1]$ and iteratively removing the open middle third of each remaining interval. That is, we start with $C_0 = [0, 1]$. At the first step, we remove the open interval $(1/3, 2/3)$ of length $1/3$ to obtain

$$C_1 = [0, 1/3] \cup [2/3, 1].$$

At the second step, we remove the open intervals $(1/9, 2/9)$ and $(7/9, 8/9)$, each of length $1/9$ to obtain

$$C_2 = [0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1].$$

Continuing in this way, at the n-th step we remove 2^{n-1} disjoint open intervals, each of length $1/3^n$, to obtain C_n. By induction, we see that for each $n \in \mathbb{Z}^+$, the set C_n consists of 2^n disjoint closed intervals, each of length $1/3^n$.

The **Cantor set** is defined to be

$$C := \bigcap_{n=0}^{\infty} C_n.$$

Clearly C is closed, as the intersection of closed sets. It can also be shown that C is uncountable and totally disconnected (i.e. the only connected subsets of C are singletons).

Moreover, we can compute the Lebesgue measure of C. For each $n \in \mathbb{Z}^+$, see that $C \subseteq C_n$ and thus $\mu(C) \leqslant \mu(C_n) = (2/3)^n$. Taking the limit as $n \to \infty$, we obtain $\mu(C) = 0$. Thus the Cantor set is an uncountable set of Lebesgue outer measure zero.   □

At this point, we have shown that the Lebesgue outer measure μ is a Borel regular outer measure on $\mathbb{R}^n$ which is finite on compact sets, (so it can measure all Borel sets), and μ computes the correct value on boxes. Thus, by 1.1.18, any measure which gives the same values on boxes as μ must agree with μ on all Borel sets. In particular, there can be no other measure which gives the correct value on boxes and is defined on all Borel sets.

Notation 1.2.14 (Lebesgue Measure)*.* From now on, we will refer to the Lebesgue outer measure μ simply as the *Lebesgue measure* on $\mathbb{R}^n$, and we will denote it by $\mathcal{L}^n$ instead of μ. That is, for each $E \subseteq \mathbb{R}^n$, we write $\mathcal{L}^n(E)$ to denote the Lebesgue measure of E.

We now prove the two remaining properties of Lebesgue measure we wanted: translation invariance and homogeneity.

Proposition 1.2.15 (Translation Invariance and Homogeneity)**.** *The Lebesgue measure* $\mathcal{L}^n$ *on* $\mathbb{R}^n$ *is translation invariant and homogeneous. That is, if* $E \subseteq \mathbb{R}^n$ *and* $y \in \mathbb{R}^n$, *then* $\mathcal{L}^n(E + y) = \mathcal{L}^n(E)$, *where* $E + y = \{x + y : x \in E\}$. *Similarly, if* $E \subseteq \mathbb{R}^n$ *and* $t > 0$, *then* $\mathcal{L}^n(tE) = t^n \mathcal{L}^n(E)$, *where* $tE = \{tx : x \in E\}$.

Proof. Translation Invariance: Let $E \subseteq \mathbb{R}^n$ and $y \in \mathbb{R}^n$. See that if $\{Q_j\}_{j=1}^\infty$ is a cover of E by closed cubes, then $\{Q_j + y\}_{j=1}^\infty$ is a cover of $E + y$ by closed cubes, and $\mathrm{vol}(Q_j + y) = \mathrm{vol}(Q_j)$ for each j. Conversely, if $\{S_j\}_{j=1}^\infty$ is a cover of $E + y$ by closed cubes, then $\{S_j - y\}_{j=1}^\infty$ is a cover of E by closed cubes, and $\mathrm{vol}(S_j - y) = \mathrm{vol}(S_j)$ for each j. Thus by definition of the Lebesgue measure $\mathcal{L}^n$, we have

$$\mathcal{L}^n(E + y) = \inf \left\{ \sum_{j=1}^\infty \mathrm{vol}(Q_j) : E + y \subseteq \bigcup_{j=1}^\infty Q_j, \text{ each } Q_j \text{ is a closed cube} \right\}$$

$$= \inf \left\{ \sum_{j=1}^\infty \mathrm{vol}(S_j) : E \subseteq \bigcup_{j=1}^\infty S_j, \text{ each } S_j \text{ is a closed cube} \right\}$$

$$= \mathcal{L}^n(E).$$

Since $E \subseteq \mathbb{R}^n$ and $y \in \mathbb{R}^n$ were arbitrary, we have shown that $\mathcal{L}^n$ is translation invariant.

Homogeneity: Let $E \subseteq \mathbb{R}^n$ and $t > 0$. See that if $\{Q_j\}_{j=1}^\infty$ is a cover of E by closed cubes, then $\{tQ_j\}_{j=1}^\infty$ is a cover of tE by closed cubes, and $\mathrm{vol}(tQ_j) = t^n \mathrm{vol}(Q_j)$ for each j. Conversely, if $\{S_j\}_{j=1}^\infty$ is a cover of tE by closed cubes, then $\{S_j/t\}_{j=1}^\infty$ is a cover of E by closed cubes, and $\mathrm{vol}(S_j/t) = t^{-n} \mathrm{vol}(S_j)$ for each j. Thus by definition of the Lebesgue measure $\mathcal{L}^n$, we have

$$\mathcal{L}^n(tE) = \inf \left\{ \sum_{j=1}^\infty \mathrm{vol}(Q_j) : tE \subseteq \bigcup_{j=1}^\infty Q_j, \text{ each } Q_j \text{ is a closed cube} \right\}$$

$$= \inf \left\{ \sum_{j=1}^\infty t^n \mathrm{vol}(S_j) : E \subseteq \bigcup_{j=1}^\infty S_j, \text{ each } S_j \text{ is a closed cube} \right\}$$

$$= t^n \mathcal{L}^n(E).$$

Since $E \subseteq \mathbb{R}^n$ and $t > 0$ were arbitrary, we have shown that $\mathcal{L}^n$ is homogeneous. $\qquad \square$

The proofs of rotation invariance and scaling under linear transformations are more involved, and we will not prove them until after we have developed a theory of integration.

1.3 Cantor Sets

1.3.1 The Standard Cantor Set and Fat Cantor Sets

The **standard Cantor set** C is constructed by starting with the closed interval $[0, 1]$ and iteratively removing the open middle third of each remaining interval. The resulting set C is compact, nonempty, perfect, totally disconnected, and has Lebesgue measure zero.

Exercise 1.3.1 (Cantor Set Properties)**.** Prove the Cantor set C is compact, nonempty, perfect, totally disconnected, and has uncountably many points.

Also show that C is nowhere dense in $[0, 1]$, and thus contains no intervals.

The fact that the Cantor set has Lebesgue measure zero was shown in Example 1.2.13.

Proof. By definition, the Cantor set C is the intersection of a decreasing sequence of nonempty closed sets, so C is itself a closed set. Since C is a closed subset of the compact set $[0, 1]$, it follows that C is compact.

To see that C is nonempty, notice that if $n \geqslant 0$, then 0 and 1 are in the set C_n obtained after n steps of the construction, so 0 and 1 are in C. More generally, for each $n \geqslant 0$, the endpoints of each of the 2^n intervals in C_n are in $[0, 1], C_1, \ldots, C_n$, because they have not been removed in any of the first n steps of the construction; because we are only removing the middle thirds open interval, these endpoints are never removed in any subsequent steps of the construction, so they are all in C. That is, if $[a_{n,k}, b_{n,k}]$ is an interval in C_n for some $n \geqslant 0$ and some $k \in \{1, \ldots, 2^n\}$, then $a_{n,k}, b_{n,k} \in C$. Thus we have shown that C actually contains at least countably many points.

Let E be the set of all endpoints of the intervals in C_n for all $n \geqslant 0$; that is,

$$E = \bigcup_{n=0}^{\infty} \{a_{n,k}, b_{n,k} : k = 1, \ldots, 2^n\}.$$

To show that C is perfect, let $x \in C$ and let $\epsilon > 0$. Fix $n \geqslant 0$ large enough that $3^{-n} < \epsilon$. Then $x \in C_n$, so x is in one of the 2^n closed intervals in C_n, say $I_{n,k} := [a_{n,k}, b_{n,k}]$ for some $k \in \{1, \ldots, 2^n\}$. Then the set $E \cap I_{n,k}$ contains infinitely many points because infinitely many intervals are removed from $I_{n,k}$ in subsequent steps of the construction of the Cantor set, and the end points of each of these removed intervals are in $E \cap I_{n,k}$. By choice of n, we have $I_{n,k} \subseteq (x - \epsilon, x + \epsilon)$, so $E \cap (x - \epsilon, x + \epsilon)$ contains infinitely many points. Since $\epsilon > 0$ was arbitrary, we conclude that x is a cluster point of E — recall this means each open set containing x contains infinitely many points of E. Since $x \in C$ was arbitrary, we conclude that every point in C is a cluster point of E, which means that C is perfect.

Finally to show that C is totally disconnected, let $x \in C$ and let $\epsilon > 0$. Fix $n \geqslant 0$ large enough that $3^{-n} < \epsilon$. Then $x \in C_n$, so x is in one of the 2^n closed intervals in C_n, say $I_{n,k} := [a_{n,k}, b_{n,k}]$ for some $k \in \{1, \ldots, 2^n\}$. The compliment $C_n \backslash I_{n,k}$ is a union of finitely many closed intervals, so it is closed. Thus $I_{n,k}$ is open in the relative topology on C_n, as its compliment in C_n is closed. As a result, the set $C \cap I_{n,k}$ is both open and closed in the relative topology on C. That is, $C \cap I_{n,k}$ is a clopen set in C which is contained in the open interval $(x - \epsilon, x + \epsilon)$. Since $\epsilon > 0$ and $x \in C$ were arbitrary, we conclude that C is totally disconnected.

Since C is compact, it must be complete. Since every complete, nonempty, perfect metric space is uncountable, we conclude that C is uncountable.

To see that C contains no intervals, suppose to the contrary that C contains an interval (a, b). Then $(a, b) \subseteq C_n$ for all $n \geqslant 0$. Let $N \geqslant 0$ be large enough that $3^{-N} < b - a$. Then (a, b) cannot be contained in any of the 2^N closed intervals in C_N, each of which has length 3^{-N}, a contradiction. Thus C contains no intervals.

To see that C is nowhere dense in $[0, 1]$, suppose that U is a nonempty open subset of $[0, 1]$ such that C is dense in U. That is, we have $\overline{C \cap U} \supseteq U$. Then U contains an interval (a, b), and

$$C = \overline{C} \supseteq \overline{C \cap U} \supseteq U \supseteq (a, b)$$

which contradicts the fact that C contains no intervals. Thus C is nowhere dense in $[0, 1]$. $\square$

We can do this construction in a slightly different way, to obtain a **fat Cantor set**, which is like the standard Cantor set but has positive Lebesgue measure. Let's give an example.

Example 1.3.1 (Fat Cantor Set). Start with the interval $[0, 1]$ and remove the open middle $1/4$ from it, leaving the two closed intervals $[0, 3/8]$ and $[5/8, 1]$, each of length $3/8$. Call the union of these two intervals C_1. Next, remove the open middle $1/16$ from each of these two intervals, leaving four closed intervals, each of length $3/16$. Call the union of these four intervals C_2. Continue this process indefinitely, removing the open middle $1/4^n$ from each of the 2^n intervals in C_n to form C_{n+1}. The resulting set

$$C^{\text{fat}} = \bigcap_{n=0}^{\infty} C_n$$

is a fat Cantor set. The set C^{fat} is closed as it is an intersection of closed sets, and it is totally disconnected (it contains no intervals) by construction. Moreover, the Lebesgue measure of C^{fat} is positive. To see this, note that the total length of the intervals removed at each step is

$$\sum_{n=1}^{\infty} 2^{n-1} \cdot \frac{1}{4^n} = \sum_{n=1}^{\infty} \frac{1}{2^{n+1}} = \frac{1}{2},$$

so the Lebesgue measure of C^{fat} is

$$\mathcal{L}^1(C^{\text{fat}}) = 1 - \frac{1}{2} = \frac{1}{2}.$$

In general, we can construct a fat Cantor set with any desired Lebesgue measure in the interval $(0, 1)$ by adjusting the lengths of the intervals removed at each step.

Definition 1.3.2 (Fat Cantor Set). Let $\{a_n\}_{n=1}^{\infty}$ be a sequence of positive numbers such that $\sum_{n=1}^{\infty} a_n < 1$. A **fat Cantor set** F can be constructed by starting with the interval

$F_0 = [0, 1]$, letting F_1 be the set obtained by removing the open middle interval of length a_1 from F_0, and letting F_2 be the set obtained by removing the open middle intervals of length a_2 from each of the two intervals in F_1, and so on. For $n \geqslant 1$, the set F_n is obtained by removing the open middle intervals of length a_n from each of the 2^{n-1} intervals in F_{n-1}; by induction, F_n consists of 2^n closed intervals of total length $1 - \sum_{k=1}^{n} a_k$. Then the fat Cantor set F is defined as

$$F = \bigcap_{n=0}^{\infty} F_n.$$

By construction, a fat Cantor set F is closed. Moreover, the Lebesgue measure of F is

$$\mathcal{L}^1(F) = 1 - \sum_{n=1}^{\infty} a_n,$$

which is positive by our assumption on the sequence $\{a_n\}_{n=1}^{\infty}$.

These sets are just like the standard Cantor set in many ways.

Exercise 1.3.2. Prove that a fat Cantor set F is compact, nonempty, perfect, totally disconnected, has uncountably many points, contains no intervals, and is nowhere dense in $[0, 1]$.

Proof. The set F is closed as it is an intersection of closed sets, and since F is a closed subset of the compact set $[0, 1]$, it follows that F is compact.

To see that F is nonempty, notice that if $n \geqslant 0$, then 0 and 1 are in the set F_n obtained after n steps of the construction, so 0 and 1 are in F. More generally, for each $n \geqslant 0$, the endpoints of each of the 2^n intervals in F_n are in $[0, 1], F_1, \ldots, F_n$, because they have not been removed in any of the first n steps of the construction; because we are only removing open middle intervals, these endpoints are never removed in any subsequent steps of the construction, so they are all in F. That is, if $[a_{n,k}, b_{n,k}]$ is an interval in F_n for some $n \geqslant 0$ and some $k \in \{1, \ldots, 2^n\}$, then $a_{n,k}, b_{n,k} \in F$. Thus we have shown that F actually contains at least countably many points.

We show that F is perfect, meaning every point of F is a limit point of F. Let $x \in F$ and let $\epsilon > 0$. Then for each $n \geqslant 0$ the set F_n consists of 2^n closed intervals of length $(1 - \sum_{k=1}^{n} a_k)/2^n$, and since $\sum_{k=1}^{\infty} a_k$ converging to a number less than 1, it follows that $(1 - \sum_{k=1}^{n} a_k)/2^n \to 0$ as $n \to \infty$. As a result, we can fix $N \geqslant 0$ so that $(1 - \sum_{k=1}^{N} a_k)/2^N < \epsilon$. That is, the length of each of the 2^N intervals in F_N is $< \epsilon$. Since $x \in F \subseteq F_N$, the point x lies in one of the 2^n intervals in F_n. Since this interval has length less than ϵ, both of its endpoints are in F (as shown above) and lie within ϵ of x. But this interval also has infinitely many other endpoints of intervals removed in subsequent steps of the construction of F, and these endpoints are also in F and lie within ϵ of x. In other words, the set $F \cap (x - \epsilon, x + \epsilon)$

contains infinitely many points. Since $\epsilon > 0$ was arbitrary, we conclude that x is a cluster point of F. Since $x \in F$ was arbitrary, we conclude that every point in F is a cluster point of F, which means that F is perfect.

Since F is compact, it must be complete. Since every complete, nonempty, perfect metric space is uncountable, we conclude that F is uncountable.

To see that F is totally disconnected, let $x \in F$ and let $\epsilon > 0$. Again fix $N \geqslant 0$ so that each interval in F_N has $< \epsilon$. Then $x \in F_N$, so x is in one of the 2^N closed intervals in F_N, say $I_{N,k} := [a_{N,k}, b_{N,k}]$ for some $k \in \{1, \ldots, 2^N\}$. The compliment $F_N \backslash I_{N,k}$ is a union of finitely many closed intervals, so it is closed. Thus $I_{N,k}$ is open in the relative topology on F_N, as its compliment in F_N is closed. As a result, the set $F \cap I_{N,k}$ is both open and closed in the relative topology on F. That is, $F \cap I_{N,k}$ is a clopen set in F which is contained in the open interval $(x - \epsilon, x + \epsilon)$. Since $\epsilon > 0$ and $x \in F$ were arbitrary, we conclude that F is totally disconnected.

Finally, to see that F contains no intervals, suppose to the contrary that F contains an interval (a, b). Then $(a, b) \subseteq F_n$ for all $n \geqslant 0$. Let $N \geqslant 0$ be large enough that each interval in F_N has length $< b - a$. Then (a, b) cannot be contained in any of the 2^N closed intervals in F_N, each of which has length $< b - a$, a contradiction. Thus F contains no intervals.

To see that F is nowhere dense in $[0, 1]$, suppose that U is a nonempty open subset of $[0, 1]$ such that F is dense in U. That is, we have $\overline{F \cap U} \supseteq U$. Then U contains an interval (a, b), and since F contains no intervals, this is a contradiction. Thus F is nowhere dense in $[0, 1]$. $\square$

1.3.2 The Devil's Staircase

We can also use the Cantor set to construct a pretty wild function.

Proposition 1.3.3. *There exist continuous increasing functions $f : [0, 1] \to [0, 1]$ such that $f(0) = 0$ and $f(1) = 1$, but $f'(x) = 0$ for almost every $x \in [0, 1]$.*

There are actually many such functions, but one of the most famous is the **Devil's Staircase**, which is constructed using the standard Cantor set. Another famous example which we do not explore here is the Minkowski question mark function.

Construction of the Devil's Staircase. Let $C = \bigcup_{n=0}^{\infty} C_n$ be the standard Cantor set, where C_n is the union of 2^n closed intervals of length 3^{-n} for each $n \geqslant 0$.

Define a sequence of functions $\{f_k\}_{k=0}^{\infty}$ as follows: Let $f_0(x) = x$ for all $x \in [0, 1]$. For

each $k \geqslant 1$, define a function f_k by

$$f_k(x) = \begin{cases} f_{k-1}(3x)/2 & \text{if } x \in [0, 1/3], \\ 1/2 & \text{if } x \in (1/3, 2/3), \\ 1/2 + f_{k-1}(3x-2)/2 & \text{if } x \in [2/3, 1]. \end{cases}$$

In other words, f_k is obtained from f_{k-1} by compressing the graph of f_{k-1} horizontally by a factor of 3, compressing it vertically by a factor of 2, and placing two copies of this compressed graph on the left and right thirds of the interval $[0, 1]$, with a flat segment at height $1/2$ in the middle third.

We claim that for each $k \geqslant 0$, the function f_k is continuous, increasing, $f_k(0) = 0$ and $f_k(1) = 1$, and thus maps $[0, 1]$ onto $[0, 1]$.

Proof of Claim. We can prove this by induction on k. Since

$$\lim_{x \to \frac{1}{3}^-} f_0(x) = \frac{1}{2} f_0(1) = \frac{1}{2}$$

and

$$\lim_{x \to \frac{2}{3}^+} f_0(x) = \frac{1}{2} + \frac{1}{2} f_0(0) = \frac{1}{2},$$

we see that f_1 is continuous. Since f_0 is increasing, we have $f_1(x) \leqslant f_1(y)$ for all $x, y \in [0, 1]$ with $x < y$. Moreover, $f_1(0) = 0$ and $f_1(1) = 1$, so f_1 maps $[0, 1]$ onto $[0, 1]$.

Now suppose that for some $k \geqslant 1$, the function f_k is continuous, increasing, and $f_k(0) = 0$ and $f_k(1) = 1$. Then we see that

$$\lim_{x \to \frac{1}{3}^-} f_{k+1}(x) = \frac{1}{2} f_k(1) = \frac{1}{2}$$

since $f_k(1) = 1$ by the induction hypothesis, and

$$\lim_{x \to \frac{2}{3}^+} f_{k+1}(x) = \frac{1}{2} + \frac{1}{2} f_k(0) = \frac{1}{2}$$

since $f_k(0) = 0$ by the induction hypothesis. Thus f_{k+1} is continuous. Since f_k is increasing, we have $f_{k+1}(x) \leqslant f_{k+1}(y)$ for all $x, y \in [0, 1]$ with $x < y$. Moreover, $f_{k+1}(0) = 0$ and $f_{k+1}(1) = 1$ by the induction hypothesis, so f_{k+1} maps $[0, 1]$ onto $[0, 1]$. By induction, we conclude that for each $k \geqslant 0$, the function f_k is continuous, increasing, and maps $[0, 1]$ onto $[0, 1]$. $\qquad\square$

Now we claim that the sequence of functions $\{f_k\}_{k=0}^{\infty}$ satisfies

$$|f_k(x) - f_{k-1}(x)| \leqslant 2^{-k} \quad \forall x \in [0, 1], \ k \geqslant 1.$$

Proof of Claim. To see this, again we use induction on k.

For $k = 1$, we have

$$|f_1(x) - f_0(x)| = \begin{cases} |f_0(3x)/2 - x| & \text{if } 0 \leqslant x \leqslant 1/3, \\ |1/2 - x| & \text{if } 1/3 < x < 2/3, \\ |1/2 + f_0(3x - 2)/2 - x| & \text{if } 2/3 \leqslant x \leqslant 1. \end{cases}$$

If $0 \leqslant x \leqslant \frac{1}{3}$, then

$$|f_1(x) - f_0(x)| = \left| \frac{3x}{2} - x \right| = \frac{x}{2} \leqslant \frac{1}{6} < \frac{1}{2}.$$

If $\frac{1}{3} < x < \frac{2}{3}$, then

$$|f_1(x) - f_0(x)| = |1/2 - x| \leqslant \max\{|1/2 - 1/3|, |1/2 - 2/3|\} = 1/6 < 1/2.$$

If $\frac{2}{3} \leqslant x \leqslant 1$, then

$$|f_1(x) - f_0(x)| = \left| \frac{1}{2} + \frac{3x - 2}{2} - x \right| = \left| \frac{1}{2} - \frac{x}{2} \right| \leqslant 1 - \frac{1}{2}.$$

Thus for all $x \in [0, 1]$, we have

$$|f_1(x) - f_0(x)| \leqslant \frac{1}{2} = 2^{-1}.$$

This establishes the base case.

Now suppose that for some $k \geqslant 1$, we have

$$|f_k(x) - f_{k-1}(x)| \leqslant 2^{-k} \quad \forall x \in [0, 1].$$

If $\frac{1}{3} < x < \frac{2}{3}$, then

$$|f_{k+1}(x) - f_k(x)| = \left| \frac{1}{2} - \frac{1}{2} \right| = 0$$

If $0 \leqslant x \leqslant \frac{1}{3}$, then

$$|f_{k+1}(x) - f_k(x)| = \left| \frac{1}{2} f_k(3x) - f_k(x) \right| = \frac{1}{2} \left| f_k(3x) - f_{k-1}(3x) \right| \leqslant \frac{1}{2} \cdot 2^{-k} = 2^{-(k+1)}.$$

If $\frac{1}{3} \leqslant x \leqslant \frac{2}{3}$, then

$$|f_{k+1}(x) - f_k(x)| = \left| \frac{1}{2} - \frac{1}{2} \right| = 0.$$

If $\frac{2}{3} \leqslant x \leqslant 1$, then

$$|f_{k+1}(x) - f_k(x)| = \left| \frac{1}{2} + \frac{1}{2} f_k(3x - 2) - f_k(x) \right| = \frac{1}{2} \left| f_k(3x - 2) - f_{k-1}(3x - 2) \right| \leqslant \frac{1}{2} \cdot 2^{-k} = 2^{-(k+1)}.$$

Thus for all $x \in [0, 1]$, we have

$$|f_{k+1}(x) - f_k(x)| \leqslant 2^{-(k+1)}.$$

By induction, we conclude that for all $k \geqslant 1$ and all $x \in [0, 1]$, we have

$$|f_k(x) - f_{k-1}(x)| \leqslant 2^{-k}.$$

$\square$

In other words, for each $k \geqslant 1$ we have

$$\|f_k - f_{k-1}\|_\infty = \sup_{x \in [0,1]} |f_k(x) - f_{k-1}(x)| \leqslant 2^{-k}.$$

By Exercise 1.3.3, this implies that the sequence of functions $\{f_k\}_{k=0}^\infty$ is a Cauchy sequence in $C^0([0, 1])$ with the supremum norm $\|\cdot\|_\infty$. Since $C^0([0, 1])$ is complete with respect to the supremum norm, there exists a continuous function $f : [0, 1] \to \mathbb{R}$ such that

$$\|f_k - f\|_\infty \to 0 \quad \text{as } k \to \infty.$$

In other words, the sequence $\{f_k\}_{k=0}^\infty$ converges uniformly to f on $[0, 1]$.

The function f is called the **Devil's staircase** or the **Cantor function**.

By construction, the function f is continuous. We claim that f is increasing and that $f(0) = 0$ and $f(1) = 1$, so f maps $[0, 1]$ onto $[0, 1]$. To see this, let $x, y \in [0, 1]$ with $x < y$. Since each f_k is increasing, we have $f_k(x) \leqslant f_k(y)$ for all $k \geqslant 0$. Taking the limit as $k \to \infty$, we get

$$f(x) = \lim_{k \to \infty} f_k(x) \leqslant \lim_{k \to \infty} f_k(y) = f(y).$$

Thus f is increasing. Moreover, since $f_k(0) = 0$ and $f_k(1) = 1$ for all $k \geqslant 0$, we have

$$f(0) = \lim_{k \to \infty} f_k(0) = 0 \quad \text{and} \quad f(1) = \lim_{k \to \infty} f_k(1) = 1.$$

Thus f maps $[0, 1]$ onto $[0, 1]$.

Our final claim is that the function f is differentiable at each point in $[0, 1] \backslash C$ with derivative zero, and that f is not differentiable at any point in the Cantor set C. To see this, let $x \in [0, 1] \backslash C$. Then x is in the interior of one of the intervals removed in the construction of C, say at the N-th step; for each $k \geqslant N$, the function f_k is constant on this interval, so f is also constant on this interval. Thus f is differentiable at x with derivative zero. That is, the derivative of f exists and is equal to zero at each point in $[0, 1] \backslash C$. $\square$
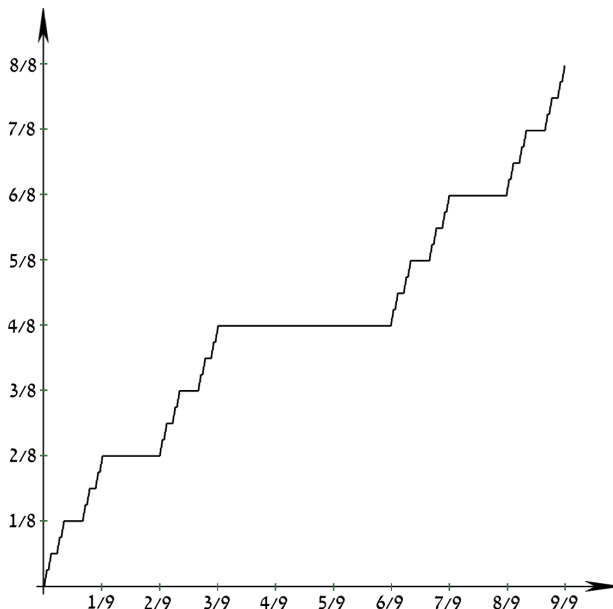
Figure 1.1: The Devil's Staircase

Exercise 1.3.3 (Sufficient Condition for Cauchy Sequence)**.** Let (X, d) be a metric space, let $\{x_k\}_{k=1}^{\infty}$ be a sequence of points in X, and let $\{r_k\}_{k=1}^{\infty}$ be a sequence of positive real numbers such that $\sum_{k=1}^{\infty} r_k < \infty$. If

$$d(x_{k+1}, x_k) \leqslant r_k \quad \forall k \in \mathbb{Z}^+,$$

then $\{x_n\}_{n=1}^{\infty}$ is a Cauchy sequence.

Proof. Let $\epsilon > 0$. Since $\sum_{k=1}^{\infty} r_k < \infty$, there exists $N \in \mathbb{Z}^+$ such that

$$\sum_{k=N+1}^{\infty} r_k < \epsilon.$$

Then for all $m \geqslant n > N$, we have

$$d(x_m, x_n) \leqslant d(x_m, x_{m-1}) + d(x_{m-1}, x_{m-2}) + \cdots + d(x_{n+1}, x_n) \leqslant \sum_{k=n}^{m-1} r_k \leqslant \sum_{k=N+1}^{\infty} r_k < \epsilon.$$

Thus $\{x_n\}_{n=1}^{\infty}$ is a Cauchy sequence. $\square$

Exercise 1.3.4 (Symmetry of the Cantor Function)**.** Show that the Cantor function f satisfies

$$f(1 - x) = 1 - f(x) \quad \forall x \in [0, 1]$$

and

$$f\left(\frac{x}{3}\right) = \frac{f(x)}{2} \quad \forall x \in [0, 1]$$

and

$$f\left(\frac{2+x}{3}\right) = \frac{1}{2} + \frac{f(x)}{2} \quad \forall x \in [0,1].$$

Proof. Let $x \in [0,1]$.

First identity: We claim that $f_k(1-x) = 1 - f_k(x)$ for all $k \geq 0$ and all $x \in [0,1]$. We prove this by induction on k.

For $k = 0$, we have

$$f_0(1-x) = 1 - x = 1 - f_0(x)$$

for each $x \in [0,1]$.

Now suppose that for some $k \geq 0$, we have $f_k(1-x) = 1 - f_k(x)$ for each $x \in [0,1]$. Then we show the same is true for f_{k+1}. We consider three cases.

If $0 \leq x \leq \frac{1}{3}$, then $1 - x \in [\frac{2}{3}, 1]$ and

$$\begin{aligned}
f_{k+1}(1-x) &= \frac{1}{2} + \frac{1}{2}f_k(3(1-x) - 2)\\
&= \frac{1}{2} + \frac{1}{2}f_k(1 - 3x)\\
&= \frac{1}{2} + \frac{1}{2}(1 - f_k(3x))\\
&= 1 - \frac{1}{2}f_k(3x)\\
&= 1 - f_{k+1}(x).
\end{aligned}$$

If $\frac{1}{3} < x < \frac{2}{3}$, then $1 - x \in (\frac{1}{3}, \frac{2}{3})$ and

$$f_{k+1}(1-x) = \frac{1}{2} = 1 - \frac{1}{2} = 1 - f_{k+1}(x).$$

If $\frac{2}{3} \leq x \leq 1$, then $1 - x \in [0, \frac{1}{3}]$ and by using the first case we have

$$f_{k+1}(x) = f_{k+1}(1 - (1-x)) = 1 - f_{k+1}(1-x)$$

which rearranges to

$$f_{k+1}(1-x) = 1 - f_{k+1}(x).$$

By induction, we conclude that for all $k \geq 0$ and all $x \in [0,1]$, we have

$$f_k(1-x) = 1 - f_k(x).$$

Taking the limit as $k \to \infty$, we get

$$f(1-x) = \lim_{k\to\infty} f_k(1-x) = \lim_{k\to\infty}(1 - f_k(x)) = 1 - f(x)$$

for each $x \in [0,1]$, so the first identity holds.

Second Identity: Suppose that $x \in [0,1]$. Then $\frac{x}{3}$ so we have

$$f\left(\frac{x}{3}\right) = \lim_{k\to\infty} f_k\left(\frac{x}{3}\right) = \lim_{k\to\infty} \frac{f_{k-1}(x)}{2} = \frac{f(x)}{2}.$$

Since $x \in [0,1]$ was arbitrary, the second identity holds.

Third Identity: Suppose that $x \in [0,1]$. Then $\frac{2+x}{3} \in [\frac{2}{3}, 1]$ so we have

$$\begin{aligned}
f\left(\frac{2+x}{3}\right) &= \lim_{k\to\infty} f_k\left(\frac{2+x}{3}\right) \\
&= \lim_{k\to\infty} \left(\frac{1}{2} + \frac{1}{2} f_{k-1}(x)\right) \\
&= \frac{1}{2} + \frac{f(x)}{2}.
\end{aligned}$$

Since $x \in [0,1]$ was arbitrary, the third identity holds.

$\square$

nonmeasurable set (appendix?)

1.4   Measure Spaces

We will use the machinery in this section to set up integration theory in the next section.

1.4.1   Measurable Spaces and Measure Spaces

Definition 1.4.1 (Measurable Space, Measure Space). A measurable space is a pair $(X, \mathcal{A})$ where X is a set and $\mathcal{A}$ is a σ-algebra of subsets of X. The sets in $\mathcal{A}$ are called *measurable* sets.

A measure space is a triple $(X, \mathcal{A}, \mu)$ where $(X, \mathcal{A})$ is a measurable space and $\mu : \mathcal{A} \to [0, \infty]$ is a measure on $\mathcal{A}$, meaining that μ satisfies the following properties:

- $\mu(\varnothing) = 0$.

- (Countable Disjoint Additivity) For any countable collection $\{A_i\}_{i=1}^{\infty}$ of pairwise disjoint sets in $\mathcal{A}$, we have

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

A measure space is called σ-finite if there exists a countable collection $\{A_i\}_{j=1}^{\infty}$ of sets in $\mathcal{A}$ such that $X = \bigcup_{j=1}^{\infty} A_j$ and $\mu(A_j) < \infty$ for all $j \in \mathbb{Z}^+$.

If X is a topological space and μ is defined on the Borel σ-algebra on X, then we say μ is a *Borel measure* on X.

Sometimes people just write (X, μ) to denote a measure space, as the σ-algebra $\mathcal{A}$ is technically the domain of the measure μ and is often clear from context. We will often do this as well, mainly because our most important examples of measure spaces will be $(X, \mathcal{M}_\mu, \mu)$ where μ is an outer measure on X and $\mathcal{M}_\mu$ is the σ-algebra of all μ-measurable sets, and so writing out the σ-algebra is redundant. If the σ-algebra is important (and sometimes it is), we will write it out explicitly.

Basically the σ-algebra $\mathcal{A}$ tells us which sets we can measure, and the measure μ assigns a size to these sets in a consistent way.

Be careful to differentiate between a Borel measure μ (which is defined only on the Borel σ-algebra) and a Borel outer measure μ (which is an outer measure such that every Borel set is μ-measurable).

Example 1.4.2 (Counting Measure). Let X be any set, and let $\mathcal{A} = 2^X$ be the power set of X. Define $\mu : \mathcal{A} \to [0, \infty]$ by

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

where $|A|$ denotes the cardinality of the set A. Then (X, μ) is a measure space and μ is called the called the *counting measure* on X.

Here is our most important example which we will use over and over again.

Example 1.4.3 (Measure induced by an Outer Measure). Let μ^* be an outer measure on a set X, and let $\mathcal{M}_\mu$ be the σ-algebra of all μ^*-measurable sets, as in Definition 1.1.3. Then $(X, \mathcal{M}_\mu, \mu)$ is a measure space, where μ is the restriction of μ^* to the σ-algebra $\mathcal{M}_\mu$. We know that μ is a measure defined on the σ-algebra $\mathcal{M}_\mu$ because it satisfies countable disjoint additivity by Proposition 1.1.6.

In particular, if X is $\mathbb{R}^n$ then the Lebesgue outer measure $\mathcal{L}^n$ is a measure on the σ-algebra of all Lebesgue measurable sets, and the Hausdorff outer measures $\mathcal{H}^\alpha$ (for $\alpha \geqslant 0$) are measures on the σ-algebra of all $\mathcal{H}^\alpha$-measurable sets.

Lemma 1.4.4 (Limits of Increasing and Decreasing Sequences of Measurable Sets). *Let (X, μ) be a measure space.*

(i) *If $\{A_j\}_{j=1}^\infty$ is an increasing sequence of measurable sets, i.e. $A_j \subseteq A_{j+1}$ for all $j \in \mathbb{Z}^+$, then*

$$\mu\left(\bigcup_{j=1}^\infty A_j \right) = \lim_{j \to \infty} \mu(A_j).$$

(ii) *If $\{A_j\}_{j=1}^{\infty}$ is a decreasing sequence of measurable sets, i.e. $A_{j+1} \subseteq A_j$ for all $j \in \mathbb{Z}^+$, and if $\mu(A_1) < \infty$, then*

$$\mu \left(\bigcap_{j=1}^{\infty} A_j \right) = \lim_{j \to \infty} \mu(A_j).$$

The proof is exactly the same as the proof of part (iii) and (iv) of Proposition 1.1.6, where this property was proved for outer measures.

We will also need the following useful result about σ-algebras.

Exercise 1.4.1 (Restriction of σ-algebra and Measure to a Measurable Set)**.** Let $(X, \mathcal{A})$ be a measurable space, and let $A \in \mathcal{A}$ be a measurable set. Then the collection $\{B \in \mathcal{A} : B \subseteq A\}$ is a σ-algebra on A.

Also if $(X, \mathcal{A}, \mu)$ is a measure space, then the restriction of μ to this σ-algebra is a measure on A.

Proof. Let $\mathcal{A}_A = \{B \in \mathcal{A} : B \subseteq A\}$. First, since $A \in \mathcal{A}$, we have $A \in \mathcal{A}_A$. Also if $B \in \mathcal{A}_A$, then $B \subseteq A$ and $B \in \mathcal{A}$, so $A \backslash B = B^c \cap A \in \mathcal{A}$ since $\mathcal{A}$ is a σ-algebra, and clearly $A \backslash B \subseteq A$ as well; hence $A \backslash B \in \mathcal{A}_A$. Finally, if $\{B_j\}_{j=1}^{\infty}$ is a countable collection of sets in $\mathcal{A}_A$, then $B_j \subseteq A$ and $B_j \in \mathcal{A}$ for all $j \in \mathbb{Z}^+$, so

$$\bigcup_{j=1}^{\infty} B_j \subseteq A$$

and

$$\bigcup_{j=1}^{\infty} B_j \in \mathcal{A}$$

since $\mathcal{A}$ is a σ-algebra; hence $\bigcup_{j=1}^{\infty} B_j \in \mathcal{A}_A$ as well. This shows that $\mathcal{A}_A$ is a σ-algebra on A.

Now if $(X, \mathcal{A}, \mu)$ is a measure space, then the restriction of μ to $\mathcal{A}_A$ is clearly a measure on $\mathcal{A}_A$ since it satisfies countable disjoint additivity by the same property for μ on $\mathcal{A}$. $\square$

1.4.2 Measurable Functions

Now that we have measurable sets, we can define measurable functions. The starting point is the following definition.

Definition 1.4.5 (Characteristic Function, Simple Function)**.** Let X be a set, and let $A \subseteq X$ be any subset. The *characteristic function* of A is the function $\mathbf{1}_A : X \to \{0, 1\}$ defined by

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

Now let (X, μ) be a measure space. A *simple function* on X is a finite linear combination of characteristic functions of sets of finite measure, i.e. a function $f : X \to \mathbb{R}$ of the form

$$f = \sum_{j=1}^{N} a_j \mathbf{1}_{A_j}$$

where $N \in \mathbb{Z}^+$, $a_j \in \mathbb{R}$ for all $j = 1, \ldots, N$, and $A_j \subseteq X$ with $\mu(A_j) < \infty$ for all $j = 1, \ldots, N$.

We now define measurable functions. We begin with the general definition, and then we will specialize to the most important cases.

Definition 1.4.6 (Measurable Function). Let $(X, \mathcal{A})$ be a measurable space, and let Y be a topological space. A map $f : X \to Y$ is called $\mathcal{A}$-*measurable* if for every open set $V \subseteq Y$, the preimage $f^{-1}(V)$ is in $\mathcal{A}$.

If (X, μ) is a measure space, we say f is μ-*measurable* if it is measurable with respect to the σ-algebra $\mathcal{A}$ which is the domain of μ.

Remark 1.4.7. Okay, we now need to note something to beware of. Let X be a topological space and let μ be a Borel outer measure on X. That is, μ is an outer measure on X such that every Borel set is μ-measurable, as in Definition 1.1.3. That means $\mathcal{B}_X \subseteq \mathcal{M}_\mu$, where $\mathcal{B}_X$ is the Borel σ-algebra on X and $\mathcal{M}_\mu$ is the σ-algebra of all μ-measurable sets.

Now let Y be another topological space, and let $f : X \to Y$ be a map. There is a difference between the following two statements:

(a) the map f is μ-measurable, i.e. for every open set $V \subseteq Y$, the preimage $f^{-1}(V)$ is in $\mathcal{M}_\mu$,

(b) the map f is Borel measurable, i.e. for every open set $V \subseteq Y$, the preimage $f^{-1}(V)$ is in $\mathcal{B}_X$.

Since $\mathcal{B}_X \subseteq \mathcal{M}_\mu$, (b) implies (a), but (a) does not necessarily imply (b). This is important to keep in mind, as many authors drop the σ-algebra from the notation and just say "f is measurable" without specifying which σ-algebra they are using. The property in (a) is more general, and what we will use most often, but the property in (b) is nice because then compositions of Borel measurable maps are Borel measurable, which is not necessarily true for maps satisfying (a).

We will do our best to be careful, and write "f is Borel measurable" when we mean (b) and "f is μ-measurable" when we mean (a).

Exercise 1.4.2. Let $(X, \mathcal{A})$ be a measurable space, and let Y be a topological space. Let $f : X \to Y$ be an $\mathcal{A}$-measurable map. Show that for every Borel set $B \subseteq Y$, the preimage $f^{-1}(B)$ is in $\mathcal{A}$.

Proof. Let $\mathcal{B}_Y$ be the Borel σ-algebra on Y, and let

$$\mathcal{C} = \{B \in \mathcal{B}_Y : f^{-1}(B) \in \mathcal{A}\}.$$

We will show that $\mathcal{C}$ is a σ-algebra containing all open sets in Y. Since $\mathcal{B}_Y$ is the smallest σ-algebra containing all open sets, this will imply that $\mathcal{B}_Y \subseteq \mathcal{C}$, which is what we want.

First, since $f^{-1}(Y) = X \in \mathcal{A}$, we have $Y \in \mathcal{C}$. Also if $B \in \mathcal{C}$, then $f^{-1}(B) \in \mathcal{A}$, so $(f^{-1}(B))^c \in \mathcal{A}$ as well since $\mathcal{A}$ is a σ-algebra; hence

$$(f^{-1}(B))^c = X \backslash f^{-1}(B) = f^{-1}(Y \backslash B) = f^{-1}(B^c) \in \mathcal{A}.$$

Thus $B^c \in \mathcal{C}$ as well. This shows that $\mathcal{C}$ is closed under complements; we have established the first two properties of a σ-algebra.

Finally, if $\{B_j\}_{j=1}^{\infty}$ is a countable collection of sets in $\mathcal{C}$, then $f^{-1}(B_j) \in \mathcal{A}$ for all $j \in \mathbb{Z}^+$, so

$$\bigcup_{j=1}^{\infty} f^{-1}(B_j) = f^{-1}\left(\bigcup_{j=1}^{\infty} B_j\right) \in \mathcal{A}$$

since $\mathcal{A}$ is a σ-algebra; hence $\bigcup_{j=1}^{\infty} B_j \in \mathcal{C}$ as well. This shows that $\mathcal{C}$ is closed under countable unions, and so $\mathcal{C}$ is a σ-algebra as desired. $\qquad\square$

Exercise 1.4.3 (Continuous Functions are Borel Measurable)**.** Let X and Y be topological spaces, and let $f : X \to Y$ be continuous. Show that f is Borel measurable.

Proof. Since f is continuous, for every open set $V \subseteq Y$, the preimage $f^{-1}(V)$ is open in X. Since every open set in X is a Borel set, we see that $f^{-1}(V)$ is a Borel set in X as well. This shows that f is Borel measurable as desired. $\qquad\square$

Now we specialize to the most important case for us.

Remark 1.4.8. The most important case for us will be when Y is either $[0, \infty]$, $[-\infty, \infty]$, $\mathbb{C}$ or $\mathbb{R}^n$ with the usual topology. In the next proposition, we prove some useful properties of measurable functions in these cases.

To be precise, we need to specify the topology on $[-\infty, \infty]$. A set $U \subseteq [-\infty, \infty]$ is defined to be open if $U \cap \mathbb{R}$ is open in $\mathbb{R}$. (It is easy to see this defines a topology on $[-\infty, \infty]$ - just distribute the $\cap$ over unions and finite intersections.) With this topology, the real line $\mathbb{R}$ is an open subset of $[-\infty, \infty]$ and has the subspace topology.

At this point, we focus on real-valued functions defined on X, but which are allowed to take on the values $\pm\infty$ as well. Thus a function $f : X \to [-\infty, \infty]$ is allowed, and takes on values

$$-\infty \leqslant f(x) \leqslant \infty, \qquad x \in X$$

in the extended real number line. We say that f is *finite valued* if $f(x) \in \mathbb{R}$ for all $x \in X$. It will be convenient for us to think of functions $X \to \mathbb{R}$ as functions $X \to [-\infty, \infty]$ which are finite valued. In applications, we almost always are in a situation where a function takes on infinite values on a set of measure zero.

1.4.3 Digression about Borel vs. Lebesgue

Remark 1.4.9. In the case that $Y = \mathbb{R}$ or $Y = [-\infty, \infty]$, we could have taken the σ-algebra of *Lebesgue* measurable sets instead of the Borel σ-algebra. You should ask: why didn't we?

Let's explore, and be precise. Let μ be the Lebesgue outer measure on $\mathbb{R}$, which we know is a Borel regular outer measure. Let $\mathcal{M}$ be the σ-algebra of all Lebesgue measurable sets, and let $\mathcal{B}$ be the Borel σ-algebra on $\mathbb{R}^n$. Then $\mathcal{B} \subseteq \mathcal{M}$. The difference is that $\mathcal{M}$ contains all subsets of measure zero, while $\mathcal{B}$ does not.

Let $(X, \mathcal{A})$ be a measurable space. Let's say a function $f : X \to \mathbb{R}$ is $\mathcal{A}$-*Borel measurable* if for every Borel set $B \subseteq \mathbb{R}$, the preimage $f^{-1}(B)$ is in $\mathcal{A}$. (This is the usual definition of measurable function.)

Let's say f is $\mathcal{A}$-*Lebesgue measurable* if for every Lebesgue measurable set $L \subseteq \mathbb{R}$, the preimage $f^{-1}(L)$ is in $\mathcal{A}$. Since $\mathcal{B} \subseteq \mathcal{M}$, we see that if f is $\mathcal{A}$-Lebesgue measurable, then it is also $\mathcal{A}$-Borel measurable. But the converse is not necessarily true, and we will prove this in a moment.

Thus for a function $f : \mathbb{R}^n \to \mathbb{R}$, there are four different notions of measurability:

(i) f is Borel-Borel measurable: for every Borel set $B \subseteq \mathbb{R}$, the preimage $f^{-1}(B)$ is a Borel set,

(ii) f is Borel-Lebesgue measurable: for every Lebesgue measurable set $L \subseteq \mathbb{R}$, the preimage $f^{-1}(L)$ is a Borel set,

(iii) f is Lebesgue-Borel measurable: for every Borel set $B \subseteq \mathbb{R}$, the preimage $f^{-1}(B)$ is a Lebesgue measurable set,

(iv) f is Lebesgue-Lebesgue measurable: for every Lebesgue measurable set $L \subseteq \mathbb{R}$, the preimage $f^{-1}(L)$ is a Lebesgue measurable set.

Since every Borel set is Lebesgue measurable, we see that (i) implies (iii). Also (ii) is useless, since the existance of Lebesgue measurable sets which are not Borel measurable implies that the identity map $\mathrm{id} : \mathbb{R} \to \mathbb{R}$ is not Lebesgue-Borel measurable.

Also the existence of non-Lebesgue measurable sets implies that that there are continuous functions for which (iv) does not hold. Let C be the Cantor set; let C^{fat} be a fat Cantor set, i.e. a closed set with positive Lebesgue measure and empty interior; and let $f : \mathbb{R} \to \mathbb{R}$

be a continuous function such that $f\left(C^{\text{fat}}\right) = C$ (that such a function exists is a classical result in topology). Let $\mathcal{N}$ be a non-Lebesgue measurable subset of C^{fat} (which exists by the Vitali construction). Then $f(\mathcal{N}) \subseteq C$ so $f(\mathcal{N})$ has Lebesgue outer measure zero, and hence is Lebesgue measurable. However $f^{-1}(f(\mathcal{N})) \supseteq \mathcal{N}$ is not Lebesgue measurable. This shows that (i) does not imply (iv).

In summary, (i) implies (iii), but neither (i) nor (iii) implies (iv) or vice versa; and (ii) is useless.

Okay, so I hope you see why we would prefer the definitions (i) and (iii) over (ii) and (iv). But how do we choose between Borel measurable functions (i) and Lebesgue measurable functions (iii)? When you are in $\mathbb{R}^n$ and say "measurable set", do you mean Borel measurable or Lebesgue measurable? Well that's a matter of personal taste.

I much prefer Lebesgue measurable sets, as all sets of measure zero are Lebesgue measurable, and this makes life easier. However, there are people who prefer Borel measurable sets, such as Axler. He writes in *Measure, Integration & Real Analysis* that "Although there exist Lebesgue measurable sets that are not Borel sets, you are unlikely to encounter one. Similarly, a Lebesgue measurable function that is not Borel measurable is unlikely to arise in anything you do."

We will try to be precise and say "Borel measurable" or "Lebesgue measurable" when talking about sets or functions, but let it be stated here that we prefer the σ-algebra of Lebesgue measurable sets over the Borel σ-algebra.

1.4.4   Properties of Measurable Functions

Proposition 1.4.10 (Equivalent Definitions of Measurable Function)**.** *Let* $(X, \mathcal{A})$ *be a measurable space, and let* $f : X \to [-\infty, \infty]$ *be a function. Then the following are equivalent:*

(i) f *is* $\mathcal{A}$-*measurable,*

(ii) *for each* $a \in \mathbb{R}$, *the set* $f^{-1}([-\infty, a))$ *is in* $\mathcal{A}$,

(iii) *for each* $a \in \mathbb{R}$, *the set* $f^{-1}([-\infty, a])$ *is in* $\mathcal{A}$,

(iv) *for each* $a \in \mathbb{R}$, *the set* $f^{-1}((a, \infty])$ *is in* $\mathcal{A}$,

(v) *for each* $a \in \mathbb{R}$, *the set* $f^{-1}([a, \infty])$ *is in* $\mathcal{A}$,

(vi) *for each interval* $I \subseteq \mathbb{R}$, *the set* $f^{-1}(I)$ *is in* $\mathcal{A}$,

Proof. (i) $\iff$ (ii). Suppose f is $\mathcal{A}$-measurable, and let $a \in \mathbb{R}$. Then $[-\infty, a) \subseteq [-\infty, \infty]$ is open, so $f^{-1}([-\infty, a)) \in \mathcal{A}$ by definition.

Conversely, suppose that for every $a \in \mathbb{R}$, the set $f^{-1}([-\infty, a))$ is in $\mathcal{A}$. We claim that the set

$$\mathcal{A}' := \{A \subseteq [-\infty, \infty] : f^{-1}(A) \in \mathcal{A}\}$$

is a σ-algebra containing all open sets in $[-\infty, \infty]$. Since $f^{-1}([-\infty, \infty]) = X \in \mathcal{A}$, we see that $[-\infty, \infty] \in \mathcal{A}'$. Similarly, we see that if $A \in \mathcal{A}'$, then $f^{-1}(A) \in \mathcal{A}$, so $(f^{-1}(A))^c = f^{-1}(A^c) \in \mathcal{A}$ as well, and hence $A^c \in \mathcal{A}'$. Finally, if $\{A_j\}_{j=1}^{\infty}$ is a countable collection of sets in $\mathcal{A}'$, then $f^{-1}(A_j) \in \mathcal{A}$ for all $j \in \mathbb{Z}^+$, so

$$\bigcup_{j=1}^{\infty} f^{-1}(A_j) = f^{-1}\left(\bigcup_{j=1}^{\infty} A_j\right) \in \mathcal{A}$$

since $\mathcal{A}$ is a σ-algebra; hence $\bigcup_{j=1}^{\infty} A_j \in \mathcal{A}'$ as well. This shows that $\mathcal{A}'$ is a σ-algebra.

Now each open set $U \subseteq [-\infty, \infty]$ can be written as a countable union of basis elements, and each basis element is either of the form (a, b), $[-\infty, b)$, or $(a, \infty]$ for some $a, b \in \mathbb{R}$. Hence $\mathcal{A}'$ is a σ-algebra containing all open sets in $[-\infty, \infty]$. Since the Borel σ-algebra on $[-\infty, \infty]$ is the smallest σ-algebra containing all open sets, we have $\mathcal{B}_{[-\infty, \infty]} \subseteq \mathcal{A}'$. Thus for every open set $U \subseteq [-\infty, \infty]$, we have $f^{-1}(U) \in \mathcal{A}$, and so f is $\mathcal{A}$-measurable.

(ii) $\iff$ (iii). Suppose that for every $a \in \mathbb{R}$, the set $f^{-1}([-\infty, a))$ is in $\mathcal{A}$. Then for each $a \in \mathbb{R}$, we have

$$[-\infty, a] = \bigcap_{j=1}^{\infty} \left[-\infty, a + \frac{1}{j}\right)$$

which is a countable intersection of open sets, so

$$f^{-1}([-\infty, a]) = \bigcap_{j=1}^{\infty} f^{-1}\left(\left[-\infty, a + \frac{1}{j}\right)\right) \in \mathcal{A}$$

since $\mathcal{A}$ is a σ-algebra.

Conversely, suppose that for every $a \in \mathbb{R}$, the set $f^{-1}([-\infty, a])$ is in $\mathcal{A}$. Then for each $a \in (-\infty, \infty)$, we have

$$[-\infty, a) = \bigcup_{j=1}^{\infty}\left[-\infty, a - \frac{1}{j}\right]$$

which is a countable union of closed sets, so

$$f^{-1}([-\infty, a)) = \bigcup_{j=1}^{\infty} f^{-1}\left(\left[-\infty, a - \frac{1}{j}\right]\right) \in \mathcal{A}$$

since $\mathcal{A}$ is a σ-algebra.

This shows that (ii) and (iii) are equivalent.

(iii) $\iff$ (iv). Note that for each $a \in \mathbb{R}$, we have

$$(a, \infty] = [-\infty, a]^c$$

so

$$f^{-1}((a, \infty]) = f^{-1}([-\infty, a]^c) = (f^{-1}([-\infty, a]))^c.$$

Since $\mathcal{A}$ is a σ-algebra, this shows that (iii) and (iv) are equivalent.

(ii) $\iff$ (v). Note that for each $a \in \mathbb{R}$, we have

$$[a, \infty] = [-\infty, a)^c$$

so

$$f^{-1}([a, \infty]) = f^{-1}([-\infty, a)^c) = (f^{-1}([-\infty, a)))^c.$$

Since $\mathcal{A}$ is a σ-algebra, this shows that (ii) and (v) are equivalent.

To finish, we show that (i) $\iff$ (vi). Suppose f is $\mathcal{A}$-measurable, and let $I \subseteq \mathbb{R}$ be an interval. Then I is one of the following forms:

$$(-\infty, a), (-\infty, a], [a, b), (a, b), (a, b], [a, b], [a, \infty), \text{ or } [a, \infty)$$

for some $a, b \in \mathbb{R}$ with $a < b$. In the case that I is of the form $(-\infty, a)$ or (a, b) or (a, ∞), then I is open in $[-\infty, \infty]$, so $f^{-1}(I) \in \mathcal{A}$ by definition and the fact that open sets in $\mathbb{R}$ are open in $[-\infty, \infty]$ as well. If I is of the form $(-\infty, a]$ or $[a, b]$ or $[a, \infty)$, then I is closed in

$[-\infty, \infty]$, so $f^{-1}(I) \in \mathcal{A}$ as well since $f^{-1}(I) = (f^{-1}(I^c))^c$ and I^c is open. Finally, if I is of the form $[a, b)$ or $(a, b]$, then we can write

$$[a, b) = [-\infty, b) \cap [a, \infty] \quad \text{and} \quad (a, b] = (-\infty, b] \cap [a, \infty)$$

so $f^{-1}(I) \in \mathcal{A}$ as well since $\mathcal{A}$ is closed under finite intersections. This shows that for every interval $I \subseteq \mathbb{R}$, the set $f^{-1}(I)$ is in $\mathcal{A}$.

Conversely, suppose that for every interval $I \subseteq \mathbb{R}$, the set $f^{-1}(I)$ is in $\mathcal{A}$. Then we have

$$f^{-1}(\mathbb{R}) = \bigcup_{\substack{a < b \\ a, b \in \mathbb{Q}}} f^{-1}([a, b]) \in \mathcal{A}$$

since $\mathbb{R} = \bigcup_{a < b, a, b \in \mathbb{Q}} [a, b]$ and $\mathcal{A}$ is closed under countable unions. Since $\mathcal{A}$ is also closed under complements, we have $f^{-1}(\{\pm\infty\}) = f^{-1}(\mathbb{R}^c) = (f^{-1}(\mathbb{R}))^c \in \mathcal{A}$.

Then for each open set $U \subseteq [-\infty, \infty]$, we can write U as

$$U = (U \cap \{\pm\infty\}) \cup (U \cap \mathbb{R})$$

where $U \cap \mathbb{R}$ is an open set in $\mathbb{R}$ and hence can be written as a countable union of open intervals in $\mathbb{R}$. Since $f^{-1}(\{\pm\infty\}) \in \mathcal{A}$ and $f^{-1}(U \cap \mathbb{R})$ is a countable union of sets of the form $f^{-1}(I)$ where I is an open interval in $\mathbb{R}$, we have $f^{-1}(U) \in \mathcal{A}$ as well. This shows that f is $\mathcal{A}$-measurable as desired. $\square$

Exercise 1.4.4. Let $(X, \mathcal{A})$ be a measurable space.

   I. Show that if $f : X \to \mathbb{R}^m$ and $g : X \to \mathbb{R}^k$ are $\mathcal{A}$-measurable, then the map $(f, g) : X \to \mathbb{R}^{m+k}$ defined by $(f, g)(x) = (f(x), g(x))$ is also $\mathcal{A}$-measurable.

   II. If $A \in \mathcal{A}$ is a measurable set, then its characteristic function $\mathbf{1}_A : X \to \mathbb{R}$ is $\mathcal{A}$-measurable.

Proof. (I) With the product topology on $\mathbb{R}^{m+k}$, a basis for the topology is given by all products $U \times V$ where $U \subseteq \mathbb{R}^m$ and $V \subseteq \mathbb{R}^k$ are open. Let $h = (f, g) : X \to \mathbb{R}^{m+k}$ be defined by $h(x) = (f(x), g(x))$. Then a nearly identical argument as in (I) shows that

$$\mathcal{A}' := \{A \subseteq \mathbb{R}^{m+k} : h^{-1}(A) \in \mathcal{A}\}$$

is a σ-algebra containing all open sets in $\mathbb{R}^{m+k}$. Since $h^{-1}(U \times V) = f^{-1}(U) \cap g^{-1}(V)$, and $f^{-1}(U), g^{-1}(V) \in \mathcal{A}$ since f, g are $\mathcal{A}$-measurable, we have $h^{-1}(U \times V) \in \mathcal{A}$ as well. Thus $\mathcal{A}'$ contains all basis elements for the topology on $\mathbb{R}^{m+k}$, and hence contains all open sets in $\mathbb{R}^{m+k}$. This shows that h is $\mathcal{A}$-measurable as desired.

(II) Let $A \in \mathcal{A}$ be a measurable set, and let $\mathbf{1}_A : X \to \mathbb{R}$ be its characteristic function. Then each we see that if $a < 0$, then $\mathbf{1}_A^{-1}([-\infty, a)) = \varnothing \in \mathcal{A}$, and if $0 \leqslant a < 1$, then $\mathbf{1}_A^{-1}([-\infty, a)) = A^c \in \mathcal{A}$, and if $a \geqslant 1$, then $\mathbf{1}_A^{-1}([-\infty, a)) = X \in \mathcal{A}$. By the equivalence of (i) and (ii) in Proposition 1.4.10, this shows that $\mathbf{1}_A$ is $\mathcal{A}$-measurable as desired. $\square$

Proposition 1.4.11 (Properties of Measurable Functions)**.** *Let* $(X, \mathcal{A})$ *be a measurable space.*

(i) *If* f, g *are* $\mathcal{A}$*-measurable functions* $X \to [-\infty, \infty]$*, then*

$$f + g, fg, |f|, \min(f, g), \max(f, g)$$

are all $\mathcal{A}$*-measurable as well, where the operations are defined pointwise. If* g *is nowhere zero, then* f/g *is* $\mathcal{A}$*-measurable as well.*

(ii) *If* $\{f_j\}_{j=1}^{\infty}$ *is a countable collection of* $\mathcal{A}$*-measurable functions* $X \to [-\infty, \infty]$*, then*

$$\sup_{j \geqslant 1} f_j, \inf_{j \geqslant 1} f_j, \limsup_{j \to \infty} f_j, \liminf_{j \to \infty} f_j$$

are all $\mathcal{A}$*-measurable as well.*

Proof. (i) Let $f, g : X \to [-\infty, \infty]$ be $\mathcal{A}$-measurable. As noted in Proposition 1.4.10 (ii), it suffices to show that for every $a \in \mathbb{R}$, the set $\{x \in X : (f + g)(x) < a\}$ is in $\mathcal{A}$, and similarly for the other functions.

Let $a \in \mathbb{R}$. Then we see that

$$(f + g)^{-1}([-\infty, a)) = \bigcup_{\substack{r, s \in \mathbb{Q} \\ r + s < a}} (f^{-1}([-\infty, r)) \cap g^{-1}([-\infty, s)))$$

which is a countable union of sets in $\mathcal{A}$ since f, g are $\mathcal{A}$-measurable, and hence is in $\mathcal{A}$ as well. Therefore $f + g$ is $\mathcal{A}$-measurable.

Since

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

which is in $\mathcal{A}$ since f is $\mathcal{A}$-measurable. Therefore f^2 is $\mathcal{A}$-measurable. As a result of the identity

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

and the fact that sums and squares of $\mathcal{A}$-measurable functions are $\mathcal{A}$-measurable, we see that fg is $\mathcal{A}$-measurable as well.

Next observe that

$$\left(\frac{1}{g}\right)^{-1}([-\infty, a)) = \begin{cases} g^{-1}((1/a, 0]) & \text{if } a < 0 \\ g^{-1}((-\infty, 0]) & \text{if } a = 0 \\ g^{-1}((-\infty, 0]) \cup g^{-1}((1/a, \infty)) & \text{if } a > 0 \end{cases}$$

which is in $\mathcal{A}$ since g is $\mathcal{A}$-measurable; hence $1/g$ is $\mathcal{A}$-measurable. By the previous paragraph, we see that f/g is $\mathcal{A}$-measurable as well.

At this point note that $\{x \in X : f(x) \geq 0\}$ is in $\mathcal{A}$ since it is equal to $f^{-1}([0, \infty])$, and hence $\mathbf{1}_{\{f \geq 0\}}$ is $\mathcal{A}$-measurable by Exercise 1.4.4(III). Then

$$f^{+} := f \cdot \mathbf{1}_{\{f \geq 0\}} = \max(f, 0)$$

is $\mathcal{A}$-measurable as the product of two $\mathcal{A}$-measurable functions, and similarly

$$f^{-} := -f \cdot \mathbf{1}_{\{f < 0\}} = \max(-f, 0)$$

is $\mathcal{A}$-measurable as well. Since $|f| = f^{+} + f^{-}$ we see that $|f|$ is $\mathcal{A}$-measurable. Since

$$\max(f, g) = (f - g)^{+} + g \quad \text{and} \quad \min(f, g) = -(f - g)^{-} + g$$

we see that $\max(f, g)$ and $\min(f, g)$ are also $\mathcal{A}$-measurable. This completes the proof of part (i).

Suppose $\{f_j\}_{j=1}^{\infty}$ is a sequence of $\mathcal{A}$-measurable functions $X \to [-\infty, \infty]$. Then for every $a \in \mathbb{R}$, we have

$$\left(\inf_{j \geq 1} f_j \right)^{-1} ([-\infty, a]) = \bigcup_{j \geq 1} \left(f_j^{-1}([-\infty, a)) \right)$$

and

$$\left(\sup_{j \geq 1} f_j \right)^{-1} ([-\infty, a]) = \bigcap_{j \geq 1} \left(f_j^{-1}([-\infty, a]) \right)$$

which are both in the σ-algebra $\mathcal{A}$ since each f_j is $\mathcal{A}$-measurable. This shows that $\inf_{j \geq 1} f_j$ and $\sup_{j \geq 1} f_j$ are $\mathcal{A}$-measurable functions.

We complete the proof by noting that

$$\liminf_{j \to \infty} f_j = \sup_{j \geq 1} \inf_{k \geq j} f_k \quad \text{and} \quad \limsup_{j \to \infty} f_j = \inf_{j \geq 1} \sup_{k \geq j} f_k$$

which shows that $\liminf_{j \to \infty} f_j$ and $\limsup_{j \to \infty} f_j$ are both $\mathcal{A}$-measurable as well. $\square$

Exercise 1.4.5 (Derivatives are Measurable)**.** Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ be a differentiable function defined on an open set U. Show that the derivative $Df : U \to \mathbb{R}^{m \times n}$ is Borel measurable.

Proof. Let $f = (f_1, f_2, \ldots, f_m)$ where each $f_k : U \to \mathbb{R}$ is differentiable. See that

$$Df(x) = \begin{pmatrix} \partial_1 f_1(x) & \partial_2 f_1(x) & \cdots & \partial_n f_1(x) \\ \partial_1 f_2(x) & \partial_2 f_2(x) & \cdots & \partial_n f_2(x) \\ \vdots & \vdots & \ddots & \vdots \\ \partial_1 f_m(x) & \partial_2 f_m(x) & \cdots & \partial_n f_m(x) \end{pmatrix}$$

for each $x \in U$. By Proposition 1.4.11(ii), it suffices to show that each partial derivative $\partial_i f_j : U \to \mathbb{R}$ is Borel measurable.

Fix $1 \leqslant j \leqslant n$ and $1 \leqslant k \leqslant m$. Since f_k is differentiable, we have

$$\partial_j f_k(x) = \lim_{h \to 0} \frac{f_k(x + he_j) - f_k(x)}{h}, \quad \forall x \in U$$

where e_j is the j-th standard basis vector in $\mathbb{R}^n$. Choose a sequence $h_l \to 0$ as $l \to \infty$ with $h_l \neq 0$ for all $l \in \mathbb{Z}^+$. Then

$$\partial_j f_k(x) = \lim_{l \to \infty} \frac{f_k(x + h_l e_j) - f_k(x)}{h_l}$$

for each $x \in U$. This expresses the partial derivative $\partial_j f_k$ as a pointwise limit of Borel measurable functions — see that

$$\partial_j f_k = \lim_{l \to \infty} \frac{f_k(\cdot + h_l e_j) - f_k}{h_l}$$

where each function in the sequence is Borel measurable since it is a sum and scalar multiple of a continuous function; the limit of Borel measurable functions is Borel measurable by Proposition 1.4.11(ii). This shows that $\partial_j f_k$ is Borel measurable as desired. $\qquad\square$

Exercise 1.4.6. Let (X, μ) be a measure space. If $f : X \to [-\infty, \infty]$ is μ-measurable then the function

$$[0, \infty) \ni t \longmapsto \mu(\{x \in X : f(x) > t\})$$

is an increasing function.

Proof. Suppose f is μ-measurable, and let $0 \leqslant s < t < \infty$. Then $\{x \in X : f(x) > t\} \subseteq \{x \in X : f(x) > s\}$ and these sets are both μ-measurable since f is μ-measurable. Thus

$$\mu(\{x \in X : f(x) > s\}) \leqslant \mu(\{x \in X : f(x) > t\})$$

which shows that the function $t \mapsto \mu(\{x \in X : f(x) > t\})$ is increasing. $\qquad\square$

Exercise 1.4.7 (Upper and Lower Semicontinuous Functions are Measurable)**.** Let (X, d) be a metric space, and let $f : X \to [-\infty, \infty]$ be a function. Recall that f is lower semicontinuous if for every $x \in X$ and every $\epsilon > 0$, there exists $\delta > 0$ such that

$$f(y) > f(x) - \epsilon \quad \text{for all } y \in B(x, \delta).$$

Similarly, f is upper semicontinuous if for every $x \in X$ and every $\epsilon > 0$, there exists $\delta > 0$ such that

$$f(y) < f(x) + \epsilon \quad \text{for all } y \in B(x, \delta).$$

Show that if f is lower semicontinuous or upper semicontinuous, then f is Borel measurable.

Proof. Suppose f is lower semicontinuous, and let $a \in \mathbb{R}$. Then for each $x \in f^{-1}((a, \infty])$, we have $f(x) > a$. By the definition of lower semicontinuity, there exists $\delta > 0$ such that $f(y) > a$ for all $y \in B(x, \delta)$. This shows that $B(x, \delta) \subseteq f^{-1}((a, \infty])$, and hence $f^{-1}((a, \infty])$ is open in X as it is a union of open balls. Since $f^{-1}((a, \infty])$ is open for each $a \in \mathbb{R}$, we see that f is Borel measurable by Proposition 1.4.10(v).

A similar argument shows that if f is upper semicontinuous, then for each $a \in \mathbb{R}$, the set $f^{-1}([-\infty, a))$ is open in X, and hence f is Borel measurable by Proposition 1.4.10(ii).   □

Let's look at one more equivalent characterization of measurable functions into $[-\infty, \infty]$.

Proposition 1.4.12. *Let (X, μ) be a measure space, and let $f : X \to [-\infty, \infty]$ be a function. Then f is μ-measurable if and only if there exists a sequence of simple functions $\{s_j\}_{j=1}^{\infty}$ such that $|s_j(x)| \leqslant |s_{j+1}(x)|$ and $\lim_{j \to \infty} s_j(x) = f(x)$ for each $x \in X$. If f is bounded, then we can choose the sequence $\{s_j\}_{j=1}^{\infty}$ to converge uniformly to f.*

Proof. One direction is clear. Simple functions are μ-measurable since they are finite linear combinations of characteristic functions of measurable sets — each characteristic function is μ-measurable by Exercise 1.4.4(II), and finite linear combinations of μ-measurable functions are μ-measurable by Proposition 1.4.11(i). Thus if $\{s_j\}_{j=1}^{\infty}$ is a sequence of simple functions converging pointwise to f, then each s_j is μ-measurable, and hence f is μ-measurable by Proposition 1.4.11(ii).

Conversely, suppose f is μ-measurable. The idea here is that for each $k \in \mathbb{Z}^+$ and each $n \in \mathbb{Z}$, we can divide the interval $[n, n + 1)$ into 2^k half-open subintervals of length 2^{-k}. Specifically, for each $k \in \mathbb{Z}^+$ we define a collection of intervals

$$\left\{ I_{m,k} := \left[\frac{m}{2^k}, \frac{m + 1}{2^k} \right) : m \in \mathbb{Z} \right\}$$

which partitions the real line $\mathbb{R}$ into half-open intervals of length 2^{-k}. Since f is μ-measurable, we see that for each $k \in \mathbb{Z}^+$ and $m \in \mathbb{Z}$, the set $f^{-1}(I_{m,k})$ is μ-measurable.

For each $k \in \mathbb{Z}^+$ we define a simple function $s_k : X \to \mathbb{R}$ by

$$s_k(x) := \begin{cases} \frac{m}{2^k} & \text{if } f(x) \in [0, k] \text{ and } m \in \mathbb{Z} \text{ is such that } f(x) \in I_{m,k}, \\ \frac{m+1}{2^k} & \text{if } f(x) \in [-k, 0) \text{ and } m \in \mathbb{Z} \text{ is such that } f(x) \in I_{m,k}, \\ k & \text{if } f(x) > k, \\ -k & \text{if } f(x) < -k. \end{cases}$$

Note that in the first case, we have $0 \leqslant m \leqslant k2^k - 1$ so that

$$0 \leqslant \frac{m}{2^k} < \frac{m + 1}{2^k} \leqslant \frac{k2^k}{2^k} = k$$

and in the second case, we have $-k2^k \leqslant m \leqslant -1$ so that

$$-k \leqslant \frac{m+1}{2^k} < \frac{m}{2^k} \leqslant -\frac{1}{2^k} < 0.$$

That implies that we can rewrite s_k as

$$s_k = k\mathbf{1}_{\{f>k\}} - k\mathbf{1}_{\{f<-k\}} + \sum_{m=0}^{k2^k-1} + \sum_{m=0}^{k2^k-1} \frac{m}{2^k}\mathbf{1}_{f^{-1}(I_{m,k})} + \sum_{m=-1}^{-k2^k} \frac{m}{2^k}\mathbf{1}_{f^{-1}(I_{m,k})}$$

which is a finite linear combination of characteristic functions of μ-measurable sets, and hence is a simple function.

We claim that by construction $|s_k(x)| \leqslant |s_{k+1}(x)| \leqslant |f(x)|$ for each $x \in X$ and $k \in \mathbb{Z}^+$. To see this, fix $x \in X$ and $k \in \mathbb{Z}^+$. We first show that $|s_k(x)| \leqslant |f(x)|$.

- If $f(x) \geqslant k$, then $s_k(x) = k \leqslant f(x)$.

- If $f(x) \leqslant -k$, then $s_k(x) = -k \geqslant f(x)$.

- If $-k < f(x) < k$, then $f(x) \in I_{m,k}$ for some $m \in \mathbb{Z}$ with $-k2^k \leqslant m < k2^k$, and hence

$$\frac{m}{2^k} \leqslant f(x) < \frac{m+1}{2^k}$$

  so that $|s_k(x)| = \left|\frac{m}{2^k}\right| \leqslant k < |f(x)|$.

This shows that $|s_k(x)| \leqslant |f(x)|$ for each $x \in X$ and $k \in \mathbb{Z}^+$.

Now we show that $|s_k(x)| \leqslant |s_{k+1}(x)|$ for each $x \in X$ and $k \in \mathbb{Z}^+$. Fix $x \in X$ and $k \in \mathbb{Z}^+$.

- If $f(x) \geqslant k+1$, then $s_k(x) = k$ and $s_{k+1}(x) = k+1$, so that $|s_k(x)| < |s_{k+1}(x)|$.

- If $k \leqslant f(x) < k+1$, then $s_k(x) = k$ but $f(x) \in I_{m,k+1}$ for some $m \geqslant k2^{k+1}$, so that

$$s_{k+1}(x) = \frac{m}{2^{k+1}} \geqslant \frac{k2^{k+1}}{2^{k+1}} = k = s_k(x).$$

- If $0 \leqslant f(x) < k$, then $f(x) \in I_{m,k}$ for some $m \in \mathbb{Z}$ with $0 \leqslant m < k2^k$. Hence $f(x) \in I_{2m,k+1}$ or $I_{2m+1,k+1}$, so that

$$s_{k+1}(x) \in \left\{ \frac{2m}{2^{k+1}}, \frac{2m+1}{2^{k+1}} \right\}$$

  which implies that

$$s_{k+1}(x) \geqslant \frac{2m}{2^{k+1}} = \frac{m}{2^k} = s_k(x).$$

- If $-k < f(x) < 0$, then $f(x) \in I_{m,k}$ for some $m \in \mathbb{Z}$ with $-k2^k \leqslant m < 0$. Hence $f(x) \in I_{2m+1,k+1}$ or $I_{2m,k+1}$, so that

$$s_{k+1}(x) \in \left\{ \frac{2m}{2^{k+1}}, \frac{2m-1}{2^{k+1}} \right\}$$

  and $s_k(x) = \frac{m+1}{2^k}$, which implies that

$$s_{k+1}(x) \leqslant \frac{2m}{2^{k+1}} = \frac{m}{2^k} < \frac{m+1}{2^k} = s_k(x)$$

- If $-k-1 < f(x) < -k$, then $s_k(x) = -k$ but $f(x) \in I_{m,k+1}$ for some $m \leqslant -k2^{k+1} - 1$, so that

$$s_{k+1}(x) = \frac{m+1}{2^{k+1}} \leqslant \frac{-k2^{k+1}}{2^{k+1}} = -k = s_k(x).$$

- If $f(x) \leqslant -k-1$, then $s_k(x) = -k$ and $s_{k+1}(x) = -k-1$, so that $s_{k+1}(x) < s_k(x)$.

In each case, we have shown that $|s_k(x)| \leqslant |s_{k+1}(x)|$.

This shows that $|s_k(x)| \leqslant |s_{k+1}(x)| \leqslant |f(x)|$ for each $x \in X$ and $k \in \mathbb{Z}^+$ as claimed.

Finally, we show that $\lim_{k \to \infty} s_k(x) = f(x)$ for each $x \in X$. By construction, for each $k \in \mathbb{Z}^+$ and each $x \in X$ such that $0 \leqslant f(x) \leqslant k$, we have $f(x) \in I_{m,k}$ for some $m \in \mathbb{Z}$ with $0 \leqslant m < k2^k$, and hence

$$s_k(x) = \frac{m}{2^k} \leqslant f(x) < \frac{m+1}{2^k} = s_k(x) + \frac{1}{2^k}$$

and for each $x \in X$ such that $-k \leqslant f(x) < 0$, we have $f(x) \in I_{m,k}$ for some $m \in \mathbb{Z}$ with $-k2^k \leqslant m < 0$, and hence

$$s_k(x) - \frac{1}{2^k} = \frac{m}{2^k} \leqslant f(x) < \frac{m+1}{2^k} = s_k(x).$$

Putting these two cases together, we see that for each $k \in \mathbb{Z}^+$ for each $x \in X$ such that $|f(x)| \leqslant k$, we have

$$|s_k(x) - f(x)| < \frac{1}{2^k}.$$

To see that $\lim_{k \to \infty} s_k(x) = f(x)$ for each $x \in X$ in the general case, fix $x \in X$. If $f(x) \in \mathbb{R}$, then there exists some $N \in \mathbb{Z}^+$ such that $|f(x)| < N$. Then for each $k \geqslant N$, we have

$$|s_k(x) - f(x)| < \frac{1}{2^k} \leqslant \frac{1}{2^N}$$

which shows that $\lim_{k \to \infty} s_k(x) = f(x)$. If $f(x) \in \{\pm\infty\}$, then the sequence $\{s_k\}_{k=1}^\infty$ is defined so that

$$s_k(x) = \pm k \quad \forall k \in \mathbb{Z}^+$$

so that $\lim_{k\to\infty} s_k(x) = f(x)$ as well. This completes the proof.

Now, if f is a bounded function, then there exists some $M > 0$ such that $|f(x)| \leqslant M$ for each $x \in X$. Choose $N \in \mathbb{Z}^+$ such that $N > M$. Then for each $k \geqslant N$ and each $x \in X$, we have

$$|s_k(x) - f(x)| < \frac{1}{2^k} \leqslant \frac{1}{2^N}$$

which shows that $s_k \to f$ uniformly as $k \to \infty$. $\qquad\square$

Now that we have developed the appropriate language, we can state and prove a famous result in measure theory.

Corollary 1.4.13 (Luzin's Theorem). *Let X be a topological space which has the property that every closed subset of X is a countable intersection of open sets, let μ be a Borel regular outer measure on X, let A be a μ-measurable set with $\mu(A) < \infty$, and let $f : A \to \mathbb{R}$ be μ-measurable. Then for every $\varepsilon > 0$, there exists a closed set $C \subseteq A$ such that $\mu(A \backslash C) < \varepsilon$ and $f|_C$ is continuous.*

In words, Luzin's theorem states that a measurable function is "nearly" continuous.

The reason for the technical assumption on X is so that we can apply Theorem 1.1.14. We remark that a large class of examples is given by separable metric spaces (X, d) where the outer measure μ is locally finite and Borel regular.

Proof. First note that by assumption, the function f is finite-valued and defined only on the set A. By assuming that f is μ-measurable on A, we really mean that f is measurable with respect to the σ-algebra on A consisting of all μ-measurable subsets of A — see Exercise 1.4.1.

For each $j \in \mathbb{Z}^+$ and each $m \in \mathbb{Z}$, let

$$A_{j,m} := f^{-1}\left(\left[\frac{m-1}{k}, \frac{m}{k}\right)\right)$$

which is a μ-measurable set since f is μ-measurable. Then for each $j \in \mathbb{Z}^+$, the sets $\{A_{j,m}\}_{m\in\mathbb{Z}}$ are disjoint and partition X.

By Lemma 1.1.13, we know that $\mu(A) < \infty$ implies $\mu \llcorner A$ is a Borel regular outer measure on A. Now we can apply Theorem 1.1.14 to the measure $\mu \llcorner A$, which says that for each $\epsilon > 0$ and each $j \in \mathbb{Z}^+, m \in \mathbb{Z}$, there exists a closed set $C_{j,m} \subseteq A_{j,m}$ such that

$$\mu(A_{j,m}\backslash C_{j,m}) = (\mu \llcorner A)(A_{j,m}\backslash C_{j,m}) < 2^{-j-|m|-2}\epsilon.$$

Therefore, for each $j \in \mathbb{Z}^+, m \in \mathbb{Z}$ we have

$$\mu\left(A_{j,m}\backslash\left(\bigcup_{n\in\mathbb{Z}} C_{j,n}\right)\right) < 2^{-j-|m|-2}\epsilon.$$

As a result, for each $k \in \mathbb{Z}^+$ we have

$$\mu\left(A \setminus \left(\bigcup_{m \in \mathbb{Z}} C_{k,m}\right)\right) < 2^{-k}\epsilon$$

so there exists an integer $N_k > 0$ such that

$$\mu\left(A \setminus \left(\bigcup_{|m| \leqslant N_k} C_{k,m}\right)\right) < 2^{-k}\epsilon.$$

Since

$$A \setminus \left(\bigcap_{k=1}^{\infty}\left(\bigcup_{|m| \leqslant N_k} C_{k,m}\right)\right) = \bigcup_{k=1}^{\infty}\left(A \setminus \left(\bigcup_{|m| \leqslant N_k} C_{k,m}\right)\right)$$

this implies that $\mu(A \setminus C) < \epsilon$ where

$$C := \bigcap_{k=1}^{\infty}\left(\bigcup_{|m| \leqslant N_k} C_{k,m}\right).$$

Note that C is a closed set since it is a countable intersection of finite unions of closed sets.

For each $k \in \mathbb{Z}^+$, define

$$g_k : \bigcup_{|m| \leqslant N_k} C_{k,m} \to \mathbb{R}$$

by setting $g_k(x) = \frac{m-1}{k}$ for each $x \in C_{k,m}$ such that $|m| \leqslant N_k$. Note that g_k is well-defined since the sets $\{C_{k,m}\}_{|m| \leqslant N_k}$ are disjoint. Also note that g_k is continuous since it is constant on each closed set $C_{k,m}$, so its restriction to C is continuous as well.

Finally, we show that $g_k \to f$ uniformly on C as $k \to \infty$. By construction of the function g_k, we see that for each $x \in C$ we have $|g_k(x) - f(x)| < 1/k$. This shows that $g_k \to f$ uniformly on C as $k \to \infty$, and hence $f|_C$ is continuous as the uniform limit of continuous functions. This completes the proof. $\qquad \square$

Egorov's theorem is another important result about measurable functions.

Theorem 1.4.14 (Egorov's Theorem). *Let (X, μ) be a measure space, and let $A \subseteq X$ be a measurable set with $\mu(A) < \infty$. If $\{f_j\}_{j=1}^{\infty}$ is a sequence of μ-measurable functions $A \to \mathbb{R}$ which converges pointwise to a function $f : A \to \mathbb{R}$, then for every $\varepsilon > 0$, there exists a measurable set $B_\varepsilon \subseteq A$ such that $\mu(A \setminus B_\varepsilon) < \varepsilon$ and $f_j \to f$ uniformly on B_ε.*

Proof. Let $\varepsilon > 0$. For the moment, fix $n \in \mathbb{Z}^+$. The definition of pointwise convergence implies that

$$X = \bigcup_{m=1}^{\infty} \bigcap_{k=m}^{\infty}\left\{x \in X : |f_k(x) - f(x)| < \frac{1}{n}\right\}.$$

For each $m \in \mathbb{Z}^+$, define

$$A_{m,n} := \bigcap_{k=m}^{\infty} \left\{ x \in X : |f_k(x) - f(x)| < \frac{1}{n} \right\}.$$

Then $\{A_{m,n}\}_{m=1}^{\infty}$ is an increasing sequence of measurable sets whose union is X. Since $\mu(A) < \infty$, we have $\mu(A) = \lim_{m \to \infty} \mu(A_{m,n})$ for each $n \in \mathbb{Z}^+$ by Proposition 1.1.6.

Thus there exists some $m_n \in \mathbb{Z}^+$ such that

$$\mu(A) - \mu(A_{m_n,n}) < \frac{\varepsilon}{2^n}.$$

We now take the union of these sets over all $n \in \mathbb{Z}^+$ and define

$$B_\varepsilon := \bigcap_{n=1}^{\infty} A_{m_n,n}.$$

Then B_ε is a measurable set since it is a countable intersection of measurable sets, and

$$\mu(A \backslash B_\varepsilon) = \mu \left(A \backslash \bigcap_{n=1}^{\infty} A_{m_n,n} \right)$$

$$= \mu \left(\bigcup_{n=1}^{\infty} (A \backslash A_{m_n,n}) \right)$$

$$\leq \sum_{n=1}^{\infty} \mu(A \backslash A_{m_n,n})$$

$$< \sum_{n=1}^{\infty} \frac{\varepsilon}{2^n} = \varepsilon.$$

Finally, we show that $f_j \to f$ uniformly on B_ε. Let $\delta > 0$ be arbitrary. Let $N \in \mathbb{Z}^+$ be such that $1/N < \delta$. Then $B_\varepsilon \subseteq A_{m_N,N}$ by construction, and hence for each $x \in B_\varepsilon$ and each $j \geq m_N$, we have

$$|f_j(x) - f(x)| < \frac{1}{N} < \delta.$$

This shows that $f_j \to f$ uniformly on B_ε as desired. $\square$

Chapter 2

Integration Theory

2.1 The Lebesgue Integral

Now that we have defined measure spaces and measurable functions, we would like to define the Lebesgue integral. The Lebesgue integral is a generalization of the Riemann integral, and it fixes many of the deficiencies of the Riemann integral.

There are many different approaches to defining the Lebesgue integral, and they can all be shown to be equivalent.

2.1.1 Integration of Nonnegative Functions with respect to a Measure

Some people take an approach of defining the Lebesgue integral with a series of steps, but we will do it all at once.

Definition 2.1.1 (Lower Lebesgue Sum, Lebesgue Integral of Nonnegative Function)**.** Let (X, μ) be a measure space, and let P be a finite collection of disjoint measurable sets such that $\bigcup_{A \in P} A = X$. We call such a collection a μ-*partition* of (X, μ).

If $f : X \to [0, \infty]$ is a nonnegative measurable function, then we define the *lower Lebesgue sum* of f with respect to the μ-partition P to be

$$L(f, P) = \sum_{A \in P} \inf_{x \in A} f(x) \mu(A).$$

We then define the *Lebesgue integral* of f with respect to μ to be

$$\int_X f \, \mathrm{d}\mu := \sup_P L(f, P),$$

where the supremum is taken over all μ-partitions P of (X, μ).

Remark 2.1.2 (Notation for Lebesgue Integral). We note that the symbol $\mathrm{d}\mu$ is used to denote integration with respect to the measure μ, and the symbol d has no independent meaning here. Sometimes people drop the X in the integral sign and just write $\int f \, \mathrm{d}\mu$ — this is fine because the domain of integration is always the largest set in the σ-algebra on which μ is defined. Sometimes people write the Lebesgue integral as $\int_X f(x) \, \mathrm{d}\mu(x)$, and we will occasionally do this as well.

We also remark that when we are integrating with respect to the Lebesgue measure on $\mathbb{R}^n$, we will usually write $\mathrm{d}x$ instead of $\mathrm{d}\mu$; this is by tradition and conveience. (There is a connection to differential forms, but we will not discuss that here.) In this case, we will usually write $\int_{\mathbb{R}^n} f(x) \, \mathrm{d}x$.

Lemma 2.1.3 (Integral of Characteristic Function). *Let (X, μ) be a measure space, and let $A \subset X$ be a measurable set. Then*

$$\int_X \mathbf{1}_A \, \mathrm{d}\mu = \mu(A).$$

Proof. Since A, $X \backslash A$ is a μ-partition of (X, μ), we clearly have $L(\mathbf{1}_A, \{A, X \backslash A\}) = \mu(A)$. Thus

$$\int_X \mathbf{1}_A \, \mathrm{d}\mu \geqslant \mu(A).$$

On the other hand, let P be a μ-partition of (X, μ) into disjoint measurable sets $A_1, \ldots, A_n$. Then for each $j = 1, 2, \ldots, n$ we see that

$$\inf_{x \in A_j} \mathbf{1}_A(x) = \begin{cases} 1 & A_j \subset A, \\ 0 & A_j \not\subset A \end{cases}$$

which implies

$$\mu(A_j) \inf_{A_j} \mathbf{1}_A = \begin{cases} \mu(A_j) & A_j \subset A, \\ 0 & A_j \not\subset A. \end{cases}$$

Then we have

$$L(\mathbf{1}_A, P) = \sum_{j:A_j \subset A} \mu(A_j) = \mu\left(\bigcup_{j:A_j \subset A} A_j\right) \leqslant \mu(A).$$

This holds for any μ-partition P, so we have

$$\int_X \mathbf{1}_A \, \mathrm{d}\mu = \sup_P L(\mathbf{1}_A, P) \leqslant \mu(A).$$

Combining the two inequalities gives the result. □

Now that we know the integral of characteristic functions, we can upgrade this to simple functions.

Lemma 2.1.4 (Integral of Simple Function). *Let (X, μ) be a measure space, and let $s : X \to [0, \infty)$ be a nonnegative simple function,*

$$s = \sum_{j=1}^{n} c_j \mathbf{1}_{A_j},$$

where $c_1, \ldots, c_n \geqslant 0$ and $A_1, \ldots, A_n$ are disjoint measurable sets. Then we have

$$\int_X s \, d\mu = \sum_{j=1}^{n} c_j \mu(A_j).$$

Some authors take this as the definition of the Lebesgue integral for simple functions, and then use this to define the Lebesgue integral for nonnegative measurable functions. We do not take this approach, because it is not clear that the right-hand side is well-defined (i.e. independent of the representation of s as a sum of characteristic functions).

Proof. Set $A_0 = X \backslash \bigcup_{i=1}^{n} A_i$. Then $A_0, A_1, \ldots, A_n$ is a μ-partition of (X, μ), and we have

$$L(s, \{A_0, A_1, \ldots, A_n\}) = \sum_{j=1}^{n} c_j \mu(A_j).$$

Thus

$$\int_X s \, d\mu \geqslant \sum_{j=1}^{n} c_j \mu(A_j).$$

Now let P be any μ-partition of (X, μ) into disjoint measurable sets $B_1, \ldots, B_m$. Then we have

$$
\begin{aligned}
L(s, P) &= \sum_{j=1}^{m} \inf_{x \in B_j} s(x) \mu(B_j) \\
&= \sum_{j=1}^{m} \min_{\{i : B_j \cap A_i \neq \varnothing\}} c_i \, \mu(B_j) \\
&\leqslant \sum_{j=1}^{m} \sum_{i=1}^{n} \mu(B_j \cap A_k) c_k \\
&= \sum_{k=1}^{n} c_k \sum_{j=1}^{m} \mu(B_j \cap A_k) \\
&= \sum_{k=1}^{n} c_k \mu(A_k).
\end{aligned}
$$

This holds for any μ-partition P, so we have

$$\int_X s \, d\mu \leqslant \sum_{j=1}^{n} c_j \mu(A_j).$$

Combining the two inequalities gives the result. $\qquad \square$

Lemma 2.1.5 (Lebesgue Integral is Order-Preserving)**.** *Let* (X, μ) *be a measure space, and let* $f, g : X \to [0, \infty]$ *be nonnegative measurable functions such that* $f(x) \leqslant g(x)$ *for all* $x \in X$. *Then*

$$\int_X f \, \mathrm{d}\mu \leqslant \int_X g \, \mathrm{d}\mu.$$

Proof. Let P be any μ-partition of (X, μ). Then for each set $A \in P$ we have

$$\inf_{x \in A} f(x) \leqslant \inf_{x \in A} g(x),$$

which implies

$$L(f, P) = \sum_{A \in P} \inf_{x \in A} f(x) \mu(A) \leqslant \sum_{A \in P} \inf_{x \in A} g(x) \mu(A) = L(g, P).$$

This holds for any μ-partition P, so we have

$$\int_X f \, \mathrm{d}\mu \leqslant \int_X g \, \mathrm{d}\mu.$$

$\square$

We finish this subsection by showing that that the integral of a general nonnegative measurable function can be approximated by integrals of simple functions.

Lemma 2.1.6 (Integrals via Simple Functions)**.** *Let* (X, μ) *be a measure space, and let* $f : X \to [0, \infty]$ *be a nonnegative measurable function. Then*

$$\int_X f \, \mathrm{d}\mu = \sup \left\{ \int_X s \, \mathrm{d}\mu : s \text{ is a simple function with } 0 \leqslant s \leqslant f \right\}$$

Proof. First note that the inequality $\geqslant$ is immediate from Lemma 2.1.5.

For the reverse inequality, we assume first that we are in the special case where $\inf_A f > 0$ for each measurable set $A \subseteq X$ with $\mu(A) > 0$. (This is a technical assumption that we will remove later.)

Let P be a μ-partition of (X, μ) into disjoint nonempty measurable sets $A_1, \ldots, A_n$. For each $j = 1, 2, \ldots, n$, set $m_j = \inf_{x \in A_j} f(x)$. Then the function

$$s = \sum_{j=1}^{n} m_j \mathbf{1}_{A_j}$$

is a simple function with $0 \leqslant s \leqslant f$, and

$$\int_X s \, \mathrm{d}\mu = \sum_{j=1}^{n} m_j \, \mu(A_j) = \sum_{j=1}^{\infty} \inf_{x \in A_j} f(x) \, \mu(A_j) = L(f, P)$$

where the first equality holds by Lemma 2.1.4. Therefore taking the supremum over all μ-partitions P gives

$$\int_X f \, \mathrm{d}\mu = \sup_P L(f, P) \leqslant \sup \left\{ \int_X s \, \mathrm{d}\mu : s \text{ is a simple function with } 0 \leqslant s \leqslant f \right\}.$$

Now we remove the technical assumption. Assume that there is a measurable set $A \subseteq X$ with $\mu(A) > 0$ and $\inf_A f = \infty$. Then we must have $f(x) = \infty$ for each $x \in A$.

In this case, for arbitrary $t > 0$ the function $t\mathbf{1}_A$ is a simple function with $0 \leqslant t\mathbf{1}_A \leqslant f$, and

$$\int_X t\mathbf{1}_A \, \mathrm{d}\mu = t\mu(A)$$

by Lemma 2.1.3. Thus the supremum on the right-hand side is greater than $t\mu(A)$ for each $t > 0$, and hence is infinite. Thus the inequality $\leqslant$ holds in this case as well, completeing the proof. $\qquad \square$

Example 2.1.7 (Sums are Integrals). Consider the counting measure μ on $\mathbb{N}$, and let $\{a_k\}_{k=0}^\infty$ be a sequence of nonnegative real numbers. Define $a : \mathbb{N} \to [0, \infty)$ by $a(n) = a_n$. Then we have

$$\int a \, \mathrm{d}\mu = \sum_{k=0}^\infty a_k.$$

Proof. for each $n \in \mathbb{N}$, the set $\{n\}$ is measurable and has measure $\mu(\{n\}) = 1$. Thus for each $n \in \mathbb{N}$ the sum

$$s_n := \sum_{j=0}^N a_j \mathbf{1}_{\{j\}} = \sum_{j=0}^N a_j$$

is a simple function; the fact that the sequence $\{a_k\}_{k=0}^\infty$ is nonnegative implies that $0 \leqslant s_n \leqslant a$. By the previous lemma, we have

$$\int a \, \mathrm{d}\mu \geqslant \int s_n \, \mathrm{d}\mu = \sum_{j=0}^N a_j$$

for each $N \in \mathbb{N}$; taking limits gives

$$\int a \, \mathrm{d}\mu \geqslant \sum_{j=0}^\infty a_j.$$

On the other hand, let P be any μ-partition of $(\mathbb{N}, \mu)$ into disjoint nonempty measurable sets $A_1, A_2, \ldots, A_m$. Then for each $j = 1, 2, \ldots, m$ we have

$$\inf_{n \in A_j} a(n) = \inf_{n \in A_j} a_n = \min_{n \in A_j} a_n$$

since A_j is a nonempty finite set. Thus we have

$$L(a, P) = \sum_{j=1}^{m} \inf_{n \in A_j} a(n)\mu(A_j) = \sum_{j=1}^{m} \min_{n \in A_j} a_n \cdot 1 \leqslant \sum_{j=1}^{m} \sum_{n \in A_j} a_n = \sum_{n=0}^{\infty} a_n.$$

Taking the supremum over all μ-partitions P gives

$$\int a \, d\mu \leqslant \sum_{n=0}^{\infty} a_n.$$

Combining the two inequalities gives the result. $\qquad\qquad\qquad\qquad\qquad$ □

2.1.2   Properties of the Lebesgue Integral

Still only working with nonnegative measurable functions, we now prove some important properties of the Lebesgue integral. One of the most important properties is the Monotone Convergence Theorem.

Theorem 2.1.8 (Monotone Convergence Theorem). *Let (X, μ) be a measure space, and let $\{f_k\}_{k=1}^{\infty}$ be a sequence of nonnegative measurable functions such that $f_k(x) \leqslant f_{k+1}(x)$ for all $x \in X$ and $k \in \mathbb{N}$. Define the function $f : X \to [0, \infty]$ by*

$$f(x) = \lim_{k \to \infty} f_k(x), \qquad \forall x \in X.$$

Then

$$\lim_{k \to \infty} \int_X f_k \, d\mu = \int_X f \, d\mu.$$

In essence, this says that one can interchange limits and integrals $\lim_k \int f_k \, d\mu = \int \lim_k f_k \, d\mu$ for monotone sequences of nonnegative measurable functions.

Proof. Note that the function f is well-defined at each $x \in X$ because the sequence $\{f_k(x)\}_{k=1}^{\infty}$ is nondecreasing and bounded below by 0; hence

$$f_k(x) \leqslant f(x) \qquad \forall x \in X, \forall k \geqslant 1.$$

Also note that f is measurable because it is the pointwise limit of measurable functions.

We remark that the limit $\lim_{k \to \infty} \int_X f_k \, d\mu$ exists because the sequence of integrals is nondecreasing (since $\{f_k\}_{k=1}^{\infty}$ is nondecreasing) and bounded below by 0. Thus everything in the statement of the theorem is well-defined.

By Lemma 2.1.5, we have

$$\int_X f_k \, d\mu \leqslant \int_X f \, d\mu \qquad \forall k \geqslant 1,$$

and thus

$$\lim_{k \to \infty} \int_X f_k \, \mathrm{d}\mu \leqslant \int_X f \, \mathrm{d}\mu.$$

To prove the reverse inequality, let $A_1, A_2, \ldots, A_n$ be disjoint measurable subsets of X, and let $c_1, c_2, \ldots, c_n \geqslant 0$ be such that

$$f(x) \geqslant \sum_{j=1}^{m} c_j \mathbf{1}_{A_j}(x) \qquad \forall x \in X.$$

Let $t \in (0, 1)$ be arbitrary. Then for each $k \in \mathbb{Z}^+$ we let

$$E_k^t := \left\{ x \in X : f_k(x) \geqslant t \sum_{j=1}^{m} c_j \mathbf{1}_{A_j}(x) \right\}.$$

Then $\{E_k^t\}_{k=1}^{\infty}$ is an increasing sequence of measurable sets such that $\bigcup_{k=1}^{\infty} E_k^t = X$. By Proposition 1.1.6, we have

$$\lim_{k \to \infty} \mu(A_j \cap E_k^t) = \mu(A_j)$$

for each $j = 1, 2, \ldots, m$.

If $k \in \mathbb{Z}^+$, then

$$f_k(x) \geqslant \sum_{j=1}^{m} t c_j \mathbf{1}_{A_j \cap E_k^t}(x)$$

for each $x \in X$, and thus by Lemma 2.1.4 we have

$$\int_X f_k \, \mathrm{d}\mu \geqslant t \sum_{j=1}^{m} c_j \mu(A_j \cap E_k^t).$$

Taking limits as $k \to \infty$ gives

$$\lim_{k \to \infty} \int_X f_k \, \mathrm{d}\mu \geqslant t \sum_{j=1}^{m} c_j \mu(A_j).$$

Taking limits as $t \to 1^-$ gives

$$\lim_{k \to \infty} \int_X f_k \, \mathrm{d}\mu \geqslant \sum_{j=1}^{m} c_j \mu(A_j).$$

Taking the supremum over all choices of m, $c_1, \ldots, c_m$, and $A_1, \ldots, A_m$ gives

$$\lim_{k \to \infty} \int_X f_k \, \mathrm{d}\mu \geqslant \int_X f \, \mathrm{d}\mu.$$

This completes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

In the Lemma 2.1.4, we computed the integral of a simple function once it had been expressed as a sum of characteristic functions of disjoint measurable sets. This expression is not unique, and the assumption that the sets are disjoint is inconvenient. It turns out that the integral of a simple function can be computed from any of its expressions and they all give the same answer.

Lemma 2.1.9 (Integral of Simple Functions, II.)**.** *Let* (X, μ) *be a measure space, and let* $s : X \to [0, \infty)$ *be a nonnegative simple function. If*

$$s = \sum_{j=1}^{m} a_j \mathbf{1}_{A_j} = \sum_{k=1}^{n} b_k \mathbf{1}_{B_k}$$

are two representations of s as nonnegative linear combinations of characteristic functions, then

$$\sum_{j=1}^{m} a_j \mu(A_j) = \sum_{k=1}^{n} b_k \mu(B_k).$$

In particular,

$$\int_X s \ \mathrm{d}\mu = \sum_{j=1}^{m} a_j \mu(A_j)$$

Proof. Let $A_0 = X \backslash \bigcup_{j=1}^{m} A_j$ so that $A_0, A_1, \ldots, A_m$ is a μ-partition of (X, μ). If $A_i \cap A_j \neq \varnothing$ for some $i \neq j$, then we can write

$$a_i \mathbf{1}_{A_i} + a_j \mathbf{1}_{A_j} = a_i \mathbf{1}_{A_i \backslash A_j} + a_j \mathbf{1}_{A_j \backslash A_i} + (a_i + a_j) \mathbf{1}_{A_i \cap A_j}$$

and the sets $A_i \backslash A_j$, $A_j \backslash A_i$, and $A_i \cap A_j$ are disjoint measurable sets.

Thus for $0 \leqslant i < j \leqslant m$ such that $A_i \cap A_j \neq \varnothing$, we have

$$A_i = (A_i \backslash A_j) \cup (A_i \cap A_j) \quad \text{and} \quad A_j = (A_j \backslash A_i) \cup (A_i \cap A_j)$$

and each of these unions is disjoint; hence

$$\mu(A_i) = \mu(A_i \backslash A_j) + \mu(A_i \cap A_j) \quad \text{and} \quad \mu(A_j) = \mu(A_j \backslash A_i) + \mu(A_i \cap A_j)$$

by disjoint additivity of μ, and it follows that

$$a_i \mu(A_i) + a_j \mu(A_j) = a_i \mu(A_i \backslash A_j) + a_j \mu(A_j \backslash A_i) + (a_i + a_j) \mu(A_i \cap A_j).$$

Thus for each $0 \leqslant i < j \leqslant m$ such that $A_i \cap A_j \neq \varnothing$, we can replace the sets A_i, A_j by the disjoint sets $A_i \backslash A_j$, $A_j \backslash A_i$, and $A_i \cap A_j$, and replace the coefficients a_i, a_j by $a_i, a_j, a_i + a_j$ respectively; the linear combinations of the characteristic functions and the sums of the measures weighted by the coefficients remain unchanged.

Repeating this process a finite number of times gives a representation of s as a sum of characteristic functions of disjoint measurable sets. By relabeling, we have finitely many disjoint measurable sets $\tilde{A}_1, \ldots, \tilde{A}_N$, with nonnegative coefficients $\tilde{a}_1, \ldots, \tilde{a}_N \geqslant 0$, such that $s = \sum_{j=1}^{N} \tilde{a}_j \mathbf{1}_{\tilde{A}_j}$.

The next step is to make the numbers $\tilde{a}_1, \ldots, \tilde{a}_N$ distinct. If $\tilde{a}_i = \tilde{a}_j$ for some $i \neq j$, then $\tilde{A}_i \cup \tilde{A}_j$ is a measurable set, and disjoint additivity of μ gives

$$\tilde{a}_i \mu(\tilde{A}_i) + \tilde{a}_j \mu(\tilde{A}_j) = \tilde{a}_i \mu(\tilde{A}_i \cup \tilde{A}_j).$$

Thus for each $i \neq j$ such that $\tilde{a}_i = \tilde{a}_j$, we can replace the sets $\tilde{A}_i, \tilde{A}_j$ by the single set $\tilde{A}_i \cup \tilde{A}_j$, and replace the coefficients $\tilde{a}_i, \tilde{a}_j$ by the single coefficient $\tilde{a}_i$.

Repeating this process a finite number of times gives a representation of s as a sum of characteristic functions of disjoint measurable sets with distinct coefficients. By relabeling, we have finitely many disjoint measurable sets $\hat{A}_1, \ldots, \hat{A}_M$, with distinct nonnegative coefficients $\hat{a}_1, \ldots, \hat{a}_M \geqslant 0$, such that $s = \sum_{j=1}^{M} \hat{a}_j \mathbf{1}_{\hat{A}_j}$.

Finally if $1 \leqslant j \leqslant M$ is such that $\hat{A}_j = \varnothing$, then we can simply remove this term from the sum without changing anything. Thus we may assume that $\hat{A}_1, \ldots, \hat{A}_M$ are nonempty disjoint measurable sets, and $\hat{a}_1, \ldots, \hat{a}_M$ are distinct nonnegative numbers. The expression $s = \sum_{j=1}^{M} \hat{a}_j \mathbf{1}_{\hat{A}_j}$ is the *standard form* of the simple function s.

We can now do the same process to the other representation $s = \sum_{k=1}^{n} b_k \mathbf{1}_{B_k}$ to obtain a standard form $s = \sum_{k=1}^{M} \hat{b}_k \mathbf{1}_{\hat{B}_k}$. By construction these standard forms must have the same number of terms (the number of distinct values that s takes), and for each $j = 1, 2, \ldots, M$ there exists $k_j \in \{1, 2, \ldots, M\}$ such that $\hat{A}_j = \hat{B}_{k_j}$ and $\hat{a}_j = \hat{b}_{k_j}$. Thus we have

$$\sum_{j=1}^{m} a_j \mu(A_j) = \sum_{j=1}^{M} \hat{a}_j \mu(\hat{A}_j) = \sum_{k=1}^{M} \hat{b}_k \mu(\hat{B}_k) = \sum_{k=1}^{n} b_k \mu(B_k)$$

as desired.

The fact that $\int_X s \, \mathrm{d}\mu = \sum_{j=1}^{m} a_j \mu(A_j)$ follows from Lemma 2.1.4. $\qquad \square$

We can now use the Monotone Convergence Theorem to prove additivity of the Lebesgue integral.

Proposition 2.1.10 (Additivity for Nonnegative Functions). *Let (X, μ) be a measure space, and let $f, g : X \to [0, \infty]$ be nonnegative measurable functions. Then*

$$\int_X (f + g) \, \mathrm{d}\mu = \int_X f \, \mathrm{d}\mu + \int_X g \, \mathrm{d}\mu.$$

Proof. Note that for simple functions the result follows from Lemma 2.1.9. Thus we approximate by such functions and use the Monotone Convergence Theorem.

Let $\{f_k\}_{k=1}^\infty$ and $\{g_k\}_{k=1}^\infty$ be increasing sequences of nonnegative simple functions such that

$$\lim_{k\to\infty} f_k = f \quad \text{and} \quad \lim_{k\to\infty} g_k = g$$

pointwise on X; such sequences exist by Proposition 1.4.12. Then the monotone convergence theorem gives

$$\begin{aligned}
\int_X (f + g) \, \mathrm{d}\mu &= \lim_{k\to\infty} \int_X (f_k + g_k) \, \mathrm{d}\mu \\
&= \lim_{k\to\infty} \int_X f_k \, \mathrm{d}\mu + \lim_{k\to\infty} \int_X g_k \, \mathrm{d}\mu \\
&= \int_X f \, \mathrm{d}\mu + \int_X g \, \mathrm{d}\mu.
\end{aligned}$$

Here we have used the Monotone Convergence Theorem in the first and last equalities, and the additivity of the integral for simple functions in the second equality. $\qquad\square$

2.1.3   Integration of Real and Complex Valued Functions

Definition 2.1.11 (f^+, f^-)**.** Let $f : X \to [-\infty, \infty]$ be a function defined on a set X. We define the functions $f^+, f^- : X \to [0, \infty]$ by

$$f^+(x) := \max\{f(x), 0\} \quad \text{and} \quad f^-(x) := \max\{-f(x), 0\} \qquad \forall x \in X.$$

Remark 2.1.12. Note that if f is measurable, then f^+ and f^- are nonnegative measurable functions, and that

$$f = f^+ - f^- \quad \text{and} \quad |f| = f^+ + f^-.$$

Definition 2.1.13 (Lebesgue Integral, Real-Valued Case)**.** Let (X, μ) be a measure space, and let $f : X \to [-\infty, \infty]$ be a measurable function such that at least one of the integrals $\int_X f^+ \, \mathrm{d}\mu$ or $\int_X f^- \, \mathrm{d}\mu$ is finite. Then we define the *Lebesgue integral* of f with respect to μ to be

$$\int_X f \, \mathrm{d}\mu := \int_X f^+ \, \mathrm{d}\mu - \int_X f^- \, \mathrm{d}\mu.$$

We say that f is *integrable* if $\int_X f \, \mathrm{d}\mu$ exists and is finite.

See that if $f \geqslant 0$, then $f^- = 0$ and $f^+ = f$, so this definition agrees with the previous definition for nonnegative measurable functions. Also notice that we allow f to take the values $\pm\infty$, but if f takes either value on a set of positive measure, then the integral will be

infinite. Also notice that

$$f \text{ is integrable} \iff \int_X f \, \mathrm{d}\mu = \int_X f^+ \, \mathrm{d}\mu - \int_X f^- \, \mathrm{d}\mu \in \mathbb{R}$$

$$\iff \int_X f^+ \, \mathrm{d}\mu < \infty \text{ and } \int_X f^- \, \mathrm{d}\mu < \infty$$

$$\iff \int_X (f^+ + f^-) \, \mathrm{d}\mu = \int_X |f| \, \mathrm{d}\mu < \infty.$$

where in the second line we have used that at most one of $\int_X f^+ \, \mathrm{d}\mu$ or $\int_X f^- \, \mathrm{d}\mu$ is infinite, since otherwise the integral of f would not be defined. Thus a real-valued measurable function is integrable if $\int_X |f| \, \mathrm{d}\mu$ is finite.

Definition 2.1.14 (Lebesgue Integral, Complex-Valued Case). Let (X, μ) be a measure space, and let $f : X \to \mathbb{C} \cup \{\infty\}$ be a measurable function such that $\int_X |f| \, \mathrm{d}\mu$ is finite. Then we define the *Lebesgue integral* of f with respect to μ to be

$$\int_X f \, \mathrm{d}\mu := \int_X \Re(f) \, \mathrm{d}\mu + i \int_X \Im(f) \, \mathrm{d}\mu.$$

If the integral $\int_X f \, \mathrm{d}\mu$ is defined, then we say that f is *integrable*.

Similarly here we allow f to take the value ∞, but if it does so on a set of positive measure, then the integral will not be defined (since $\int_X |f| \, \mathrm{d}\mu$ will be infinite). We should take note here that for complex-valued functions, we require absolute integrability in order for the integral to be defined. Also if f is real-valued, then this definition agrees with the previous definition for real-valued functions.

We would like to state and prove properties of the Lebesgue integral in the real-valued and complex-valued cases we have just defined. For efficiency, we let $\mathbb{F}$ denote either $\mathbb{R}$ or $\mathbb{C}$, and we let $\mathbb{F}_\infty$ denote either $[-\infty, \infty]$ or $\mathbb{C} \cup \{\infty\}$ respectively.

Proposition 2.1.15 (Properties of the Lebesgue Integral). *Let (X, μ) be a measure space.*

(i) *(Homogeneity) If $f : X \to \mathbb{F}_\infty$ is a measurable function such that the integral $\int_X f \, \mathrm{d}\mu$ is defined, and if $c \in \mathbb{F}$, then*

$$\int_X cf \, \mathrm{d}\mu = c \int_X f \, \mathrm{d}\mu.$$

(ii) *(Additivity) If $f, g : X \to \mathbb{F}_\infty$ are measurable functions such that $\int_X |f| \, \mathrm{d}\mu$ and $\int_X |g| \, \mathrm{d}\mu$ are both finite, then*

$$\int_X (f + g) \, \mathrm{d}\mu = \int_X f \, \mathrm{d}\mu + \int_X g \, \mathrm{d}\mu.$$

(iii) (Monotonicity) If $f, g : X \to [-\infty, \infty]$ are measurable functions such that $f \leqslant g$ and the integrals $\int_X f \, d\mu$ and $\int_X g \, d\mu$ are both defined, then

$$\int_X f \, d\mu \leqslant \int_X g \, d\mu.$$

(iv) (Triangle Inequality) If $f : X \to \mathbb{F}_\infty$ is a measurable function such that $\int_X f \, d\mu$ is defined, then

$$\left| \int_X f \, d\mu \right| \leqslant \int_X |f| \, d\mu.$$

Notice the third property only applies to real-valued functions, since the complex numbers are not an ordered field.

Proof. (i) *Homogeneity.*

Case $\mathbb{F} = \mathbb{R}$. First we consider the special case of nonnegative measurable functions and constants. Let $f : X \to [0, \infty]$ be a nonnegative measurable function, and let $c \geqslant 0$. Then for each μ-partition P of (X, μ) we have

$$L(cf, P) = \sum_{A \in P} \inf_{x \in A} cf(x)\mu(A) = c \sum_{A \in P} \inf_{x \in A} f(x)\mu(A) = cL(f, P).$$

Taking the supremum over all μ-partitions P gives

$$\int_X cf \, d\mu = c \int_X f \, d\mu.$$

Now we consider the general case. Let $f : X \to [-\infty, \infty]$ be a measurable function such that the integral $\int_X f \, d\mu$ is defined. If $c \geqslant 0$, then we have

$$\begin{aligned}
\int_X cf \, d\mu &= \int_X (cf)^+ \, d\mu - \int_X (cf)^- \, d\mu \\
&= \int_X cf^+ \, d\mu - \int_X cf^- \, d\mu \\
&= c \left(\int_X f^+ \, d\mu - \int_X f^- \, d\mu \right) \\
&= c \int_X f \, d\mu
\end{aligned}$$

where we have used the result for nonnegative measurable functions in the third equality.

If $c < 0$ then $-c > 0$, and we have

$$
\begin{aligned}
\int_X cf \ \mathrm{d}\mu &= \int_X (cf)^+ \ \mathrm{d}\mu - \int_X (cf)^- \ \mathrm{d}\mu \\
&= \int_X (-c)f^- \ \mathrm{d}\mu - \int_X (-c)f^+ \ \mathrm{d}\mu \\
&= -c \left(\int_X f^- \ \mathrm{d}\mu - \int_X f^+ \ \mathrm{d}\mu \right) \\
&= (-c) \cdot \left(- \int_X f \ \mathrm{d}\mu \right)
\end{aligned}
$$

which completes the proof of homogeneity.

Case $\mathbb{F} = \mathbb{C}$. Let $f : X \to \mathbb{C} \cup \{\infty\}$ be a measurable function such that $\int_X |f| \ \mathrm{d}\mu$ is finite, and let $c \in \mathbb{C}$. Then

$$
\begin{aligned}
\int_X cf \ \mathrm{d}\mu &= \int_X \Re(cf) \ \mathrm{d}\mu + i \int_X \Im(cf) \ \mathrm{d}\mu \\
&= \Re(c) \int_X \Re(f) \ \mathrm{d}\mu - \Im(c) \int_X \Im(f) \ \mathrm{d}\mu + i \left(\Re(c) \int_X \Im(f) \ \mathrm{d}\mu + \Im(c) \int_X \Re(f) \ \mathrm{d}\mu \right) \\
&= c \left(\int_X \Re(f) \ \mathrm{d}\mu + i \int_X \Im(f) \ \mathrm{d}\mu \right) \\
&= c \int_X f \ \mathrm{d}\mu
\end{aligned}
$$

where we have used homogeneity for real-valued functions in the second equality. This completes the proof of homogeneity.

(ii) Additivity.

Case $\mathbb{F} = \mathbb{R}$. Let $f, g : X \to [-\infty, \infty]$ be measurable functions such that $\int_X |f| \ \mathrm{d}\mu$ and $\int_X |g| \ \mathrm{d}\mu$ are both finite. Then

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

which implies

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

Both sides of this equation are nonnegative measurable functions, so we can integrate both sides to obtain

$$
\int_X (f + g)^+ \ \mathrm{d}\mu + \int_X f^- \ \mathrm{d}\mu + \int_X g^- \ \mathrm{d}\mu = \int_X (f + g)^- \ \mathrm{d}\mu + \int_X f^+ \ \mathrm{d}\mu + \int_X g^+ \ \mathrm{d}\mu
$$

by additivity of the Lebesgue integral for nonnegative measurable functions.

Rearranging gives

$$\int_X (f+g)^+ \, \mathrm{d}\mu - \int_X (f+g)^- \, \mathrm{d}\mu = \int_X f^+ \, \mathrm{d}\mu - \int_X f^- \, \mathrm{d}\mu + \int_X g^+ \, \mathrm{d}\mu - \int_X g^- \, \mathrm{d}\mu$$

The left side is not of the form $\infty - \infty$ since $(f+g)^+ \leqslant f^+ + g^+$ and $(f+g)^- \leqslant f^- + g^-$, and the assumption that $\int_X |f| \, \mathrm{d}\mu$ and $\int_X |g| \, \mathrm{d}\mu$ are both finite to ensure that the integrals $\int_X f^\pm \, \mathrm{d}\mu$ and $\int_X g^\pm \, \mathrm{d}\mu$ are all finite, and thus both of the integrals on the left side are also finite.

This last equation is exactly the desired result, by definition of the Lebesgue integral for general measurable functions.

Case $\mathbb{F} = \mathbb{C}$. Let $f, g : X \to \mathbb{C} \cup \{\infty\}$ be measurable functions such that $\int_X |f| \, \mathrm{d}\mu$ and $\int_X |g| \, \mathrm{d}\mu$ are both finite. Then

$$\begin{aligned}
\int_X (f+g) \, \mathrm{d}\mu &= \int_X \Re(f+g) \, \mathrm{d}\mu + i \int_X \Im(f+g) \, \mathrm{d}\mu \\
&= \int_X \Re(f) \, \mathrm{d}\mu + \int_X \Re(g) \, \mathrm{d}\mu + i \left(\int_X \Im(f) \, \mathrm{d}\mu + \int_X \Im(g) \, \mathrm{d}\mu \right) \\
&= \left(\int_X \Re(f) \, \mathrm{d}\mu + i \int_X \Im(f) \, \mathrm{d}\mu \right) + \left(\int_X \Re(g) \, \mathrm{d}\mu + i \int_X \Im(g) \, \mathrm{d}\mu \right) \\
&= \int_X f \, \mathrm{d}\mu + \int_X g \, \mathrm{d}\mu
\end{aligned}$$

where we have used additivity for real-valued functions in the second equality. This completes the proof of additivity.

(iii) Monotonicity.

Let $f, g : X \to [-\infty, \infty]$ be measurable functions such that $f \leqslant g$ almost everywhere with respect to μ and the integrals $\int_X f \, \mathrm{d}\mu$ and $\int_X g \, \mathrm{d}\mu$ are both defined. If $\int_X f \, \mathrm{d}\mu = \pm\infty$ or $\int_X g \, \mathrm{d}\mu = \pm\infty$, then we check this case by case:

- If $\int_X f \, \mathrm{d}\mu = -\infty$ or $\int_X g \, \mathrm{d}\mu = \infty$, then trivially $\int_X f \, \mathrm{d}\mu \leqslant \int_X g \, \mathrm{d}\mu$.

- If $\int_X f \, \mathrm{d}\mu = \infty$, then we must have either $\int_X f^+ \, \mathrm{d}\mu = \infty$ and $\int_X f^- \, \mathrm{d}\mu < \infty$, or $\int_X f^+ \, \mathrm{d}\mu < \infty$ and $\int_X f^- \, \mathrm{d}\mu = \infty$. In the first subcase, we have $\int_X g^+ \, \mathrm{d}\mu \geqslant \int_X f^+ \, \mathrm{d}\mu = \infty$ and thus $\int_X g \, \mathrm{d}\mu = \infty$; hence $\int_X f \, \mathrm{d}\mu \leqslant \int_X g \, \mathrm{d}\mu$. In the second subcase, we have $\int_X g^- \, \mathrm{d}\mu \leqslant \int_X f^- \, \mathrm{d}\mu = \infty$ and thus $\int_X g \, \mathrm{d}\mu = -\infty$; hence $\int_X f \, \mathrm{d}\mu \leqslant \int_X g \, \mathrm{d}\mu$.

- If $\int_X g \, \mathrm{d}\mu = -\infty$, then this is similar to the previous case.

Now assume that both $\int_X f \, \mathrm{d}\mu$ and $\int_X g \, \mathrm{d}\mu$ are finite. Then additivity and homogeneity with $c = -1$ give

$$\int_X (g-f) \, \mathrm{d}\mu = \int_X g \, \mathrm{d}\mu - \int_X f \, \mathrm{d}\mu.$$

Since $g - f \geqslant 0$, we have $\int_X (g - f) \, d\mu \geqslant 0$. Rearranging gives

$$\int_X g \, d\mu \geqslant \int_X f \, d\mu$$

as desired.

(iv) Triangle Inequality.

Case $\mathbb{F} = \mathbb{R}$. Let $f : X \to [-\infty, \infty]$ be a measurable function such that $\int_X f \, d\mu$ is defined. Then one of $\int_X f^+ \, d\mu$ or $\int_X f^- \, d\mu$ is finite. Thus

$$
\begin{aligned}
\left| \int_X f \, d\mu \right| &= \left| \int_X f^+ \, d\mu - \int_X f^- \, d\mu \right| \\
&\leqslant \int_X f^+ \, d\mu + \int_X f^- \, d\mu \\
&= \int_X (f^+ + f^-) \, d\mu \\
&= \int_X |f| \, d\mu
\end{aligned}
$$

where we have used the triangle inequality for real numbers in the second line, and additivity of the Lebesgue integral for nonnegative measurable functions in the third line.

Case $\mathbb{F} = \mathbb{C}$. Let $f : X \to \mathbb{C} \cup \{\infty\}$ be a measurable function such that $\int_X |f| \, d\mu$ is finite. If $\int_X f \, d\mu = 0$ then the inequality is trivial, so assume that $\int_X f \, d\mu \neq 0$. Then let

$$\alpha := \frac{\left| \int_X f \, d\mu \right|}{\int_X f \, d\mu}.$$

We compute that

$$
\begin{aligned}
\left| \int_X f \, d\mu \right| = \alpha \int_X f \, d\mu &= \int_X \alpha f \, d\mu \\
&= \int_X \Re(\alpha f) \, d\mu + i \int_X \Im(\alpha f) \, d\mu \\
&= \int_X \Re(\alpha f) \, d\mu \\
&\leqslant \int_X |\alpha f| \, d\mu \\
&= \int_X |f| \, d\mu
\end{aligned}
$$

where we have use homogeneity for complex-valued functions in the second equality, and the fact that $\alpha \int_X f \, d\mu$ is a nonnegative real number in the fourth equality, and the triangle inequality for real-valued functions in the inequality. This completes the proof of the triangle inequality. $\square$

2.1.4   Integration of $\mathbb{R}^m$-Valued Functions

Sometimes we wish to integrate functions that take values in $\mathbb{R}^m$ or some vector space other than $\mathbb{R}$ or $\mathbb{C}$. This can be done component-wise using the Lebesgue integral we have already defined.

Definition 2.1.16 (Lebesgue Integral, Vector-Valued Case). Let (X, μ) be a measure space, and let $f : X \to \mathbb{R}^m$ be a measurable function. We define the *Lebesgue integral* of f with respect to μ as the vector

$$\int_X f \, \mathrm{d}\mu := \left(\int_X f_1 \, \mathrm{d}\mu, \int_X f_2 \, \mathrm{d}\mu, \ldots, \int_X f_m \, \mathrm{d}\mu \right) \in \mathbb{R}^m$$

where $f_j : X \to \mathbb{R}$ is the j-th component function of f for each $1 \leqslant j \leqslant m$.

Proposition 2.1.17 (Properties of the Lebesgue Integral, Vector-Valued Case). *Let (X, μ) be a measure space, and let $f, g : X \to \mathbb{R}^m$ be measurable functions. Then the following properties hold:*

(i) *(Linearity) For each $\lambda \in \mathbb{R}$ we have*

$$\int_X (f + \lambda g) \, \mathrm{d}\mu = \int_X f \, \mathrm{d}\mu + \lambda \int_X g \, \mathrm{d}\mu.$$

(ii) *(Triangle Inequality) We have*

$$\left\| \int_X f \, \mathrm{d}\mu \right\| \leqslant \int_X \|f\| \, \mathrm{d}\mu.$$

Proof. Letting $f_1, f_2, \ldots, f_m : X \to \mathbb{R}$ and $g_1, g_2, \ldots, g_m : X \to \mathbb{R}$ be the component functions of f and g respectively, we have

$$\begin{aligned}
\int_X (f + \lambda g) \, \mathrm{d}\mu &= \left(\int_X (f_1 + \lambda g_1) \, \mathrm{d}\mu, \ldots, \int_X (f_m + \lambda g_m) \, \mathrm{d}\mu \right) \\
&= \left(\int_X f_1 \, \mathrm{d}\mu + \lambda \int_X g_1 \, \mathrm{d}\mu, \ldots, \int_X f_m \, \mathrm{d}\mu + \lambda \int_X g_m \, \mathrm{d}\mu \right) \qquad \text{by linearity for real-valued} \\
&= \left(\int_X f_1 \, \mathrm{d}\mu, \ldots, \int_X f_m \, \mathrm{d}\mu \right) + \lambda \left(\int_X g_1 \, \mathrm{d}\mu, \ldots, \int_X g_m \, \mathrm{d}\mu \right) \\
&= \int_X f \, \mathrm{d}\mu + \lambda \int_X g \, \mathrm{d}\mu.
\end{aligned}$$

If $\int_X f \, \mathrm{d}\mu = 0 \in \mathbb{R}^m$, the inequality is trivial, so we may assume that $\int_X f \, \mathrm{d}\mu \neq 0$.

Let $v \in \mathbb{R}^m \backslash \{0\}$ be arbitrary. Then

$$\left\langle v, \int_X f \, \mathrm{d}\mu \right\rangle = \sum_j v_j \int_X f_j \, \mathrm{d}\mu$$

$$= \int_X \sum_j v_j f_j(x) \, \mathrm{d}\mu(x)$$

$$= \int_X \langle v, f(x) \rangle \, \mathrm{d}\mu(x)$$

$$\leqslant \int_X \|v\| \|f(x)\| \, \mathrm{d}\mu(x) \qquad \text{by Cauchy-Schwarz Inequality}$$

$$= \|v\| \int_X \|f(x)\| \, \mathrm{d}\mu(x).$$

By taking

$$v := \frac{\int_X f \, \mathrm{d}\mu}{\| \int_X f \, \mathrm{d}\mu \|} \in \mathbb{R}^m \backslash \{0\}$$

which has norm 1, we get

$$\left\langle \frac{\int_X f \, \mathrm{d}\mu}{\| \int_X f \, \mathrm{d}\mu \|}, \int_X f \, \mathrm{d}\mu \right\rangle = \left\| \int_X f \, \mathrm{d}\mu \right\|$$

and thus

$$\left\| \int_X f \, \mathrm{d}\mu \right\| \leqslant \int_X \|f(x)\| \, \mathrm{d}\mu(x).$$

$\square$

2.1.5 Almost Everywhere

Definition 2.1.18 (Almost Everywhere). Let (X, μ) be a measure space. We say that a property holds *almost everywhere* (a.e.) if the set of points where the property fails to hold has μ measure zero.

If the property depends on a point $x \in X$, we sometimes say that the property $P(x)$ holds for μ-almost every $x \in X$.

Exercise 2.1.1. Let (X, μ) be a measure space such that μ is complete, i.e. every subset of μ-measure zero set is measurable. (This is true for all outer measures, for instance.) Let Y be a topological space with Borel σ-algebra $\mathcal{B}_Y$.

Suppose that $f : X \to Y$ is a measurable function and that $g : X \to Y$ is a function such that $f(x) = g(x)$ for μ-almost every $x \in X$. Then g is measurable.

Proof. Let $N := \{x \in X : f(x) \neq g(x)\}$, which has measure zero by assumption, and is hence a measurable subset of X by completeness of μ. Let $E \in \mathcal{B}_Y$ be a Borel set. Then

$$g^{-1}(E) = \left(g^{-1}(E)\backslash N\right) \cup \left(g^{-1}(E) \cap N\right) = \left(f^{-1}(E)\backslash N\right) \cup \left(g^{-1}(E) \cap N\right).$$

Since f is measurable, $f^{-1}(E)$ is a measurable subset of X, so the set $f^{-1}(E)\backslash N$ is also a measurable subset of X. Since N has μ-measure zero, every subset of N also has μ-measure zero by monotonicity, so $g^{-1}(E) \cap N$ is a measurable subset of X by completeness of μ. Thus $g^{-1}(E)$ is a union of two measurable sets, and hence is measurable. Since $E \in \mathcal{B}_Y$ was arbitrary, we have that g is measurable. $\qquad\square$

We again use the notation $\mathbb{F}$ to denote either $\mathbb{R}$ or $\mathbb{C}$, and we let $\mathbb{F}_\infty$ be $[-\infty, \infty]$ if $\mathbb{F} = \mathbb{R}$ and $\mathbb{C} \cup \{\infty\}$ if $\mathbb{F} = \mathbb{C}$.

Definition 2.1.19 (Lebesgue Integral over a Measurable Set). Let (X, μ) be a measure space, let $E \subseteq X$ be a measurable set, and let f be a measurable function $X \to \mathbb{F}_\infty$. Then we define the *Lebesgue integral* of f *over* E with respect to μ to be

$$\int_E f \, \mathrm{d}\mu := \int_X f\mathbf{1}_E \, \mathrm{d}\mu$$

if the integral on the right-hand side is defined.

Remark 2.1.20 (Notation for Integrals over Intervals). One should be careful with this notation in the case that E is an interval $[a, b]$ and $X = [-\infty, \infty]$. Following tradition, we would like to write $\int_a^b f(x) \, \mathrm{d}x$ to denote the Lebesgue integral of f over $[a, b]$, but this notation suggests that the integral is "oriented" from a to b, so that

$$\int_a^b f(x) \, \mathrm{d}x = -\int_b^a f(x) \, \mathrm{d}x.$$

In the Lebesgue theory, this is not the case; there is no orientation, and so while it is proper to write $\int_{[a,b]} f(x) \, \mathrm{d}x$, we will take the more traditional notation $\int_a^b f(x) \, \mathrm{d}x$ and just be careful to always have $a \leqslant b$ and never swap the limits of integration.

Remark 2.1.21 (All Functions are Integrable over Sets of Measure Zero). If $\mu(E) = 0$ then every measurable function is integrable over E with integral 0. We check this — let $f : X \to [-\infty, \infty]$ be a measurable function. Then f^+ and f^- are nonnegative measurable functions, and $\mathbf{1}_E f^+$ and $\mathbf{1}_E f^-$ are also nonnegative measurable functions. Notice that each simple function $s : X \to [0, \infty)$ satisfying $s \leqslant \mathbf{1}_E f^+$ must have standard form $s = \sum_{j=0}^N a_j \mathbf{1}_{A_j}$ where $A_j \subseteq E$ for each $j = 1, 2, \ldots, N$ and $A_0 = X\backslash E$; hence $\mu(A_j) = 0$ for each $j = 1, 2, \ldots, N$. Thus by Lemma 2.1.9 for each such simple function s we have

$$\int_X s \, \mathrm{d}\mu = \sum_{j=0}^N a_j \mu(A_j) = a_0 \mu(A_0) + \sum_{j=1}^N a_j \mu(A_j) = 0.$$

Taking the supremum over all such simple functions s gives

$$\int_X \mathbf{1}_E f^+ \, \mathrm{d}\mu = 0.$$

Similarly, we have

$$\int_X \mathbf{1}_E f^- \, \mathrm{d}\mu = 0.$$

Thus the integral $\int_E f \, \mathrm{d}\mu = \int_X \mathbf{1}_E f \, \mathrm{d}\mu$ is defined and equal to 0.

In the case that $f : X \to \mathbb{C} \cup \{\infty\}$ is a complex-valued measurable function, we can write $f = u + iv$ where $u, v : X \to [-\infty, \infty]$ are measurable functions. Then by the above argument, the integrals $\int_E u \, \mathrm{d}\mu$ and $\int_E v \, \mathrm{d}\mu$ are both defined and equal to 0, so the integral $\int_E f \, \mathrm{d}\mu = \int_E u \, \mathrm{d}\mu + i \int_E v \, \mathrm{d}\mu$ is also defined and equal to 0.

Exercise 2.1.2 (Bounding an Integral)**.** Let (X, μ) be a measure space, let $E \subseteq X$ be a measurable set, and let f be a measurable function $X \to \mathbb{F}_\infty$, such that the integral $\int_E |f| \, \mathrm{d}\mu$ is defined. Then

$$\left| \int_E f \, \mathrm{d}\mu \right| \leqslant \mu(E) \sup_{x \in E} |f(x)|.$$

Proof. Let $c := \sup_{x \in E} |f(x)| \in [0, \infty]$. If $c = \infty$, then the inequality is trivial. If $c < \infty$, then

$$\left| \int_E f \, \mathrm{d}\mu \right| = \left| \int_X f \mathbf{1}_E \, \mathrm{d}\mu \right|$$
$$\leqslant \int_X |f| \mathbf{1}_E \, \mathrm{d}\mu$$
$$\leqslant \int_X c \mathbf{1}_E \, \mathrm{d}\mu$$
$$= c\mu(E)$$

where we have used the triangle inequality for integrals in the second line, the definition of c in the third line, and Lemma 2.1.4 in the fourth line. $\square$

Exercise 2.1.3 (Integral of Functions Equal Almost Everywhere)**.** Let (X, μ) be a measure space, and let $f, g : X \to \mathbb{F}_\infty$ be measurable functions such that $f(x) = g(x)$ for μ-almost every $x \in X$. If the integral $\int_X f \, \mathrm{d}\mu$ is defined, then the integral $\int_X g \, \mathrm{d}\mu$ is also defined, and

$$\int_X f \, \mathrm{d}\mu = \int_X g \, \mathrm{d}\mu.$$

Proof. First we assume that $\mathbb{F} = \mathbb{R}$. Then $f^+, f^-, g^+, g^- : X \to [0, \infty]$ are nonnegative measurable functions. Let $E := \{x \in X : f(x) \neq g(x)\}$ so that $\mu(E) = 0$. Then $f(x) = g(x)$ for each $x \in X \backslash E$. Note that $f^+(x) = g^+(x)$ and $f^-(x) = g^-(x)$ for each $x \in X \backslash E$.

Since the integral $\int_X f \, d\mu$ is defined, at least one of the integrals $\int_X f^+ \, d\mu$ or $\int_X f^- \, d\mu$ is finite. Assume without loss of generality that $\int_X f^+ \, d\mu < \infty$. Then

$$\int_E f^+ \, d\mu \leqslant \mu(E) \sup_E f^+ = 0$$

by Exercise 2.1.2; we have that

$$\int_X f^+ \, d\mu = \int_{X \setminus E} f^+ \, d\mu + \int_E f^+ \, d\mu$$

$$= \int_{X \setminus E} g^+ \, d\mu + 0$$

$$= \int_X g^+ \, d\mu - \int_E g^+ \, d\mu$$

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

where we have used the fact that $f^+ = g^+$ on $X \setminus E$ in the second line, and that $\int_E g^+ \, d\mu$ is defined and equal to 0 by remark 2.1.21 in the fourth line. Similarly, we have

$$\int_X f^- \, d\mu = \int_X g^- \, d\mu.$$

Thus the integral of g is defined, and

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

In the case that $\mathbb{F} = \mathbb{C}$, we can write $f = u + iv$ and $g = s + it$ where $u, v, s, t : X \to [-\infty, \infty]$ are measurable functions. Then $u(x) = s(x)$ and $v(x) = t(x)$ for μ-almost every $x \in X$, so by the above argument we have

$$\int_X u \, d\mu = \int_X s \, d\mu, \qquad \int_X v \, d\mu = \int_X t \, d\mu.$$

Thus the integral of g is defined, and

$$\int_X g \, d\mu = \int_X s \, d\mu + i \int_X t \, d\mu = \int_X u \, d\mu + i \int_X v \, d\mu = \int_X f \, d\mu.$$

$\square$

Exercise 2.1.4 (Almost Everywhere Monotonicity). Let (X, μ) be a measure space, and let $f, g : X \to [-\infty, \infty]$ be measurable functions such that $f(x) \leqslant g(x)$ for μ-almost every $x \in X$. If the integrals $\int_X f \, d\mu$ and $\int_X g \, d\mu$ are both defined, then

$$\int_X f \, d\mu \leqslant \int_X g \, d\mu.$$

Proof. Let $E := \{x \in X : f(x) > g(x)\}$ so that $\mu(E) = 0$. Then $f(x) \leqslant g(x)$ for each $x \in X \backslash E$. Define the function $h : X \to [-\infty, \infty]$ by

$$h(x) := \begin{cases} -\infty, & x \in E \\ f(x), & x \in X \backslash E \end{cases}$$

so that $h(x) \leqslant g(x)$ for each $x \in X$, and $h(x) = f(x)$ for μ-almost every $x \in X$. Since the integral $\int_X f \, d\mu$ is defined, the integral $\int_X h \, d\mu$ is also defined and equal to $\int_X f \, d\mu$ by Exercise 2.1.3. Thus by monotonicity of the Lebesgue integral we have

$$\int_X f \, d\mu = \int_X h \, d\mu \leqslant \int_X g \, d\mu$$

as desired. $\square$

Lemma 2.1.22 (Integral on Small Sets is Small). *Let (X, μ) be a measure space, and let $f : X \to [0, \infty]$ be an integrable function. Then for each $\epsilon > 0$, there exists $\delta > 0$ such that for each measurable set $E \subseteq X$ with $\mu(E) < \delta$, we have*

$$\left| \int_E f \, d\mu \right| < \epsilon.$$

Proof. **Exercise.** Recall f is integrable means that f is measurable and $\int_X f \, d\mu < \infty$. Let $\epsilon > 0$. Let $s : X \to [0, \infty)$ be a simple function such that $0 \leqslant h \leqslant f$ and

$$\int_X f \, d\mu - \int_X s \, d\mu < \frac{\epsilon}{2}$$

which exists by Lemma 2.1.6. Let $M := \sup_X s < \infty$. Now choose $\delta < \frac{\epsilon}{2M}$.

Let $E \subseteq X$ be a measurable set with $\mu(E) < \delta$. Then

$$\begin{aligned} \int_E f \, d\mu &= \int_E (f - s) \, d\mu + \int_E s \, d\mu \\ &\leqslant \frac{\epsilon}{2} + M\mu(E) \\ &< \frac{\epsilon}{2} + M\frac{\epsilon}{2M} = \epsilon \end{aligned}$$

as desired. $\square$

Lemma 2.1.23 (Absolute Continuity of the Integral). *Let (X, μ) be a measure space, and let $f : X \to [0, \infty]$ be an integrable function. Then for each $\epsilon > 0$, there exists a measurable set $E \subseteq X$ such that $\mu(E) < \infty$ and*

$$\int_{X \backslash E} f \, d\mu < \epsilon.$$

Proof. Let $\epsilon > 0$. Let P be a μ-partition of X into disjoint measurable sets $A_1, A_2, \ldots, A_n$ such that

$$\int_X g \ \mathrm{d}\mu < \epsilon + L(f, P).$$

Let $E = \bigcup \{A_j : \inf_{A_j} f > 0, 1 \leqslant j \leqslant n\}$. Then $\mu(E) < \infty$, because other wise we would have $L(f, P) = \infty$ which would imply $\int_X f \ \mathrm{d} mu = \infty$ which contradicts the integrability of f.

Now

$$\int_{X \setminus E} f \ \mathrm{d}\mu = \int_X f \ \mathrm{d}\mu - \int_E f \ \mathrm{d}\mu$$

$$< (\epsilon + L(f, P)) - L(\mathbf{1}_E f, P)$$

$$= \epsilon$$

where the last line holds because $\inf_{A_j} f = 0$ for each A_j that is not contained in E.   □

Exercise 2.1.5 (Integral of a Positive Function is Zero iff Function is Zero a.e.)**.** Let (X, μ) be a measure space, and let $f : X \to [0, \infty]$ be a nonnegative measurable function. Then $\int_X f \ \mathrm{d}\mu = 0$ if and only if $f(x) = 0$ for μ-almost every $x \in X$.

Proof. The reverse direction is easy: if $f(x) = 0$ for μ-almost every $x \in X$, then by Remark **??** we have $\int_X f \ \mathrm{d}\mu = 0$.

We prove the reverse direction in a few steps. If f is a simple function

$$f = \sum_{j=1}^{n} a_j \mathbf{1}_{E_j}$$

with $a_j \geqslant 0$ for each $j = 1, 2, \ldots, n$, then by Lemma 2.1.9 we have

$$\int_X f \ \mathrm{d}\mu = \sum_{j=1}^{n} a_j \mu(E_j).$$

Thus the integral of f is zero if and only if for eah $j = 1, 2, \ldots, n$ either $a_j = 0$ or $\mu(E_j) = 0$. In this case, $f(x) = 0$ for μ-almost every $x \in X$.

In general, we prove the contrapositive. Let $f : X \to [0, \infty]$ be a nonnegative measurable function. Then we have

$$\{x \in X : f(x) > 0\} = \bigcup_{j=1}^{\infty} \{x \in X : f(x) > \frac{1}{j}\}.$$

Since f is not zero almost everywhere, then there exists $j \geqslant 1$ such that the set $E_j := \{x \in X : f(x) > \frac{1}{j}\}$ has positive measure. Then $f > \frac{1}{j} \mathbf{1}_{E_j}$, so by monotonicity of the Lebesgue integral we have

$$\int_X f \ \mathrm{d}\mu \geqslant \int_X \frac{1}{j} \mathbf{1}_{E_j} \ \mathrm{d}\mu = \frac{1}{j} \mu(E_j) > 0.$$

Thus if f is not zero almost everywhere, then $\int_X f \ \mathrm{d}\mu > 0$.   □

We end this section with a few comments, which will be useful later.

Note that if $f : X \to \mathbb{F}_\infty$ is an integrable function, i.e. the integral $\int_X f \, \mathrm{d}\mu$ is defined and finite, then the definitions require that the integral $\int_X |f| \, \mathrm{d}\mu$ is also defined and finite (go and check). In particular, this implies that f is finite μ-almost everywhere, because if f were infinite on a set of positive measure, then $|f|$ would also be infinite on that set, and hence the integral $\int_X |f| \, \mathrm{d}\mu$ would be infinite. Thus if $f : X \to \mathbb{F}_\infty$ is an integrable function, then f is finite μ-almost everywhere.

In view of this comment and the results about functions that are equal almost everywhere, we can often treat integrable functions as if they were finite-valued functions. In particular, it is really no loss of generality to say things like "let $f : X \to \mathbb{F}$ be an integrable function" instead of the more correct "let $f : X \to \mathbb{F}_\infty$ be an integrable function". This is becuase if $f : X \to \mathbb{F}_\infty$ is an integrable function, then f is finite μ-almost everywhere, so we can define a function $g : X \to \mathbb{F}$ which is equal to f on the set where f is finite and equal to 0 (or any other finite value) on the set where f is infinite. Then g is a finite-valued function which is equal to f almost everywhere, so by Exercise 2.1.3 the integral of g is defined and equal to the integral of f. We will be slightly lazy and conform to the convention of saying "let $f : X \to \mathbb{F}$ be an integrable function" and proving theorems in that case. However, we should remember that this is just a notational convenience, and that the theorems about integrable functions still hold for integrable functions $f : X \to \mathbb{F}_\infty$, as they must be finite-valued almost everywhere.

2.2 Limits and Integration

In this section, we prove some classical results about limits and integration. In the previous section, we proved the Monotone Convergence Theorem (Theorem 2.1.8). Now we prove the famous Dominated Convergence Theorem.

2.2.1 Dominated Convergence Theorem

Exercise 2.2.1 (Fatou's Lemma, Pointwise a.e. Version)**.** Let (X, μ) be a measure space, and let $\{f_k\}_{k=1}^\infty$ be a sequence of nonnegative measurable functions. Suppose that f is a measurable function such that

$$f(x) = \liminf_{k \to \infty} f_k(x), \qquad \text{for } \mu\text{-a.e. } x \in X.$$

Then

$$\int_X f \, \mathrm{d}\mu \leqslant \liminf_{k \to \infty} \int_X f_k \, \mathrm{d}\mu.$$

Give an example where equality does not hold.

Proof. First we see that f is measurable — the limit inferior of measurable functions $\liminf_{k\to\infty} f_k$ is measurable by Proposition 1.4.11, and changing a function on a set of measure zero does not affect measurability, so f is measurable. Also f must be nonnegative μ-almost everywhere, so the integrals in the statement are well-defined.

For each $k \in \mathbb{Z}^+$ and each $n \geqslant k$, we have

$$\inf_{n \geqslant k} f_n(x) \leqslant f_k(x) \qquad \forall x \in X$$

which implies

$$\int_X \inf_{n \geqslant k} f_n \ \mathrm{d}\mu \leqslant \int_X f_k \ \mathrm{d}\mu$$

for each $n \geqslant k$. Since the left side does not depend on n, we have

$$\int_X \inf_{n \geqslant k} f_n \ \mathrm{d}\mu \leqslant \inf_{m \geqslant k} \int_X f_m \ \mathrm{d}\mu. \qquad (*)$$

The sequence of functions $\{\inf_{n \geqslant k} f_n\}_{k=1}^{\infty}$ is increasing in k and converges pointwise to f as $k \to \infty$. Thus by the Monotone Convergence Theorem (Theorem 2.1.8), we have

$$\begin{aligned}
\int_X f \ \mathrm{d}\mu &= \int_X \lim_{k\to\infty} \inf_{n \geqslant k} f_n \ \mathrm{d}\mu \\
&= \lim_{k\to\infty} \int_X \inf_{n \geqslant k} f_n \ \mathrm{d}\mu \\
&\leqslant \lim_{k\to\infty} \inf_{m \geqslant k} \int_X f_m \ \mathrm{d}\mu \\
&= \liminf_{k\to\infty} \int_X f_k \ \mathrm{d}\mu
\end{aligned}$$

where the inequality follows from $(*)$.

For an example where equality does not hold, let $X = [0,1]$ and consider the Lebesgue outer measure on X. For each $k \in \mathbb{Z}^+$, define the function $f_k : [0,1] \to [0,\infty)$ by

$$f_k(x) = \begin{cases} k, & x \in (0, 1/k) \\ 0, & \text{otherwise} \end{cases}.$$

Then for each $x \in [0,1]$, we have $\liminf_{k\to\infty} f_k(x) = 0$. Thus $f(x) = 0$ for all $x \in [0,1]$, so $\int_{[0,1]} f \ \mathrm{d}x = 0$. However, for each $k \in \mathbb{Z}^+$, we have

$$\int_{[0,1]} f_k \ \mathrm{d}x = \int_0^{1/k} k \ \mathrm{d}x = 1.$$

$\square$

Our limit results are not as general as we would like — the Monotone Convergence Theorem requires the sequence of functions to be nonnegative and increasing, and Fatou's Lemma requires the functions to be nonnegative (or at least uniformly bounded below) and we only get an inequality. The dominated convergence theorem removes these restrictions.

Theorem 2.2.1 (Dominated Convergence Theorem, Pointwise a.e. Version). *Let (X, μ) be a measure space, and let $\{f_k\}_{k=1}^{\infty}$ be a sequence of measurable functions that converge pointwise μ-almost everywhere to a function f. Suppose that there exists an integrable function $g : X \to [0, \infty]$ such that $|f_k(x)| \leqslant g(x)$ for all $k \in \mathbb{Z}^+$ and μ-almost every $x \in X$. Then f is integrable and*

$$\lim_{k \to \infty} \int_X |f_k - f| \ \mathrm{d}\mu = 0,$$

and in particular

$$\lim_{k \to \infty} \int_X f_k \ \mathrm{d}\mu = \int_X f \ \mathrm{d}\mu.$$

Before the proof, we remark that there are several versions of the Dominated Convergence Theorem, as well as Fatou's Lemma. We will not attempt to prove the most general versions, but we will state several versions of practical importance. See the appendix about Modes of Convergence for details.

Proof. See that by assumption that $|f_k(x)| \leqslant g(x)$ for all $k \in \mathbb{Z}^+$ and μ-almost every $x \in X$, and that $\{f_k\}_{k=1}^{\infty}$ converges pointwise μ-almost everywhere to f, so we must have $|f(x)| \leqslant g(x)$ for μ-almost every $x \in X$ as well. As a result, f is measurable and integrable by Exercise 2.1.3 and Exercise 2.1.4. Also the sequence of functions $\{2g - |f_k - f|\}_{k=1}^{\infty}$ is a sequence of nonnegative measurable functions that converges pointwise μ-almost everywhere to the function $2g$.

By Fatou's Lemma (Exercise 2.2.1), we have

$$\int_X 2g \ \mathrm{d}\mu = \int_X \liminf_{k \to \infty} (2g - |f_k - f|) \ \mathrm{d}\mu$$

$$\leqslant \liminf_{k \to \infty} \int_X (2g - |f_k - f|) \ \mathrm{d}\mu$$

$$= \int_X 2g \ \mathrm{d}\mu - \limsup_{k \to \infty} \int_X |f_k - f| \ \mathrm{d}\mu$$

which implies

$$\limsup_{k \to \infty} \int_X |f_k - f| \ \mathrm{d}\mu \leqslant 0.$$

Since $\int_X |f_k - f| \ \mathrm{d}\mu \geqslant 0$ for each $k \in \mathbb{Z}^+$, we have

$$\lim_{k \to \infty} \int_X |f_k - f| \ \mathrm{d}\mu = 0$$

as desired.

By the triangle inequality for the integral, we have that

$$\lim_{k \to \infty} \int_X f_k \ \mathrm{d}\mu = \lim_{k \to \infty} \int_X f \ \mathrm{d}\mu$$

which is the second desired conclusion. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Exercise 2.2.2 (Dominated Convergence Theorem Variant). Let (X, μ) be a measure space, and let $\{f_k\}_{k=1}^{\infty}$ be a sequence of measurable functions that converge pointwise μ-almost everywhere to a function f. Let $\{g_k\}_{k=1}^{\infty}$ be a sequence of integrable functions that converge pointwise μ-almost everywhere to an integrable function $g : X \to [0, \infty]$. Suppose that $|f_k(x)| \leqslant g_k(x)$ for all $k \in \mathbb{Z}^+$ and μ-almost every $x \in X$, and that

$$\lim_{k \to \infty} \int_X g_k \ \mathrm{d}\mu = \int_X g \ \mathrm{d}\mu.$$

Then f is integrable and

$$\lim_{k \to \infty} \int_X |f_k - f| \ \mathrm{d}\mu = 0,$$

and in particular

$$\lim_{k \to \infty} \int_X f_k \ \mathrm{d}\mu = \int_X f \ \mathrm{d}\mu.$$

Proof. See that by assumption that $|f_k(x)| \leqslant g_k(x)$ for all $k \in \mathbb{Z}^+$ and μ-almost every $x \in X$, and that $\{f_k\}_{k=1}^{\infty}$ converges pointwise μ-almost everywhere to f, and that $\{g_k\}_{k=1}^{\infty}$ converges pointwise μ-almost everywhere to g, so we must have $|f(x)| \leqslant g(x)$ for μ-almost every $x \in X$ as well. As a result, f is measurable and integrable by Exercise 2.1.3 and Exercise 2.1.4. Also the sequence of functions $\{g + g_k - |f_k - f|\}_{k=1}^{\infty}$ is a sequence of nonnegative measurable functions that converges pointwise μ-almost everywhere to the function $2g$.

By Fatou's Lemma (Exercise 2.2.1), we have

$$\int_X 2g \ \mathrm{d}\mu = \int_X \liminf_{k \to \infty}(g + g_k - |f_k - f|) \ \mathrm{d}\mu$$

$$\leqslant \liminf_{k \to \infty} \int_X (g + g_k - |f_k - f|) \ \mathrm{d}\mu$$

$$= \int_X g \ \mathrm{d}\mu + \liminf_{k \to \infty} \int_X g_k \ \mathrm{d}\mu - \limsup_{k \to \infty} \int_X |f_k - f| \ \mathrm{d}\mu$$

$$= \int_X 2g \ \mathrm{d}\mu - \limsup_{k \to \infty} \int_X |f_k - f| \ \mathrm{d}\mu$$

where the last equality follows from the assumption that $\lim_{k \to \infty} \int_X g_k \ \mathrm{d}\mu = \int_X g \ \mathrm{d}\mu$. This implies

$$\limsup_{k \to \infty} \int_X |f_k - f| \ \mathrm{d}\mu \leqslant 0.$$

Since $\int_X |f_k - f| \, d\mu \geqslant 0$ for each $k \in \mathbb{Z}^+$, we have

$$\lim_{k \to \infty} \int_X |f_k - f| \, d\mu = 0$$

as desired, and the conclusion $\lim_{k \to \infty} \int_X f_k \, d\mu = \int_X f \, d\mu$ follows from the triangle inequality for the integral. $\qquad\square$

Exercise 2.2.3 (Bounded Convergence Theorem). Let (X, μ) be a measure space with $\mu(X) < \infty$, and let $\{f_k\}_{k=1}^{\infty}$ be a sequence of measurable functions that converge pointwise μ-almost everywhere to a function f. Suppose that there exists $M > 0$ such that $|f_k(x)| \leqslant M$ for all $k \in \mathbb{Z}^+$ and μ-almost every $x \in X$. Then f is integrable and

$$\lim_{k \to \infty} \int_X |f_k - f| \, d\mu = 0,$$

and in particular

$$\lim_{k \to \infty} \int_X f_k \, d\mu = \int_X f \, d\mu.$$

Proof. This is a special case of the Dominated Convergence Theorem — since $\mu(X) < \infty$, the constant function $x \longmapsto M$ is integrable, so we can apply Theorem 2.2.1 with $g(x) = M$ for each $x \in X$. $\qquad\square$

2.2.2   Connection to Riemann Integral

In this section, we review the result that a function on a closed interval is Riemann integrable if and only if it is bounded on that interval and continuous almost everywhere, and we will prove that in this case the Riemann integral equals the Lebesgue integral. We then extend this result to unbounded intervals and unbounded positive functions.

Lemma 2.2.2 (Upper and Lower Riemann Sums via Dyadic Partitions). *Let $f : [a, b] \to \mathbb{R}$ be a bounded function. For each $n \in \mathbb{Z}^+$, let P_n be the partition of $[a, b]$ into 2^n closed almost disjoint subintervals of equal length $\frac{b-a}{2^n}$. Then*

$$\lim_{n \to \infty} L(f, P_n) = L(f, [a, b]) \quad and \quad \lim_{n \to \infty} U(f, P_n) = U(f, [a, b]).$$

Here we are using the notation $L(f, P)$ and $U(f, P)$ for the lower and upper Riemann sums of f with respect to the partition P of $[a, b]$. The notation $L(f, [a, b])$ and $U(f, [a, b])$ denotes the lower and upper Riemann integrals of f on $[a, b]$, i.e. the supremum of the lower Riemann sums and the infimum of the upper Riemann sums, respectively, over all partitions of $[a, b]$.

Proof. Let $n \in \mathbb{Z}^+$. Then P_n is a refinement of P_{n-1}, so we have

$$L(f, P_{n-1}) \leqslant L(f, P_n) \leqslant L(f, [a, b]) \quad \text{and} \quad U(f, [a, b]) \leqslant U(f, P_n) \leqslant U(f, P_{n-1}).$$

Thus the sequence $\{L(f, P_n)\}_{n=1}^\infty$ is increasing and bounded above by $L(f, [a, b])$, so it converges to a limit L_* with $L_* \leqslant L(f, [a, b])$. Similarly, the sequence $\{U(f, P_n)\}_{n=1}^\infty$ is decreasing and bounded below by $U(f, [a, b])$, so it converges to a limit U_* with $U_* \geqslant U(f, [a, b])$.

Let

$$B := \sup_{x \in [a,b]} |f(x)| < \infty$$

so that $|f(x)| \leqslant B$ for each $x \in [a, b]$. For each $n \geqslant 1$ we let $\delta_n := \frac{b-a}{2^n}$ be the length of each subinterval in the partition P_n. Then for each $n \geqslant 1$, each subinterval in the partition P_n is given by

$$[t_{k-1}, t_k] = \left[a + (k-1)\delta_n, a + k\delta_n\right]$$

for some $k = 1, 2, \ldots, 2^n$.

Let $\epsilon > 0$ and choose a partition P_ϵ of $[a, b]$ given by

$$a = x_0 < x_1 < \cdots < x_m = b$$

such that

$$L(f, [a, b]) - \frac{\epsilon}{2} < L(f, P_\epsilon).$$

Let $N \in \mathbb{Z}^+$ be such that

$$3mB \cdot \frac{(b-a)}{2^N} < \frac{\epsilon}{2}.$$

Fix an interval $[x_{j-1}, x_j]$ in the partition P_ϵ for some $1 \leqslant j \leqslant m$. Let

$$j_L := \min\{k : x_{j-1} \leqslant t_k\} \quad \text{and} \quad j_R := \max\{k : t_k \leqslant x_j\}.$$

Notice these sets are nonempty since $t_0 = a \leqslant x_{j-1}$ and $t_{2^N} = b \geqslant x_j$ for each $j = 1, 2, \ldots, m$. The index j_L is the index of the leftmost point in P_N that is in $[x_{j-1}, x_j]$, and j_R is the index of the rightmost point in P_N that is in $[x_{j-1}, x_j]$. By choice of N, there must be at least one subinterval in P_N that is strictly contained in $[x_{j-1}, x_j]$, so we have $j_L < j_R$. Then we have $[t_{j_L}, t_{j_R}] \subset [x_{j-1}, x_j]$ and we claim that

$$t_{j_L} - x_{j-1} \leqslant \delta_N \quad \text{and} \quad x_j - t_{j_R} \leqslant \delta_N.$$

To see this, note that by definition of j_L we have $t_{j_L-1} < x_{j-1} \leqslant t_{j_L}$, so

$$t_{j_L} - x_{j-1} < t_{j_L} - t_{j_L-1} = \delta_N.$$

Similarly, by definition of j_R we have $t_{j_R} \leqslant x_j < t_{j_R+1}$, so

$$x_j - t_{j_R} < t_{j_R+1} - t_{j_R} = \delta_N.$$

Therefore

$$\mathcal{L}^1([x_{j-1}, x_j]\backslash[t_{j_L}, t_{j_R}]) = (x_j - x_{j-1}) - (t_{j_R} - t_{j_L}) = (x_j - t_{j_R}) + (t_{j_L} - x_{j-1}) \leqslant 2\delta_N. \quad (\star)$$

Let P' be the partition of $[a, b]$ given by

$$P' := \{t_{j_L}, t_{j_R} : j = 1, 2, \ldots, m\}.$$

(Sanity check - see that $t_{1_L} = t_0 = a$ and $t_{m_R} = t_{2^N} = b$, so P' includes the endpoints of $[a, b]$. Also, since $j_L < j_R$ and

$$j_R = (j+1)_L \iff x_j \in P_N \quad \text{and} \quad j_R + 1 = (j+1)_L \text{ otherwise}$$

we see that the points in P' are in increasing order, after removing the duplicate points $t_{j_R} = t_{(j+1)_L}$ when $x_j \in P_N$.)

We claim that

$$L(f, P_\epsilon) \leqslant L(f, P') + \frac{\epsilon}{2}.$$

By definition we have

$$L(f, P') = \sum_{j=1}^m \inf_{x \in [t_{j_L}, t_{j_R}]} f(x) \cdot (t_{j_R} - t_{j_L}) + \sum_{j=1}^{m-1} \inf_{x \in [t_{j_R}, t_{(j+1)_L}]} f(x) \cdot (t_{(j+1)_L} - t_{j_R})$$

and the terms $(t_{(j+1)_L} - t_{j_R})$ in the sum on the right are either 0 if $x_j \in P_n$ or equal to δ_N otherwise. Thus we have

$$L(f, P') \geqslant \sum_{j=1}^m \inf_{x \in [t_{j_L}, t_{j_R}]} f(x) \cdot (t_{j_R} - t_{j_L}) - (m-1)B\delta_N \quad (\star\star)$$

since $\inf_{x \in [t_{j_R}, t_{(j+1)_L}]} f(x) \geqslant -B$ for each $j = 1, 2, \ldots, m-1$. Now we can estimate

$$L(f, P_\epsilon) - L(f, P') \leqslant \sum_{j=1}^m \inf_{x \in [x_{j-1}, x_j]} f(x) \cdot (x_j - x_{j-1})$$

$$-\sum_{j=1}^m \inf_{x \in [t_{j_L}, t_{j_R}]} f(x) \cdot (t_{j_R} - t_{j_L}) + (m-1)B\delta_N \quad \text{by } (\star\star)$$

$$\leqslant B \sum_{j=1}^m (x_j - x_{j-1}) - B \sum_{j=1}^m (t_{j_R} - t_{j_L}) + (m-1)B\delta_n \quad \text{by definition of } B$$

$$= B \sum_{j=1}^m \left((x_j - x_{j-1}) - (t_{j_R} - t_{j_L})\right) + (m-1)B\delta_n$$

$$= B \sum_{j=1}^m 2\delta_N + (m-1)B\delta_N \quad \text{by } (\star)$$

$$= 2mB\delta_N + (m-1)B\delta_N$$

$$< 3mB\delta_N$$

$$< \frac{\epsilon}{2} \quad \text{by choice of } N.$$

This implies

$$L(f, P_\epsilon) \leqslant L(f, P') + \frac{\epsilon}{2}$$

as claimed.

Now we notice that P' only consists of points in P_N, so the partition P_N is a refinement of P'. Thus we have the estimate

$$L(f, P_\epsilon) \leqslant L(f, P') + \frac{\epsilon}{2} \leqslant L(f, P_N) + \frac{\epsilon}{2}$$

By choice of the partition P_ϵ we have that

$$L(f, [a, b]) - \frac{\epsilon}{2} < L(f, P_\epsilon) < L(f, P_N) + \frac{\epsilon}{2}$$

which implies

$$L(f, [a, b]) - \epsilon < L(f, P_N).$$

Since $\epsilon > 0$ was arbitrary, we have $L(f, [a, b]) \leqslant L_*$. Since we already had $L_* \leqslant L(f, [a, b])$, we conclude that $L_* = L(f, [a, b])$. A similar argument shows that $U_* = U(f, [a, b])$. This proves

$$\lim_{n \to \infty} L(f, P_n) = L(f, [a, b]) \quad \text{and} \quad \lim_{n \to \infty} U(f, P_n) = U(f, [a, b]).$$

$\square$

Theorem 2.2.3 (Lebesgue Criterion for Riemann Integrability). *Let $f : [a, b] \to \mathbb{R}$ be a function. Then f is Riemann integrable if and only if f is bounded on $[a, b]$ and the set of discontinuities of f has Lebesgue measure zero.*

This fact was proven in Analysis 1, so we do not repeat it here. As a corollary, we remark that every Riemann integrable function is measurable, since every function that is continuous almost everywhere is measurable.

Theorem 2.2.4 (Riemann-Lebesgue Theorem). *Let $f : [a, b] \to \mathbb{R}$ be a bounded measurable function. If f is Riemann integrable, then the Riemann integral equals the Lebesgue integral,*

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

Proof. First note that by assumption, f is a bounded measurable function on a set of finite measure, so f is Lebesgue integrable by Exercise 2.1.2.

Let $n \in \mathbb{Z}^+$. Consider the partition P_n of $[a, b]$ into 2^n closed almost disjoint subintervals $I_1, I_2, \ldots, I_{2^n}$ of equal length $\frac{b-a}{2^n}$. Let

$$g_n := \sum_{j=1}^{2^n} \inf_{x \in I_j} f(x) \mathbf{1}_{I_j} \quad \text{and} \quad h_n := \sum_{j=1}^{2^n} \sup_{x \in I_j} f(x) \mathbf{1}_{I_j}.$$

See that for each g_n and h_n are simple functions, and that $g_n(x) \leqslant f(x) \leqslant h_n(x)$ for each $x \in [a, b]$. Furthermore, the Lebesgue integrals of g_n and h_n are the lower and upper Riemann sums of f with respect to the partition P_n, since

$$\int_{[a,b]} g_n \; \mathrm{d}x = \sum_{j=1}^{2^n} \inf_{x \in I_j} f(x) \mu(I_j) = \sum_{j=1}^{2^n} \inf_{x \in I_j} f(x) \frac{b-a}{2^n} = L(f, P_n, [a, b])$$

and similarly

$$\int_{[a,b]} h_n \; \mathrm{d}x = \sum_{j=1}^{2^n} \sup_{x \in I_j} f(x) \mu(I_j) = \sum_{j=1}^{2^n} \sup_{x \in I_j} f(x) \frac{b-a}{2^n} = U(f, P_n, [a, b]).$$

Now we see that the sequence of functions $\{g_n\}_{n=1}^{\infty}$ is increasing, and the sequence of functions $\{h_n\}_{n=1}^{\infty}$ is decreasing; since $\{g_n\}_{n=1}^{\infty}$ is bounded above by $\sup_{x \in [a,b]} |f(x)|$, it converges pointwise to a function g, and since $\{h_n\}_{n=1}^{\infty}$ is bounded below by $\inf_{x \in [a,b]} |f(x)|$, it also converges pointwise to a function h. Since $g_n(x) \leqslant f(x) \leqslant h_n(x)$ for each $n \in \mathbb{Z}^{+}$ and each $x \in [a, b]$, monotonicity of the Lebesgue integral implies that

$$L(f, P_n, [a, b]) = \int_{[a,b]} g_n \; \mathrm{d}x \leqslant \int_{[a,b]} f \; \mathrm{d}x \leqslant \int_{[a,b]} h_n \; \mathrm{d}x = U(f, P_n, [a, b]).$$

Now Lemma 2.2.2 implies that

$$L(f, [a, b]) \leqslant \int_{[a,b]} f \; \mathrm{d}x \leqslant U(f, [a, b]).$$

Since $\{g_n\}_{n=1}^{\infty}$ converges pointwise to g and $\{h_n\}_{n=1}^{\infty}$ converges pointwise to h and the sequences $\{g_n\}_{n=1}^{\infty}$ and $\{h_n\}_{n=1}^{\infty}$ are bounded above by $\sup_{x \in [a,b]} |f(x)|$, we take limits as $n \to \infty$ and apply the Bounded Convergence Theorem to interchange the limits and the integrals to obtain

$$\lim_{n \to \infty} L(f, P_n, [a, b]) = \int_{[a,b]} g \; \mathrm{d}x \leqslant \int_{[a,b]} f \; \mathrm{d}x \leqslant \int_{[a,b]} h \; \mathrm{d}x = \lim_{n \to \infty} U(f, P_n, [a, b]).$$

Since f is Riemann integrable if and only if $L(f, [a, b]) = U(f, [a, b])$, and in this case the common value equals the Riemann integral, we have that f is Riemann integrable if and only if $L(f, [a, b]) = U(f, [a, b]) = \int_{[a,b]} f \; \mathrm{d}x$.

This proves that if f is a bounded measurable function on $[a, b]$, then f is also Lebesgue integrable on $[a, b]$ and the Riemann integral equals the Lebesgue integral. $\square$

Corollary 2.2.5 (Improper Riemann Integral of Nonnegative Function is Lebesgue Integral).
Let $f : [0, \infty) \to [0, \infty)$ be a nonnegative function such that the improper Riemann integral $\int_0^{\infty} f(x) \; \mathrm{d}x$ converges. That is, we require that the limit

$$\int_0^{\infty} f(x) \; \mathrm{d}x := \lim_{R \to \infty} \int_0^R f(x) \; \mathrm{d}x$$

exists and is finite. Then f is Lebesgue integrable over $[0, \infty)$, and the improper Riemann integral equals the Lebesgue integral,

$$\int_0^\infty f(x) \, \mathrm{d}x = \int_{[0,\infty)} f \, \mathrm{d}x.$$

A similar result holds if the domain $[0, \infty)$ is replaced by $(-\infty, 0]$ or $\mathbb{R}$.

Remark 2.2.6 (Other Improper Riemann Integrals). In the above, we have only defined Riemann integrals in the special case where we take the limit of one or both endpoints of a closed interval to infinity. There are more cases in which we could do something similar.

Say we have a function $f : (a, b) \to \mathbb{R}$ where $-\infty \leqslant a < b \leqslant \infty$, so f is not defined at one or both endpoints. We define the improper Riemann integral $\int_a^b f(x) \, \mathrm{d}x$ to be convergent if there exists $c \in (a, b)$ such that both improper Riemann integrals

$$\int_a^c f(x) \, \mathrm{d}x := \lim_{t \to a^+} \int_t^c f(x) \, \mathrm{d}x \quad \text{and} \quad \int_c^b f(x) \, \mathrm{d}x := \lim_{t \to b^-} \int_c^t f(x) \, \mathrm{d}x$$

converge, and in this case we define

$$\int_a^b f(x) \, \mathrm{d}x := \int_a^c f(x) \, \mathrm{d}x + \int_c^b f(x) \, \mathrm{d}x.$$

(The choice of c does not matter, as can be seen by splitting the integral at a different point $d \in (a, b)$ — see that if $c < d$, then

$$\int_a^c f(x) \, \mathrm{d}x + \int_c^d f(x) \, \mathrm{d}x = \int_a^d f(x) \, \mathrm{d}x$$

which implies

$$\int_a^c f(x) \, \mathrm{d}x + \int_c^b f(x) \, \mathrm{d}x = \int_a^c f(x) \, \mathrm{d}x + \int_c^d f(x) \, \mathrm{d}x + \int_d^b f(x) \, \mathrm{d}x = \int_a^d f(x) \, \mathrm{d}x + \int_d^b f(x) \, \mathrm{d}x.$$

The case $d < c$ is similar.)

Of course, for these limits to make sense, we need to assume that f is Riemann integrable on each closed subinterval of (a, b). Otherwise, these limits will not be defined.

The same result as in Corollary 2.2.5 holds in this more general setting, with basically the same proof. That is, if $f : (a, b) \to [0, \infty)$ is a nonnegative function such that the improper Riemann integral $\int_a^b f(x) \, \mathrm{d}x$ converges, then f is Lebesgue integrable over (a, b), and the improper Riemann integral equals the Lebesgue integral,

$$\int_a^b f(x) \, \mathrm{d}x = \int_{(a,b)} f \, \mathrm{d}x.$$

Proof. Let $f : [0, \infty) \to [0, \infty)$ be such that the limit

$$\lim_{R \to \infty} \int_0^R f(x) \, \mathrm{d}x$$

exists and is finite. Then for each $R > 0$, the function f is Riemann integrable on the closed interval $[0, R]$. By Theorem 2.2.4, we know that f is continuous almost everywhere on $[0, R]$, and that the Riemann integral equals the Lebesgue integral, i.e.

$$\int_0^R f(x) \, \mathrm{d}x = \int_{[0,R]} f \, \mathrm{d}x.$$

Since f is nonnegative, the sequence of function $\{\mathbf{1}_{[0,k]} f\}_{k=1}^\infty$ is an increasing sequence of nonnegative measurable functions that converges pointwise to f as $k \to \infty$. Thus by the Monotone Convergence Theorem (Theorem 2.1.8), we have

$$\int_{[0,\infty)} f \, \mathrm{d}x = \int_{[0,\infty)} \lim_{k \to \infty} \mathbf{1}_{[0,k]} f \, \mathrm{d}x$$

$$= \lim_{k \to \infty} \int_{[0,k]} f \, \mathrm{d}x$$

$$= \lim_{k \to \infty} \int_0^k f(x) \, \mathrm{d}x$$

which is finite by assumption.

(The Monotone Convergence Theorem is used in the first line, the definition of the Lebesgue integral over $[0, k]$ is used in the second line, and the equality of the Riemann and Lebesgue integrals over $[0, k]$ is used in the third line.)

The other cases are similar, and follow from the first. If $f : (-\infty, 0] \to [0, \infty)$ is such that the improper Riemann integral $\int_{-\infty}^0 f(x) \, \mathrm{d}x$ converges, then define $g : [0, \infty) \to [0, \infty)$ by $g(x) := f(-x)$. Then the improper Riemann integral $\int_0^\infty g(x) \, \mathrm{d}x$ converges, so by the above result g is Lebesgue integrable over $[0, \infty)$ and

$$\int_0^\infty g(x) \, \mathrm{d}x = \int_{[0,\infty)} g \, \mathrm{d}x.$$

By a change of variables, we have

$$\int_{-\infty}^0 f(x) \, \mathrm{d}x = \int_0^\infty g(x) \, \mathrm{d}x = \int_{[0,\infty)} g \, \mathrm{d}x = \int_{(-\infty,0]} f \, \mathrm{d}x$$

as desired.

If $f : \mathbb{R} \to [0, \infty)$ is such that the improper Riemann integral $\int_{-\infty}^\infty f(x) \, \mathrm{d}x$ converges, then both improper Riemann integrals $\int_{-\infty}^0 f(x) \, \mathrm{d}x$ and $\int_0^\infty f(x) \, \mathrm{d}x$ converge; thus by the above results, f is Lebesgue integrable over both $(-\infty, 0]$ and $[0, \infty)$, and

$$\int_{-\infty}^0 f(x) \, \mathrm{d}x = \int_{(-\infty,0]} f \, \mathrm{d}x, \qquad \int_0^\infty f(x) \, \mathrm{d}x = \int_{[0,\infty)} f \, \mathrm{d}x.$$

Since f is nonnegative, it is integrable over $\mathbb{R}$, and

$$\int_{-\infty}^{\infty} f(x)\ \mathrm{d}x = \int_{-\infty}^{0} f(x)\ \mathrm{d}x + \int_{0}^{\infty} f(x)\ \mathrm{d}x$$

$$= \int_{(-\infty,0]} f\ \mathrm{d}x + \int_{[0,\infty)} f\ \mathrm{d}x$$

$$= \int_{\mathbb{R}} f\ \mathrm{d}x$$

as desired.   □

Exercise 2.2.4 (Example where Improper Riemann Integral Converges but not Lebesgue Integral)**.** Show that the conclusion of Corollary 2.2.5 can fail if f is allowed to take negative values. In particular, there is a continuous function $f : [0, \infty) \to \mathbb{R}$ such that the improper Riemann integral $\int_{0}^{\infty} f(x)\ \mathrm{d}x$ converges, but f is not Lebesgue integrable over $[0, \infty)$.

Proof. Let

$$f(x) = \frac{\sin x}{x}$$

for $x > 1$ and $f(x) = \sin(1)$ for $0 \leqslant x \leqslant 1$. Then f is continuous on $[0, \infty)$, and for each $R > 1$ we have

$$\int_{1}^{R} f(x)\ \mathrm{d}x = \int_{1}^{R} \frac{\sin x}{x}\ \mathrm{d}x$$

$$= -\frac{\cos x}{x}\Big|_{1}^{R} - \int_{1}^{R} \frac{\cos(x)}{x^2}\ \mathrm{d}x$$

$$= -\frac{\cos(R)}{R} + \cos(1) - \int_{1}^{R} \frac{\cos(x)}{x^2}\ \mathrm{d}x$$

by using integration by parts. Thus for each $R > 1$ we estimate

$$\left| \int_{1}^{R} \frac{\sin x}{x}\ \mathrm{d}x \right| \leqslant \left| \frac{\cos(R)}{R} \right| + \cos(1) + \left| \int_{1}^{R} \frac{\cos(x)}{x^2}\ \mathrm{d}x \right|$$

$$\leqslant \frac{1}{R} + \cos(1) + \int_{1}^{R} \frac{1}{x^2}\ \mathrm{d}x$$

$$\leqslant 1 + \cos(1) + \int_{1}^{\infty} \frac{1}{x^2}\ \mathrm{d}x.$$

In particular, the function

$$[1, \infty) \ni R \mapsto \left| \int_{1}^{R} \frac{\sin x}{x}\ \mathrm{d}x \right|$$

is increasing and bounded above, so the limit as $R \to \infty$ exists and is finite. Thus the improper Riemann integral $\int_1^\infty f(x) \, \mathrm{d}x$ converges. The (proper) Riemann integral $\int_0^1 f(x) \, \mathrm{d}x$ also converges, so the improper Riemann integral $\int_0^\infty f(x) \, \mathrm{d}x$ converges.

However, the Lebesgue integral $\int_{[0,\infty)} f \, \mathrm{d}x$ does not converge. To see this, note that for each integer $k \geqslant 1$ we have

$$\int_{k\pi}^{(k+1)\pi} \frac{|\sin x|}{x} \, \mathrm{d}x \geqslant \frac{1}{(k+1)\pi} \int_{k\pi}^{(k+1)\pi} |\sin x| \, \mathrm{d}x = \frac{1}{(k+1)\pi} \int_0^\pi \sin t \, \mathrm{d}t = \frac{2}{(k+1)\pi}.$$

Thus we have

$$\int_{[0,\infty)} \frac{|\sin x|}{x} \, \mathrm{d}x \geqslant \int_{[\pi,\infty)} \frac{|\sin x|}{x} \, \mathrm{d}x \geqslant \sum_{k=1}^\infty \int_{k\pi}^{(k+1)\pi} \frac{|\sin x|}{x} \, \mathrm{d}x \geqslant \sum_{k=1}^\infty \frac{2}{(k+1)\pi} = \infty$$

since the last series on the right is a constant multiple of the harmonic series and thus diverges. Thus the Lebesgue integral $\int_{[0,\infty)} f \, \mathrm{d}x$ does not converge, so f is not Lebesgue integrable over $[0, \infty)$. □

    Differentiation under integral sign
    Complex and Vector Valued Functions

2.3 The Lebesgue L^p Spaces

2.3.1 Introduction

2.3.2 L^1 Functions

Definition 2.3.1 (L^1 Norm). Let (X, μ) be a measure space. The L^1-*norm* of a measurable function $f : X \to [-\infty, \infty]$ or $f : X \to \mathbb{C} \cup \{\infty\}$ is defined to be

$$\|f\|_{L^1(X,\mu)} = \int_X |f| \, \mathrm{d}\mu.$$

We define the space $L^1(X, \mu)$ to be the set of all equivalence classes of measurable functions $f : X \to [-\infty, \infty]$ or $f : X \to \mathbb{C} \cup \{\infty\}$ such that $\|f\|_{L^1} < \infty$, where two functions are considered equivalent if they are equal μ-almost everywhere.

    We will commit the usual sin and write $f \in L^1(X, \mu)$ to mean that f is a representative of an equivalence class in $L^1(X, \mu)$. That is, we will think about $L^1(X, \mu)$ as a set of functions for the most part. We must be careful to remember that if $f = g$ as equivalence classes in $L^1(X, \mu)$, then f and g may differ on a set of measure zero; however, this will not cause any issues in practice, as we only care about properties that hold almost everywhere.

We also remark that by the Lebesgue philosophy we follow, we can allow function f which take on the values $\pm\infty$ on sets of measure zero without causing any issues. As a result, for each equivalence class in $L^1(X,\mu)$, there exists a representative $f : X \to \mathbb{R}$ or $f : X \to \mathbb{C}$ such that $\|f\|_{L^1} < \infty$. Because of this fact, we usually state our result for real-valued or complex-valued functions, and the reader should understand that we are really working with equivalence classes in $L^1(X,\mu)$, and the results also hold for functions in $L^1(X,\mu)$ that take on the values $\pm\infty$ on sets of measure zero.

Depending on context and what we want to emphasize, we also may omit the μ or both X and μ in the notation when the measure space is clear from context. Also it should either not matter or be clear from context whether we are considering real-valued or complex-valued functions, so we may omit the field $\mathbb{R}$ or $\mathbb{C}$ in the notation as well.

Proposition 2.3.2 ($\|\cdot\|_{L^1}$ is a Norm on L^1). *Let (X,μ) be a measure space. Then $\|\cdot\|_{L^1(X,\mu)}$ is a norm on $L^1(X,\mu)$.*

Proof. We verify the three properties of a norm. Note that $\|\cdot\|_{L^1}$ is well-defined on equivalence classes in $L^1(X,\mu)$ by Exercise 2.1.3.

Suppose that f is a representative of an equivalence class in $L^1(X,\mu)$. Then f is measurable and $\|f\|_{L^1} < \infty$, and we have $\|f\|_{L^1} \geqslant 0$ by definition of the integral.

Positive Definiteness. See that $\|f\|_{L^1} = 0$ if and only if $\int_X |f|\,\mathrm{d}\mu = 0$, which holds if and only if every simple function s with $0 \leqslant s \leqslant |f|$ is zero μ-almost everywhere. This holds if and only if $|f| = 0$ μ-almost everywhere by Exercise 2.1.5. Since $|f| = 0$ μ-almost everywhere if and only if $f = 0$ μ-almost everywhere, we have $\|f\|_{L^1} = 0$ if and only if f is in the same equivalence class as the zero function in $L^1(X,\mu)$.

Homogeneity. For each $\alpha \in \mathbb{R}$ we see that

$$\|\alpha f\|_{L^1} = \int_X |\alpha f|\,\mathrm{d}\mu = |\alpha| \int_X |f|\,\mathrm{d}\mu = |\alpha|\|f\|_{L^1}$$

by homogeneity property of the integral.

Triangle Inequality. Now let g be another representative of an equivalence class in $L^1(X,\mu)$. Then g is measurable and $\|g\|_{L^1} < \infty$. See that $|f + g| \leqslant |f| + |g|$ pointwise, so by the additivity of the integral we have

$$\|f + g\|_{L^1} = \int_X |f + g|\,\mathrm{d}\mu \leqslant \int_X (|f| + |g|)\,\mathrm{d}\mu = \int_X |f|\,\mathrm{d}\mu + \int_X |g|\,\mathrm{d}\mu = \|f\|_{L^1} + \|g\|_{L^1}.$$

$\square$

The above proof is one of the few times we will explicitly view $L^1(X,\mu)$ as a set of equivalence classes, but shows why we need to do so — if we did not identify functions that are equal almost everywhere, then the positive definiteness property would fail.

From now on, we will always furnish $L^1(X, \mu)$ with the L^1-norm and view it as a normed vector space. In particular, if $\{f_k\}_{k=1}^\infty$ is a sequence of functions in $L^1(X, \mu)$, we say that $\{f_k\}_{k=1}^\infty$ converges in L^1 to a function $f \in L^1(X, \mu)$ if

$$\lim_{k \to \infty} \|f_k - f\|_{L^1} = 0.$$

Theorem 2.3.3 (Restatement of Dominated Convergence Theorem). *Let (X, μ) be a measure space, and let $\{f_k\}_{k=1}^\infty \subseteq L^1(X, \mu)$ be a sequence of functions that converge pointwise μ-almost everywhere to a function f. Suppose that there exists an integrable function $g : X \to [0, \infty]$ such that $|f_k(x)| \leqslant g(x)$ for all $k \in \mathbb{Z}^+$ and μ-almost every $x \in X$. Then $f \in L^1(X, \mu)$ and $\{f_k\}_{k=1}^\infty$ converges in L^1-norm to f.*

Proposition 2.3.4 (L^1 Convergence implies a Subsequence Converges a.e.). *Let (X, μ) be a measure space, and let $\{f_k\}_{k=1}^\infty \subseteq L^1(X, \mu)$ be a sequence of functions that converges in L^1-norm to a function $f \in L^1(X, \mu)$. Then there exists a subsequence $\{f_{k_j}\}_{j=1}^\infty$ that converges pointwise μ-almost everywhere to f.*

Proof. Since $\{f_k\}_{k=1}^\infty$ converges in L^1-norm to f, we have

$$\lim_{k \to \infty} \|f_k - f\|_{L^1} = 0$$

which means

$$\lim_{k \to \infty} \int_X |f_k - f| \, \mathrm{d}\mu = 0.$$

For each $j \in \mathbb{Z}^+$, we can find $k_j \in \mathbb{Z}^+$ such that

$$\int_X |f_{k_j} - f| \, \mathrm{d}\mu < 2^{-j}.$$

Then we have

$$\sum_{j=0}^\infty \int_X |f_{k_j} - f| \, \mathrm{d}\mu < \sum_{j=0}^\infty 2^{-j} = 1 < \infty$$

so the Monotone Convergence Theorem (Theorem 2.1.8) implies that

$$\int_X \sum_{j=0}^\infty |f_{k_j} - f| \, \mathrm{d}\mu < \infty.$$

Thus the function $\sum_{j=0}^\infty |f_{k_j} - f|$ is finite μ-almost everywhere, so the sequence $\{|f_{k_j} - f|\}_{j=1}^\infty$ converges to zero μ-almost everywhere; that is, the sequence $\{f_{k_j}\}_{j=1}^\infty$ converges pointwise μ-almost everywhere to f. $\square$

Proposition 2.3.5 (L^1 Approximation by Simple Functions). *Let (X, μ) be a measure space, and let $f \in L^1(X, \mu)$. Then there exists a sequence of simple functions $\{s_k\}_{k=1}^\infty$ such that*

$$\lim_{k \to \infty} \|f - s_k\|_{L^1} = 0.$$

For an approximation by continuous functions, we need LCH spaces — look ahead to that section.

Proof. First consider the case where the field is $\mathbb{R}$. Then $f : X \to [-\infty, \infty]$ is a measurable function such that $\|f\|_{L^1} < \infty$.

Let $\epsilon > 0$. Then by Lemma 2.1.6, there exist simple functions $s_1, s_2 : X \to [0, \infty)$ such that $0 \leqslant s_1 \leqslant f^+$, $0 \leqslant s_2 \leqslant f^-$, and

$$\int_X (f^+ - s_1) \, \mathrm{d}\mu < \epsilon/2 \quad \text{and} \quad \int_X (f^- - s_2) \, \mathrm{d}\mu < \epsilon/2.$$

Define the simple function $s : X \to [-\infty, \infty]$ by $s = s_1 - s_2$. Then we have

$$\begin{aligned}
\|f - s\|_{L^1} &= \|f^+ - f^- - (s_1 - s_2)\|_{L^1} \\
&\leqslant \|f^+ - s_1\|_{L^1} + \|f^- - s_2\|_{L^1} \\
&= \int_X |f^+ - s_1| \, \mathrm{d}\mu + \int_X |f^- - s_2| \, \mathrm{d}\mu \\
&= \int_X (f^+ - s_1) \, \mathrm{d}\mu + \int_X (f^- - s_2) \, \mathrm{d}\mu < \epsilon.
\end{aligned}$$

In particular, we see that for each $k \in \mathbb{Z}^+$, there exists a simple function $s_k : X \to [-\infty, \infty]$ such that $\|f - s_k\|_{L^1} < 1/k$. Then the sequence of simple functions $\{s_k\}_{k=1}^\infty$ satisfies

$$\lim_{k \to \infty} \|f - s_k\|_{L^1} = 0$$

as desired.

Now consider the case where the field is $\mathbb{C}$. Then $f : X \to \mathbb{C} \cup \{\infty\}$ is a measurable function such that $\|f\|_{L^1} < \infty$. Write $f = \Re(f) + i\Im(f)$, where $\Re(f), \Im(f) : X \to [-\infty, \infty]$ are measurable functions such that $\|\Re(f)\|_{L^1}, \|\Im(f)\|_{L^1} < \infty$. By the above case, there exist sequences of simple functions $\{s_k\}_{k=1}^\infty$ and $\{t_k\}_{k=1}^\infty$ such that

$$\lim_{k \to \infty} \|\Re(f) - s_k\|_{L^1} = 0 \quad \text{and} \quad \lim_{k \to \infty} \|\Im(f) - t_k\|_{L^1} = 0.$$

Define the sequence of simple functions $\{u_k\}_{k=1}^\infty$ by $u_k = s_k + it_k$ for each $k \in \mathbb{Z}^+$. Then we have

$$\begin{aligned}
\|f - u_k\|_{L^1} &\leqslant \|\Re(f) - s_k| + |\Im(f) - t_k| \\
&\leqslant \|\Re(f) - s_k\|_{L^1} + \|\Im(f) - t_k\|_{L^1}
\end{aligned}$$

for each $k \in \mathbb{Z}^+$, so

$$\lim_{k \to \infty} \|f - u_k\|_{L^1} = 0$$

as desired. $\square$

Exercise 2.3.1 (Chebyshev's Inequality)**.** Let (X, μ) be a measure space, and let $f \in L^1(X, \mu)$. Then for each $t > 0$, we have

$$\mu(\{x \in X : |f(x)| \geqslant t\}) \leqslant \frac{1}{t}\|f\|_{L^1}.$$

Proof. Fix $t > 0$. Then

$$\begin{aligned}
\mu(\{x \in X : |f(x)| \geqslant t\}) &= \frac{1}{t}\int_{\{|f|\geqslant t\}} t \; \mathrm{d}\mu \\
&\leqslant \frac{1}{t}\int_{\{|f|\geqslant t\}} |f| \; \mathrm{d}\mu \\
&\leqslant \frac{1}{t}\int_X |f| \; \mathrm{d}\mu = \frac{1}{t}\|f\|_{L^1}
\end{aligned}$$

as desired. □

2.3.3 L^∞ Functions

Definition 2.3.6 (L^∞ Norm)**.** Let (X, μ) be a measure space. The L^∞-*norm* of a measurable function $f : X \to [-\infty, \infty]$ or $f : X \to \mathbb{C} \cup \{\infty\}$ is defined to be

$$\|f\|_{L^\infty(X,\mu)} = \inf\{M \geqslant 0 : |f(x)| \leqslant M \text{ for } \mu\text{-almost every } x \in X\}.$$

We define the space $L^\infty(X, \mu)$ to be the set of all equivalence classes of measurable functions $f : X \to [-\infty, \infty]$ such that $\|f\|_{L^\infty} < \infty$, where two functions are considered equivalent if they are equal μ-almost everywhere.

As with $L^1(X, \mu)$, we will often write $f \in L^\infty(X, \mu)$ to mean that f is a representative of an equivalence class in $L^\infty(X, \mu)$. We may also omit the μ or both X and μ in the notation when the measure space is clear from context.

Remark 2.3.7 (Essential Supremum)**.** The L^∞ norm is also called the essential supremum norm. The essential supremum of a set $A \subseteq \mathbb{R}$ is defined to be the infimum of all $M \in \mathbb{R}$ such that the set $\{x \in A : x > M\}$ has Lebesgue measure zero. With this terminology, we can rewrite the L^∞ norm as

$$\|f\|_{L^\infty(X,\mu)} = \operatorname{ess\,sup}_{x \in X} |f(x)|.$$

Because of this language, sometimes people say that L^∞ is the space of essentially bounded functions.

Proposition 2.3.8 ($\|\cdot\|_{L^\infty}$ is a Norm on L^∞)**.** *Let (X, μ) be a measure space. Then $\|\cdot\|_{L^\infty(X,\mu)}$ is a norm on $L^\infty(X, \mu)$.*

Proof. We verify the three properties of a norm. Note that $\|\cdot\|_{L^\infty}$ is well-defined on equivalence classes in $L^\infty(X,\mu)$ since the essential supremum of two functions that are equal almost everywhere is the same.

Suppose that f is a representative of an equivalence class in $L^\infty(X,\mu)$. Then f is measurable and $\|f\|_{L^\infty} < \infty$, and we have $\|f\|_{L^\infty} \geqslant 0$ by definition of the essential supremum.

Positive Definiteness. See that $\|f\|_{L^\infty} = 0$ if and only if $|f(x)| \leqslant 0$ for μ-almost every $x \in X$, which holds if and only if $f = 0$ μ-almost everywhere, so f is in the same equivalence class as the zero function in $L^\infty(X,\mu)$.

Homogeneity. For each $\alpha \in \mathbb{R}$ we see that

$$\|\alpha f\|_{L^\infty} = \inf\{M \geqslant 0 : |\alpha f(x)| \leqslant M \text{ for } \mu\text{-almost every } x \in X\}$$
$$= |\alpha| \inf\{M \geqslant 0 : |f(x)| \leqslant M \text{ for } \mu\text{-almost every } x \in X\}$$
$$= |\alpha| \|f\|_{L^\infty}.$$

Triangle Inequality. Now let g be another representative of an equivalence class in $L^\infty(X,\mu)$. Then g is measurable and $\|g\|_{L^\infty} < \infty$. See that $|f(x) + g(x)| \leqslant |f(x)| + |g(x)|$ for each $x \in X$, and $|f(x)| \leqslant \|f\|_{L^\infty}$ and $|g(x)| \leqslant \|g\|_{L^\infty}$ for μ-almost every $x \in X$, so

$$|f(x) + g(x)| \leqslant \|f\|_{L^\infty} + \|g\|_{L^\infty}$$

for μ-almost every $x \in X$. Thus by definition of the L^∞ norm, we have

$$\|f + g\|_{L^\infty} = \inf\{M \geqslant 0 : |f(x) + g(x)| \leqslant M \text{ for } \mu\text{-almost every } x \in X\} \leqslant \|f\|_{L^\infty} + \|g\|_{L^\infty}.$$

This completes the verification that $\|\cdot\|_{L^\infty}$ is a norm on $L^\infty(X,\mu)$.

As before, this is one of the few times we will explicitly view $L^\infty(X,\mu)$ as a set of equivalence classes, but shows why we need to do so — if we did not identify functions that are equal almost everywhere, then the positive definiteness property would fail. $\qquad\square$

From now on, we will always furnish $L^\infty(X,\mu)$ with the L^∞-norm and view it as a normed vector space. In particular, if $\{f_k\}_{k=1}^\infty$ is a sequence of functions in $L^\infty(X,\mu)$, we say that $\{f_k\}_{k=1}^\infty$ converges in L^∞ to a function $f \in L^\infty(X,\mu)$ if

$$\lim_{k\to\infty} \|f_k - f\|_{L^\infty} = 0.$$

2.3.4 L^p Spaces

We have already defined the L^1 and L^∞ norms and spaces in Definitions 2.3.1 and 2.3.6. We now define the L^p norms and spaces for $1 < p < \infty$.

Definition 2.3.9 (L^p Norm, L^p Space). Let (X, μ) be a measure space and let $1 \leqslant p < \infty$. For a measurable function $f : X \to [-\infty, \infty]$ or $f : X \to \mathbb{C} \cup \{\infty\}$, we define the L^p-*norm* of f to be

$$\|f\|_{L^p(X,\mu)} = \left(\int_X |f|^p \, \mathrm{d}\mu \right)^{1/p},$$

where we interpret the integral as $+\infty$ if $|f|^p$ is not integrable. We define the L^p *space* on (X, μ) to be the set of all equivalence classes of measurable functions $f : X \to [-\infty, \infty]$ such that $\|f\|_{L^p(X,\mu)} < \infty$, where two functions are considered equivalent if they are equal μ-almost everywhere.

For $p = 1$ we see that this definition agrees with Definition 2.3.1. At this point, you should question whether the L^p norm is actually a norm. The hard thing to verify is the triangle inequality, which in this case is known as Minkowski's Inequality. We will prove this in Theorem 2.4.9 in the appendix on inequalities.

Also we note that since we are working with equivalence classes of functions, we can just look at functions which take values in $\mathbb{R}$ or $\mathbb{C}$ instead of $[-\infty, \infty]$ or $\mathbb{C} \cup \{\infty\}$, since if a function takes the value $\pm\infty$ or ∞ on a set of positive measure, then its L^p norm is infinite. Thus we can think of $L^p(X, \mu)$ as a space of equivalence classes of functions $f : X \to \mathbb{R}$ or $f : X \to \mathbb{C}$ with finite L^p norm. In reality, these equivalence classes also contain functions which are infinite on sets of measure zero, but this does not really matter.

Lemma 2.3.10 ($\| \cdot \|_{L^p}$ is a Norm on L^p). *Let (X, μ) be a measure space and let $1 \leqslant p < \infty$. Then $\| \cdot \|_{L^p(X,\mu)}$ is a norm on $L^p(X, \mu)$.*

Proof. Like we said, we defer the proof of the triangle inequality to Theorem 2.4.9 in the appendix on inequalities. The other properties of a norm are easy to verify. First note that the L^p norm is well-defined on equivalence classes of functions since if f and g are two representatives of the same equivalence class, then $f = g$ μ-almost everywhere, so $|f|^p = |g|^p$ μ-almost everywhere, so $\int_X |f|^p \, \mathrm{d}\mu = \int_X |g|^p \, \mathrm{d}\mu$, so $\|f\|_{L^p(X,\mu)} = \|g\|_{L^p(X,\mu)}$.

Let $f : X \to [-\infty, \infty]$ or $f : X \to \mathbb{C} \cup \{\infty\}$ be a representative of an equivalence class in $L^p(X, \mu)$. Then $\|f\|_{L^p(X,\mu)} \geqslant 0$ since the integral of a nonnegative function is nonnegative. Also, $\|f\|_{L^p(X,\mu)} = 0$ if and only if $\int_X |f|^p \, \mathrm{d}\mu = 0$, which holds if and only if $|f|^p = 0$ μ-almost everywhere, which holds if and only if $f = 0$ μ-almost everywhere. Thus $\|f\|_{L^p(X,\mu)} = 0$ if and only if f is the zero equivalence class in $L^p(X, \mu)$. This shows that the L^p norm is positive definite.

Now we also let $\alpha \in \mathbb{R}$ and let $f : X \to [-\infty, \infty]$ or $f : X \to \mathbb{C} \cup \{\infty\}$ be a representative of an equivalence class in $L^p(X, \mu)$. Then

$$\|\alpha f\|_{L^p(X,\mu)} = \left(\int_X |\alpha f|^p \, \mathrm{d}\mu \right)^{1/p} = \left(|\alpha|^p \int_X |f|^p \, \mathrm{d}\mu \right)^{1/p} = |\alpha| \left(\int_X |f|^p \, \mathrm{d}\mu \right)^{1/p} = |\alpha| \|f\|_{L^p(X,\mu)}.$$

This shows that the L^p norm is absolutely homogeneous. $\square$

This is one of the few times we will work with equivalence classes of functions instead of functions themselves, but the proof of the above theorem shows why we must do so — if we do not identify functions that are equal almost everywhere, then the L^p norm is not positive definite. From now on, we will commit the usual sin of writing $f \in L^p(X, \mu)$ to mean that f is a function from X to $\mathbb{R}$ or $\mathbb{C}$ which has finite L^p norm.

Also from this point on, we will think of $L^p(X, \mu)$ as a normed vector space with the L^p norm.

Definition 2.3.11 (Convergence in L^p)**.** Let (X, μ) be a measure space and let $1 \leqslant p < \infty$. Let $\{f_n\}_{n=1}^{\infty}$ be a sequence of measurable functions and let f be a measurable function. We say that $\{f_n\}_{n=1}^{\infty}$ *converges to f in L^p norm* if

$$\lim_{n \to \infty} \|f_n - f\|_{L^p(X,\mu)} = 0.$$

We will study the properties of convergence in L^p norm in later sections.

2.4   Inequality Party

Get ready to party.

2.4.1   Hölder's Inequality

Lemma 2.4.1 (Young's Inequality)**.** *Let $p, q \in (1, \infty)$ be such that $\frac{1}{p} + \frac{1}{q} = 1$. Then for all $a, b \geqslant 0$, we have*

$$ab \leqslant \frac{a^p}{p} + \frac{b^q}{q},$$

with equality if and only if $a^p = b^q$.

Proof. If $a = 0$ or $b = 0$, then the result is trivial, as the left side is zero and the right side is nonnegative; the right side is zero if and only if $a = b = 0$, which agrees with the equality condition.

Fix $b > 0$. Define the function $f : (0, \infty) \to \mathbb{R}$ by

$$f(a) = \frac{a^p}{p} + \frac{b^q}{q} - ab.$$

Then f is differentiable and

$$f'(a) = a^{p-1} - b \qquad \forall a > 0.$$

See that $f'(a) = 0$ if and only if $a^{p-1} = b$, or equivalently, $a^p = b^q$. Since

$$f''(a) = (p-1)a^{p-2} > 0, \qquad \forall a > 0$$

this critical point is a global minimum. Thus, for all $a > 0$, we have

$$f(a) \geqslant f(b^{1/(p-1)}) = \frac{b^{p/(p-1)}}{p} + \frac{b^q}{q} - b^{(p-1)/(p-1)}b = \frac{b^q}{p} + \frac{b^q}{q} - b^q = 0$$

which is equivalent to

$$\frac{a^p}{p} + \frac{b^q}{q} - ab \geqslant 0$$

with equality if and only if $a^p = b^q$.

Since $b > 0$ was arbitrary, the result follows. $\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Proposition 2.4.2 (Hölder's Inequality)**.** *Let (X, μ) be a measure space, and let $p, q \in [1, \infty]$ be such that $\frac{1}{p} + \frac{1}{q} = 1$. If $f \in L^p(X, \mu)$ and $g \in L^q(X, \mu)$, then $fg \in L^1(X, \mu)$ and*

$$\|fg\|_{L^1} \leqslant \|f\|_{L^p} \|g\|_{L^q}$$

with equality if and only if there are constants $c_1, c_2 \geqslant 0$, not both zero, such that $c_1|f|^p = c_2|g|^q$ almost everywhere.

Proof. Step 1: Suppose first that $1 < p, q < \infty$.

We will prove the result in the special case that both functions have unit norm. Suppose that $f \in L^p(X, \mu)$ and $g \in L^q(X, \mu)$ are such that $\|f\|_{L^p} = \|g\|_{L^q} = 1$. Then by 2.4.1, we have

$$|f(x)g(x)| \leqslant \frac{|f(x)|^p}{p} + \frac{|g(x)|^q}{q} \qquad \text{for each } x \in X.$$

By monotonicity of the integral, we have

$$\begin{aligned}
\|fg\|_{L^1} = \int_X |f(x)g(x)| \, \mathrm{d}\mu(x) &\leqslant \int_X \left(\frac{|f(x)|^p}{p} + \frac{|g(x)|^q}{q} \right) \mathrm{d}\mu(x) \\
&= \frac{1}{p} \int_X |f(x)|^p \, \mathrm{d}\mu(x) + \frac{1}{q} \int_X |g(x)|^q \, \mathrm{d}\mu(x) \\
&= \frac{1}{p}\|f\|_{L^p}^p + \frac{1}{q}\|g\|_{L^q}^q = \frac{1}{p} + \frac{1}{q} = 1.
\end{aligned}$$

Thus, in this special case, we have $\|fg\|_{L^1} \leqslant 1 = \|f\|_{L^p}\|g\|_{L^q}$. Moreover, equality holds if and only if $|f(x)|^p = |g(x)|^q$ for μ-almost every $x \in X$. This proves the result in the special case that both functions have unit norm.

Now, let $f \in L^p(X, \mu)$ and $g \in L^q(X, \mu)$ be arbitrary. If either f or g equals zero μ-almost everywhere, then the result is trivial as both sides of the inequality are zero, so suppose that neither function is zero μ-almost everywhere. Then $\|f\|_{L^p}, \|g\|_{L^q} > 0$, and so the functions

$\frac{f}{\|f\|_{L^p}}$ and $\frac{g}{\|g\|_{L^q}}$ both have unit norm in their respective spaces. By the special case already proven, we have

$$\left\| \frac{f}{\|f\|_{L^p}} \frac{g}{\|g\|_{L^q}} \right\|_{L^1} \leqslant 1,$$

which is equivalent to

$$\|fg\|_{L^1} \leqslant \|f\|_{L^p} \|g\|_{L^q}$$

by homogeneity of the L^1 norm. Moreover, equality holds if and only if there is a constant $c > 0$ such that $|f|^p = c|g|^q$ μ-almost everywhere. This proves the result in the case that $1 < p, q < \infty$.

   Step 2: Now, suppose that $p = 1$ and $q = \infty$ (the case $p = \infty$ and $q = 1$ is symmetric). Then we recall that the L^∞-norm is defined as the essential supremum,

$$\|g\|_{L^\infty} := \operatorname{ess\,sup}_{x \in X} |g(x)| := \inf\{C : |g(x)| \leqslant C \text{ for } \mu\text{-almost every } x \in X\} \qquad \forall g \in L^\infty(X, \mu).$$

   Let $f \in L^1(X, \mu)$ and $g \in L^\infty(X, \mu)$ be arbitrary. Then

$$\|fg\|_{L^1} = \int_X |f(x)g(x)| \, \mathrm{d}\mu(x) \leqslant \int_X |f(x)| \|g\|_{L^\infty} \, \mathrm{d}\mu(x)$$
$$= \|g\|_{L^\infty} \int_X |f(x)| \, \mathrm{d}\mu(x) = \|f\|_{L^1} \|g\|_{L^\infty}$$

by monotonicity and homogeneity of the integral. Moreover, equality holds if and only if there is a constant $c \geqslant 0$ such that $|f| = c|g|$ μ-almost everywhere. This proves the result in the case that $p = 1$ and $q = \infty$. $\square$

Proposition 2.4.3 (Reverse Hölder Inequality). *Let (X, μ) be a measure space, and let $p < 1$ be nonzero, and let q be such that $\frac{1}{p} + \frac{1}{q} = 1$. Also suppose that $\mu(X) > 0$. If $f, g : X \to \mathbb{C}$ are measurable functions such that $g \neq 0$ μ-almost everywhere, $fg \in L^1(X, \mu)$, and $\int_X |g|^q \, \mathrm{d}\mu < \infty$, then*

$$\|fg\|_{L^1} \geqslant \left(\int_X |f|^p \, \mathrm{d}\mu \right)^{\frac{1}{p}} \left(\int_X |g|^q \, \mathrm{d}\mu \right)^{\frac{1}{q}}.$$

Proof. Notice that

$$0 < p < 1 \iff q < 0 \qquad \text{and} \qquad p < 0 \iff 0 < q < 1.$$

Thus by symmetry we may assume without loss of generality that $0 < p < 1$ and $q < 0$.

   Now let $f, g : X \to \mathbb{C}$ be measurable functions such that $g \neq 0$ μ-almost everywhere,

$$\int_X |fg| \, \mathrm{d}\mu < \infty \qquad \text{and} \qquad \int_X |g|^q \, \mathrm{d}\mu < \infty.$$

Define $r := \frac{1}{p} > 1$ and $r' := \frac{r}{r-1} > 1$ so that we have

$$\frac{1}{r} + \frac{1}{r'} = 1.$$

We estimate

$$\||fg|^{1/r}\|_{L^r} = \left(\int_X |fg|^{1/r \cdot r} \, d\mu\right)^{1/r}$$

$$= \left(\int_X |fg| \, d\mu\right)^{1/r} = \left(\int_X |fg| \, d\mu\right)^p = \|fg\|_{L^1}^p < \infty$$

and

$$\||g|^{-1/r}\|_{L^{r'}} = \left(\int_X |g|^{-1/r \cdot r'} \, d\mu\right)^{1/r'}$$

$$= \left(\int_X |g|^{-\frac{1}{r-1}} \, d\mu\right)^{1/r'} = \left(\int_X |g|^q \, d\mu\right)^{1-p} < \infty.$$

by using our assumptions on f and g. From this we can use Hölder's inequality (**??**) to estimate

$$\int_X |f|^p \, d\mu = \int_X |fg|^p |g|^{-p} \, d\mu = \int_X |fg|^{1/r} |g|^{-1/r} \, d\mu$$

$$= \||fg|^{1/r} |g|^{-1/r}\|_{L^1}$$

$$\leqslant \||fg|^{1/r}\|_{L^r} \||g|^{-1/r}\|_{L^{r'}}$$

$$= \|fg\|_{L^1}^p \left(\int_X |g|^q \, d\mu\right)^{1-p}.$$

Because $\mu(X) > 0$ and $g \neq 0$ μ-almost everywhere, we have $\int_X |g|^q \, d\mu > 0$, and so we can divide both sides by $\left(\int_X |g|^q \, d\mu\right)^{1-p}$ to obtain

$$\int_X |f|^p \, d\mu \left(\int_X |g|^q \, d\mu\right)^{-(1-p)} \leqslant \|fg\|_{L^1}^p$$

which is equivalent to

$$\left(\int_X |f|^p \, d\mu\right)^{1/p} \left(\int_X |g|^q \, d\mu\right)^{\frac{p-1}{p}} \leqslant \|fg\|_{L^1}$$

by taking p-th roots of both sides. Since $\frac{p-1}{p} = \frac{1}{q}$, the desired result follows. $\qquad\square$

Proposition 2.4.4. *Let (X, μ) be a measure space, and let $0 < r, p, q \leqslant \infty$ be such that*

$$\frac{1}{r} = \frac{1}{p} + \frac{1}{q}.$$

If $f \in L^p(X, \mu)$ and $g \in L^q(X, \mu)$, then $fg \in L^r(X, \mu)$ and

$$\|fg\|_{L^r} \leqslant \|f\|_{L^p} \|g\|_{L^q}.$$

Proof. Case 1: First suppose that $r = \infty$. Then

$$\frac{1}{p} + \frac{1}{q} = 0$$

so we must have $p = q = \infty$. Thus if $f, g \in L^\infty(X, \mu)$, then we check that

$$\|fg\|_{L^\infty} = \operatorname{ess\,sup}_{x \in X} |f(x)g(x)|$$
$$\leqslant \operatorname{ess\,sup}_{x \in X} |f(x)| \cdot \operatorname{ess\,sup}_{x \in X} |g(x)| = \|f\|_{L^\infty} \|g\|_{L^\infty}.$$

This proves the result in the case that $r = \infty$.

Case 2: Now suppose that $0 < r < \infty$. Then we consider two subcases.

$$\text{a)} \quad q < \infty \text{ and } p = \infty \qquad \qquad \text{b)} \quad p, q < \infty.$$

Note that both p and q cannot be infinite, as this would imply that $r = \infty$, which is covered in Case 1. *Subcase 2a:* Suppose that $q < \infty$ and $p = \infty$. Then we have

$$\frac{1}{r} = \frac{1}{\infty} + \frac{1}{q} = 0 + \frac{1}{q} = \frac{1}{q}$$

so that $r = q$. Let $f \in L^q(X, \mu)$ and $g \in L^\infty(X, \mu)$ be arbitrary. Then we estimate

$$\|fg\|_{L^r}^r = \int_X |f(x)g(x)|^r \, d\mu(x) = \int_X |f(x)|^r |g(x)|^r \, d\mu(x)$$
$$\leqslant \int_X |f(x)|^r \|g\|_{L^\infty}^r \, d\mu(x) = \|g\|_{L^\infty}^r \|f\|_{L^r}^r.$$

By taking r-th roots and using that $r = q$, we obtain

$$\|fg\|_{L^r} \leqslant \|f\|_{L^r} \|g\|_{L^\infty}.$$

Subcase 2b: Now suppose that $p, q < \infty$. Then $\frac{p}{r}, \frac{q}{r} > 1$ and

$$\frac{1}{p/r} + \frac{1}{q/r} = 1$$

so we can apply Hölder's inequality (**??**) with exponents $\frac{p}{r}$ and $\frac{q}{r}$.

Let $f \in L^p(X, \mu)$ and $g \in L^q(X, \mu)$ be arbitrary. Then $|f|^r \in L^{p/r}(X, \mu)$ and $|g|^r \in L^{q/r}(X, \mu)$, so we can use Hölder's inequality to estimate

$$\|fg\|_{L^r}^r = \||f|^r |g|^r\|_{L^1}$$
$$\leqslant \||f|^r\|_{L^{p/r}} \||g|^r\|_{L^{q/r}}$$
$$= \left(\int_X |f(x)|^{r \cdot p/r} \, d\mu(x) \right)^{r/p} \left(\int_X |g(x)|^{r \cdot q/r} \, d\mu(x) \right)^{r/q}$$
$$= \|f\|_{L^p}^r \|g\|_{L^q}^r.$$

By taking r-th roots of both sides, we obtain the desired result. $\qquad \square$

Theorem 2.4.5 (Generalized Holder's Inequality)**.** *Let (X, μ) be a measure space, and let $k \in \mathbb{Z}^+$. Also let $0 < p, p_1, \ldots, p_k \leqslant \infty$ be such that*

$$\frac{1}{p} = \sum_{j=1}^{k} \frac{1}{p_j}.$$

If $f_j \in L^{p_j}(X, \mu)$ for each $j = 1, \ldots, k$, then $\prod_{j=1}^{k} f_j \in L^p(X, \mu)$ and

$$\left\| \prod_{j=1}^{k} f_j \right\|_{L^p} \leqslant \prod_{j=1}^{k} \|f_j\|_{L^{p_j}}.$$

Proof. We proceed by induction on $k \in \mathbb{Z}^+$. The base case $k = 1$ is trivial — if $0 < p, p_1 \leqslant \infty$ are such that $\frac{1}{p} = \frac{1}{p_1}$, then $p = p_1$, and so if $f_1 \in L^{p_1}(X, \mu)$, then $f_1 \in L^p(X, \mu)$ and $\|f_1\|_{L^p} = \|f_1\|_{L^{p_1}}$.

Now suppose that the result holds for some $k \in \mathbb{Z}^+$, and let $0 < p, p_1, \ldots, p_{k+1} \leqslant \infty$ be such that

$$\frac{1}{p} = \sum_{j=1}^{k+1} \frac{1}{p_j}.$$

Define

$$q := \left(\sum_{j=1}^{k+1} \frac{1}{p_j} \right)^{-1}$$

so that $0 < q \leqslant \infty$ and

$$\frac{1}{p} = \frac{1}{q} + \frac{1}{p_{k+1}}.$$

Let $f_j \in L^{p_j}(X, \mu)$ for each $j = 1, \ldots, k+1$ be arbitrary. Then by the induction hypothesis, we have $\prod_{j=1}^{k} f_j \in L^q(X, \mu)$, so the previous proposition (2.4.4) implies that

$$\left\| \left(\prod_{j=1}^{k} f_j \right) \cdot f_{k+1} \right\| \leqslant \left\| \prod_{j=1}^{k} f_j \right\|_{L^q} \|f_{k+1}\|_{L^{p_{k+1}}}$$

$$\leqslant \left(\prod_{j=1}^{k} \|f_j\|_{L^{p_j}} \right) \|f_{k+1}\|_{L^{p_{k+1}}} = \prod_{j=1}^{k+1} \|f_j\|_{L^{p_j}}.$$

That is,

$$\left\| \prod_{j=1}^{k+1} f_j \right\|_{L^p} \leqslant \prod_{j=1}^{k+1} \|f_j\|_{L^{p_j}}$$

which completes the induction. $\square$

2.4.2   Minkowski's Inequality

Lemma 2.4.6 (Convexity Lemma). *Suppose that $f : [0, \infty) \to [0, \infty)$ is a convex function such that $f(0) = 0$. Then for all $t_1, t_2 \geqslant 0$ we have*

$$f(t_1) + f(t_2) \leqslant f(t_1 + t_2).$$

Proof. Since f is convex with $f(0) = 0$, for each $\lambda \in [0, 1]$ and each $x \geqslant 0$ we have

$$f(\lambda x) = f(\lambda x + (1 - \lambda)0) \leqslant \lambda f(x) + (1 - \lambda)f(0) = \lambda f(x)$$

Let $t_1, t_2 \geqslant 0$ be arbitrary. If $t_1 + t_2 = 0$, then the result is trivial, so suppose that $t_1 + t_2 > 0$. Define $\lambda_0 := \frac{t_1}{t_1 + t_2} \in (0, 1)$. Then we estimate

$$f(t_1) = f(\lambda_0(t_1 + t_2)) \leqslant \lambda_0 f(t_1 + t_2) \quad \text{and} \quad f(t_2) = f((1 - \lambda_0)(t_1 + t_2)) \leqslant (1 - \lambda_0)f(t_1 + t_2).$$

By adding these two inequalities, we obtain

$$f(t_1) + f(t_2) \leqslant (\lambda_0 + (1 - \lambda_0))f(t_1 + t_2) = f(t_1 + t_2)$$

which is the desired result. $\square$

Lemma 2.4.7 (Concavity Lemma). *Suppose that $f : [0, \infty) \to [0, \infty)$ is a concave function such that $f(0) = 0$. Then for all $t_1, t_2 \geqslant 0$ we have*

$$f(t_1 + t_2) \leqslant f(t_1) + f(t_2).$$

Proof. Note that $-f$ is convex and $-f(0) = 0$, so the result follows from 2.4.6. $\square$

Proposition 2.4.8. *Let $\{a_j\}_{j=1}^{\infty} \subseteq [0, \infty)$ be a sequence of nonnegative real numbers. Then for all $0 \leqslant p \leqslant 1$, we have*

$$\left(\sum_{j=1}^{\infty} a_j \right)^p \leqslant \sum_{j=1}^{\infty} a_j^p.$$

If $p > 1$, then the reverse inequality holds.

Proof. First note that if $p = 1$, then both sides are equal, so the result is trivial. Also if $p = 0$, then the left side is 1 if the sequence is not identically zero, and the right side is the number of nonzero terms in the sequence, so the result holds.

Let $N \in \mathbb{Z}^+$ be arbitrary. See that the function

$$f : [0, \infty) \to [0, \infty), \quad f(x) = x^p$$

is twice differentiable with

$$f'(x) = px^{p-1} \quad \text{and thus} \quad f''(x) = p(p-1)x^{p-2}.$$

Thus

$$f'(x) > 0 \iff p > 1 \quad \text{and} \quad f''(x) < 0 \iff 0 < p < 1.$$

That is, f is convex if $p > 1$ and concave if $0 < p < 1$, and in both cases we have $f(0) = 0$. Thus by 2.4.6, 2.4.7, we have

$$f\left(\sum_{j=1}^{N} a_j\right) \leqslant \sum_{j=1}^{N} f(a_j) \quad \text{if } 0 < p < 1$$

and

$$f\left(\sum_{j=1}^{N} a_j\right) \geqslant \sum_{j=1}^{N} f(a_j) \quad \text{if } p > 1.$$

That is,

$$\left(\sum_{j=1}^{N} a_j\right)^p \leqslant \sum_{j=1}^{N} a_j^p \quad \text{if } 0 < p < 1$$

and

$$\left(\sum_{j=1}^{N} a_j\right)^p \geqslant \sum_{j=1}^{N} a_j^p \quad \text{if } p > 1.$$

Since $N \in \mathbb{Z}^+$ was arbitrary, the result follows by taking limits as $N \to \infty$. $\qquad \square$

Theorem 2.4.9 (Minkowski's Inequality). *Let (X, μ) be a measure space, and let $1 \leqslant p \leqslant \infty$. If $f, g \in L^p(X, \mu)$, then $f + g \in L^p(X, \mu)$ and*

$$\|f + g\|_{L^p} \leqslant \|f\|_{L^p} + \|g\|_{L^p}$$

with equality if and only if there is a constant $c \geqslant 0$ such that $|f| = c|g|$ almost everywhere.

Proof. We remark that this has already been proven in the case that $p = 1$ as part of the proof that $\|\cdot\|_{L^1}$ is a norm on $L^1(X, \mu)$. Also, the case that $p = \infty$ is easy to check directly — if $f, g \in L^\infty(X, \mu)$, then

$$\|f + g\|_{L^\infty} = \operatorname{ess\,sup}_{x \in X} |f(x) + g(x)| \leqslant \operatorname{ess\,sup}_{x \in X} |f(x)| + \operatorname{ess\,sup}_{x \in X} |g(x)|$$
$$= \|f\|_{L^\infty} + \|g\|_{L^\infty}$$

by the triangle inequality for complex numbers.

Now suppose that $1 < p < \infty$ and let $f, g \in L^p(X, \mu)$ be arbitrary. It is not even obvious that $f + g \in L^p(X, \mu)$, so we will prove this first, with a more elementary inequality.

See that for each $x \in X$, we have

$$|f(x) + g(x)|^p \leqslant (|f(x)| + |g(x)|)^p \leqslant (2 \max\{|f(x)|, |g(x)|\})^p \leqslant 2^p(|f(x)|^p + |g(x)|^p)$$

by using the triangle inequality for complex numbers, the fact that $t \mapsto t^p$ is increasing on $[0, \infty)$, and the fact that $\max\{a, b\} \leqslant a + b$ for all $a, b \geqslant 0$. By monotonicity of the integral, we have

$$\|f + g\|_{L^p}^p = \int_X |f(x) + g(x)|^p \, \mathrm{d}\mu(x) \leqslant \int_X 2^p(|f(x)|^p + |g(x)|^p) \, \mathrm{d}\mu(x)$$
$$= 2^p \left(\|f\|_{L^p}^p + \|g\|_{L^p}^p \right) < \infty$$

since $f, g \in L^p(X, \mu)$. Thus $f + g \in L^p(X, \mu)$.

We can now improve this estimate. Let $1 < q < \infty$ be such that $\frac{1}{p} + \frac{1}{q} = 1$. See that

$$\|f + g\|_{L^p}^p = \int_X |f + g|^p \, \mathrm{d}\mu = \int_X |f + g||f + g|^{p-1} \, \mathrm{d}\mu$$
$$\leqslant \int_X (|f| + |g|)|f + g|^{p-1} \, \mathrm{d}\mu \qquad \text{by the triangle inequality}$$
$$= \int_X |f||f + g|^{p-1} \, \mathrm{d}\mu + \int_X |g||f + g|^{p-1} \, \mathrm{d}\mu$$
$$= \||f||f + g|^{p-1}\|_{L^1} + \||g||f + g|^{p-1}\|_{L^1}$$
$$\leqslant \|f\|_{L^p}\||f + g|^{p-1}\|_{L^q} + \|g\|_{L^p}\||f + g|^{p-1}\|_{L^q}$$

by Hölder's inequality. We should check that $|f + g|^{p-1} \in L^q(X, \mu)$ to justify the use of Hölder's inequality above — we compute

$$\||f + g|^{p-1}\|_{L^q}^q = \int_X |f + g|^{(p-1)q} \, \mathrm{d}\mu = \int_X |f + g|^{p-1 \cdot \frac{p}{p-1}} \, \mathrm{d}\mu = \int_X |f + g|^p \, \mathrm{d}\mu = \|f + g\|_{L^p}^p < \infty$$

since we have already shown that $f + g \in L^p(X, \mu)$. Thus our use of Hölder's inequality is justified.

Also the above computation shows that

$$\||f + g|^{p-1}\|_{L^q} = \|f + g\|_{L^p}^{p/q} = \|f + g\|_{L^p}^{p-1}$$

Thus our use of Hölder's inequality gives

$$\|f + g\|_{L^p}^p \leqslant \|f\|_{L^p}\|f + g\|_{L^p}^{p-1} + \|g\|_{L^p}\|f + g\|_{L^p}^{p-1}.$$

If $\|f + g\|_{L^p} = 0$, then the result is trivial, so suppose that $\|f + g\|_{L^p} > 0$. Then we can divide both sides by $\|f + g\|_{L^p}^{p-1}$ to obtain

$$\|f + g\|_{L^p} \leqslant \|f\|_{L^p} + \|g\|_{L^p}$$

which is the desired result.

Finally, we check when equality holds. If there is a constant $c \geqslant 0$ such that $|f| = c|g|$ μ-almost everywhere, then it is easy to check that equality holds in the above inequality.

Conversely, if equality holds, then we must have equality in the application of Hölder's inequality above, which by 2.4.2 implies that there is a constant $c \geqslant 0$ such that

$$|f(x)| = c|f(x) + g(x)|^{p-1} \quad \text{and} \quad |g(x)| = \frac{1}{c}|f(x) + g(x)|^{p-1}$$

for μ-almost every $x \in X$. If $c = 0$, then $f = 0$ μ-almost everywhere, so we may assume that $c > 0$. From this we see that

$$|f(x)|^p = c^p|f(x) + g(x)|^{p(p-1)} \quad \text{and} \quad |g(x)|^p = c^{-p}|f(x) + g(x)|^{p(p-1)}$$

for μ-almost every $x \in X$. By dividing these two equations, we obtain

$$|f(x)|^p = c^{2p}|g(x)|^p$$

for μ-almost every $x \in X$, which is the desired result. □

Proposition 2.4.10 (Minkowski's Triangle Inequality)**.**

Proof. □

Proposition 2.4.11 (Reverse Minkowski's Inequality)**.**

Proof. □

Proposition 2.4.12 (Minkowski's Integral Inequality)**.**

Proof. □

Proposition 2.4.13 (Jensen's Inequality)**.**

2.4.3   Hanner's Inequalities

Lemma 2.4.14 (Hanner's Inequality Lemma $2 \leqslant p < \infty$)**.** *Suppose $2 \leqslant p < \infty$. Then*

$$2(|w|^p + |z|^p) \leqslant |w + z|^p + |w - z|^p \leqslant 2^{p-1}(|w|^p + |z|^p)$$

for all $w, z \in \mathbb{C}$.

Proof. Notice that the inequalities are symmetric in the variables w and z, and that if either $w = 0$ or $z = 0$, then both inequalities are trivially true since $2 \leqslant 2 \leqslant 2^{p-1}$.

   Step 1: We prove the right inequality first.

   Let $w, z \in \mathbb{C}$ be arbitrary nonzero complex numbers. Since the inequalities are symmetric in w and z, we may assume without loss of generality that $|w| \geqslant |z| > 0$.

Since $p \geqslant 2$, the function

$$\phi : [0, \infty) \to [0, \infty), \quad \phi(t) = t^{p/2}$$

is convex and has $\phi(0) = 0$, so by 2.4.6, we have

$$\phi(t_1) + \phi(t_2) \leqslant \phi(t_1 + t_2) \qquad \forall t_1, t_2 \geqslant 0$$

which says that

$$t_1^{p/2} + t_2^{p/2} \leqslant (t_1 + t_2)^{p/2} \tag{$\star$}$$

for all $t_1, t_2 \geqslant 0$. But also convexity of ϕ implies that

$$\phi\left(\frac{1}{2}t_1 + \frac{1}{2}t_2\right) \leqslant \frac{1}{2}\phi(t_1) + \frac{1}{2}\phi(t_2) \qquad \forall t_1, t_2 \geqslant 0$$

which says that

$$\left(\frac{1}{2}t_1 + \frac{1}{2}t_2\right)^{p/2} \leqslant \frac{1}{2}t_1^{p/2} + \frac{1}{2}t_2^{p/2} \tag{$\star\star$}$$

for all $t_1, t_2 \geqslant 0$. Now we see that

$$
\begin{aligned}
\left|\frac{w+z}{2}\right|^p + \left|\frac{w-z}{2}\right|^p &= \left(\left|\frac{w+z}{2}\right|^2\right)^{p/2} + \left(\left|\frac{w-z}{2}\right|^2\right)^{p/2} \\
&\leqslant \left(\left|\frac{w+z}{2}\right|^2 + \left|\frac{w-z}{2}\right|^2\right)^{p/2} && \text{by } (\star) \\
&= \left(\frac{2|w|^2 + 2|z|^2}{4}\right)^{p/2} \\
&= \left(\frac{1}{2}|w|^2 + \frac{1}{2}|z|^2\right)^{p/2} \\
&\leqslant \frac{1}{2}|w|^p + \frac{1}{2}|z|^p. && \text{by } (\star\star)
\end{aligned}
$$

Multiplying both sides by 2^p gives us

$$|w+z|^p + |w-z|^p \leqslant 2^{p-1}(|w|^p + |z|^p)$$

which is the right inequality.

Since z and w were arbitrary nonzero complex numbers, and the inequality is trivially true if either $z = 0$ or $w = 0$, the right inequality holds for all $w, z \in \mathbb{C}$.

Step 2: Now we prove the left inequality. Let $w, z \in \mathbb{C}$ be arbitrary nonzero complex numbers.

Again, since the inequalities are symmetric in w and z, we may assume without loss of generality that $|w| \geqslant |z| > 0$. By using the right hand inequality we just proved with the complex numbers

$$a := \frac{w+z}{2}, \quad b := \frac{w-z}{2}$$

we obtain

$$|a+b|^p + |a-b|^p \leqslant 2^{p-1}(|a|^p + |b|^p).$$

But

$$a+b = w, \quad a-b = z, \quad |a|^p + |b|^p = \frac{|w+z|^p + |w-z|^p}{2^p}.$$

Substituting these back into the inequality gives

$$|w|^p + |z|^p \leqslant 2^{p-1}\frac{|w+z|^p + |w-z|^p}{2^p},$$

which simplifies to

$$2(|w|^p + |z|^p) \leqslant |w+z|^p + |w-z|^p.$$

This completes the proof of the left inequality. $\qquad\square$

Lemma 2.4.15 (Hanner's Inequality Lemma $1 < p \leqslant 2$). *Suppose $1 < p \leqslant 2$. Then*

$$2^{p-1}(|w|^p + |z|^p) \leqslant |w+z|^p + |w-z|^p \leqslant 2(|w|^p + |z|^p)$$

for all $w, z \in \mathbb{C}$.

Proof. Notice that this inequality is symmetric in the variables w and z, and that if either $w = 0$ or $z = 0$, then both inequalities are trivially true since $2^{p-1} \leqslant 2 \leqslant 2$.

Step 1: We prove the left inequality first. Let $w, z \in \mathbb{C}$ be arbitrary nonzero complex numbers.

Since the inequalities are symmetric in w and z, we may assume without loss of generality that $|w| \geqslant |z| > 0$. Since $1 < p \leqslant 2$, the function

$$\psi : [0, \infty) \to [0, \infty), \quad \psi(t) = t^{p/2}$$

is concave and has $\psi(0) = 0$, so by 2.4.7, we have

$$\psi(t_1) + \psi(t_2) \geqslant \psi(t_1 + t_2) \qquad \forall t_1, t_2 \geqslant 0$$

which says that

$$t_1^{p/2} + t_2^{p/2} \geqslant (t_1 + t_2)^{p/2} \qquad (\star)$$

for all $t_1, t_2 \geqslant 0$. But also concavity of ψ implies that

$$\psi\left(\frac{1}{2}t_1 + \frac{1}{2}t_2\right) \geqslant \frac{1}{2}\psi(t_1) + \frac{1}{2}\psi(t_2) \qquad \forall t_1, t_2 \geqslant 0$$

which says that

$$\left(\frac{1}{2}t_1 + \frac{1}{2}t_2\right)^{p/2} \geqslant \frac{1}{2}t_1^{p/2} + \frac{1}{2}t_2^{p/2} \tag{$\star\star$}$$

for all $t_1, t_2 \geqslant 0$. Now we see that

$$\begin{aligned}
\left|\frac{w+z}{2}\right|^p + \left|\frac{w-z}{2}\right|^p &= \left(\left|\frac{w+z}{2}\right|^2\right)^{p/2} + \left(\left|\frac{w-z}{2}\right|^2\right)^{p/2} \\
&\geqslant \left(\left|\frac{w+z}{2}\right|^2 + \left|\frac{w-z}{2}\right|^2\right)^{p/2} &&\text{by } (\star) \\
&= \left(\frac{2|w|^2 + 2|z|^2}{4}\right)^{p/2} \\
&= \left(\frac{1}{2}|w|^2 + \frac{1}{2}|z|^2\right)^{p/2} \\
&\geqslant \frac{1}{2}|w|^p + \frac{1}{2}|z|^p. &&\text{by } (\star\star)
\end{aligned}$$

Multiplying both sides by 2^p gives us

$$|w+z|^p + |w-z|^p \geqslant 2^{p-1}(|w|^p + |z|^p)$$

which is the left inequality.

Since z and w were arbitrary nonzero complex numbers, and the inequality is trivially true if either $z = 0$ or $w = 0$, the left inequality holds for all $w, z \in \mathbb{C}$.

Step 2: Now we prove the right inequality. Let $w, z \in \mathbb{C}$ be arbitrary nonzero complex numbers.

Again, since the inequalities are symmetric in w and z, we may assume without loss of generality that $|w| \geqslant |z| > 0$. By using the left hand inequality we just proved with the complex numbers

$$a := \frac{w+z}{2}, \quad b := \frac{w-z}{2}$$

we obtain

$$|a+b|^p + |a-b|^p \geqslant 2^{p-1}(|a|^p + |b|^p)$$

which is equivalent to

$$2^{-p}|w|^p + 2^{-p}|z|^p \geqslant 2^{p-1}\frac{|w+z|^p + |w-z|^p}{2^p}$$

which simplifies to

$$2(|w|^p + |z|^p) \geqslant |w+z|^p + |w-z|^p$$

which is the desired right inequality.

Since z and w were arbitrary nonzero complex numbers, and the inequality is trivially true if either $z = 0$ or $w = 0$, the right inequality holds for all $w, z \in \mathbb{C}$. □

Theorem 2.4.16 (Hanner's Inequalities). *Let (X, μ) be a measure space. Suppose that $2 \leqslant p < \infty$. Then for all $f, g \in L^p(X, \mu)$, we have*

$$2(\|f\|_{L^p}^p + \|g\|_{L^p}^p) \leqslant \|f + g\|_{L^p}^p + \|f - g\|_{L^p}^p \leqslant 2^{p-1}(\|f\|_{L^p}^p + \|g\|_{L^p}^p).$$

Suppose that $1 < p \leqslant 2$. Then for all $f, g \in L^p(X, \mu)$, we have

$$2^{p-1}(\|f\|_{L^p}^p + \|g\|_{L^p}^p) \leqslant \|f + g\|_{L^p}^p + \|f - g\|_{L^p}^p \leqslant 2(\|f\|_{L^p}^p + \|g\|_{L^p}^p).$$

Proof. Apply the Lemma 2.4.14 or 2.4.15 pointwise, integrate over X, and use the monotonicity and linearity of the integral.

If you don't believe me ...

Let $2 \leqslant p < \infty$ and let $f, g \in L^p(X, \mu)$ be arbitrary. Then by 2.4.14, we have

$$2(|f(x)|^p + |g(x)|^p) \leqslant |f(x) + g(x)|^p + |f(x) - g(x)|^p \leqslant 2^{p-1}(|f(x)|^p + |g(x)|^p)$$

for all $x \in X$. By integrating over X and using the linearity and monotonicity of the integral, we obtain

$$2 \left(\int_X |f(x)|^p \, \mathrm{d}\mu(x) + \int_X |g(x)|^p \, \mathrm{d}\mu(x) \right) \leqslant \int_X |f(x) + g(x)|^p \, \mathrm{d}\mu(x) + \int_X |f(x) - g(x)|^p \, \mathrm{d}\mu(x)$$

$$\leqslant 2^{p-1} \left(\int_X |f(x)|^p \, \mathrm{d}\mu(x) + \int_X |g(x)|^p \, \mathrm{d}\mu(x) \right)$$

which says exactly that

$$2(\|f\|_{L^p}^p + \|g\|_{L^p}^p) \leqslant \|f + g\|_{L^p}^p + \|f - g\|_{L^p}^p \leqslant 2^{p-1}(\|f\|_{L^p}^p + \|g\|_{L^p}^p)$$

as desired.

The case that $1 < p \leqslant 2$ is similar, using 2.4.15 instead. □

2.4.4 Hardy's Inequality

Theorem 2.4.17 (Hardy's Inequality for Integrals). *Let $f : (0, \infty) \to [0, \infty)$ be a measurable function and let $1 \leqslant p < \infty$. Then*

$$\int_0^\infty \left(\frac{1}{x} \int_0^x f(t) \, \mathrm{d}t \right)^p \mathrm{d}x \leqslant \left(\frac{p}{p-1} \right)^p \int_0^\infty f(x)^p \, \mathrm{d}x.$$

with equality if and only if $f = 0$ almost everywhere.

Corollary 2.4.18 (Hardy's Inequality for Sums). *Let $\{a_n\}_{n=1}^\infty \subseteq [0, \infty)$ be a sequence of nonnegative real numbers, and let $1 \leqslant p < \infty$. Then*

$$\sum_{n=1}^\infty \left(\frac{a_1 + a_2 + \cdots + a_n}{n} \right)^p \leqslant \left(\frac{p}{p-1} \right)^p \sum_{n=1}^\infty a_n^p.$$

with equality if and only if $a_n = 0$ for all $n \in \mathbb{Z}^+$.

2.5   Modes of Convergence

2.5.1   Modes of Convergence

If one has a sequence of real or complex numbers $\{a_n\}_{n=1}^{\infty}$, then it is unambiguous what it means for $\{a_n\}_{n=1}^{\infty}$ to converge to a limit a — it means that for each $\epsilon > 0$, there exists $N \in \mathbb{Z}^+$ such that for all $n \geqslant N$, we have $|a_n - a| < \epsilon$. The same works in $\mathbb{R}^n$ or $\mathbb{C}^n$ if we just replace the absolute value with a norm.

The same cannot be said for sequences of functions though. There are many different ways for a sequence of functions $\{f_n\}_{n=1}^{\infty}$ to converge to a function f, and each has its own utility. In this note, we define and study several different modes of convergence for sequences of functions. We hope to do this in a unified manner for functions which take values in $\mathbb{R}$, $\mathbb{C}$, or even $[-\infty, \infty]$.

To this end, let $\mathbb{F}$ be either $\mathbb{R}$ or $\mathbb{C}$, and let $\mathbb{F}_\infty$ be either $[-\infty, \infty]$ if $\mathbb{F} = \mathbb{R}$, or $\mathbb{C} \cup \{\infty\}$ if $\mathbb{F} = \mathbb{C}$.

We now define several different modes of convergence for sequences of functions. The first two modes are standard in analysis, and are reviewed here for completeness. The next five modes are specific to measure theory and integration.

Definition 2.5.1 (Modes of Convergence). Let X be a set, and let $\{f_n\}_{n=1}^{\infty}$ be a sequence of functions $f_n : X \to \mathbb{F}$, and let $f : X \to \mathbb{F}$ be another function.

1. (Pointwise Convergence) We say that $\{f_n\}_{n=1}^{\infty}$ converges **pointwise** to f if for each $x \in X$ and each $\epsilon > 0$, there exists $N \in \mathbb{Z}^+$ (which depends on both x and ϵ) such that for all $n \geqslant N$, we have $|f_n(x) - f(x)| < \epsilon$.

2. (Uniform Convergence) We say that $\{f_n\}_{n=1}^{\infty}$ converges **uniformly** to f if for each $\epsilon > 0$, there exists $N \in \mathbb{Z}^+$ (which depends only on ϵ) such that for all $n \geqslant N$ and all $x \in X$, we have $|f_n(x) - f(x)| < \epsilon$.

Now suppose that (X, μ) is a measure space, and let $\{f_n\}_{n=1}^{\infty}$ be a sequence of measurable functions $X \to \mathbb{F}_\infty$ which are finite almost everywhere, and let f be a measurable function $X \to \mathbb{F}_\infty$ which is finite almost everywhere. Then we define the following five modes of convergence.

3. (Pointwise a.e. Convergence) We say that $\{f_n\}_{n=1}^{\infty}$ converges **pointwise almost everywhere** to f if for μ-almost every $x \in X$, we have $f_n(x) \to f(x)$ as $n \to \infty$.

4. (L^∞ Convergence) We say that $\{f_n\}_{n=1}^{\infty}$ converges **in L^∞ norm** to f if for each $\epsilon > 0$, there exists N (which depends only on ϵ) such that for all $n \geqslant N$, we have $|f_n(x) - f(x)| < \epsilon$ for μ-almost every $x \in X$.

5. (Almost Uniform Convergence) We say that $\{f_n\}_{n=1}^{\infty}$ converges **almost uniformly** to f if for each $\epsilon > 0$, there exists a measurable set $E \subseteq X$ with $\mu(E) < \epsilon$ such that $\{f_n\}_{n=1}^{\infty}$ converges uniformly to f on $X \backslash E$.

6. (L^1 Convergence) We say that $\{f_n\}_{n=1}^{\infty}$ converges **in L^1 norm** to f if

$$\lim_{n \to \infty} \int_X |f_n - f| \, \mathrm{d}\mu = 0.$$

7. (Convergence in Measure) We say that $\{f_n\}_{n=1}^{\infty}$ converges **in measure** to f if for each $\epsilon > 0$,

$$\lim_{n \to \infty} \mu(\{x \in X : |f_n(x) - f(x)| \geq \epsilon\}) = 0.$$

Instead of saying that $\{f_n\}_{n=1}^{\infty}$ is defined as a sequence of functions $X \to \mathbb{F}_{\infty}$ which are finite almost everywhere, we will often just say that $\{f_n\}_{n=1}^{\infty}$ is a sequence of measurable functions $X \to \mathbb{F}$. Similarly for the proposed limit function f. This is because we can always redefine the functions on sets of measure zero to make them finite-valued without changing any of the modes of convergence defined above, or any integrals involving these functions, or any properties of these functions that hold almost everywhere. Our mantra is that in measure theory, we only care about what happens almost everywhere.

The utility of the almost everywhere modes of convergence is that we can modify functions on sets of measure zero without changing their convergence. Pointwise and uniform convergence are too rigid for this purpose.

Proposition 2.5.2 (Linearity of Convergence). *Let (X, μ) be a measure space. Let $\{f_n\}_{n=1}^{\infty}$ and $\{g_n\}_{n=1}^{\infty}$ be sequences of measurable functions and let f and g be measurable functions.*

1. *The sequence $\{f_n\}_{n=1}^{\infty}$ converges to f in one of the seven modes of convergence if and only if $\{|f_n - f|\}_{n=1}^{\infty}$ converges to 0 in the same mode of convergence.*

2. *If $\{f_n\}_{n=1}^{\infty}$ converges to f in one of the seven modes of convergence, and $\{g_n\}_{n=1}^{\infty}$ converges to g in the same mode of convergence, then $\{f_n + g_n\}_{n=1}^{\infty}$ converges to $f + g$ in the same mode of convergence.*

3. *(Squeeze Theorem.) If $\{f_n\}_{n=1}^{\infty}$ converges to 0 in one of the seven modes of convergence, and $|g_n| \leq f_n$ for all n, then $\{g_n\}_{n=1}^{\infty}$ converges to 0 in the same mode of convergence.*

Proof. (i).

Pointwise Convergence. Assume that $\{f_n\}_{n=1}^{\infty}$ converges to f pointwise. Let $x \in X$ and let $\epsilon > 0$. Then there exists $N \in \mathbb{Z}^+$ such that for all $n \geq N$, we have $|f_n(x) - f(x)| < \epsilon$. But

this is equivalent to saying that for all $n \geqslant N$, we have $||f_n(x) - f(x)| - 0| < \epsilon$. Since x and ϵ were arbitrary, we have shown that $\{|f_n - f|\}_{n=1}^{\infty}$ converges pointwise to 0. The converse is similar.

Uniform Convergence. Assume that $\{f_n\}_{n=1}^{\infty}$ converges to f uniformly. Let $\epsilon > 0$. Then there exists $N \in \mathbb{Z}^+$ such that for all $n \geqslant N$ and all $x \in X$, we have $|f_n(x) - f(x)| < \epsilon$. But this is equivalent to saying that for all $n \geqslant N$ and all $x \in X$, we have $||f_n(x) - f(x)| - 0| < \epsilon$. Since ϵ was arbitrary, we have shown that $\{|f_n - f|\}_{n=1}^{\infty}$ converges uniformly to 0. The converse is similar.

Pointwise a.e. Convergence. Assume that $\{f_n\}_{n=1}^{\infty}$ converges to f pointwise a.e. Then there exists a measurable set $E \subseteq X$ with $\mu(E) = 0$ such that for each $x \in X \backslash E$, we have $f_n(x) \to f(x)$ as $n \to \infty$. As we have already shown, this is equivalent to saying that for each $x \in X \backslash E$, we have $|f_n(x) - f(x)| \to 0$ as $n \to \infty$. Since E is a measurable set with $\mu(E) = 0$, we have shown that $\{|f_n - f|\}_{n=1}^{\infty}$ converges pointwise a.e. to 0.

Assume that $\{|f_n - f|\}_{n=1}^{\infty}$ converges to 0 pointwise a.e. Then there exists a measurable set $E \subseteq X$ with $\mu(E) = 0$ such that for each $x \in X \backslash E$, we have $|f_n(x) - f(x)| \to 0$ as $n \to \infty$. As we have already shown, this is equivalent to saying that for each $x \in X \backslash E$, we have $f_n(x) \to f(x)$ as $n \to \infty$. Since E is a measurable set with $\mu(E) = 0$, we have shown that $\{f_n\}_{n=1}^{\infty}$ converges pointwise a.e. to f.

L^{∞} Convergence. Assume that $\{f_n\}_{n=1}^{\infty}$ converges to f in L^{∞} norm. Let $\epsilon > 0$. Then there exists $N \in \mathbb{Z}^+$ such that for all $n \geqslant N$, we have $|f_n(x) - f(x)| < \epsilon$ for μ-almost every $x \in X$. But this is equivalent to saying that for all $n \geqslant N$, we have $||f_n(x) - f(x)| - 0| < \epsilon$ for μ-almost every $x \in X$. Since ϵ was arbitrary, we have shown that $\{|f_n - f|\}_{n=1}^{\infty}$ converges to 0 in L^{∞} norm. The converse is similar.

Almost Uniform Convergence. Assume that $\{f_n\}_{n=1}^{\infty}$ converges to f almost uniformly. Let $\epsilon > 0$. Then there exists a measurable set $E \subseteq X$ with $\mu(E) < \epsilon$ such that $\{f_n\}_{n=1}^{\infty}$ converges uniformly to f on $X \backslash E$. As we have already shown, this is equivalent to saying that $\{|f_n - f|\}_{n=1}^{\infty}$ converges uniformly to 0 on $X \backslash E$. Since E is a measurable set with $\mu(E) < \epsilon$, we have shown that $\{|f_n - f|\}_{n=1}^{\infty}$ converges almost uniformly to 0. The converse is similar.

L^1 Convergence. Assume that $\{f_n\}_{n=1}^{\infty}$ converges to f in L^1 norm. Then

$$\lim_{n \to \infty} \int_X |f_n - f| \ \mathrm{d}\mu = 0.$$

But this is equivalent to saying that

$$\lim_{n \to \infty} \int_X ||f_n - f| - 0| \ \mathrm{d}\mu = 0.$$

Thus $\{|f_n - f|\}_{n=1}^{\infty}$ converges to 0 in L^1 norm. The converse is similar.

Convergence in Measure. Assume that $\{f_n\}_{n=1}^{\infty}$ converges to f in measure. Let $\epsilon > 0$. Then

$$\lim_{n \to \infty} \mu(\{x \in X : |f_n(x) - f(x)| \geq \epsilon\}) = 0.$$

But this is equivalent to saying that

$$\lim_{n \to \infty} \mu(\{x \in X : ||f_n(x) - f(x)| - 0| \geq \epsilon\}) = 0.$$

Thus $\{|f_n - f|\}_{n=1}^{\infty}$ converges to 0 in measure. The converse is similar.

(ii). For pointwise and uniform convergence, this is known, so we prove it for the other five modes.

Pointwise a.e. Convergence. Assume that $\{f_n\}_{n=1}^{\infty}$ converges to f pointwise a.e. and $\{g_n\}_{n=1}^{\infty}$ converges to g pointwise a.e. Then there exist measurable sets $E_f, E_g \subseteq X$ with $\mu(E_f) = 0$ and $\mu(E_g) = 0$ such that for each $x \in X \backslash E_f$, we have $f_n(x) \to f(x)$ as $n \to \infty$, and for each $x \in X \backslash E_g$, we have $g_n(x) \to g(x)$ as $n \to \infty$. Let $E = E_f \cup E_g$, which is a measurable set with $\mu(E) = 0$. For each $x \in X \backslash E$, we have $x \in X \backslash E_f$ and $x \in X \backslash E_g$, so we have $f_n(x) \to f(x)$ and $g_n(x) \to g(x)$ as $n \to \infty$. Thus for each $x \in X \backslash E$, we have

$$f_n(x) + g_n(x) \to f(x) + g(x)$$

as $n \to \infty$. Since E is a measurable set with $\mu(E) = 0$, we have shown that $\{f_n + g_n\}_{n=1}^{\infty}$ converges pointwise a.e. to $f + g$.

L^{∞} *Convergence.* Assume that $\{f_n\}_{n=1}^{\infty}$ converges to f in L^{∞} norm and $\{g_n\}_{n=1}^{\infty}$ converges to g in L^{∞} norm. Then for each $\epsilon > 0$, there exist $N_f, N_g \in \mathbb{Z}^+$ such that for all $n \geq N_f$, we have $|f_n(x) - f(x)| < \epsilon/2$ for μ-almost every $x \in X$, and for all $n \geq N_g$, we have $|g_n(x) - g(x)| < \epsilon/2$ for μ-almost every $x \in X$. Let $N = \max(N_f, N_g)$. Then for all $n \geq N$ and for μ-almost every $x \in X$, we have

$$|(f_n(x) + g_n(x)) - (f(x) + g(x))| \leq |f_n(x) - f(x)| + |g_n(x) - g(x)| < \epsilon/2 + \epsilon/2 = \epsilon.$$

Since $\epsilon > 0$ was arbitrary, we have shown that $\{f_n + g_n\}_{n=1}^{\infty}$ converges to $f + g$ in L^{∞} norm.

Almost Uniform Convergence. Assume that $\{f_n\}_{n=1}^{\infty}$ converges to f almost uniformly and $\{g_n\}_{n=1}^{\infty}$ converges to g almost uniformly. Let $\epsilon > 0$. Then there exist measurable sets $E_f, E_g \subseteq X$ with $\mu(E_f) < \epsilon/2$ and $\mu(E_g) < \epsilon/2$ such that $\{f_n\}_{n=1}^{\infty}$ converges uniformly to f on $X \backslash E_f$, and $\{g_n\}_{n=1}^{\infty}$ converges uniformly to g on $X \backslash E_g$. Let $E = E_f \cup E_g$, which is a measurable set with $\mu(E) < \epsilon$. Since $\{f_n\}_{n=1}^{\infty}$ converges uniformly to f on $X \backslash E_f$ and $\{g_n\}_{n=1}^{\infty}$ converges uniformly to g on $X \backslash E_g$, we have that $\{f_n + g_n\}_{n=1}^{\infty}$ converges uniformly to $f + g$ on $X \backslash E$. Since $\epsilon > 0$ was arbitrary, we have shown that $\{f_n + g_n\}_{n=1}^{\infty}$ converges almost uniformly to $f + g$.

L^1 *Convergence.* Assume that $\{f_n\}_{n=1}^{\infty}$ converges to f in L^1 norm and $\{g_n\}_{n=1}^{\infty}$ converges to g in L^1 norm. Then

$$\lim_{n\to\infty} \int_X |f_n - f| \; \mathrm{d}\mu = \lim_{n\to\infty} \int_X |g_n - g| \; \mathrm{d}\mu = 0.$$

By the triangle inequality for integrals (Proposition **??**), we have

$$\int_X |(f_n + g_n) - (f + g)| \; \mathrm{d}\mu \leqslant \int_X |f_n - f| \; \mathrm{d}\mu + \int_X |g_n - g| \; \mathrm{d}\mu \to 0 \quad \text{as } n \to \infty.$$

Thus $\{f_n + g_n\}_{n=1}^{\infty}$ converges to $f + g$ in L^1 norm.

Convergence in Measure. Assume that $\{f_n\}_{n=1}^{\infty}$ converges to f in measure and $\{g_n\}_{n=1}^{\infty}$ converges to g in measure. Let $\epsilon > 0$. See that for each $n \in \mathbb{Z}^+$, we have

$$\mu(\{x \in X : |(f_n(x)+g_n(x))-(f(x)+g(x))| \geqslant \epsilon\}) \leqslant \mu(\{x \in X : |f_n(x)-f(x)| \geqslant \epsilon/2\})+\mu(\{x \in X : |g_n(x)-g(x)|$$

Taking limits as $n \to \infty$ and using the fact that $\{f_n\}_{n=1}^{\infty}$ converges to f in measure and $\{g_n\}_{n=1}^{\infty}$ converges to g in measure, we have we see that the right-hand side tends to 0 as $n \to \infty$. Thus

$$\lim_{n\to\infty} \mu(\{x \in X : |(f_n(x) + g_n(x)) - (f(x) + g(x))| \geqslant \epsilon\}) = 0.$$

Since $\epsilon > 0$ was arbitrary, we have shown that $\{f_n + g_n\}_{n=1}^{\infty}$ converges to $f + g$ in measure.

(iii).

Pointwise Convergence. Assume that $\{f_n\}_{n=1}^{\infty}$ converges to 0 pointwise and $|g_n| \leqslant f_n$ for all n. Let $x \in X$ and let $\epsilon > 0$. Then there exists $N \in \mathbb{Z}^+$ such that for all $n \geqslant N$, we have $|f_n(x)| < \epsilon$. Since $|g_n(x)| \leqslant |f_n(x)|$, we have $|g_n(x)| < \epsilon$ for all $n \geqslant N$. Since x and ϵ were arbitrary, we have shown that $\{g_n\}_{n=1}^{\infty}$ converges pointwise to 0.

Uniform Convergence. Assume that $\{f_n\}_{n=1}^{\infty}$ converges to 0 uniformly and $|g_n| \leqslant f_n$ for all n. Let $\epsilon > 0$. Then there exists $N \in \mathbb{Z}^+$ such that for all $n \geqslant N$ and all $x \in X$, we have $|f_n(x)| < \epsilon$. Since $|g_n(x)| \leqslant |f_n(x)|$, we have $|g_n(x)| < \epsilon$ for all $n \geqslant N$ and all $x \in X$. Since ϵ was arbitrary, we have shown that $\{g_n\}_{n=1}^{\infty}$ converges uniformly to 0.

Pointwise a.e. Convergence. Assume that $\{f_n\}_{n=1}^{\infty}$ converges to 0 pointwise a.e. and $|g_n| \leqslant f_n$ for all n and almost everywhere on X. Then there exists a measurable set $E_f \subseteq X$ with $\mu(E_f) = 0$ such that for each $x \in X \backslash E_f$, we have $f_n(x) \to 0$ as $n \to \infty$. There is also a measurable set $E_g \subseteq X$ with $\mu(E_g) = 0$ such that for each n, we have $|g_n(x)| \leqslant f_n(x)$ for each $x \in X \backslash E_g$. Let $E = E_f \cup E_g$, which is a measurable set with $\mu(E) = 0$. For each $x \in X \backslash E$, we have $x \in X \backslash E_f$ and $x \in X \backslash E_g$, so we have $f_n(x) \to 0$ as $n \to \infty$ and $|g_n(x)| \leqslant f_n(x)$ for all n. Thus we have $|g_n(x)| \to 0$ as $n \to \infty$ by the case for pointwise convergence. Since E is a measurable set with $\mu(E) = 0$, we have shown that $\{g_n\}_{n=1}^{\infty}$ converges pointwise a.e. to 0.

L^∞ *Convergence.* Assume that $\{f_n\}_{n=1}^\infty$ converges to 0 in L^∞ norm and $|g_n| \leqslant f_n$ for all n and almost everywhere on X. Let $\epsilon > 0$. Then there exists $N \in \mathbb{Z}^+$ such that for all $n \geqslant N$, we have $|f_n(x)| < \epsilon$ for μ-almost every $x \in X$. There is also a measurable set $E_g \subseteq X$ with $\mu(E_g) = 0$ such that for each n, we have $|g_n(x)| \leqslant f_n(x)$ for each $x \in X\backslash E_g$. Let $E = E_g$, which is a measurable set with $\mu(E) = 0$. For all $n \geqslant N$ and for μ-almost every $x \in X$, we have $x \in X\backslash E$, so we have $|g_n(x)| \leqslant |f_n(x)| < \epsilon$. Since $\epsilon > 0$ was arbitrary, we have shown that $\{g_n\}_{n=1}^\infty$ converges to 0 in L^∞ norm.

Almost Uniform Convergence. Assume that $\{f_n\}_{n=1}^\infty$ converges to 0 almost uniformly and $|g_n| \leqslant f_n$ for all n and almost everywhere on X. Let $\epsilon > 0$. Then there exists a measurable set $E_f \subseteq X$ with $\mu(E_f) < \epsilon$ such that $\{f_n\}_{n=1}^\infty$ converges uniformly to 0 on $X\backslash E_f$. There is also a measurable set $E_g \subseteq X$ with $\mu(E_g) = 0$ such that for each n, we have $|g_n(x)| \leqslant f_n(x)$ for each $x \in X\backslash E_g$. Let $E = E_f \cup E_g$, which is a measurable set with $\mu(E) < \epsilon$. Since $\{f_n\}_{n=1}^\infty$ converges uniformly to 0 on $X\backslash E_f$ and $|g_n(x)| \leqslant |f_n(x)|$ for each $x \in X\backslash E_g$, we have that $\{g_n\}_{n=1}^\infty$ converges uniformly to 0 on $X\backslash E$. Since $\epsilon > 0$ was arbitrary, we have shown that $\{g_n\}_{n=1}^\infty$ converges almost uniformly to 0.

L^1 *Convergence.* Assume that $\{f_n\}_{n=1}^\infty$ converges to 0 in L^1 norm and $|g_n| \leqslant f_n$ for all n and almost everywhere on X. Then

$$\lim_{n\to\infty} \int_X |f_n| \; \mathrm{d}\mu = 0.$$

Then monotonicity of the Lebesgue integral (Proposition **??**) implies that

$$\int_X |g_n| \; \mathrm{d}\mu \leqslant \int_X |f_n| \; \mathrm{d}\mu \to 0 \quad \text{as } n \to \infty.$$

Thus $\{g_n\}_{n=1}^\infty$ converges to 0 in L^1 norm.

Convergence in Measure. Assume that $\{f_n\}_{n=1}^\infty$ converges to 0 in measure and $|g_n| \leqslant f_n$ for all n and almost everywhere on X. Let $\epsilon > 0$. See that for each $n \in \mathbb{Z}^+$, we have

$$\mu(\{x \in X : |g_n(x)| \geqslant \epsilon\}) \leqslant \mu(\{x \in X : |f_n(x)| \geqslant \epsilon\}).$$

Taking limits as $n \to \infty$ and using the fact that $\{f_n\}_{n=1}^\infty$ converges to 0 in measure, we have we see that the right-hand side tends to 0 as $n \to \infty$. Thus

$$\lim_{n\to\infty} \mu(\{x \in X : |g_n(x)| \geqslant \epsilon\}) = 0.$$

Since $\epsilon > 0$ was arbitrary, we have shown that $\{g_n\}_{n=1}^\infty$ converges to 0 in measure. $\qquad\square$

Proposition 2.5.3 (Easy Implications). *Let (X, μ) be a measure space. Let $\{f_n\}_{n=1}^\infty$ and let f be measurable functions. Then the following implications hold:*

  1. Uniform convergence $\implies$ pointwise convergence.

$$\text{Uniform} \Longrightarrow \text{Pointwise}$$

$$L^1 \qquad L^\infty$$

$$\text{Almost Uniform} \Longrightarrow \text{Pointwise a.e}$$
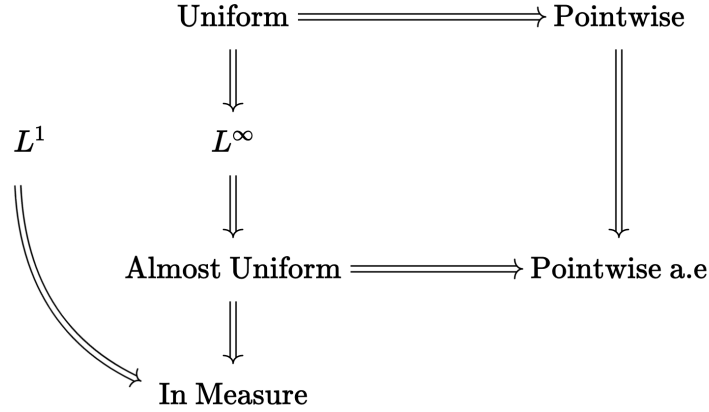
$$\text{In Measure}$$

2. *Uniform convergence $\implies L^\infty$ convergence.*

3. *L^∞ convergence $\implies$ almost uniform convergence.*

4. *Almost uniform convergence $\implies$ pointwise a.e. convergence.*

5. *Pointwise convergence $\implies$ pointwise a.e. convergence.*

6. *L^1 convergence $\implies$ convergence in measure.*

7. *Almost uniform convergence $\implies$ convergence in measure.*

Proof. (1). Suppose that $\{f_n\}_{n=1}^\infty$ converges uniformly to f. Then for each $\epsilon > 0$, there exists $N \in \mathbb{Z}^+$ such that for all $n \geqslant N$ and all $x \in X$, we have $|f_n(x) - f(x)| < \epsilon$. Let $x \in X$ and let $\epsilon > 0$. Then for all $n \geqslant N$, we have $|f_n(x) - f(x)| < \epsilon$. Since x and ϵ were arbitrary, we have shown that $\{f_n\}_{n=1}^\infty$ converges pointwise to f.

(2). Suppose that $\{f_n\}_{n=1}^\infty$ converges uniformly to f. Then for each $\epsilon > 0$, there exists $N \in \mathbb{Z}^+$ such that for all $n \geqslant N$ and all $x \in X$, we have $|f_n(x) - f(x)| < \epsilon$. Since this holds for all $x \in X$, it holds for μ-almost every $x \in X$. Since $\epsilon > 0$ was arbitrary, we have shown that $\{f_n\}_{n=1}^\infty$ converges to f in L^∞ norm.

(3). Suppose that $\{f_n\}_{n=1}^\infty$ converges to f in L^∞ norm. Let $\epsilon > 0$. Then there exists $N \in \mathbb{Z}^+$ such that for all $n \geqslant N$, we have $|f_n(x) - f(x)| < \epsilon$ for μ-almost every $x \in X$. Let $E = \{x \in X : |f_N(x) - f(x)| \geqslant \epsilon\}$. Then E is a measurable set with $\mu(E) = 0$. For all $n \geqslant N$ and all $x \in X \backslash E$, we have $|f_n(x) - f(x)| < \epsilon$. Since $\epsilon > 0$ was arbitrary, we have shown that $\{f_n\}_{n=1}^\infty$ converges almost uniformly to f.

(4). Suppose that $\{f_n\}_{n=1}^\infty$ converges almost uniformly to f. Then for each $\epsilon > 0$, there exists a measurable set $E_\epsilon \subseteq X$ with $\mu(E_\epsilon) < \epsilon$ such that $\{f_n\}_{n=1}^\infty$ converges uniformly to f on $X \backslash E_\epsilon$. Let $E = \bigcap_{m=1}^\infty E_{1/m}$. Then E is a measurable set with $\mu(E) = 0$ by countable
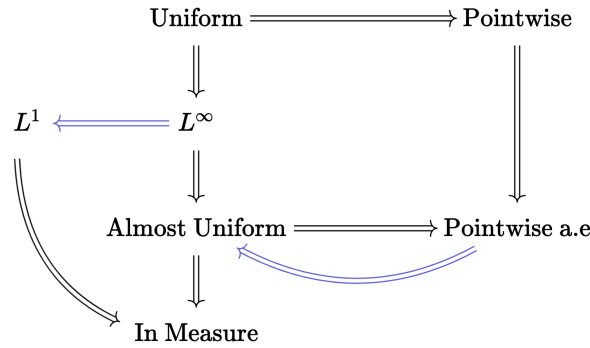
subadditivity. For each $x \in X\backslash E$, we have $x \in X\backslash E_{1/m}$ for some m, and since $\{f_n\}_{n=1}^{\infty}$ converges uniformly to f on $X\backslash E_{1/m}$, we have $f_n(x) \to f(x)$ as $n \to \infty$. Thus $\{f_n\}_{n=1}^{\infty}$ converges pointwise a.e. to f.

(5). Suppose that $\{f_n\}_{n=1}^{\infty}$ converges pointwise to f. Then it is trivially true that $\{f_n\}_{n=1}^{\infty}$ converges pointwise a.e. to f.

(6). Suppose that $\{f_n\}_{n=1}^{\infty}$ converges to f in L^1 norm. Let $\epsilon > 0$. By Chebyshev's Inequality (Lemma 2.3.1), we have

$$\mu(\{x \in X : |f_n(x) - f(x)| \geqslant \epsilon\}) \leqslant \frac{1}{\epsilon} \int_X |f_n - f| \, \mathrm{d}\mu.$$

Since $\{f_n\}_{n=1}^{\infty}$ converges to f in L^1 norm, the right-hand side tends to 0 as $n \to \infty$. Thus $\mu(\{|f_n - f| \geqslant \epsilon\}) \to 0$ as $n \to \infty$, so $\{f_n\}_{n=1}^{\infty}$ converges to f in measure.

(7). Suppose that $\{f_n\}_{n=1}^{\infty}$ converges almost uniformly to f. Let $\epsilon > 0$ and $\delta > 0$. By L^∞ convergence, there exists a measurable set $E \subseteq X$ with $\mu(E) < \delta$ such that $\{f_n\}_{n=1}^{\infty}$ converges uniformly to f on $X\backslash E$. Thus there exists $N \in \mathbb{Z}^+$ such that for all $n \geqslant N$ and all $x \in X\backslash E$, we have $|f_n(x) - f(x)| < \epsilon$. Thus for all $n \geqslant N$, we have

$$\{x \in X : |f_n(x) - f(x)| \geqslant \epsilon\} \subseteq E,$$

so $\mu(\{|f_n - f| \geqslant \epsilon\}) \leqslant \mu(E) < \delta$ for all $n \geqslant N$. Since $\delta > 0$ was arbitrary, we have shown that $\mu(\{|f_n - f| \geqslant \epsilon\}) \to 0$ as $n \to \infty$, so $\{f_n\}_{n=1}^{\infty}$ converges to f in measure. $\qquad\square$

Remark 2.5.4 (Extra Implications in Finite Measure Spaces). We remark that if the measure of the whole space is finite, then uniform convergence also implies L^1 convergence, and L^∞ convergence also implies L^1 convergence. Also pointwise almost everywhere convergence implies almost uniform convergence by Egorov's Theorem (Theorem 1.4.14).

We prove these claims below.

Let (X, μ) be a measure space with $\mu(X) < \infty$. Let $\{f_n\}_{n=1}^{\infty}$ and let f be measurable

functions such that $\{f_n\}_{n=1}^\infty$ converges to f uniformly or in L^∞ norm. Then we see that

$$\int_X |f_n - f|\, d\mu \leqslant \mu(X)\|f_n - f\|_\infty$$

for each $n \in \mathbb{Z}^+$, and taking limits as $n \to \infty$ and using the uniform convergence or L^∞ convergence of $\{f_n\}_{n=1}^\infty$ to f, we have

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

Thus $\{f_n\}_{n=1}^\infty$ converges to f in L^1 norm.

Now suppose that $\{f_n\}_{n=1}^\infty$ converges to f pointwise a.e. Then by Egorov's Theorem (Theorem 1.4.14), for each $\epsilon > 0$, there exists a measurable set $E_\epsilon \subseteq X$ with $\mu(E_\epsilon) < \epsilon$ such that $\{f_n\}_{n=1}^\infty$ converges uniformly to f on $X\backslash E_\epsilon$. Since $\epsilon > 0$ was arbitrary, we have shown that $\{f_n\}_{n=1}^\infty$ converges almost uniformly to f. This completes the proof of the claims. Note that these implications do not hold in general if $\mu(X) = \infty$.

Example 2.5.5 (Escape to Horizontal Infinity)**.**

Example 2.5.6 (Escape to Width Infinity)**.**

Example 2.5.7 (Escape to Vertical Infinity)**.**

Example 2.5.8 (Typewriter Sequence)**.**

2.5.2   Uniqueness of Limits

Even though the modes of convergence all differ from each other, they are all compatible in the sense that they never disagree about which function f is the limit of a sequence $\{f_n\}_{n=1}^\infty$, outside of a set of measure zero.

Theorem 2.5.9 (Uniqueness of Limits)**.** *Let (X, μ) be a measure space. Let $\{f_n\}_{n=1}^\infty$ be a sequence of measurable functions and let f and g be measurable functions. If $\{f_n\}_{n=1}^\infty$ converges to f in one of the seven modes of convergence, and $\{f_n\}_{n=1}^\infty$ converges to g in another of the seven modes of convergence, then $f = g$ almost everywhere.*

height width tail

2.5.3   The Case of Step Functions

height width tail

2.5.4 Topologies of Convergence

Ok so as analysts, we are mainly interested in limits of sequences, but occasionally we want to talk topology. It is useful after all.

By definition, for each $1 \leqslant p \leqslant \infty$, convergence in L^p norm is the convergence on $L^p(X, \mu)$ induced by the topology of the L^p norm. Pointwise convergence is by definition the convergence induced by the product topology on $[-\infty, \infty]^X$. Prove this last sentence?

There is also a topology for uniform convergence, the uniform norm topology.

The others are more mysterious.

There is no topology for pointwise a.e. convergence.

is there a topology for almost uniform convergence?

convergence in measure see Folland exercise

2.5.5 Variants of Convergence Theorems

Chapter 3

Fubini and Tonelli Theorems

3.1 Fubini and Tonelli Theorems

In this section, we prove the Fubini and Tonelli theorems, which allow us to compute integrals on product spaces as iterated integrals. In a later section, we will prove a curvilinear version of this theorem for computing integrals called the *Area Formula* and the *Coarea Formula*.

We follow the treatment given by Axler.

3.1.1 Products of Measures

Definition 3.1.1 (Product of σ-algebras)**.** Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be measurable spaces. We define the *product* of the σ-algebras $\mathcal{A}$ and $\mathcal{B}$ as the smallest σ-algebra on $X \times Y$ containing all products $A \times B$ with $A \in \mathcal{A}$ and $B \in \mathcal{B}$, and we denote this σ-algebra by $\mathcal{A} \otimes \mathcal{B}$. A *measurable product* is a set of the form $A \times B$ with $A \in \mathcal{A}$ and $B \in \mathcal{B}$.

Definition 3.1.2 (Cross Sections of a Set)**.** Let X and Y be sets, and let $E \subseteq X \times Y$. For each $x \in X$, we define the *cross section* of E at x as

$$E_x := \{y \in Y : (x, y) \in E\} \subseteq Y.$$

For each $y \in Y$, we define the *cross section* of E at y as

$$E^y := \{x \in X : (x, y) \in E\} \subseteq X.$$

Lemma 3.1.3 (Cross Sections of Measurable Products are Measurable)**.** *Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be measurable spaces, and let $E = A \times B$ be a measurable product with $A \in \mathcal{A}$ and $B \in \mathcal{B}$. Then for each $x \in X$ we have $E_x \in \mathcal{B}$, and for each $y \in Y$ we have $E^y \in \mathcal{A}$.*

Proof. Let $\mathcal{E}$ be the collection of all subsets $E \subseteq X \times Y$ such that $E_x \in \mathcal{B}$ for each $x \in X$ and $E^y \in \mathcal{A}$ for each $y \in Y$. Then for all $A \in \mathcal{A}$ and all $B \in \mathcal{B}$, we have $(A \times B)_x = B$ if

$x \in A$ and $(A \times B)_x = \varnothing$ otherwise, so $(A \times B)_x \in \mathcal{B}$ for each $x \in X$. Similarly, we have $(A \times B)^y = A$ if $y \in B$ and $(A \times B)^y = \varnothing$ otherwise, so $(A \times B)^y \in \mathcal{A}$ for each $y \in Y$. Thus $A \times B \in \mathcal{E}$ for all $A \in \mathcal{A}$ and $B \in \mathcal{B}$.

See that $\mathcal{E}$ is closed under complements because for each $E \in \mathcal{E}$ we have

$$((X \times Y) \backslash E)_x = Y \backslash E_x \in \mathcal{B}$$

for all $x \in X$ and similarly

$$((X \times Y) \backslash E)^y = X \backslash E^y \in \mathcal{A}$$

for all $y \in Y$.

Also see that $\mathcal{E}$ is closed under countable unions because for each sequence $\{E_j\}_{j=1}^{\infty}$ in $\mathcal{E}$ we have

$$\left(\bigcup_{j=1}^{\infty} E_j \right)_x = \bigcup_{j=1}^{\infty} (E_j)_x \in \mathcal{B}$$

for all $x \in X$ and similarly

$$\left(\bigcup_{j=1}^{\infty} E_j \right)^y = \bigcup_{j=1}^{\infty} (E_j)^y \in \mathcal{A}$$

for all $y \in Y$.

Since $\mathcal{E}$ contains all measurable products and is a σ-algebra, we have $\mathcal{A} \otimes \mathcal{B} \subseteq \mathcal{E}$. $\square$

Definition 3.1.4 (Cross Section of a Function). Let X and Y be sets, and let $f : X \times Y \to [-\infty, \infty]$ be a function. For each $x \in X$, we define the *cross section* of f at x as the function $f_x : Y \to [-\infty, \infty]$ given by

$$f_x(y) := f(x, y) \quad \text{for all } y \in Y.$$

For each $y \in Y$, we define the *cross section* of f at y as the function $f^y : X \to [-\infty, \infty]$ given by

$$f^y(x) := f(x, y) \quad \text{for all } x \in X.$$

Basically if f is a function defined on a product $X \times Y$, then fixing $x \in X$ gives a function f_x defined on Y, and fixing $y \in Y$ gives a function f^y defined on X.

Lemma 3.1.5 (Cross Sections of Measurable Functions are Measurable). *Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be measurable spaces, and let $f : X \times Y \to [-\infty, \infty]$ be an $(\mathcal{A} \otimes \mathcal{B})$-measurable function. Then for each $x \in X$, the cross section $f_x : Y \to [-\infty, \infty]$ is a $\mathcal{B}$-measurable function, and for each $y \in Y$, the cross section $f^y : X \to [-\infty, \infty]$ is an $\mathcal{A}$-measurable function.*

Proof. Let D be a Borel subset of $[-\infty, \infty]$ and let $x \in X$. Then we have

$$
\begin{aligned}
y = (f_x)^{-1}(D) &\iff f_x(y) \in D \\
&\iff f(x, y) \in D \\
&\iff (x, y) \in f^{-1}(D). \qquad\qquad \iff y \in (f^{-1}(D))_x
\end{aligned}
$$

which shows that

$$
(f_x)^{-1}(D) = (f^{-1}(D))_x \in \mathcal{B}.
$$

Since f is $(\mathcal{A}\otimes\mathcal{B})$-measurable, we have $f^{-1}(D) \in \mathcal{A}\otimes\mathcal{B}$, so by Lemma **??** we have $(f^{-1}(D))_x \in \mathcal{B}$. Thus f_x is $\mathcal{B}$-measurable. Since $x \in X$ was arbitrary, we have that f_x is $\mathcal{B}$-measurable for each $x \in X$.

The same idea shows that for each $y \in Y$, the cross section $f^y : X \to [-\infty, \infty]$ is an $\mathcal{A}$-measurable function. $\qquad\square$

3.1.2   Monotone Class Lemma

Definition 3.1.6 (Algebra (of Sets))**.** Let X be a set. A collection $\mathcal{A}$ of subsets of X is an *algebra* if

- $\varnothing \in \mathcal{A}$,

- if $A \in \mathcal{A}$, then $X\backslash A \in \mathcal{A}$, and

- if $A, B \in \mathcal{A}$, then $A \cup B \in \mathcal{A}$.

Notice that an algebra is also closed under finite intersections, by DeMorgan's laws.

An algebra is like a σ-algebra, except that it is only closed under finite unions instead of countable unions.

Lemma 3.1.7 (Finite Unions of Measurable Products Form an Algebra)**.** *Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be measurable spaces, and let $\mathcal{E}$ be the collection of all finite unions of measurable products $A \times B$ with $A \in \mathcal{A}$ and $B \in \mathcal{B}$. Then $\mathcal{E}$ is an algebra on $X \times Y$.*

Moreover, every finite union of measurable products in $\mathcal{A}\otimes\mathcal{B}$ can be written as a countable union of disjoint measurable products in $\mathcal{A}\otimes\mathcal{B}$.

Proof. It is clear that $\varnothing \in \mathcal{E}$ since $\varnothing = \varnothing \times \varnothing$. Furthermore it is clear that $\mathcal{E}$ is closed under finite unions by definition. The set $\mathcal{E}$ is also closed under finite intersections because if $(A_1 \times B_1) \cup (A_2 \times B_2) \cup \cdots \cup (A_m \times B_m)$ and $(C_1 \times D_1) \cup (C_2 \times D_2) \cup \cdots \cup (C_n \times D_n)$

are in $\mathcal{E}$, then

$$\bigcup_{j=1}^{m}(A_j \times B_j) \cap \bigcup_{k=1}^{n}(C_k \times D_k) = \bigcup_{j=1}^{m}\bigcup_{k=1}^{n}\left((A_j \times B_j) \cap (C_k \times D_k)\right)$$

$$= \bigcup_{j=1}^{m}\bigcup_{k=1}^{n}\left((A_j \cap C_k) \times (B_j \cap D_k)\right) \in \mathcal{E}.$$

To see that the set $\mathcal{E}$ is closed under compliments, let $A \in \mathcal{A}$ and $B \in \mathcal{B}$. Then

$$(X \times Y)\backslash(A \times B) = \left((X\backslash A) \times Y\right) \cup \left(A \times (Y\backslash B)\right).$$

Hence the compliment of $A \times B$ is in $\mathcal{E}$. That is, the compliemnt of each measurable product is in $\mathcal{E}$. Since $\mathcal{E}$ is closed under finite unions, we see that the compliment of any finite union of measurable products is in $\mathcal{E}$ by DeMorgan's laws and the fact that $\mathcal{E}$ is closed under finite intersections.

Therefore $\mathcal{E}$ is an algebra on $X \times Y$.

For the second part, note that if $A \times B$ and $C \times D$ are measurable products, then

$$(A \times B) \cup (C \times D) = (A \times B) \cup (C \times (D\backslash B)) \cup (C\backslash A) \times (B \cap D)$$

is a union of three disjoint measurable products.

Now consider a finite union of measurable products

$$E = \bigcup_{j=1}^{m}(A_j \times B_j)$$

with $A_j \in \mathcal{A}$ and $B_j \in \mathcal{B}$ for each $j = 1, 2, \ldots, m$. If there is some $j \in \{1, 2, \ldots, m\}$ such that $A_j \times B_j$ is not disjoint from $\bigcup_{k \neq j}(A_k \times B_k)$, then we can replace $A_j \times B_j$ in the union with a union of three disjoint measurable products as above. Doing this for all such j gives E as a finite union of disjoint measurable products. $\square$

Definition 3.1.8 (Monotone Class)**.** Let X be a set and let $\mathcal{M}$ be a collection of subsets of X. We say that $\mathcal{M}$ is a *monotone class* if

- if $\{A_j\}_{j=1}^{\infty}$ is an increasing sequence of sets in $\mathcal{M}$, then $\bigcup_{j=1}^{\infty} A_j \in \mathcal{M}$, and

- if $\{A_j\}_{j=1}^{\infty}$ is a decreasing sequence of sets in $\mathcal{M}$, then $\bigcap_{j=1}^{\infty} A_j \in \mathcal{M}$.

Clearly every σ-algebra is a monotone class, since σ-algebras are closed under countable unions and countable intersections. Monotone classes don't have to be algebras or σ-algebras, as the following example shows.

Example 3.1.9 (Monotone Class of Intervals). Let $\mathcal{M}$ be the collection of all intervals (closed, open, half-open, degenerate) in $\mathbb{R}$. Then $\mathcal{M}$ is a monotone class since increasing unions and decreasing intersections of intervals are intervals.

However, $\mathcal{M}$ is not an algebra since it is not closed under finite unions.

Exercise 3.1.1 (Smallest Monotone Class). Let X be a set, and let $\mathcal{E}$ be a collection of subsets of X. Show that

$$\mathcal{M}_{\mathcal{E}} := \bigcap\{\mathcal{M} : \mathcal{M} \text{ is a monotone class containing } \mathcal{E}\}$$

is the smallest monotone class containing $\mathcal{E}$.

Proof. There are two things to show here — first that $\mathcal{M}_{\mathcal{E}}$ is a monotone class, and second that for each monotone class $\mathcal{M}$ containing $\mathcal{E}$, we have $\mathcal{M}_{\mathcal{E}} \subseteq \mathcal{M}$.

To see that $\mathcal{M}_{\mathcal{E}}$ is a monotone class, let $\{A_j\}_{j=1}^{\infty}$ be an increasing sequence of sets in $\mathcal{M}_{\mathcal{E}}$. Then for each monotone class $\mathcal{M}$ containing $\mathcal{E}$, we have $A_j \in \mathcal{M}$ for each $j \in \mathbb{Z}^+$, so $\bigcup_{j=1}^{\infty} A_j \in \mathcal{M}$ since $\mathcal{M}$ is a monotone class. Since this is true for each monotone class $\mathcal{M}$ containing $\mathcal{E}$, we have $\bigcup_{j=1}^{\infty} A_j \in \mathcal{M}_{\mathcal{E}}$. A similar argument shows that if $\{A_j\}_{j=1}^{\infty}$ is a decreasing sequence of sets in $\mathcal{M}_{\mathcal{E}}$, then $\bigcap_{j=1}^{\infty} A_j \in \mathcal{M}_{\mathcal{E}}$. Thus $\mathcal{M}_{\mathcal{E}}$ is a monotone class.

Now let $\mathcal{M}$ be a monotone class containing $\mathcal{E}$. Then by definition of $\mathcal{M}_{\mathcal{E}}$, we have $\mathcal{M}_{\mathcal{E}} \subseteq \mathcal{M}$. $\square$

Theorem 3.1.10 (Monotone Class Lemma). *Let X be a set, and let $\mathcal{A}$ be an algebra of subsets of X. Then the smallest monotone class containing $\mathcal{A}$ is the smallest σ-algebra containing $\mathcal{A}$.*

Proof. Let $\mathcal{M}_{\mathcal{A}}$ be the smallest monotone class containing $\mathcal{A}$. Since every σ-algebra is a monotone class, we have $\mathcal{M}_{\mathcal{A}} \subseteq \sigma(\mathcal{A})$, where $\sigma(\mathcal{A})$ is the smallest σ-algebra containing $\mathcal{A}$. To show the reverse inclusion, fix a set $A \in \mathcal{A}$ and let

$$\mathcal{E}_A := \{E \in \mathcal{M}_{\mathcal{A}} : A \cup E \in \mathcal{M}_{\mathcal{A}}\}.$$

Then $\mathcal{A} \subseteq \mathcal{E}_A$ because $\mathcal{A}$ is closed under finite unions.

We want to show that $\mathcal{E}_A$ is a monotone class. Let $\{E_j\}_{j=1}^{\infty}$ be an increasing sequence of sets in $\mathcal{E}_A$. Then for each $j \in \mathbb{Z}^+$ we have $A \cup E_j \in \mathcal{M}_{\mathcal{A}}$, so

$$A \cup \left(\bigcup_{j=1}^{\infty} E_j\right) = \bigcup_{j=1}^{\infty}(A \cup E_j) \in \mathcal{M}_{\mathcal{A}}.$$

Thus $\bigcup_{j=1}^{\infty} E_j \in \mathcal{E}_A$. Since $\{E_j\}_{j=1}^{\infty}$ was arbitrary, we have shown that $\mathcal{E}_A$ is closed under countable increasing unions. Now let $\{E_j\}_{j=1}^{\infty}$ be a decreasing sequence of sets in $\mathcal{E}_A$. Then

for each $j \in \mathbb{Z}^+$ we have $A \cup E_j \in \mathcal{M}_\mathcal{A}$, so

$$A \cup \left(\bigcap_{j=1}^{\infty} E_j \right) = \bigcap_{j=1}^{\infty} (A \cup E_j) \in \mathcal{M}_\mathcal{A}.$$

Thus $\bigcap_{j=1}^{\infty} E_j \in \mathcal{E}_A$. Since $\{E_j\}_{j=1}^{\infty}$ was arbitrary, we have shown that $\mathcal{E}_A$ is closed under countable decreasing intersections. Therefore $\mathcal{E}_A$ is a monotone class containing $\mathcal{A}$, so $\mathcal{M}_\mathcal{A} \subseteq \mathcal{E}_A$ by Exercise 3.1.1. Hence we have shown that $A \cup E \in \mathcal{M}_\mathcal{A}$ for each $E \in \mathcal{M}_\mathcal{A}$.

Now define

$$\mathcal{F} := \{F \in \mathcal{M}_\mathcal{A} : F \cup E \text{ for each } E \in \mathcal{M}_\mathcal{A}\}.$$

Then $\mathcal{A} \subseteq \mathcal{F}$ since we have just shown that $A \cup E \in \mathcal{M}_\mathcal{A}$ for each $A \in \mathcal{A}$ and each $E \in \mathcal{M}_\mathcal{A}$. Again, one shows that $\mathcal{F}$ is a monotone class by a similar argument as above. As before, we have $\mathcal{M}_\mathcal{A} \subseteq \mathcal{F}$, so we have shown that $F \cup E \in \mathcal{M}_\mathcal{A}$ for each $E, F \in \mathcal{M}_\mathcal{A}$.

We claim that $\mathcal{M}_\mathcal{A}$ is a σ-algebra. The previous paragraph shows that $\mathcal{M}_\mathcal{A}$ is closed under finite unions. If $\{E_j\}_{j=1}^{\infty}$ is a sequence of sets in $\mathcal{M}_\mathcal{A}$, then we can define a new sequence $\{F_j\}_{j=1}^{\infty}$ by

$$F_j := \bigcup_{k=1}^{j} E_k.$$

Then $\{F_j\}_{j=1}^{\infty}$ is an increasing sequence of sets in $\mathcal{M}_\mathcal{A}$, so

$$\bigcup_{j=1}^{\infty} E_j = \bigcup_{j=1}^{\infty} F_j \in \mathcal{M}_\mathcal{A}.$$

Thus $\mathcal{M}_\mathcal{A}$ is closed under countable unions.

To see that $\mathcal{M}_\mathcal{A}$ is closed under complements, let

$$\mathcal{D} := \{D \in \mathcal{M}_\mathcal{A} : D^c \in \mathcal{M}_\mathcal{A}\}.$$

Then $\mathcal{A} \subseteq \mathcal{D}$ since $\mathcal{A}$ is closed under compliments. Once again, one shows that $\mathcal{D}$ is a monotone class by a similar argument as above. As before, we have $\mathcal{M}_\mathcal{A} \subseteq \mathcal{D}$, so we have shown that $D^c \in \mathcal{M}_\mathcal{A}$ for each $D \in \mathcal{M}_\mathcal{A}$. Therefore $\mathcal{M}_\mathcal{A}$ is closed under complements. This shows that $\mathcal{M}_\mathcal{A}$ is a σ-algebra containing $\mathcal{A}$, so $\sigma(\mathcal{A}) \subseteq \mathcal{M}_\mathcal{A}$. $\square$

3.1.3   Products of σ-finite Measures

Definition 3.1.11 (σ-finite)**.** Let $(X, \mathcal{A}, \mu)$ be a measure space. We say that $(X, \mathcal{A}, \mu)$ is *σ-finite* if there exists a countable collection of μ-measurable sets $\{E_j\}_{j=1}^{\infty}$ such that $X = \bigcup_{j=1}^{\infty} E_j$ and $\mu(E_j) < \infty$ for each $j \in \mathbb{Z}^+$.

A subset $A \subseteq X$ is *σ-finite* if there exists a countable collection of μ-measurable sets $\{A_j\}_{j=1}^{\infty}$ such that $A = \bigcup_{j=1}^{\infty} A_j$ and $\mu(A_j) < \infty$ for each $j \in \mathbb{Z}^+$.

A function $f : X \to [-\infty, \infty]$ is *σ-finite* if f is μ-measurable and the set $\{x \in X : f(x) \neq 0\}$ is σ-finite.

Lemma 3.1.12 (The Measure of Cross Sections is a Measurable Function). *Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be σ-finite measure spaces. If $E \in \mathcal{A} \otimes \mathcal{B}$, then the function*

$$x \longmapsto \nu(E_x)$$

is $\mathcal{A}$-measurable, and the function

$$y \longmapsto \mu(E^y)$$

is $\mathcal{B}$-measurable.

Proof. If $E \in \mathcal{A} \otimes \mathcal{B}$, we know that $E_x \in \mathcal{B}$ for each $x \in X$. Thus the function $x \mapsto \nu(E_x)$ is well-defined on X.

Step 1: Assume first that $\nu(Y) < \infty$.

Let

$$\mathcal{M} := \{E \in \mathcal{A} \otimes \mathcal{B} : x \mapsto \nu(E_x) \text{ is } \mathcal{A}\text{-measurable on } X\}.$$

We want to show that $\mathcal{M} = \mathcal{A} \otimes \mathcal{B}$.

If $A \in \mathcal{A}$ and $B \in \mathcal{B}$, then for each $x \in X$ we have

$$(A \times B)_x = \begin{cases} B & \text{if } x \in A \\ \varnothing & \text{if } x \notin A \end{cases}$$

so that

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

In particular, the function $x \longmapsto \nu((A \times B)_x)$ is $\mathcal{A}$-measurable on X since χ_A is $\mathcal{A}$-measurable. Therefore $A \times B \in \mathcal{M}$ for all $A \in \mathcal{A}$ and $B \in \mathcal{B}$.

Let $\mathcal{E}$ be the collection of all finite unions of measurable products $A \times B$ with $A \in \mathcal{A}$ and $B \in \mathcal{B}$, and let $E \in \mathcal{E}$. Then by Lemma 3.1.7, we can write E as a finite union of disjoint

measurable products $E = \bigcup_{j=1}^{m} E_j$. Then

$$\nu(E_x) = \nu\left(\left(\bigcup_{j=1}^{m} E_j\right)_x\right)$$

$$= \nu\left(\bigcup_{j=1}^{m} (E_j)_x\right)$$

$$= \sum_{j=1}^{m} \nu((E_j)_x)$$

by disjoint additivity of ν. Since each E_j is a measurable product, we have $E_j \in \mathcal{M}$, so the function $x \mapsto \nu((E_j)_x)$ is $\mathcal{A}$-measurable on X for each $j = 1, 2, \ldots, m$. Thus the function $x \mapsto \nu(E_x)$ is a finite sum of $\mathcal{A}$-measurable functions, so $x \mapsto \nu(E_x)$ is $\mathcal{A}$-measurable on X. Therefore $E \in \mathcal{M}$ for all $E \in \mathcal{E}$. Since $E \in \mathcal{E}$ was arbitrary, we have shown that $\mathcal{E} \subseteq \mathcal{M}$.

Our next goal is to show that $\mathcal{M}$ is a monotone class on $X \times Y$. Let $\{E_j\}_{j=1}^{\infty}$ be an increasing sequence of sets in $\mathcal{M}$, i.e.

$$E_1 \subseteq E_2 \subseteq E_3 \subseteq \cdots .$$

Then for each $x \in X$, we have

$$\left(\bigcup_{j=1}^{\infty} E_j\right)_x = \bigcup_{j=1}^{\infty} (E_j)_x$$

which implies that

$$\nu\left(\left(\bigcup_{j=1}^{\infty} E_j\right)_x\right) = \nu\left(\bigcup_{j=1}^{\infty} (E_j)_x\right) = \lim_{j \to \infty} \nu((E_j)_x)$$

since the sequence of sets $\{(E_j)_x\}_{j=1}^{\infty}$ is increasing. Since the pointwise limit of measurable functions is measurable, we have that the function $x \mapsto \nu\left(\left(\bigcup_{j=1}^{\infty} E_j\right)_x\right)$ is $\mathcal{A}$-measurable on X. Thus $\bigcup_{j=1}^{\infty} E_j \in \mathcal{M}$. This shows that $\mathcal{M}$ is closed under countable increasing unions.

Now let $\{E_j\}_{j=1}^{\infty}$ be a decreasing sequence of sets in $\mathcal{M}$, i.e.

$$E_1 \supseteq E_2 \supseteq E_3 \supseteq \cdots .$$

Then for each $x \in X$, we have

$$\left(\bigcap_{j=1}^{\infty} E_j\right)_x = \bigcap_{j=1}^{\infty} (E_j)_x$$

which implies that

$$\nu\left(\left(\bigcap_{j=1}^{\infty} E_j\right)_x\right) = \nu\left(\bigcap_{j=1}^{\infty}(E_j)_x\right) = \lim_{j\to\infty} \nu((E_j)_x)$$

since the sequence of sets $\{(E_j)_x\}_{j=1}^{\infty}$ is decreasing and $\nu(Y) < \infty$. Since the pointwise limit of measurable functions is measurable, we have that the function $x \mapsto \nu\left(\left(\bigcap_{j=1}^{\infty} E_j\right)_x\right)$ is $\mathcal{A}$-measurable on X. Thus $\bigcap_{j=1}^{\infty} E_j \in \mathcal{M}$. This shows that $\mathcal{M}$ is closed under countable decreasing intersections; hence $\mathcal{M}$ is a monotone class.

We have shown that $\mathcal{M}$ is a monotone class that contains the algebra $\mathcal{E}$ of all finite unions of measurable products in $\mathcal{A} \times [B]$. By the Monotone Class Lemma (Theorem 3.1.10), the smallest σ-algebra containing $\mathcal{E}$ is the smallest monotone class containing $\mathcal{E}$, so we have $\mathcal{A} \otimes \mathcal{B} \subset \mathcal{M}$. This completes the proof of the fact that $x \mapsto \nu(E_x)$ is $\mathcal{A}$-measurable on X for each $E \in \mathcal{A} \otimes \mathcal{B}$ in the case that $\nu(Y) < \infty$.

Step 2: Now we prove the general case.

Since $(Y, \mathcal{B}, \nu)$ is σ-finite, there exists an increasing sequence $\{Y_j\}_{j=1}^{\infty}$ of ν-measurable subsets of Y such that $Y = \bigcup_{j=1}^{\infty} Y_j$ and $\nu(Y_j) < \infty$ for each $j \in \mathbb{Z}^+$. If $E \in \mathcal{A} \otimes \mathcal{B}$, then

$$\nu(E_x) = \lim_{j\to\infty} \nu((E \cap (X \times Y_j))_x)$$

for each $x \in X$ since $\{(E \cap (X \times Y_j))_x\}_{j=1}^{\infty}$ is an increasing sequence of sets whose union is E_x. By Step 1, for each $j \in \mathbb{Z}^+$ the function $x \mapsto \nu((E \cap (X \times Y_j))_x)$ is $\mathcal{A}$-measurable on X since $E \cap (X \times Y_j) \in \mathcal{A} \otimes \mathcal{B}$ and $\nu(Y_j) < \infty$. Since the pointwise limit of measurable functions is measurable, we have that the function $x \mapsto \nu(E_x)$ is $\mathcal{A}$-measurable on X. This completes the proof of the first part of the lemma.

The proof of the second part is similar. $\qquad\square$

At last, we are ready to define the product of two measures.

Definition 3.1.13 (Product of Measures)**.** Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be σ-finite measure spaces. We define their *product* $\mu \otimes \nu : \mathcal{A} \otimes \mathcal{B} \to [0, \infty]$ by

$$(\mu \otimes \nu)(E) = \int_X \int_Y \mathbf{1}_E(x, y) \, d\nu(y) \, d\mu(x).$$

Notice that we need the σ-finiteness assumption to ensure that the function $x \mapsto \nu(E_x)$ is $\mathcal{A}$-measurable, so that the outer integral is well-defined.

Lemma 3.1.14. *Let* $(X, \mathcal{A}, \mu)$ *and* $(Y, \mathcal{B}, \nu)$ *be* σ*-finite measure spaces. Then the product* $\mu \otimes \nu : \mathcal{A} \otimes \mathcal{B} \to [0, \infty]$ *is a measure on* $X \times Y$.

Proof. Clearly we have $(\mu \otimes \nu)(\varnothing) = 0$ since $\chi_\varnothing(x, y) = 0$ for all $(x, y) \in X \times Y$. Now let $\{E_j\}_{j=1}^\infty$ be a sequence of disjoint sets in $\mathcal{A} \otimes \mathcal{B}$. Then we have

$$(\mu \otimes \nu)\left(\bigcup_{j=1}^\infty E_j\right) = \int_X \int_Y \mathbf{1}_{\bigcup_{j=1}^\infty E_j}(x, y) \; \mathrm{d}\nu(y) \; \mathrm{d}\mu(x)$$

$$= \int_X \nu\left(\bigcup_{j=1}^\infty (E_j)_x\right) \mathrm{d}\mu(x)$$

$$= \int_X \sum_{j=1}^\infty \nu((E_j)_x) \; \mathrm{d}\mu(x)$$

$$= \sum_{j=1}^\infty \int_X \nu((E_j)_x) \; \mathrm{d}\mu(x)$$

$$= \sum_{j=1}^\infty \int_X \int_Y \mathbf{1}_{E_j}(x, y) \; \mathrm{d}\nu(y) \; \mathrm{d}\mu(x)$$

$$= \sum_{j=1}^\infty (\mu \otimes \nu)(E_j)$$

where the interchange of the sum and integral is justified by the Monotone Convergence Theorem (2.1.8). Thus $\mu \otimes \nu$ is countably additive, so $\mu \otimes \nu$ is a measure on $X \times Y$.   □

3.1.4   Tonelli's Theorem and Fubini's Theorem

We have built up all the machinery we need to state and prove Tonelli's theorem and Fubini's theorem.

Theorem 3.1.15 (Tonelli's Theorem). *Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be σ-finite measure spaces, and let $f : X \times Y \to [0, \infty]$ be a nonnegative $\mathcal{A} \otimes \mathcal{B}$-measurable function. Then*

$$x \longmapsto \int_Y f(x, y) \; \mathrm{d}\nu(y) \qquad \textit{is a μ-measurable function on X,}$$

$$y \longmapsto \int_X f(x, y) \; \mathrm{d}\mu(x) \qquad \textit{is a ν-measurable function on Y,}$$

and

$$\int_{X \times Y} f \; \mathrm{d}(\mu \otimes \nu) = \int_X \int_Y f(x, y) \; \mathrm{d}\nu(y) \; \mathrm{d}\mu(x) = \int_Y \int_X f(x, y) \; \mathrm{d}\mu(x) \; \mathrm{d}\nu(y).$$

We want to notice a few things — first, the spaces we integrate over must be σ-finite. Second, the function f must be nonnegative. Third is that we make *no* integrability assumptions on f. This is a very powerful theorem.

Proof. Step 0: First consider the special case where $f = \mathbf{1}_E$ for some $E \in \mathcal{A} \otimes \mathcal{B}$, and assume that $\mu(X) < \infty$ and $\nu(Y) < \infty$.

Let $E \in \mathcal{A} \otimes \mathcal{B}$ be arbitrary. Then we have

$$\int_Y \mathbf{1}_E(x, y) \; \mathrm{d}\nu(y) = \nu(E_x)$$

for each $x \in X$, and similarly

$$\int_X \mathbf{1}_E(x, y) \; \mathrm{d}\mu(x) = \mu(E^y)$$

for each $y \in Y$. By Lemma 3.1.12, the functions $x \mapsto \nu(E_x)$ and $y \mapsto \mu(E^y)$ are measurable. Thus, in this case, the first two conclusions of Tonelli's theorem hold.

Now let

$$\mathcal{M} := \left\{ E \in \mathcal{A} \otimes \mathcal{B} : \int_X \int_Y \mathbf{1}_E(x, y) \; \mathrm{d}\nu(y) \; \mathrm{d}\mu(x) = \int_Y \int_X \mathbf{1}_E(x, y) \; \mathrm{d}\mu(x) \; \mathrm{d}\nu(y) \right\}.$$

If $A \in \mathcal{A}$ and $B \in \mathcal{B}$, then we have $A \times B \in \mathcal{M}$ since

$$
\begin{aligned}
\int_X \int_Y \mathbf{1}_{A \times B}(x, y) \; \mathrm{d}\nu(y) \; \mathrm{d}\mu(x) &= \int_X \int_Y \mathbf{1}_A(x) \mathbf{1}_B(y) \; \mathrm{d}\nu(y) \; \mathrm{d}\mu(x) \\
&= \int_X \mathbf{1}_A(x) \left(\int_Y \mathbf{1}_B(y) \; \mathrm{d}\nu(y) \right) \mathrm{d}\mu(x) \\
&= \nu(B) \int_X \mathbf{1}_A(x) \; \mathrm{d}\mu(x) \\
&= \nu(B)\mu(A)
\end{aligned}
$$

and similarly the other integral equals $\mu(A)\nu(B)$. Thus $A \times B \in \mathcal{M}$ for all $A \in \mathcal{A}$ and $B \in \mathcal{B}$.

Let $\mathcal{E}$ be the collection of all finite unions of measurable products $A \times B$ with $A \in \mathcal{A}$ and $B \in \mathcal{B}$. Then by Lemma 3.1.7, we can write an arbitrary $E \in \mathcal{E}$ as a finite union of disjoint measurable products $E = \bigcup_{j=1}^m E_j$. Then by disjoint additivity of the integral, we have

$$
\begin{aligned}
\int_X \int_Y \mathbf{1}_E(x, y) \; \mathrm{d}\nu(y) \; \mathrm{d}\mu(x) &= \int_X \int_Y \mathbf{1}_{\bigcup_{j=1}^m E_j}(x, y) \; \mathrm{d}\nu(y) \; \mathrm{d}\mu(x) \\
&= \int_X \int_Y \sum_{j=1}^m \mathbf{1}_{E_j}(x, y) \; \mathrm{d}\nu(y) \; \mathrm{d}\mu(x) \\
&= \sum_{j=1}^m \int_X \int_Y \mathbf{1}_{E_j}(x, y) \; \mathrm{d}\nu(y) \; \mathrm{d}\mu(x) \\
&= \sum_{j=1}^m \int_Y \int_X \mathbf{1}_{E_j}(x, y) \; \mathrm{d}\mu(x) \; \mathrm{d}\nu(y) \\
&= \int_Y \int_X \mathbf{1}_{\bigcup_{j=1}^m E_j}(x, y) \; \mathrm{d}\mu(x) \; \mathrm{d}\nu(y) \\
&= \int_Y \int_X \mathbf{1}_E(x, y) \; \mathrm{d}\mu(x) \; \mathrm{d}\nu(y).
\end{aligned}
$$

Thus $E \in \mathcal{M}$. Since $E \in \mathcal{E}$ was arbitrary, we have shown that $\mathcal{E} \subseteq \mathcal{M}$.

Now the Monotone Convergence Theorem (2.1.8) shows that $\mathcal{M}$ is closed under countable increasing unions. The Bounded Convergence Theorem (**??**) shows that $\mathcal{M}$ is closed under countable decreasing intersections — note that *this* is where we need the finiteness assumptions on $\mu(X)$ and $\nu(Y)$. Thus $\mathcal{M}$ is a monotone class containing the algebra $\mathcal{E}$ of all finite unions of measurable products in $\mathcal{A} \times \mathcal{B}$. By the Monotone Class Lemma (Theorem 3.1.10), the smallest σ-algebra containing $\mathcal{E}$ is the smallest monotone class containing $\mathcal{E}$, so we have $\mathcal{A} \otimes \mathcal{B} \subset \mathcal{M}$. That is, for each $E \in \mathcal{A} \otimes \mathcal{B}$ we have

$$\int_X \int_Y \mathbf{1}_E(x,y) \ \mathrm{d}\nu(y) \ \mathrm{d}\mu(x) = \int_Y \int_X \mathbf{1}_E(x,y) \ \mathrm{d}\mu(x) \ \mathrm{d}\nu(y).$$

Since the integral on the left-hand side is precisely $(\mu \otimes \nu)(E)$ by Definition 3.1.13, we have

$$(\mu \otimes \nu)(E) = \int_{X \times Y} \mathbf{1}_E \ \mathrm{d}(\mu \otimes \nu) = \int_X \int_Y \mathbf{1}_E(x,y) \ \mathrm{d}\nu(y) \ \mathrm{d}\mu(x) = \int_Y \int_X \mathbf{1}_E(x,y) \ \mathrm{d}\mu(x) \ \mathrm{d}\nu(y).$$

This completes the proof of Tonelli's theorem in the special case where $f = \mathbf{1}_E$ for some $E \in \mathcal{A} \otimes \mathcal{B}$ and $\mu(X) < \infty$ and $\nu(Y) < \infty$.

Step 1: Now consider the case where $f = \mathbf{1}_E$ for some $E \in \mathcal{A} \otimes \mathcal{B}$, but we do not assume that $\mu(X) < \infty$ or $\nu(Y) < \infty$.

Since $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ are σ-finite, there exist increasing sequences $\{X_j\}_{j=1}^{\infty}$ and $\{Y_k\}_{k=1}^{\infty}$ of measurable subsets of X and Y, respectively, such that $X = \bigcup_{j=1}^{\infty} X_j$ and $Y = \bigcup_{k=1}^{\infty} Y_k$ and $\mu(X_j) < \infty$ and $\nu(Y_k) < \infty$ for each $j, k \in \mathbb{Z}^+$.

Let $E \in \mathcal{A} \otimes \mathcal{B}$. For each $j, k \in \mathbb{Z}^+$, we apply the previous case to the finite measures $\mu \restriction_{X_j}$ and $\nu \restriction_{Y_k}$ to conclude that

$$\int_{X_j} \int_{Y_k} \mathbf{1}_E(x,y) \ \mathrm{d}\nu(y) \ \mathrm{d}\mu(x) = \int_{Y_k} \int_{X_j} \mathbf{1}_E(x,y) \ \mathrm{d}\mu(x) \ \mathrm{d}\nu(y).$$

Now by the Monotone Convergence Theorem (2.1.8), we have

$$\int_X \int_Y \mathbf{1}_E(x,y) \ \mathrm{d}\nu(y) \ \mathrm{d}\mu(x) = \lim_{j \to \infty} \lim_{k \to \infty} \int_{X_j} \int_{Y_k} \mathbf{1}_E(x,y) \ \mathrm{d}\nu(y) \ \mathrm{d}\mu(x)$$

$$= \lim_{j \to \infty} \lim_{k \to \infty} \int_{Y_k} \int_{X_j} \mathbf{1}_E(x,y) \ \mathrm{d}\mu(x) \ \mathrm{d}\nu(y)$$

$$= \int_Y \int_X \mathbf{1}_E(x,y) \ \mathrm{d}\mu(x) \ \mathrm{d}\nu(y).$$

Now the definition of the product measure (Definition 3.1.13) shows that

$$(\mu \otimes \nu)(E) = \int_X \int_Y \mathbf{1}_E(x,y) \ \mathrm{d}\nu(y) \ \mathrm{d}\mu(x).$$

Putting this together, we have

$$\int_{X \times Y} \mathbf{1}_E \ \mathrm{d}(\mu \otimes \nu) = \int_X \int_Y \mathbf{1}_E(x, y) \ \mathrm{d}\nu(y) \ \mathrm{d}\mu(x) = \int_Y \int_X \mathbf{1}_E(x, y) \ \mathrm{d}\mu(x) \ \mathrm{d}\nu(y).$$

Since $E \in \mathcal{A} \otimes \mathcal{B}$ was arbitrary, this completes the proof of Tonelli's theorem in the case where $f = \mathbf{1}_E$ for some $E \in \mathcal{A} \otimes \mathcal{B}$.

Step 2: Now consider the general case.

Let $f : X \times Y \to [0, \infty]$ be a nonnegative $\mathcal{A} \otimes \mathcal{B}$-measurable function. Then we define a sequence of nonnegative simple functions $\{s_k\}_{k=1}^{\infty}$ on $X \times Y$ by

$$s_k(x, y) := \begin{cases} \frac{m}{2^k}, & \text{if } f(x, y) < k \text{ and } \frac{m}{2^k} \leqslant f(x, y) < \frac{m+1}{2^k} \text{ for some } m = 0, 1, 2, \ldots, 2^{2k}k - 1 \\ k, & \text{if } f(x, y) \geqslant k \end{cases}$$

for each $k \in \mathbb{Z}^+$.

Note that by construction, we have

$$0 \leqslant s_k(x, y) \leqslant s_{k+1}(x, y) \leqslant f(x, y)$$

for all $(x, y) \in X \times Y$ and $k \in \mathbb{Z}^+$, and

$$\lim_{k \to \infty} s_k(x, y) = f(x, y)$$

for all $(x, y) \in X \times Y$. Notice that each s_k is a finite linear combination of characteristic functions of measurable sets, so by the previous step we know Tonelli's theorem holds for each s_k. Now the Monotone Convergence Theorem (2.1.8) shows that

$$\int_Y f(x, y) \ \mathrm{d}\nu(y) = \lim_{k \to \infty} \int_Y s_k(x, y) \ \mathrm{d}\nu(y)$$

for each $x \in X$. Thus the function

$$x \longmapsto \int_Y f(x, y) \ \mathrm{d}\nu(y)$$

is the pointwise limit of a sequence of $\mathcal{A}$-measurable functions, so it is $\mathcal{A}$-measurable on X. Similarly the map

$$y \longmapsto \int_X f(x, y) \ \mathrm{d}\mu(x)$$

is $\mathcal{B}$-measurable on Y.

Finally, another application of the Monotone Convergence Theorem (2.1.8) shows that

$$\begin{aligned} \int_{X \times Y} f \ \mathrm{d}(\mu \otimes \nu) &= \lim_{k \to \infty} \int_{X \times Y} s_k \ \mathrm{d}(\mu \otimes \nu) \\ &= \lim_{k \to \infty} \int_X \int_Y s_k(x, y) \ \mathrm{d}\nu(y) \ \mathrm{d}\mu(x) \\ &= \int_X \int_Y f(x, y) \ \mathrm{d}\nu(y) \ \mathrm{d}\mu(x) \end{aligned}$$

and similarly

$$\int_{X \times Y} f \; \mathrm{d}(\mu \otimes \nu) = \int_Y \int_X f(x,y) \; \mathrm{d}\mu(x) \; \mathrm{d}\nu(y).$$

This completes the proof of Tonelli's theorem. $\square$

Corollary 3.1.16. *Let* $\{a_{i,j}\}_{i,j=1}^\infty$ *be a doubly-indexed collection of nonnegative numbers. Then*

$$\sum_{i,j=1}^\infty a_{i,j} = \sum_{i=1}^\infty \sum_{j=1}^\infty a_{i,j} = \sum_{j=1}^\infty \sum_{i=1}^\infty a_{i,j}.$$

Proof. Apply Tonelli's theorem with $X = Y = \mathbb{Z}^+$, $\mu = \nu$ the counting measure, and $f(i,j) = a_{ij}$. $\square$

Theorem 3.1.17 (Fubini's Theorem). *Let* $(X, \mathcal{A}, \mu)$ *and* $(Y, \mathcal{B}, \nu)$ *be* σ-*finite measure spaces, and let* $f : X \times Y \to [-\infty, \infty]$ *be an* $\mathcal{A} \otimes \mathcal{B}$-*integrable function. Then*

$$x \longmapsto \int_Y f(x,y) \; \mathrm{d}\nu(y) \qquad \text{is a } \mu\text{-measurable function on } X,$$

$$y \longmapsto \int_X f(x,y) \; \mathrm{d}\mu(x) \qquad \text{is a } \nu\text{-measurable function on } Y,$$

and

$$\int_Y |f(x,y)| \; \mathrm{d}\nu(y) < \infty \text{ for } \mu\text{-almost every } x \in X \qquad \text{and} \qquad \int_X |f(x,y)| \; \mathrm{d}\mu(x) < \infty \text{ for } \nu\text{-almost every } y \in Y$$

and

$$\int_{X \times Y} f \; \mathrm{d}(\mu \otimes \nu) = \int_X \int_Y f(x,y) \; \mathrm{d}\nu(y) \; \mathrm{d}\mu(x) = \int_Y \int_X f(x,y) \; \mathrm{d}\mu(x) \; \mathrm{d}\nu(y).$$

We remark a few things — first, the spaces we integrate over must be σ-finite. Second, the function f can take on negative values, but we must assume that f is integrable, i.e. that $\int_{X \times Y} |f| \; \mathrm{d}(\mu \otimes \nu) < \infty$. Third is that we have the same conclusion as Tonelli's theorem, but we also have that the iterated integrals of $|f|$ are finite almost everywhere.

Remark 3.1.18 (Using Tonelli and Fubini in Practice). Here's the trick. Assume that you have a function f defined on a product space $X \times Y$ and you want to compute its integral. What you should do is apply Tonelli's theorem to $|f|$ and show that the integral $\int_{X \times Y} |f| \; \mathrm{d}(\mu \otimes \nu)$ is finite by evaluating one of the iterated integrals given in Tonelli's theorem. This shows that f is integrable on $X \times Y$. Then you can apply Fubini's theorem to f to compute the integral $\int_{X \times Y} f \; \mathrm{d}(\mu \otimes \nu)$ by evaluating either of the iterated integrals given in Fubini's theorem.

In practice, this combo is referred to as the Fubini-Tonelli theorem or just Fubini's theorem.

Proof. Applying Tonelli's theorem (Theorem 3.1.15) to the nonnegative $\mathcal{A} \otimes \mathcal{B}$-measurable function $|f|$, we have

$$x \longmapsto \int_Y |f(x,y)| \, d\nu(y) \qquad \text{is a } \mu\text{-measurable function on } X,$$

so the set

$$\left\{ x \in X : \int_Y |f(x,y)| \, d\nu(y) = \infty \right\}$$

is in $\mathcal{A}$. Since f is integrable on $X \times Y$, Tonelli's theorem and the assumption f is integrable shows that

$$\int_{X \times Y} |f| \, d(\mu \otimes \nu) = \int_X \int_Y |f(x,y)| \, d\nu(y) \, d\mu(x) < \infty.$$

Thus we must have

$$\mu\left(\left\{ x \in X : \int_Y |f(x,y)| \, d\nu(y) = \infty \right\} \right) = 0.$$

Recall the definitions

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

which satisfy $f = f^+ - f^-$ and $|f| = f^+ + f^-$. Applying Tonelli's theorem to the nonnegative $\mathcal{A} \otimes \mathcal{B}$-measurable functions f^+ and f^- shows that

$$x \longmapsto \int_Y f^+(x,y) \, d\nu(y) \qquad \text{and} \qquad x \longmapsto \int_Y f^-(x,y) \, d\nu(y)$$

are μ-measurable functions on X. Since $f^+ \leqslant |f|$ and $f^- \leqslant |f|$, the sets

$$\left\{ x \in X : \int_Y f^+(x,y) \, d\nu(y) = \infty \right\} \quad \text{and} \quad \left\{ x \in X : \int_Y f^-(x,y) \, d\nu(y) = \infty \right\}$$

must also have μ-measure zero. The intersection of these two sets is

$$\left\{ x \in X : \int_Y |f(x,y)| \, d\nu(y) = \infty \right\},$$

which is exacly the set on which the integral $\int_Y f(x,y) \, d\nu(y)$ fails to be defined. Thus we have shown that

$$\int_Y |f(x,y)| \, d\nu(y) < \infty \text{ for } \mu\text{-almost every } x \in X.$$

Now the function

$$x \longmapsto \int_Y f(x,y) \, d\nu(y) = \int_Y f^+(x,y) \, d\nu(y) - \int_Y f^-(x,y) \, d\nu(y)$$

which must be modified to take the valued 0 on if the right side is of the form $\infty - \infty$ (but this happens on a set of μ-measure zero, as we just showed), is the difference of two μ-measurable functions, so it is μ-measurable on X. We compute that

$$
\begin{aligned}
\int_{X \times Y} f \, \mathrm{d}(\mu \otimes \nu) &= \int_{X \times Y} f^+ \, \mathrm{d}(\mu \otimes \nu) - \int_{X \times Y} f^- \, \mathrm{d}(\mu \otimes \nu) \\
&= \int_X \int_Y f^+(x,y) \, \mathrm{d}\nu(y) \, \mathrm{d}\mu(x) - \int_X \int_Y f^-(x,y) \, \mathrm{d}\nu(y) \, \mathrm{d}\mu(x) \\
&= \int_X \int_Y (f^+(x,y) - f^-(x,y)) \, \mathrm{d}\nu(y) \, \mathrm{d}\mu(x) \\
&= \int_X \int_Y f(x,y) \, \mathrm{d}\nu(y) \, \mathrm{d}\mu(x)
\end{aligned}
$$

where we have used the definition of the integral of a real-valued function and Tonelli's theorem on f^+ and f^-.

This proves the first iterated integral formula in Fubini's theorem.

The assertions about the ν-measurability of the function $y \mapsto \int_X f(x,y) \, \mathrm{d}\mu(x)$ and the second iterated integral formula in Fubini's theorem follow by a symmetric argument. $\square$

We summarize the technique described in Remark 3.1.18 in the following result.

Corollary 3.1.19 (The Fubini-Tonelli Theorem). *Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be σ-finite measure spaces, and let $f : X \times Y \to [-\infty, \infty]$ be an $\mathcal{A} \otimes \mathcal{B}$-measurable function. If either*

$$
\int_X \int_Y |f(x,y)| \, \mathrm{d}\nu(y) \, \mathrm{d}\mu(x) < \infty \quad or \quad \int_Y \int_X |f(x,y)| \, \mathrm{d}\mu(x) \, \mathrm{d}\nu(y) < \infty,
$$

then f is integrable on $X \times Y$, and

$$
\int_{X \times Y} f \, \mathrm{d}(\mu \otimes \nu) = \int_X \int_Y f(x,y) \, \mathrm{d}\nu(y) \, \mathrm{d}\mu(x) = \int_Y \int_X f(x,y) \, \mathrm{d}\mu(x) \, \mathrm{d}\nu(y).
$$

Proof. The function $|f|$ is nonnegative and $\mathcal{A} \otimes \mathcal{B}$-measurable, so we can apply Tonelli's theorem to $|f|$ to conclude that

$$
\int_{X \times Y} |f| \, \mathrm{d}(\mu \otimes \nu) = \int_X \int_Y |f(x,y)| \, \mathrm{d}\nu(y) \, \mathrm{d}\mu(x) = \int_Y \int_X |f(x,y)| \, \mathrm{d}\mu(x) \, \mathrm{d}\nu(y).
$$

If either of the iterated integrals on the right-hand side is finite, then $\int_{X \times Y} |f| \, \mathrm{d}(\mu \otimes \nu) < \infty$, so f is integrable on $X \times Y$. We can then apply Fubini's theorem to f to conclude the desired result. $\square$

Exercise 3.1.2 (Area Under Graph Formula). Let $(X, \mathcal{A}, \mu)$ be a σ-finite measure space and let $f : X \to [0, \infty]$ be a μ-measurable function. Then

$$
\int_X f \, d\mu = \int_0^\infty \mu(\{x \in X : t < f(x)\}) \, \mathrm{d}t.
$$

Proof. Let $\mathcal{B}_{\mathbb{R}}$ be the Borel σ-algebra on $\mathbb{R}$.

For each $k \in \mathbb{Z}^+$ let

$$E_k := \bigcup_{m=0}^{k^2-1} \left(f^{-1}\left(\left[\frac{m}{k}, \frac{m+1}{k} \right) \right) \right) \quad \text{and} \quad F_k := f^{-1}\left([k, \infty) \right) \times (0, k).$$

The for each $k \in \mathbb{Z}^+$ the set E_k is a finite union of measurable products in $\mathcal{A} \times \mathcal{B}_{\mathbb{R}}$, and F_k is a measurable product in $\mathcal{A} \times \mathcal{B}_{\mathbb{R}}$, so both E_k and F_k are in $\mathcal{A} \otimes \mathcal{B}_{\mathbb{R}}$. Since

$$\{(x,t) \in X \times \mathbb{R} : 0 < t < f(x)\} = \bigcup_{k=1}^{\infty}(E_k \cup F_k),$$

the set $\{(x,t) \in X \times \mathbb{R} : 0 < t < f(x)\}$ is in $\mathcal{A} \otimes \mathcal{B}_{\mathbb{R}}$.

The definition of the product measure $\mu \otimes \mathcal{L}^1$ then gives

$$\begin{aligned}
(\mu \otimes \mathcal{L}^1)\left(\{(x,t) \in X \times \mathbb{R} : 0 < t < f(x)\} \right) &= \int_X \int_{\mathbb{R}} \mathbf{1}_{\{(x,t):0<t<f(x)\}}(x,t) \, \mathrm{d}\mathcal{L}^1(t) \, \mathrm{d}\mu(x) \\
&= \int_X \int_0^{f(x)} 1 \, \mathrm{d}t \, \mathrm{d}\mu(x) \\
&= \int_X f(x) \, \mathrm{d}\mu(x).
\end{aligned}$$

This formula shows that the measure of the set $\{(x,t) \in X \times \mathbb{R} : 0 < t < f(x)\}$ with respect to the product measure $\mu \otimes \mathcal{L}^1$ is equal to the integral of f over X. That is, the "area under the graph" of f is equal to the integral of f.

On the other hand, Tonelli's theorem (Theorem 3.1.15) says that we chan interchange the order of integration to get

$$\begin{aligned}
(\mu \otimes \mathcal{L}^1)\left(\{(x,t) \in X \times \mathbb{R} : 0 < t < f(x)\} \right) &= \int_0^{\infty} \int_X \mathbf{1}_{\{(x,t):0<t<f(x)\}}(x,t) \, \mathrm{d}\mu(x) \, \mathrm{d}\mathcal{L}^1(t) \\
&= \int_0^{\infty} \mu(\{x \in X : t < f(x)\}) \, \mathrm{d}t.
\end{aligned}$$

Combinging the two formulas gives

$$\int_X f(x) \, \mathrm{d}\mu(x) = \int_0^{\infty} \mu(\{x \in X : t < f(x)\}) \, \mathrm{d}t,$$

which is the desired result. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\square$

3.1.5 Notation for Multiple Integrals in $\mathbb{R}^n$

Lemma 3.1.20 (Another Definition of Lebesgue Measure). *For each $n > 1$, the Lebesgue measure $\mathcal{L}^n$ on $\mathbb{R}^n$ is the product measure of $(\mathbb{R}^{n-1}, \mathcal{L}^{n-1})$ and $(\mathbb{R}, \mathcal{L}^1)$, i.e.*

$$\mathcal{L}^n = \mathcal{L}^{n-1} \otimes \mathcal{L}^1.$$

Proof. Fix $n > 1$. Then the measure spaces $(\mathbb{R}^{n-1}, \mathcal{L}^{n-1})$ and $(\mathbb{R}, \mathcal{L}^1)$ are σ-finite since $\mathbb{R}^{n-1}$ and $\mathbb{R}$ can both be written as a countable union of finite measure sets. Thus the product measure $\mathcal{L}^{n-1} \otimes \mathcal{L}^1$ is well-defined on $\mathbb{R}^{n-1} \times \mathbb{R} = \mathbb{R}^n$.

By Cor 1.1.18, we only need to show that $\mathcal{L}^n$ and $\mathcal{L}^{n-1} \otimes \mathcal{L}^1$ agree on all boxes in $\mathbb{R}^n$. Let $R = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$ be a box in $\mathbb{R}^n$. Then

$$
\begin{aligned}
(\mathcal{L}^{n-1} \otimes \mathcal{L}^1)(R) &= \int_{\mathbb{R}^{n-1}} \int_{\mathbb{R}} \mathbf{1}_R(x_1, x_2, \ldots, x_n) \, d\mathcal{L}^1(x_n) \, d\mathcal{L}^{n-1}(x_1, x_2, \ldots, x_{n-1}) \\
&= \int_{\mathbb{R}^{n-1}} \int_{a_n}^{b_n} \mathbf{1}_{[a_1,b_1] \times [a_2,b_2] \times \cdots \times [a_{n-1},b_{n-1}]}(x_1, x_2, \ldots, x_{n-1}) \, dx_n \, d\mathcal{L}^{n-1}(x_1, x_2, \ldots, x_{n-1}) \\
&= \int_{a_n}^{b_n} dx_n \cdot \int_{\mathbb{R}^{n-1}} \mathbf{1}_{[a_1,b_1] \times [a_2,b_2] \times \cdots \times [a_{n-1},b_{n-1}]}(x_1, x_2, \ldots, x_{n-1}) \, d\mathcal{L}^{n-1}(x_1, x_2, \ldots, x_{n-1}) \\
&= (b_n - a_n) \int_{\mathbb{R}^{n-1}} \mathbf{1}_{[a_1,b_1] \times [a_2,b_2] \times \cdots \times [a_{n-1},b_{n-1}]}(x_1, x_2, \ldots, x_{n-1}) \, d\mathcal{L}^{n-1}(x_1, x_2, \ldots, x_{n-1}) \\
&= (b_n - a_n) \mathcal{L}^{n-1}([a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_{n-1}, b_{n-1}]) \\
&= (b_n - a_n)(b_1 - a_1)(b_2 - a_2) \cdots (b_{n-1} - a_{n-1}) \\
&= (b_1 - a_1)(b_2 - a_2) \cdots (b_n - a_n) = \mathcal{L}^n(R)
\end{aligned}
$$

so we are done. $\square$

By this result and induction, if $n > 1$ and we have an integer $1 \leqslant k < n$, then

$$
\mathcal{L}^n = \mathcal{L}^{n-k} \otimes \mathcal{L}^k.
$$

Then we can apply Tonelli's theorem and Fubini's theorem to functions defined on $\mathbb{R}^n$ to get formulas such as

$$
\int_{\mathbb{R}^k \times \mathbb{R}^{n-k}} f \, dx = \int_{\mathbb{R}^k} \int_{\mathbb{R}^{n-k}} f(x_1, x_2) \, d\mathcal{L}^{n-k}(x_2) \, d\mathcal{L}^k(x_1) = \int_{\mathbb{R}^{n-k}} \int_{\mathbb{R}^k} f(x_1, x_2) \, d\mathcal{L}^k(x_1) \, d\mathcal{L}^{n-k}(x_2).
$$

for integrable $f : \mathbb{R}^n \to [-\infty, \infty]$. We remark that you should be careful in using the above formula, since we have $x_1 \in \mathbb{R}^k$ and $x_2 \in \mathbb{R}^{n-k}$ which is at odds with the usual notation.

To avoid confusion, we usually write $x \in \mathbb{R}^n$ as $x = (x_1, x_2, \ldots, x_n)$ where each $x_j \in \mathbb{R}$. Then Fubini-Tonelli takes the form

$$
\int_{\mathbb{R}^n} f(x) \, dx = \int_{\mathbb{R}^k} \int_{\mathbb{R}^{n-k}} f(x_1, x_2, \ldots, x_n) \, dx_{k+1} \cdots dx_n \, dx_1 \cdots dx_k = \int_{\mathbb{R}^{n-k}} \int_{\mathbb{R}^k} f(x_1, x_2, \ldots, x_n) \, dx_1 \cdots dx_k
$$

for integrable $f : \mathbb{R}^n \to [-\infty, \infty]$, which we think is better (because it is more explicit and conforms to the usual notation), even if it is longer to write.

The next section will be devoted to examples and applications of Fubini-Tonelli theorems in $\mathbb{R}^n$.

3.2 Applications of The Fubini-Tonelli Theorem to Integrals in $\mathbb{R}^n$

With the Fubini-Tonelli theorem, we finally have the tools to prove the following important result about how Lebesgue measure behaves under linear transformations. This completes our verification that Lebesgue measure is the "right" notion of volume in $\mathbb{R}^n$, and behaves as expected, at least for measurable sets.

This first exercise does not use the Fubini-Tonelli theorem, but it is a useful property of the Lebesgue integral that will be used in the following Proposition about Lebesgue measure and linear maps.

Exercise 3.2.1 (Translation Invariance of the Lebesgue Integral on $\mathbb{R}^n$). Let $f \in L^1(\mathbb{R}^n)$ and let $y \in \mathbb{R}^n$. Then

$$\int_{\mathbb{R}^n} f(x - y) \, \mathrm{d}x = \int_{\mathbb{R}^n} f(x) \, \mathrm{d}x.$$

Proof. Step 1: We prove the result for characteristic functions of measurable sets of finite measure.

Let $E \subseteq \mathbb{R}^n$ be a measurable set with $\mathcal{L}^n(E) < \infty$, and let $y \in \mathbb{R}^n$. Then we compute

$$\mathbf{1}_{E+y}(x) = \begin{cases} 1 & \text{if } x \in E + y \\ 0 & \text{if } x \notin E + y \end{cases} = \begin{cases} 1 & \text{if } x - y \in E \\ 0 & \text{if } x - y \notin E \end{cases} = \mathbf{1}_E(x - y)$$

for each $x \in \mathbb{R}^n$, which implies that

$$\int_{\mathbb{R}^n} \mathbf{1}_{E+y}(x) \, \mathrm{d}x = \mathcal{L}^n(E + y) = \mathcal{L}^n(E) = \int_{\mathbb{R}^n} \mathbf{1}_E(x) \, \mathrm{d}x$$

by translation invariance of Lebesgue measure (Proposition 1.2.15).

Thus the result holds for characteristic functions of measurable sets with finite measure.

Step 2: We prove the result for simple functions.

Let $\varphi : \mathbb{R}^n \to [0, \infty)$ be a simple function. Then there exist measurable sets $E_1, \ldots, E_m \subseteq \mathbb{R}^n$ with $\mathcal{L}^n(E_j) < \infty$ for each $j = 1, \ldots, m$, and nonnegative numbers $a_1, \ldots, a_m \geq 0$ such that

$$\varphi = \sum_{j=1}^m a_j \mathbf{1}_{E_j}.$$

Then we compute

$$\int_{\mathbb{R}^n} \varphi(x-y) \; \mathrm{d}x = \int_{\mathbb{R}^n} \sum_{j=1}^{m} a_j \mathbf{1}_{E_j}(x-y) \; \mathrm{d}x$$

$$= \sum_{j=1}^{m} a_j \int_{\mathbb{R}^n} \mathbf{1}_{E_j}(x-y) \; \mathrm{d}x$$

$$= \sum_{j=1}^{m} a_j \int_{\mathbb{R}^n} \mathbf{1}_{E_j}(x) \; \mathrm{d}x$$

$$= \int_{\mathbb{R}^n} \varphi(x) \; \mathrm{d}x,$$

where we have used Step 1 in the third line. Thus the result holds for simple functions.

Step 3: We prove the result for nonnegative measurable functions.

Let $f : \mathbb{R}^n \to [0, \infty]$ be a nonnegative measurable function. Then there exists a sequence of simple functions $\{\varphi_k\}_{k=1}^{\infty}$ such that $0 \leqslant \varphi_1 \leqslant \varphi_2 \leqslant \cdots \leqslant f$ and $\varphi_k(x) \to f(x)$ for each $x \in \mathbb{R}^n$ as $k \to \infty$ by Proposition **??**. Then $\{\varphi_k(\cdot - y)\}_{k=1}^{\infty}$ is a sequence of simple functions such that $0 \leqslant \varphi_1(\cdot - y) \leqslant \varphi_2(\cdot - y) \leqslant \cdots \leqslant f(\cdot - y)$ and $\varphi_k(x-y) \to f(x-y)$ for each $x \in \mathbb{R}^n$ as $k \to \infty$. Now we can use the Monotone Convergence Theorem to compute

$$\int_{\mathbb{R}^n} f(x-y) \; \mathrm{d}x = \int_{\mathbb{R}^n} \lim_{k\to\infty} \varphi_k(x-y) \; \mathrm{d}x$$

$$= \lim_{k\to\infty} \int_{\mathbb{R}^n} \varphi_k(x-y) \; \mathrm{d}x$$

$$= \lim_{k\to\infty} \int_{\mathbb{R}^n} \varphi_k(x) \; \mathrm{d}x = \int_{\mathbb{R}^n} f(x) \; \mathrm{d}x,$$

where we have used Step 2 in the third line. Thus the result holds for nonnegative measurable functions.

Step 4: We prove the result for general integrable functions.

Let $f \in L^1(\mathbb{R}^n)$ and let $y \in \mathbb{R}^n$. Then we can write $f = f^+ - f^-$ where $f^+, f^- : \mathbb{R}^n \to [0, \infty]$ are nonnegative measurable functions with $\int_{\mathbb{R}^n} f^+ \; \mathrm{d}x < \infty$ and $\int_{\mathbb{R}^n} f^- \; \mathrm{d}x < \infty$. Then we compute

$$\int_{\mathbb{R}^n} f(x-y) \; \mathrm{d}x = \int_{\mathbb{R}^n} f^+(x-y) \; \mathrm{d}x - \int_{\mathbb{R}^n} f^-(x-y) \; \mathrm{d}x$$

$$= \int_{\mathbb{R}^n} f^+(x) \; \mathrm{d}x - \int_{\mathbb{R}^n} f^-(x) \; \mathrm{d}x$$

$$= \int_{\mathbb{R}^n} f(x) \; \mathrm{d}x,$$

where we have used Step 3 in the second line. Thus the result holds for general integrable functions. $\qquad\square$

Proposition 3.2.1 (Lebesgue Measure and Linear Maps). *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear map. Then for each measurable set $E \subseteq \mathbb{R}^n$ we have*

$$\mathcal{L}^n(T(E)) = |\det(T)| \cdot \mathcal{L}^n(E).$$

Proof. Step 1: Suppose we are in the special case of a box and a diagonal linear map.

Let $R = [a_1, b_1] \times \cdots \times [a_n, b_n] \subseteq \mathbb{R}^n$ be a box, and let $T : \mathbb{R}^n \to \mathbb{R}^n$ be the diagonal linear map defined by

$$T(x_1, \ldots, x_n) = (\lambda_1 x_1, \ldots, \lambda_n x_n)$$

for some $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$. If $\lambda_j = 0$ for some $j \in \{1, \ldots, n\}$, then $\det(T) = 0$ and $T(R)$ is contained in the hyperplane $\{x_j = 0\}$, which has Lebesgue measure zero; the result holds in this case.

Thus we may assume that $\lambda_j \neq 0$ for each $j = 1, \ldots, n$. Then $T(R)$ is the box

$$T(R) = [\lambda_1 a_1, \lambda_1 b_1] \times \cdots \times [\lambda_n a_n, \lambda_n b_n].$$

By direct computation we have

$$\operatorname{vol}(T(R)) = \prod_{j=1}^n |\lambda_j b_j - \lambda_j a_j| = \prod_{j=1}^n |\lambda_j| \cdot \prod_{j=1}^n |b_j - a_j|$$
$$= |\det(T)| \cdot \operatorname{vol}(R).$$

This proves the result in this special case.

Step 2: We prove the special case of a measurable set and a diagonal linear map.

Let $E \subseteq \mathbb{R}^n$ be a measurable set, and let $T : \mathbb{R}^n \to \mathbb{R}^n$ be the diagonal linear map defined by

$$T(x_1, \ldots, x_n) = (\lambda_1 x_1, \ldots, \lambda_n x_n)$$

for some $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$. If $\lambda_j = 0$ for some $j \in \{1, \ldots, n\}$, then $\det(T) = 0$ and $T(E)$ is contained in the hyperplane $\{x_j = 0\}$, which has Lebesgue measure zero; the result holds in this case. Thus we may assume that $\lambda_j \neq 0$ for each $j = 1, \ldots, n$. Also if $\mathcal{L}^n(E) = \infty$, then the result holds trivially, so we may assume that $\mathcal{L}^n(E) < \infty$.

Let $\epsilon > 0$. Then there exists a countable collection of boxes $\{R_k\}_{k=1}^\infty$ such that $E \subseteq \bigcup_{k=1}^\infty R_k$ and

$$\sum_{k=1}^\infty \operatorname{vol}(R_k) < \mathcal{L}^n(E) + \epsilon.$$

Then the collection $\{T(R_k)\}_{k=1}^{\infty}$ is a countable cover of $T(E)$ by boxes, so

$$
\begin{aligned}
\mathcal{L}^n(T(E)) \leqslant \mathcal{L}^n \left(\bigcup_{k=1}^{\infty} T(R_k) \right) &\leqslant \sum_{k=1}^{\infty} \mathrm{vol}(T(R_k)) \\
&= \sum_{k=1}^{\infty} |\det(T)| \cdot \mathrm{vol}(R_k) = |\det(T)| \cdot \sum_{k=1}^{\infty} \mathrm{vol}(R_k) \\
&< |\det(T)| \cdot (\mathcal{L}^n(E) + \epsilon)
\end{aligned}
$$

by Step 1. Since $\epsilon > 0$ was arbitrary, we have

$$
\mathcal{L}^n(T(E)) \leqslant |\det(T)| \cdot \mathcal{L}^n(E).
$$

But T is invertible since $\lambda_j \neq 0$ for each $j = 1, \ldots, n$, so because T^{-1} is also a diagonal linear map with determinant $\det(T^{-1}) = 1/\det(T)$, we can apply the same argument to T^{-1} to get

$$
\mathcal{L}^n(T^{-1}(T(E))) \leqslant |\det(T^{-1})| \cdot \mathcal{L}^n(T(E)) = \frac{1}{|\det(T)|} \cdot \mathcal{L}^n(T(E)).
$$

That is,

$$
|\det(T)| \cdot \mathcal{L}^n(E) \leqslant \mathcal{L}^n(T(E)).
$$

Therefore we have equality,

$$
\mathcal{L}^n(T(E)) = |\det(T)| \cdot \mathcal{L}^n(E)
$$

and the result holds in this case.

Step 3: We prove the special case of a measurable set and a special class of invertible linear transformations.

Let $E \subseteq \mathbb{R}^n$ be a measurable set with $\mathcal{L}^n(E) < \infty$, and let $T : \mathbb{R}^n \to \mathbb{R}^n$ be an invertible linear map given by an upper triangular matrix with 1's on the diagonal; that is, there exist numbers $\{a_{j,k}\}_{1 \leqslant j,k \leqslant n}$ such that $a_{j,j} = 1$ for each $j = 1, \ldots, n$, and $a_{j,k} = 0$ for each $j > k$, and

$$
T \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n \\ x_2 + a_{2,3}x_3 + \cdots + a_{2,n}x_n \\ \vdots \\ x_j + \sum_{j<k\leqslant n} a_{j,k}x_k \\ \vdots \\ x_n \end{pmatrix}
$$

for each $(x_1, \ldots, x_n) \in \mathbb{R}^n$.

Note that $\det(T) = 1$. Also note that

$$\mathbf{1}_{T(E)}(y) = \mathbf{1}_E(T^{-1}(y))$$

$$= \mathbf{1}_E \begin{pmatrix} y_1 - a_{1,2}y_2 - \cdots - a_{1,n}y_n \\ y_2 - a_{2,3}y_3 - \cdots - a_{2,n}y_n \\ \vdots \\ y_j - \sum_{j<k\leqslant n} a_{j,k}y_k \\ \vdots \\ y_n \end{pmatrix}$$

since T is a linear bijection. If $E \subseteq \mathbb{R}^n$ is a measurable set, then we compute

$$\mathcal{L}^n(T(E)) = \int_{\mathbb{R}^n} \mathbf{1}_{T(E)}(y) \, dy$$

$$= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \mathbf{1}_{T(E)}(y_1, \ldots, y_n) \, dy_1 \cdots dy_n$$

$$= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \mathbf{1}_E \left(y_1 - \sum_{1<k\leqslant n} a_{1,k}y_k, \ldots, y_j - \sum_{j<k\leqslant n} a_{j,k}y_k, \ldots, y_n \right) dy_1 \cdots dy_n$$

$$= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \mathbf{1}_E(y_1, y_2 - \sum_{2<k\leqslant n} a_{2,k}y_k, \ldots, y_j - \sum_{j<k\leqslant n} a_{j,k}y_k, \ldots, y_n) \, dy_1 \, dy_2 \cdots dy_n$$

$$= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \mathbf{1}_E(y_1, y_2 - \sum_{2<k\leqslant n} a_{2,k}y_k, \ldots, y_{n-1} - a_{n-1,n}y_n, y_n) \, dy_2 \, dy_1 \cdots dy_n$$

where we have used Tonelli's theorem in the second and last line, and translation invariance of the Lebesgue integral on $\mathbb{R}^n$ in the fourth line.

If we repeat these last two steps $n - 2$ more times, we eventually get

$$\mathcal{L}^n(T(E)) = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \mathbf{1}_E(y_1, y_2, \ldots, y_n) \, dy_{n-1} \cdots dy_2 \, dy_1 \, dy_n$$

$$= \int_{\mathbb{R}^n} \mathbf{1}_E(y) \, dy = \mathcal{L}^n(E)$$

where we have again used Tonelli's theorem in the second line.

Since $\det(T) = 1$, we have proven the result in this case.

Step 4: We prove the general case.

Let $E \subseteq \mathbb{R}^n$ be a measurable set, and let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear map. If $\det(T) = 0$, then $T(E)$ is contained in a hyperplane, which has Lebesgue measure zero; the result holds in this case. Thus we may assume that $\det(T) \neq 0$. Also if $\mathcal{L}^n(E) = \infty$, then the result holds trivially, so we may assume that $\mathcal{L}^n(E) < \infty$.

By the LDU decomposition from linear algebra, we can write $T = LDU$ where L is a lower triangular matrix with 1's on the diagonal, D is a diagonal matrix, and U is an upper triangular matrix with 1's on the diagonal. Then $\det(T) = \det(L)\det(D)\det(U) = \det(D)$ since $\det(L) = \det(U) = 1$. We compute that

$$
\begin{aligned}
\mathcal{L}^n(T(E)) &= \mathcal{L}^n(L(D(U(E)))) \\
&= \mathcal{L}^n(D(U(E))) && \text{by Step 3} \\
&= |\det(D)| \cdot \mathcal{L}^n(U(E)) && \text{by Step 2} \\
&= |\det(D)| \cdot \mathcal{L}^n(E) && \text{by Step 3} \\
&= |\det(T)| \cdot \mathcal{L}^n(E)
\end{aligned}
$$

at last. This proves the result in general. $\qquad\square$

Exercise 3.2.2 (Linear Change of Variables). Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear isomorphism, and let $f \in L^1(\mathbb{R}^n)$. Then

$$
\int_{\mathbb{R}^n} f(T(x))|\det(T)|\ \mathrm{d}x = \int_{\mathbb{R}^n} f(y)\ \mathrm{d}y.
$$

Proof. Define

$$
\mathcal{F} := \{f \in L^1(\mathbb{R}^n) : \int_{\mathbb{R}^n} f(T(x))|\det(T)|\ \mathrm{d}x = \int_{\mathbb{R}^n} f(y)\ \mathrm{d}y\}.
$$

We argue that $\mathcal{F} = L^1(\mathbb{R}^n)$ — we break the proof into three steps.

Step 1: We prove that $\mathcal{F}$ is a vector subspace of $L^1(\mathbb{R}^n)$.

Let $f, g \in \mathcal{F}$ and let $\alpha \in \mathbb{R}$. Then

$$
\begin{aligned}
\int_{\mathbb{R}^n} (f + \alpha g)\ \mathrm{d}y &= \int_{\mathbb{R}^n} f(y)\ \mathrm{d}y + \alpha \int_{\mathbb{R}^n} g(y)\ \mathrm{d}y \\
&= \int_{\mathbb{R}^n} f(T(x))|\det(T)|\ \mathrm{d}x + \alpha \int_{\mathbb{R}^n} g(T(x))|\det(T)|\ \mathrm{d}x \\
&= \int_{\mathbb{R}^n} (f(T(x)) + \alpha g(T(x)))|\det(T)|\ \mathrm{d}x \\
&= \int_{\mathbb{R}^n} (f + \alpha g)(T(x))|\det(T)|\ \mathrm{d}x.
\end{aligned}
$$

Thus $f + \alpha g \in \mathcal{F}$, and $\mathcal{F}$ is a vector subspace of $L^1(\mathbb{R}^n)$.

Step 2: We prove that $\mathcal{F}$ contains all simple functions.

Since every simple function is a finite linear combination of characteristic functions of measurable sets of finite measure, it suffices to show that $\mathbf{1}_E \in \mathcal{F}$ for each measurable set $E \subseteq \mathbb{R}^n$ with $\mathcal{L}^n(E) < \infty$ by Step 1.

If $E \subseteq \mathbb{R}^n$ is a measurable set with $\mathcal{L}^n(E) < \infty$, then we compute

$$\mathbf{1}_{T^{-1}(E)}(x) = \begin{cases} 1 & \text{if } x \in T^{-1}(E) \\ 0 & \text{if } x \notin T^{-1}(E) \end{cases} = \begin{cases} 1 & \text{if } T(x) \in E \\ 0 & \text{if } T(x) \notin E \end{cases} = \mathbf{1}_E(T(x))$$

for each $x \in \mathbb{R}^n$. Thus we use the result in Proposition 3.2.1 to compute

$$\int_{\mathbb{R}^n} \mathbf{1}_E(T(x)) \ \mathrm{d}x = \int_{\mathbb{R}^n} \mathbf{1}_{T^{-1}(E)}(x) \ \mathrm{d}x = \mathcal{L}^n(T^{-1}(E))$$

$$= \frac{1}{|\det(T)|} \cdot \mathcal{L}^n(E) = \frac{1}{|\det(T)|} \cdot \int_{\mathbb{R}^n} \mathbf{1}_E(y) \ \mathrm{d}y$$

and multiplying both sides by $|\det(T)|$ shows that $\mathbf{1}_E \in \mathcal{F}$.

Since E was an arbitrary measurable set with $\mathcal{L}^n(E) < \infty$, we have shown that $\mathcal{F}$ contains all characteristic functions of measurable sets of finite measure, and hence $\mathcal{F}$ contains all simple functionsby Step 1.

Step 3: We prove that $\mathcal{F} = L^1(\mathbb{R}^n)$.

Let $f \in L^1(\mathbb{R}^n)$ be a nonnegative function. Then there exists a sequence of simple functions $\{f_k\}_{k=1}^{\infty}$ such that $0 \leqslant f_1 \leqslant f_2 \leqslant \cdots \leqslant f$ and $f_k(x) \to f(x)$ for each $x \in \mathbb{R}^n$ as $k \to \infty$ by Proposition **??**. Then $\{f_k \circ T\}_{k=1}^{\infty}$ is a sequence of simple functions such that $0 \leqslant f_1 \circ T \leqslant f_2 \circ T \leqslant \cdots \leqslant f \circ T$ and $f_k(T(x)) \to f(T(x))$ for each $x \in \mathbb{R}^n$ as $k \to \infty$. Now we can use the Monotone Convergence Theorem to compute

$$\int_{\mathbb{R}^n} f(T(x))|\det(T)| \ \mathrm{d}x = \lim_{k \to \infty} \int_{\mathbb{R}^n} f_k(T(x))|\det(T)| \ \mathrm{d}x = \lim_{k \to \infty} \int_{\mathbb{R}^n} f_k(y) \ \mathrm{d}y = \int_{\mathbb{R}^n} f(y) \ \mathrm{d}y$$

since each $f_k \in \mathcal{F}$ by Step 2. Thus $f \in \mathcal{F}$.

Let $f \in L^1(\mathbb{R}^n)$ be a real-valued integrable function, then we can write $f = f^+ - f^-$ where $f^+, f^- \geqslant 0$ are integrable functions. Since $f^+, f^- \in \mathcal{F}$ and $\mathcal{F}$ is a vector subspace of $L^1(\mathbb{R}^n)$ by Step 1, we have $f \in \mathcal{F}$. $\qquad\square$

Exercise 3.2.3 (Baby Gauss-Green Theorem)**.** If $\varphi \in C_c^1(\mathbb{R}^n)$ then

$$\int_{\mathbb{R}^n} \nabla \varphi(x) \ \mathrm{d}x = 0.$$

Proof. Let $\varphi \in C_c^1(\mathbb{R}^n)$. Then the support of φ is compact, so there exists $R > 0$ such that $\varphi(x) = 0$ for all $x \in \mathbb{R}^n$ with $\|x\|_\infty \geqslant R$. That is, the function φ vanishes outside the cube $[-R, R]^n$.

Thus $\nabla \varphi$ also vanishes outside $[-R, R]^n$, so

$$\int_{\mathbb{R}^n} \nabla \varphi \ \mathrm{d}x = \int_{[-R,R]^n} \nabla \varphi \ \mathrm{d}x.$$

As $\nabla\varphi = (\partial_1\varphi, \ldots, \partial_n\varphi)$ is vector-valued, we have

$$\int_{[-R,R]^n} \partial_j\varphi(x) \ dx = \int_{[-R,R]^n} \frac{\partial\varphi}{\partial x_j}(x) \ dx$$

for each $j = 1, \ldots, n$.

Fix $j \in \{1, \ldots, n\}$. We can compute this integral using iterated integrals — use the Fubini-Tonelli theorem to see that

$$\int_{\mathbb{R}^n} \partial_j \ dx = \int_{[-R,R]^{n-1}} \left(\int_{-R}^{R} \partial_j\varphi(x) \ dx_j \right) dx_1 \cdots \widehat{dx_j} \cdots dx_n.$$

By the fundamental theorem of calculus, for each fixed $(x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n) \in [-R, R]^{n-1}$, we have

$$\int_{-R}^{R} \partial_j\varphi(x) \ dx_j = \varphi(x_1, \ldots, x_{j-1}, R, x_{j+1}, \ldots, x_n) - \varphi(x_1, \ldots, x_{j-1}, -R, x_{j+1}, \ldots, x_n) = 0$$

since φ vanishes on $\mathbb{R}^n \backslash (-R, R)^n$.

Thus

$$\int_{\mathbb{R}^n} \partial_j\varphi(x) \ dx = \int_{[-R,R]^{n-1}} 0 \ dx_1 \cdots \widehat{dx_j} \cdots dx_n = 0.$$

Since this holds for each $j = 1, \ldots, n$, we have

$$\int_{\mathbb{R}^n} \nabla\varphi(x) \ dx = 0.$$

$\square$

3.2.1   Volume of the Unit Ball in $\mathbb{R}^n$

In this section, we compute the volume of the unit ball in $\mathbb{R}^n$ from scratch. In particular, we will not assume the formula for the volume (area) of the unit ball (disk) in $\mathbb{R}^2$ or use polar coordinates.

This is important to be totally rigorous, as I do not want to assume any results from multivariable Riemann integration. All we will need is the Fubini-Tonelli theorem and some single variable Riemann integration results like the Fundamental Theorem of Calculus, the Change of Variables Theorem, and Integration by Parts.

Exercise 3.2.4 (Warm-up)**.** Show that

$$\int_{0}^{\infty} ye^{-ay^2} \ dy = \frac{1}{2a}$$

for each $a > 0$.

Proof. Let $a > 0$. Then we can compute this as a Riemann integral using the change of variable

$$t(y) = -ay^2 \quad \Rightarrow \mathrm{d}t = -2ay \, \mathrm{d}y$$

which gives

$$
\begin{aligned}
\int_0^\infty ye^{-ay^2} \, \mathrm{d}y &= \lim_{R \to \infty} \int_0^R ye^{-ay^2} \, \mathrm{d}y \\
&= \lim_{R \to \infty} \int_{t(0)}^{t(R)} -\frac{1}{2a}e^t \, \mathrm{d}t \\
&= \lim_{R \to \infty} -\frac{1}{2a} \int_0^{-aR^2} e^t \, \mathrm{d}t \\
&= \lim_{R \to \infty} \frac{1}{2a} \int_{-aR^2}^0 e^t \, \mathrm{d}t \\
&= \lim_{R \to \infty} \frac{1}{2a} \left(e^0 - e^{-aR^2} \right) \qquad \text{by the Fundamental Theorem of Calculus} \\
&= \lim_{R \to \infty} \frac{1}{2a}(1 - e^{-aR^2}) = \frac{1}{2a}.
\end{aligned}
$$

We note that because the integrand is nonnegative and the improper Riemann integral converges, the Lebesgue integral also converges and agrees with the improper Riemann integral. This completes the proof. $\qquad \square$

Proposition 3.2.2 (Gaussian Integral Identity).

$$\int_{\mathbb{R}} e^{-x^2} \, \mathrm{d}x = \sqrt{\pi}.$$

Proof. Notice that

$$\int_{\mathbb{R}} e^{-x^2} \, \mathrm{d}x = \int_{-\infty}^\infty e^{-x^2} \, \mathrm{d}x = 2 \int_0^\infty e^{-x^2} \, \mathrm{d}x.$$

Set

$$J := \int_0^\infty e^{-x^2} \, \mathrm{d}x$$

and notice that

$$
\begin{aligned}
J^2 &= \left(\int_0^\infty e^{-x^2} \, \mathrm{d}x \right) \left(\int_0^\infty e^{-y^2} \, \mathrm{d}y \right) \\
&= \int_0^\infty \int_0^\infty e^{-x^2} e^{-y^2} \, \mathrm{d}x \, \mathrm{d}y \qquad \text{by the Fubini-Tonelli theorem} \\
&= \int_0^\infty \int_0^\infty e^{-(x^2+y^2)} \, \mathrm{d}x \, \mathrm{d}y.
\end{aligned}
$$

Define

$$\phi(y) = \int_0^\infty e^{-(x^2+y^2)} \, \mathrm{d}x$$

for each $y \in [0, \infty)$. For each fixed $y \in [0, \infty)$, we compute this integral with the change of variable

$$t(x) = \frac{x}{y} \quad \Rightarrow \quad x = yt, \quad dx = y \, dt$$

to see that

$$\phi(y) = \int_0^\infty e^{-(y^2 t^2 + y^2)} y \, dt = \int_0^\infty e^{-y^2(t^2 + 1)} y \, dt$$

Returning to the original computation for J^2 we have that

$$
\begin{aligned}
J^2 &= \int_0^\infty \phi(y) \, dy \\
&= \int_0^\infty \int_0^\infty e^{-y^2(t^2+1)} y \, dt \, dy \\
&= \int_0^\infty \left(\int_0^\infty e^{-y^2(t^2+1)} y \, dy \right) dt && \text{by the Fubini-Tonelli theorem} \\
&= \int_0^\infty \frac{1}{2(t^2+1)} \, dt && \text{by the Warm-up Exercise} \\
&= \frac{1}{2} \left[\tan^{-1}(t) \right] \Big|_{t=0}^{t=\infty} = \frac{1}{2} \left(\frac{\pi}{2} - 0 \right) = \frac{\pi}{4}.
\end{aligned}
$$

We note that in this computation, we have used the Fubini-Tonelli theorem twice, two single-variable substitutions (in the Warm-up Exercise and in the definition of ϕ), the single variable Fundamental Theorem of Calculus, and the fact that $\lim_{t \to \infty} \tan^{-1}(t) = \frac{\pi}{2}$. We have also used the fact that all the functions we have integrated are nonnegative and have convergent improper Riemann integrals, so that their Lebesgue integrals are also finite and agree with the improper Riemann integrals.

Thus $J^2 = \frac{\pi}{4}$, so $J = \frac{\sqrt{\pi}}{2}$. Therefore

$$\int_{\mathbb{R}} e^{-x^2} \, dx = 2J = \sqrt{\pi}.$$

$\square$

Proposition 3.2.3 (Volume of the n-Dimensional Ball)**.** *For each $n \geqslant 2$, the volume of the n-dimensional ball of unit radius in $\mathbb{R}^n$ satisfies*

$$\mathcal{L}^n(B_n(0,1)) = \frac{2\pi}{n} \mathcal{L}^{n-2}(B_{n-2}(0,1)).$$

Proof. Step 0: We compute the volume of the 0-dimensional, 1-dimensional, and 2-dimensional unit balls to start the induction.

Since $\mathcal{L}^0$ is the counting measure, we have

$$B_0(0,1) = \{0\} \quad \Rightarrow \quad \mathcal{L}^0(B_0(0,1)) = 1.$$

In one dimension we have

$$B_1(0,1) = (-1,1) \quad \Rightarrow \quad \mathcal{L}^1(B_1(0,1)) = 2.$$

In two dimensions, we have the unit disk

$$B_2(0,1) = \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 < 1\}.$$

We compute its volume using Tonelli's theorem.

Note that if $x_2 \in \mathbb{R}$ is such that $|x_2| > 1$, then $\mathbf{1}_{B_2(0,1)}(x_1, x_2) = 0$ for each $x_1 \in \mathbb{R}$. If $x_2 \in \mathbb{R}$ is such that $|x_2| \leqslant 1$, then

$$\mathbf{1}_{B_2(0,1)}(x_1, x_2) = \begin{cases} 1, & \text{if } x_1^2 + x_2^2 < 1 \\ 0, & \text{if } x_1^2 + x_2^2 \geqslant 1 \end{cases}$$

$$= \begin{cases} 1, & \text{if } -\sqrt{1 - x_2^2} < x_1 < \sqrt{1 - x_2^2} \\ 0, & \text{otherwise} \end{cases} = \mathbf{1}_{\left[-\sqrt{1-x_2^2}, \sqrt{1-x_2^2}\right]}(x_1)$$

for each $x_1 \in \mathbb{R}$.

Now can can compute

$$\begin{aligned}
\mathcal{L}^2(B_2(0,1)) &= \int_{\mathbb{R}^2} \mathbf{1}_{B_2(0,1)} \, \mathrm{d}x \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbf{1}_{B_2(0,1)}(x_1, x_2) \, \mathrm{d}x_1 \, \mathrm{d}x_2 && \text{by Fubini's theorem} \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbf{1}_{\left[-\sqrt{1-x_2^2}, \sqrt{1-x_2^2}\right]}(x_1) \, \mathrm{d}x_1 \, \mathrm{d}x_2 && \text{by previous computation} \\
&= \int_{-1}^{1} \left(\int_{-\sqrt{1-x_1^2}}^{\sqrt{1-x_1^2}} 1 \, \mathrm{d}x_2 \right) \mathrm{d}x_1 \\
&= \int_{-1}^{1} 2\sqrt{1 - x_1^2} \, \mathrm{d}x_1 \\
&= 4 \int_{0}^{1} \sqrt{1 - x_1^2} \, \mathrm{d}x_1 && \text{by symmetry} \\
&= 4 \int_{0}^{\pi/2} \cos^2(\theta) \, \mathrm{d}\theta && \text{by the change of variables } x_1 = \sin(\theta) \\
&= 4 \int_{0}^{\pi/2} \frac{1 + \cos(2\theta)}{2} \, \mathrm{d}\theta && \text{by the double-angle formula} \\
&= 2 \left[\theta + \frac{\sin(2\theta)}{2} \right] \Big|_{0}^{\pi/2} = 2 \left(\frac{\pi}{2} - 0 + 0 - 0 \right) = \pi.
\end{aligned}$$

Thus we have

$$\mathcal{L}^0(B_0(0,1)) = 1, \quad \mathcal{L}^1(B_1(0,1)) = 2, \quad \mathcal{L}^2(B_2(0,1)) = \pi$$

as we expect.

Step 1: Now we prove the result for $n > 2$ by induction. Assume that the result holds for $n - 1$ and $n - 2$ where $n > 2$. We compute the volume of the n-dimensional unit ball using Tonelli's theorem.

Note that if $(x_1, x_2) \in \mathbb{R}^2$ is such that $x_1^2 + x_2^2 > 1$, then $\mathbf{1}_{B_n(0,1)}(x_1, x_2, y) = 0$ for each $y \in \mathbb{R}^{n-2}$. If $(x_1, x_2) \in \mathbb{R}^2$ is such that $x_1^2 + x_2^2 \leqslant 1$, then

$$
\mathbf{1}_{B_n(0,1)}(x_1, x_2, y) = \begin{cases} 1, & \text{if } x_1^2 + x_2^2 + \|y\|^2 < 1 \\ 0, & \text{if } x_1^2 + x_2^2 + \|y\|^2 \geqslant 1 \end{cases}
$$

$$
= \begin{cases} 1, & \text{if } \|y\|^2 < 1 - (x_1^2 + x_2^2) \\ 0, & \text{if } \|y\|^2 \geqslant 1 - (x_1^2 + x_2^2) \end{cases}
$$

$$
= \mathbf{1}_{B_{n-2}(0,\sqrt{1-(x_1^2+x_2^2)})}(y)
$$

for each $y \in \mathbb{R}^{n-2}$.

Now we can compute

$$
\mathcal{L}^n(B_n(0, 1)) = \int_{\mathbb{R}^n} \mathbf{1}_{B_n(0,1)} \, \mathrm{d}x
$$

$$
= \int_{\mathbb{R}^2} \int_{\mathbb{R}^{n-2}} \mathbf{1}_{B_n(0,1)}(x_1, x_2, y) \, \mathrm{d}y \, \mathrm{d}x_1 \, \mathrm{d}x_2 \qquad \text{by Tonelli's theorem}
$$

$$
= \int_{\mathbb{R}^2} \int_{\mathbb{R}^{n-2}} \mathbf{1}_{B_{n-2}(0,\sqrt{1-(x_1^2+x_2^2)})}(y) \, \mathrm{d}y \, \mathrm{d}x_1 \, \mathrm{d}x_2 \qquad \text{by previous computation}
$$

$$
= \int_{x_1^2+x_2^2 \leqslant 1} \left(\int_{\mathbb{R}^{n-2}} \mathbf{1}_{B_{n-2}(0,\sqrt{1-(x_1^2+x_2^2)})}(y) \, \mathrm{d}y \right) \mathrm{d}x_1 \, \mathrm{d}x_2
$$

$$
= \int_{x_1^2+x_2^2 \leqslant 1} \mathcal{L}^{n-2}(B_{n-2}(0, \sqrt{1 - (x_1^2 + x_2^2)})) \, \mathrm{d}x_1 \, \mathrm{d}x_2
$$

$$
= \int_{x_1^2+x_2^2 \leqslant 1} (1 - (x_1^2 + x_2^2))^{\frac{n-2}{2}} \mathcal{L}^{n-2}(B_{n-2}(0, 1)) \, \mathrm{d}x_1 \, \mathrm{d}x_2 \qquad \text{by 3.2.1}
$$

$$
= \mathcal{L}^{n-2}(B_{n-2}(0, 1)) \int_{x_1^2+x_2^2 \leqslant 1} (1 - (x_1^2 + x_2^2))^{\frac{n-2}{2}} \, \mathrm{d}x_1 \, \mathrm{d}x_2
$$

$$
= \mathcal{L}^{n-2}(B_{n-2}(0, 1)) \int_{-1}^{1} \int_{-\sqrt{1-x_1^2}}^{\sqrt{1-x_1^2}} (1 - (x_1^2 + x_2^2))^{\frac{n-2}{2}} \, \mathrm{d}x_2 \, \mathrm{d}x_1 \qquad \text{by Tonelli's theorem}
$$

$$
= 4 \, \mathcal{L}^{n-2}(B_{n-2}(0, 1)) \int_{0}^{1} \int_{0}^{\sqrt{1-x_1^2}} (1 - (x_1^2 + x_2^2))^{\frac{n-2}{2}} \, \mathrm{d}x_2 \, \mathrm{d}x_1 \qquad \text{by symmetry and homogeneity}
$$

For each $s \in [-1, 1]$ we define

$$
\psi(s) := \int_{0}^{\sqrt{1-s^2}} (1 - (s^2 + x_2^2))^{\frac{n-2}{2}} \, \mathrm{d}x_2.
$$

Then the big computation above shows that

$$\mathcal{L}^n(B_n(0,1)) = 4\,\mathcal{L}^{n-2}(B_{n-2}(0,1))\int_0^1 \psi(s)\ \mathrm{d}s. \tag{$\star$}$$

We claim That

$$\psi(s) = (1-s^2)^{\frac{n}{2}}\int_0^1 (1-t^2)^{\frac{n-2}{2}}\ \mathrm{d}t$$

for each $s \in (-1,1)$. To see this, we use the change of variables

$$x(t) = t\sqrt{1-s^2} \quad \Rightarrow \quad \mathrm{d}x = \sqrt{1-s^2}\ \mathrm{d}t$$

to see that for $0 \leqslant x \leqslant \sqrt{1-s^2}$, we have

$$(1 - s^2 - x^2) = (1 - s^2 - t^2(1-s^2)) = (1-s^2)(1-t^2)$$

which implies

$$\int_0^{\sqrt{1-s^2}} (1 - s^2 - x^2)^{\frac{n-2}{2}}\ \mathrm{d}x = (1-s^2)^{\frac{n-2}{2}}\int_0^1 (1-t^2)^{\frac{n-2}{2}}\sqrt{1-s^2}\ \mathrm{d}t$$

$$= (1-s^2)^{\frac{n-1}{2}}\int_0^1 (1-t^2)^{\frac{n-2}{2}}\ \mathrm{d}t$$

proving our claim.

Returning to the computation in ($\star$), we have

$$\mathcal{L}^n(B_n(0,1)) = 4\,\mathcal{L}^{n-2}(B_{n-2}(0,1))\int_0^1 (1-s^2)^{\frac{n-1}{2}}\ \mathrm{d}s \int_0^1 (1-t^2)^{\frac{n-2}{2}}\ \mathrm{d}t \tag{$\star\star$}$$

by our claim and another application of Tonelli's theorem.

Step 2: We finish the computation.

Let

$$C_m := \int_0^{\pi/2} \cos^m\theta\ \mathrm{d}\theta$$

for each integer $m \geqslant 0$. We notice that with the change of variables $s = \sin\theta$ and $t = \sin\phi$, we have

$$\int_0^1 (1-s^2)^{\frac{n-1}{2}}\ \mathrm{d}s = \int_0^{\pi/2} \cos^n\theta\ \mathrm{d}\theta = C_n$$

and

$$\int_0^1 (1-t^2)^{\frac{n-2}{2}}\ \mathrm{d}t = \int_0^{\pi/2} \cos^{n-1}\phi\ \mathrm{d}\phi = C_{n-1}.$$

Going back to ($\star\star$), we have

$$\mathcal{L}^n(B_n(0,1)) = 4\,\mathcal{L}^{n-2}(B_{n-2}(0,1))C_n C_{n-1}.$$

By exercise 3.2.5, we see that $2C_n C_{n-1} = \frac{\pi}{n}$ so that

$$\mathcal{L}^n(B_n(0,1)) = \frac{2\pi}{n}\mathcal{L}^{n-2}(B_{n-2}(0,1))$$

as desired.   □

You see, it's a bit of work to avoid computing in polar coordinates.

Exercise 3.2.5 (Recurrance Relation for Cosine Powers)**.** For each integer $m \geqslant 0$ we define

$$C_m := \int_0^{\pi/2} \cos^m \theta \; d\theta.$$

Show that

$$C_0 = \frac{\pi}{2}, \quad C_1 = 1, \quad \text{and} \quad C_m = \frac{m-1}{m} C_{m-2} \text{ for } m \geqslant 2.$$

From this, deduce that

$$2C_m C_{m-1} = \frac{\pi}{m} \text{ for each integer } m \geqslant 1.$$

Proof. The base cases are straightforward:

$$C_0 = \int_0^{\pi/2} 1 \; d\theta = \frac{\pi}{2}, \quad \text{and} \quad C_1 = \int_0^{\pi/2} \cos\theta \; d\theta = \sin\left(\frac{\pi}{2}\right) - \sin(0) = 1.$$

Let's prove the recurrence relation for $m \geqslant 2$ using integration by parts. See that

$$
\begin{aligned}
C_m &= \int_0^{\pi/2} \cos^m \theta \; d\theta \\
&= \int_0^{\pi/2} \cos^{m-1} \theta \cdot \cos\theta \; d\theta \\
&= \int_0^{\pi/2} \cos^{m-1} \theta \; d(\sin\theta) \\
&= \left[\cos^{m-1}\theta \sin\theta\right]\Big|_0^{\pi/2} - \int_0^{\pi/2} \sin\theta \; d(\cos^{m-1}\theta) \\
&= 0 - (m-1)\int_0^{\pi/2} \sin\theta \cdot \cos^{m-2}\theta(-\sin\theta) \; d\theta \\
&= (m-1)\int_0^{\pi/2} \sin^2\theta \cdot \cos^{m-2}\theta \; d\theta \\
&= (m-1)\int_0^{\pi/2} (1 - \cos^2\theta) \cdot \cos^{m-2}\theta \; d\theta \\
&= (m-1)\left(\int_0^{\pi/2} \cos^{m-2}\theta \; d\theta - \int_0^{\pi/2} \cos^m \theta \; d\theta\right) \\
&= (m-1)(C_{m-2} - C_m)
\end{aligned}
$$

which implies
$$C_m = (m-1)C_{m-2} - (m-1)C_m$$

and hence

$$C_m + (m-1)C_m = (m-1)C_{m-2}$$
$$\implies mC_m = (m-1)C_{m-2}$$
$$\implies C_m = \frac{m-1}{m}C_{m-2}.$$

This proves the recurrence relation.

Now we prove that
$$2C_mC_{m-1} = \frac{\pi}{m}$$
for each integer $m \geqslant 1$ by induction on m. The base case $m = 1$ is true since
$$2C_1C_0 = 2 \cdot 1 \cdot \frac{\pi}{2} = \pi = \frac{\pi}{1}.$$

For the induction hypothesis, assume that there is an integer $m - 1 \geqslant 1$ such that for each integer k with $1 \leqslant k \leqslant m - 1$, we have
$$2C_kC_{k-1} = \frac{\pi}{k}.$$

Then

$$\begin{aligned} 2C_mC_{m-1} &= 2 \cdot \frac{m-2}{m}C_{m-2}C_{m-3} \\ &= \frac{m-2}{m} \cdot 2C_{m-2}C_{m-3} \\ &= \frac{m-2}{m} \cdot \frac{\pi}{m-2} \qquad\qquad \text{by the induction hypothesis} \\ &= \frac{\pi}{m} \end{aligned}$$

completing the induction. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Exercise 3.2.6. Define the Gamma function $\Gamma : (0, \infty) \to (0, \infty)$ by

$$\Gamma(s) := \int_0^\infty t^{s-1}e^{-t} \, \mathrm{d}t \qquad \text{for } s > 0.$$

Show that $\Gamma(1) = 1$ and $\Gamma(s+1) = s\Gamma(s)$ for each $s > 0$. Deduce that $\Gamma(n) = (n-1)!$ for each $n \in \mathbb{Z}^+$.

Proof. We compute $\Gamma(1)$ directly:

$$\Gamma(1) = \int_0^\infty t^{1-1}e^{-t} \, \mathrm{d}t = \int_0^\infty e^{-t} \, \mathrm{d}t = \lim_{R\to\infty} \int_0^R e^{-t} \, \mathrm{d}t = \lim_{R\to\infty} \left(-e^{-R} - (-e^0)\right) = 0 + 1 = 1.$$

Now let $s > 0$. We compute $\Gamma(s + 1)$ using integration by parts:

$$\Gamma(s + 1) = \int_0^\infty t^{(s+1)-1} e^{-t} \, dt = \int_0^\infty t^s e^{-t} \, dt$$

$$= \lim_{R \to \infty} \int_0^R t^s e^{-t} \, dt$$

$$= \lim_{R \to \infty} \left(\left[-t^s e^{-t} \right] \Big|_0^R + \int_0^R s t^{s-1} e^{-t} \, dt \right)$$

$$= 0 + \int_0^\infty s t^{s-1} e^{-t} \, dt$$

$$= s \int_0^\infty t^{s-1} e^{-t} \, dt = s\Gamma(s).$$

This proves the desired identity.

Finally, note that $\Gamma(1) = 1 = 0!$, which is the base case of an induction argument. Now let $n \in \mathbb{Z}^+$ and assume that $\Gamma(k) = (k-1)!$ for each $1 \leqslant k < n$. Then

$$\Gamma(n) = (n-1)\Gamma(n-1) = (n-1)(n-2)! = (n-1)!$$

by the induction hypothesis. Thus by induction, we conclude that $\Gamma(n) = (n-1)!$ for each $n \in \mathbb{Z}^+$. $\square$

Exercise 3.2.7. Show that for each integer $n \geqslant 1$, we have

$$\mathcal{L}^n(B_n(0,1)) = \frac{\pi^{n/2}}{\Gamma\left(\frac{n}{2} + 1\right)}.$$

Proof. We prove this by induction on n. The base case $n = 1$ is true since

$$\mathcal{L}^1(B_1(0,1)) = \mathcal{L}^1((-1,1)) = 2$$

and

$$\Gamma\left(\frac{1}{2} + 1\right) = \Gamma\left(\frac{3}{2}\right) = \frac{1}{2}\Gamma\left(\frac{1}{2}\right)$$

$$= \frac{1}{2} \int_0^\infty t^{\frac{1}{2}-1} e^{-t} \, dt = \frac{1}{2} \int_0^\infty t^{-\frac{1}{2}} e^{-t} \, dt$$

$$= \frac{1}{2} \int_0^\infty \frac{e^{-t}}{\sqrt{t}} \, dt$$

$$= \frac{1}{2} \int_0^\infty \frac{e^{-x^2}}{x} 2x \, dx \qquad\qquad \text{by the change of variables } t = x^2$$

$$= \int_0^\infty e^{-x^2} \, dx = \frac{\sqrt{\pi}}{2} \qquad\qquad \text{by the Gaussian Identity 3.2.2.}$$

Thus

$$\frac{\pi^{1/2}}{\Gamma\left(\frac{1}{2}+1\right)} = \frac{\pi^{1/2}}{\frac{\sqrt{\pi}}{2}} = 2 = \mathcal{L}^1(B_1(0,1)).$$

Now let $n \geqslant 2$ and assume that the result holds for each integer $1 \leqslant k < n$. Then by the induction hypothesis, we have

$$\mathcal{L}^{n-2}(B_{n-2}(0,1)) = \frac{\pi^{(n-2)/2}}{\Gamma\left(\frac{n-2}{2}+1\right)}.$$

By the volume recurrence relation 3.2.3, we have

$$
\begin{aligned}
\mathcal{L}^n(B_n(0,1)) &= \frac{2\pi}{n}\mathcal{L}^{n-2}(B_{n-2}(0,1)) \\
&= \frac{2\pi}{n} \cdot \frac{\pi^{(n-2)/2}}{\Gamma\left(\frac{n-2}{2}+1\right)} && \text{by the induction hypothesis} \\
&= \frac{2\pi^{n/2}}{n\Gamma\left(\frac{n-2}{2}+1\right)} \\
&= \frac{\pi^{n/2}}{\left(\frac{n}{2}\right)\Gamma\left(\frac{n-2}{2}+1\right)} \\
&= \frac{\pi^{n/2}}{\Gamma\left(\frac{n}{2}+1\right)} && \text{by exercise 3.2.6.}
\end{aligned}
$$

$\square$

Chapter 4

Differentiation in Euclidean Space

4.0.1 Introduction

hhhh

4.1 Vitali Covering Theorems

In this section we present some important covering theorems which will be useful in later sections.

Notation 4.1.1. Let (X, d) be a metric space. For each ball $B \subset X$ and each $\lambda > 0$, we denote by λB the ball with the same center as B and radius λ times the radius of B.

Lemma 4.1.2 (Vitali's Finite 3-times Covering Lemma)**.** *Let (X, d) be a metric space. Let $\{B_j\}_{j=1}^N$ be a finite collection of closed balls in X. Then there exists a disjoint subcollection $\{B_{j_k}\}_{k=1}^M$ such that*

$$\bigcup_{j=1}^N B_j \subseteq \bigcup_{k=1}^M 3B_{j_k}.$$

Basically if we are given a finite collection of balls, there exists a disjoint subcollection which, when we triple the radii of the balls in the subcollection, covers all the original balls.

Proof. If the number of balls $N = 0$, then the result is trivial. Assume $N \geqslant 1$. Let B_{j_1} be a ball of largest radius in the collection $\{B_j\}_{j=1}^N$. If there are multiple such balls, choose one arbitrarily.

Now let B_{j_2} be a ball of largest radius in the collection $\{B_j\}_{j=1}^N$ which is disjoint from B_{j_1}. If there is no such ball, then we stop without defining B_{j_2}. If there are multiple such balls, choose one arbitrarily.

Continuing in this manner, at step $k \geqslant 2$, assume that we have already defined disjoint balls $B_{j_1}, B_{j_2}, \ldots, B_{j_{k-1}}$. Let B_{j_k} be a ball of largest radius in the collection $\{B_j\}_{j=1}^N$ which is

disjoint from each of the previously defined balls $B_{j_1}, B_{j_2}, \ldots, B_{j_{k-1}}$. If there is no such ball, then we stop without defining B_{j_k}. If there are multiple such balls, choose one arbitrarily. Since the original collection $\{B_j\}_{j=1}^{N}$ is finite, this process must stop after a finite number of steps. Suppose the process stops after M steps, so that we have defined disjoint balls $B_{j_1}, B_{j_2}, \ldots, B_{j_M}$.

It remains to prove the desired inclusion. Let B_j be an arbitrary ball in the original collection $\{B_j\}_{j=1}^{N}$. If B_j is one of the selected balls $B_{j_1}, B_{j_2}, \ldots, B_{j_M}$, then clearly $B_j \subseteq 3B_j$. Otherwise, if B_j is not one of the selected balls, then by construction B_j must intersect at least one of the selected balls — let B_{j_k} be the selected ball which B_j intersects with minimal possible index $k \in \{1, 2, \ldots, M\}$. Note that the radius of B_j is at most the radius of B_{j_k} by construction — otherwise we would have selected B_j instead of B_{j_k} at step k. Now for each point $x \in B_j$, we have

$$d(x, \text{center}(B_{j_k})) \leqslant d(x, \text{center}(B_j)) + d(\text{center}(B_j), \text{center}(B_{j_k})) \leqslant r_j + (r_j + r_{j_k}) \leqslant 3r_{j_k},$$

where r_j and r_{j_k} denote the radii of B_j and B_{j_k} respectively. This shows that $x \in 3B_{j_k}$, and hence $B_j \subseteq 3B_{j_k}$. Since B_j was arbitrary, this completes the proof.

A pedantic point about centers and radii of balls.

Okay extremely pedantic point here: the center and radius of a closed ball is not necessarily unique in an arbitrary metric space. That is, there exists metric spaces (X, d) and points $x, y \in X$ with $x \neq y$ such that

$$\overline{B}(x, r) = \overline{B}(y, r), \qquad \forall r > 0.$$

The p-adic numbers (for any prime p) have this property. Very weird, I know.

Here's why you might think this is a problem. We have not defined the balls by reference to their centers and radii, so by "a ball in X" we mean a set $B \subseteq X$ such that there exists $x \in X$ and $r > 0$ with $B = \overline{B}(x, r)$. These x and r are not necessarily unique, but we have used them in the proof above — in the estimates — and have been speaking of "the center" and "the radius" of a ball as if they were well-defined. If you go through the proof with a particular choice of center and radius for each ball, you will see that the proof works out fine regardless of which choices you make. □

Exercise 4.1.1. Show that in Lemma 4.1.2, the constant 3 is the best possible.

Proof. Let $(X, d) = (\mathbb{R}, |\cdot|)$ be the real line with the usual metric, and let

$$\mathcal{F} := \{[-1, 1], [1, 3]\}.$$

Then any disjoint subcollection $\mathcal{C} \subseteq \mathcal{F}$ can contain at most one interval, since the two intervals in $\mathcal{F}$ are disjoint. If $\mathcal{C} = \{[-1, 1]\}$, then for each $r < 3$, we have

$$[1, 3] \nsubseteq r[-1, 1] = [-r, r]$$

Similarly, if $\mathcal{C} = \{[1,3]\}$, then for each $r < 3$, we have

$$[-1,1] \nsubseteq r[1,3] = [2-r, 2+r].$$

Draw the picture on a number line.

This shows that the conclusion of Lemma 4.1.2 fails if we replace the constant 3 by any smaller constant. $\qquad\square$

This is cool, but what if we have an infinite collection of balls?

Lemma 4.1.3 (Vitali's Infinite 5-times Covering Lemma). *Let (X,d) be a seperable metric space. Let $\mathcal{F}$ be an arbitrary collection of closed balls in X with uniformly bounded radii, i.e.,*

$$R := \sup\{radius(B) : B \in \mathcal{F}\} < \infty.$$

Then there exists a countable disjoint subcollection $\{B_j\}_{j=1}^{\infty} \subseteq \mathcal{F}$ such that

$$\bigcup_{B \in \mathcal{F}} B \subseteq \bigcup_{j=1}^{\infty} 5B_j.$$

Note that the seperability assumption on (X,d) is necessary to ensure that we can extract a countable subcollection of balls (as no uncountable disjoint collection of balls can exist in a seperable metric space).

Proof. For each integer $k \geq 0$, define

$$\mathcal{F}_k := \{B \in \mathcal{F} : 2^{-k-1}R < \text{radius}(B) \leq 2^{-k}R\}$$

so that $\mathcal{F}_k$ is the collection of balls in $\mathcal{F}$ whose radii lie in the interval $(2^{-k-1}R, 2^{-k}R]$.

This defines a partition of $\mathcal{F}$, i.e., we have

$$\mathcal{F} = \bigcup_{k=0}^{\infty} \mathcal{F}_k.$$

We define a sequence $\{\mathcal{C}_k\}_{k=0}^{\infty}$ as follows. First, we set $\mathcal{C}_0$ to be a maximal disjoint subcollection of $\mathcal{F}_0$ (which exists by Zorn's Lemma). Let $\mathcal{C}_1$ be a maximal disjoint subcollection of

$$\{B \in \mathcal{F}_1 : B \cap C = \varnothing \text{ for all } C \in \mathcal{C}_0\}.$$

Continuing in this manner, for each integer $k \geq 1$, we define $\mathcal{C}_k$ to be a maximal disjoint subcollection of

$$\{B \in \mathcal{F}_k : B \cap C = \varnothing \text{ for all } C \in \bigcup_{j=0}^{k-1} \mathcal{C}_j\}.$$

Finally, we define

$$\mathcal{C} := \bigcup_{k=0}^{\infty} \mathcal{C}_k.$$

By construction, $\mathcal{C}$ is a disjoint collection of balls in $\mathcal{F}$. Since (X, d) is seperable, this implies that $\mathcal{C}$ is countable, so we may write $\mathcal{C} = \{B_j\}_{j=1}^{\infty}$.

It remains to prove the desired inclusion. Let $B \in \mathcal{F}$ be arbitrary. Then there exists a unique integer $k \geqslant 0$ such that $B \in \mathcal{F}_k$. There are two possibilities — either B satisfies

$$B \cap \hat{B} = \varnothing \quad \text{for all} \quad \hat{B} \in \bigcup_{j=0}^{k-1} \mathcal{C}_j$$

or not.

If $B \cap \hat{B} = \varnothing$ for all $\hat{B} \in \bigcup_{j=0}^{k-1} \mathcal{C}_j$, then by maximality of $\mathcal{C}_k$, we must have $B \in \mathcal{C}_k$. Otherwise, if there exists some $\hat{B} \in \bigcup_{j=0}^{k-1} \mathcal{C}_j$ such that $B \cap \hat{B} \neq \varnothing$, then B intersects a ball from the union of $\mathcal{C}_0, \mathcal{C}_1, \ldots, \mathcal{C}_{k-1}$. In either case, the ball B intersects some ball $\hat{B}$ that belongs to the union of the collections $\mathcal{C}_0, \mathcal{C}_1, \ldots, \mathcal{C}_k$. Note that such a ball $\hat{B}$ must have radius at least $2^{-k-1}R$ since it belongs to one of the collections $\mathcal{C}_0, \mathcal{C}_1, \ldots, \mathcal{C}_k$. Meanwhile, the radius of B is at most $2^{-k}R$ since $B \in \mathcal{F}_k$. Thus for each point $x \in B$, we have

$$
\begin{aligned}
d(x, \text{center}(\hat{B})) &\leqslant d(x, \text{center}(B)) + d(\text{center}(B), \text{center}(\hat{B})) \\
&\leqslant r_B + (r_B + r_{\hat{B}}) \\
&= 2r_B + r_{\hat{B}} \\
&\leqslant 2 \cdot 2^{-k}R + r_{\hat{B}} \\
&= 4 \cdot 2^{-k-1}R + r_{\hat{B}} \\
&\leqslant 4r_{\hat{B}} + r_{\hat{B}} = 5r_{\hat{B}},
\end{aligned}
$$

where r_B and $r_{\hat{B}}$ denote the radii of B and $\hat{B}$ respectively. This shows that $x \in 5\hat{B}$, and hence $B \subseteq 5\hat{B}$. Since $B \in \mathcal{F}$ was arbitrary, this completes the proof. $\qquad \square$

Exercise 4.1.2 (Best Constant in Infinite Vitali Covering Lemma)**.** Show that in Lemma 4.1.3, the constant 5 can be replaced by any constant larger than 3, but not by 3 itself.

Why 5? We like 5. 5 is cool.

Proof. Fix $b > 1$. For each integer $k \geqslant 0$, define

$$\mathcal{F}_k := \{B \in \mathcal{F} : b^{-k-1}R < \text{radius}(B) \leqslant b^{-k}R\}$$

instead of the dyadic definition in the proof of Lemma 4.1.3.

Following the exact same proof as in Lemma 4.1.3, we can construct a countable disjoint subcollection $\mathcal{C} \subseteq \mathcal{F}$. We claim that

$$\bigcup_{B \in \mathcal{F}} B \subseteq \bigcup_{B \in \mathcal{C}} (1 + 2b)B.$$

Let $B \in \mathcal{F}$ be arbitrary. Then there exists a unique integer $k \geqslant 0$ such that $B \in \mathcal{F}_k$. There are two possibilities — either B satisfies

$$B \cap \hat{B} = \varnothing \quad \text{for all} \quad \hat{B} \in \bigcup_{j=0}^{k-1} \mathcal{C}_j$$

or not.

If $B \cap \hat{B} = \varnothing$ for all $\hat{B} \in \bigcup_{j=0}^{k-1} \mathcal{C}_j$, then by maximality of $\mathcal{C}_k$, we must have $B \in \mathcal{C}_k$. Otherwise, if there exists some $\hat{B} \in \bigcup_{j=0}^{k-1} \mathcal{C}_j$ such that $B \cap \hat{B} \neq \varnothing$, then B intersects a ball from the union of $\mathcal{C}_0, \mathcal{C}_1, \ldots, \mathcal{C}_{k-1}$. In either case, the ball B intersects some ball $\hat{B}$ that belongs to the union of the collections $\mathcal{C}_0, \mathcal{C}_1, \ldots, \mathcal{C}_k$. Note that such a ball $\hat{B}$ must have radius at least $c^{-k-1}R$ since it belongs to one of the collections $\mathcal{C}_0, \mathcal{C}_1, \ldots, \mathcal{C}_k$. Meanwhile, the radius of B is at most $c^{-k}R$ since $B \in \mathcal{F}_k$. Thus for each point $x \in B$, we have

$$\begin{aligned}
d(x, \text{center}(\hat{B})) &\leqslant d(x, \text{center}(B)) + d(\text{center}(B), \text{center}(\hat{B})) \\
&\leqslant r_B + (r_B + r_{\hat{B}}) \\
&= 2r_B + r_{\hat{B}} \\
&\leqslant 2 \cdot c^{-k}R + r_{\hat{B}} \\
&= 2c \cdot c^{-k-1}R + r_{\hat{B}} \\
&\leqslant 2cr_{\hat{B}} + r_{\hat{B}} = (1 + 2c)r_{\hat{B}},
\end{aligned}$$

where r_B and $r_{\hat{B}}$ denote the radii of B and $\hat{B}$ respectively. This shows that $x \in (1 + 2c)\hat{B}$, and hence $B \subseteq (1 + 2c)\hat{B}$. Since $B \in \mathcal{F}$ was arbitrary, this proves the claim.

This also proves that the constant 5 in Lemma 4.1.3 can be replaced by any constant larger than 3 by choosing $b > 1$ such that $1 + 2b$ equals the desired constant.

Now we show that the constant 3 cannot be achieved. Let $(X, d) = (\mathbb{R}, |\cdot|)$ be the real line with the usual metric, and let

$$\mathcal{F} := \{(x - r, x + r) : |x| < 1/2, 0 < r < (|x| + 1)/3\}.$$

Then each interval in $\mathcal{F}$ contains 0 so any disjoint subcollection $\mathcal{C} \subseteq \mathcal{F}$ can contain at most one interval. However see that

$$(-1, 1) = \bigcup_{|x| < 1/2} (x - (|x| + 1)/3, x + (|x| + 1)/3)$$

but for each interval $(x - r, x + r) \in \mathcal{F}$, we have

$$(-1, 1) \nsubseteq 3(x - r, x + r) = (x - 3r, x + 3r).$$

Let's check this — if $0 \leqslant x < 1/2$, then $x - 3r > x - (|x| + 1) = -1$, so $(x - 3r, x + 3r)$ does not contain $(-1, 1)$, and if $-1/2 < x < 0$, then $x + 3r < x + (|x| + 1) = 1$, so again $(x - 3r, x + 3r)$ does not contain $(-1, 1)$. This shows that the conclusion of Lemma 4.1.3 fails if we replace the constant 5 by 3. $\qquad\square$

Exercise 4.1.3 (Unbounded Radii in Infinite Vitali Covering Lemma)**.** Show that if we remove the assumption that the radii of the balls in $\mathcal{F}$ are uniformly bounded in Lemma 4.1.3, then the conclusion of the lemma may fail.

Proof. Let $(X, d) = (\mathbb{R}, |\cdot|)$ be the real line with the usual metric, and let

$$\mathcal{F} := \{[-r, r] : r > 0\}.$$

Then $\mathcal{F}$ is the collection of all closed intervals centered at the origin. Then any disjoint subcollection $\mathcal{C} \subseteq \mathcal{F}$ can contain at most one interval, since any two intervals in $\mathcal{F}$ intersect. But for each interval $[-R, R] \in \mathcal{F}$ with $R > 0$, we have

$$\mathbb{R} = \bigcup_{r > 0} [-r, r] \nsubseteq 5[-R, R] = [-5R, 5R].$$

This shows that the conclusion of Lemma 4.1.3 fails in this case. $\qquad\square$

Exercise 4.1.4 (Non-Separability in Infinite Vitali Covering Lemma)**.** Show that if we drop the assumption that (X, d) is separable in Lemma 4.1.3, then the set $\mathcal{C}$ may not be countable, but retains the other properties stated in the lemma.

Proof. The only place in the proof of Lemma 4.1.3 where we used the separability of (X, d) was to conclude that the disjoint collection $\mathcal{C}$ is countable. If we drop the separability assumption, then $\mathcal{C}$ may be uncountable, but the rest of the proof remains unchanged. We still have that $\mathcal{C}$ is a disjoint collection of balls in $\mathcal{F}$ such that

$$\bigcup_{B \in \mathcal{F}} B \subseteq \bigcup_{B \in \mathcal{C}} 5B.$$

$\qquad\square$

Lemma 4.1.4 (Vitali Covering Lemma Technical Variant)**.** *Let (X, d) be a separable metric space and let $A \subset X$. Let $\mathcal{F}$ be a collection of closed balls in X which covers A and has uniformly bounded radii, i.e.,*

$$R := \sup\{radius(B) : B \in \mathcal{F}\} < \infty.$$

Also assume that for each point $x \in A$ and each $\delta > 0$, there exists a ball $B \in \mathcal{F}$ such that $x \in B$ and radius$(B) < \delta$. Then there exists a countable disjoint subcollection $\mathcal{C} \subseteq \mathcal{F}$ such that for each finite collection of balls $\{B_j\}_{j=1}^N \subseteq \mathcal{F}$, we have

$$A \backslash \bigcup_{j=1}^N B_j \subseteq \bigcup_{B \in \mathcal{C} \backslash \{B_1,\dots,B_N\}} 5B.$$

Proof. Let $\mathcal{C}$ be the countable disjoint subcollection of balls constructed in the proof of Lemma 4.1.3. Let $\{B_j\}_{j=1}^N \subseteq \mathcal{F}$ be an arbitrary finite collection of balls.

If $A \subset \bigcup_{j=1}^N B_j$, then the desired inclusion is trivial as the set on the left side is empty. Thus we suppose A is not contained in $\bigcup_{j=1}^N B_j$, and let $x \in A \backslash \bigcup_{j=1}^N B_j$ be arbitrary. By assumption that there exist balls in $\mathcal{F}$ of arbitrarily small radius covering each point in A, there exists a ball $B \in \mathcal{F}$ such that $x \in B$ and the radius of B is less than the distance from x to the set $\bigcup_{j=1}^N B_j$. In particular, this implies that B does not intersect any of the balls $B_1, B_2, \dots, B_N$, as all these balls are closed. But from the proof of Lemma 4.1.3, we know that there exists some ball $\hat{B} \in \mathcal{C}$ which intersects B and hence $B \subseteq 5\hat{B}$. That is, we have found some ball $\hat{B} \in \mathcal{C}$ such that $x \in B \subseteq 5\hat{B}$ and $\hat{B} \notin \{B_1, B_2, \dots, B_N\}$. Since $x \in A \backslash \bigcup_{j=1}^N B_j$ was arbitrary, this completes the proof. $\square$

Lemma 4.1.5 (Filling Open Sets with Balls). *Let $\delta > 0$. For each open set $U \subset \mathbb{R}^n$, there exists a disjoint collection of closed balls $\{B_j\}_{j=1}^\infty$ such that $\bigcup_{j=1}^\infty B_j \subseteq U$, diam$(B_j) \leqslant \delta$ for each $j \geqslant 1$, and*

$$\mathcal{L}^n \left(U \backslash \bigcup_{j=1}^\infty B_j \right) = 0.$$

In other words, for each $\delta > 0$ we can fill up an open set U with disjoint balls of diameter at most δ so that the leftover set has Lebesgue measure zero.

Proof. Fix $\theta \in (1 - \frac{1}{5^n}, 1)$. Assume first that $\mathcal{L}^n(U) < \infty$.

Step 1: We claim there exists a finite collection of disjoint balls $\{B_j\}_{j=1}^{N_1}$ such that diam$(B_j) \leqslant \delta$ for each $j = 1, 2, \dots, N_1$ and

$$\mathcal{L}^n \left(U \backslash \bigcup_{j=1}^{N_1} B_j \right) \leqslant \theta \mathcal{L}^n(U). \tag{$\Diamond$}$$

To see this, let $\mathcal{F} := \{B \subseteq U : B \text{ is a closed ball with diam } B < \delta\}$. Then $\mathcal{F}$ is a collection of closed balls which covers U and has uniformly bounded radii, so by Lemma 4.1.3, there exists a countable disjoint subcollection $\{B_j\}_{j=1}^\infty \subseteq \mathcal{F}$ such that

$$U \subseteq \bigcup_{j=1}^\infty 5B_j.$$

Thus we have

$$\mathcal{L}^n(U) \leqslant \mathcal{L}^n\left(\bigcup_{j=1}^{\infty} 5B_j\right) \leqslant \sum_{j=1}^{\infty} \mathcal{L}^n(5B_j)$$

$$= \sum_{j=1}^{\infty} 5^n \mathcal{L}^n(B_j)$$

$$= 5^n \sum_{j=1}^{\infty} \mathcal{L}^n(B_j)$$

$$= 5^n \mathcal{L}^n\left(\bigcup_{j=1}^{\infty} B_j\right)$$

since the balls $\{B_j\}_{j=1}^{\infty}$ are disjoint. Rearranging gives

$$\frac{1}{5^n}\mathcal{L}^n(U) \leqslant \mathcal{L}^n\left(\bigcup_{j=1}^{\infty} B_j\right)$$

which implies that

$$\mathcal{L}^n\left(U\backslash \bigcup_{j=1}^{\infty} B_j\right) \leqslant \left(1 - \frac{1}{5^n}\right)\mathcal{L}^n(U)$$

Since $\theta > 1 - \frac{1}{5^n}$, we can choose N_1 sufficiently large so that the finite collection of disjoint balls $\{B_j\}_{j=1}^{N_1}$ satisfies ($\Diamond$). This proves the claim in Step 1.

   Step 2: Now let

$$U_1 := U\backslash \bigcup_{j=1}^{N_1} B_j$$

and

$$\mathcal{F}_1 := \{B \subseteq U_1 : B \text{ is a closed ball with } \operatorname{diam} B < \delta\}.$$

As in Step 1, there exists finitely many disjoint balls $\{B_j\}_{j=N_1+1}^{N_2} \subseteq \mathcal{F}_1$ such that $\operatorname{diam}(B_j) \leqslant \delta$ for each $j = N_1 + 1, N_1 + 2, \ldots, N_2$ and

$$\mathcal{L}^n\left(U_1\backslash \bigcup_{j=N_1+1}^{N_2} B_j\right) \leqslant \theta\mathcal{L}^n(U_1).$$

Thus we obtain

$$\mathcal{L}^n\left(U\backslash\bigcup_{j=1}^{N_2}B_j\right) = \mathcal{L}^n\left(U_1\backslash\bigcup_{j=N_1+1}^{N_2}B_j\right)$$
$$\leqslant \theta\mathcal{L}^n(U_1)$$
$$= \theta\mathcal{L}^n\left(U\backslash\bigcup_{j=1}^{N_1}B_j\right)$$
$$\leqslant \theta^2\mathcal{L}^n(U).$$

By iterating this process, we obtain for each integer $m \geqslant 1$ a finite collection of disjoint balls $\{B_j\}_{j=N_{m-1}+1}^{N_m}$ such that $\mathrm{diam}(B_j) \leqslant \delta$ for each $j = N_{m-1}+1, N_{m-1}+2, \ldots, N_m$ and

$$\mathcal{L}^n\left(U\backslash\bigcup_{j=1}^{N_m}B_j\right) \leqslant \theta^m\mathcal{L}^n(U).$$

Since $\theta \in (1-\frac{1}{5^n}, 1)$, we have $\theta^m \to 0$ as $m \to \infty$, which implies that

$$\mathcal{L}^n\left(U\backslash\bigcup_{j=1}^{\infty}B_j\right) = 0.$$

This completes the proof when $\mathcal{L}^n(U) < \infty$.

Step 3: Finally, if $\mathcal{L}^n(U) = \infty$, we can write U as a countable union of bounded open sets

$$U = \bigcup_{j=1}^{\infty}(U \cap \{x \in \mathbb{R}^n : (j-1) \leqslant |x| < j\}).$$

Applying the previous case to each bounded open set $U \cap \{x \in \mathbb{R}^n : (j-1) \leqslant |x| < j\}$ for $j = 1, 2, \ldots$ and taking the union of all the resulting collections of balls gives the desired result. $\square$

This ends the section on Vitali Covering Lemmas. We will use these in the next section to investigate the Hardy-Littlewood Maximal Function, and later on to study Hausdorff measure.

4.2 Hardy-Littlewood Maximal Function and Lebesgue Differentiation Theorem

4.2.1 Differentiation of Integrals

Suppose $f : [a, b] \to \mathbb{R}$ is an integrable function, and we let

$$F(x) := \int_a^x f(t) \, \mathrm{d}t \qquad \forall x \in [a, b].$$

If we want to differentiate the function F at a point $x \in (a, b)$, we must use the difference quotient

$$\frac{F(x + h) - F(x)}{h} = \frac{1}{h} \int_x^{x+h} f(t) \, \mathrm{d}t.$$

We pause for a moment to note that this is precisely the average value of f on the interval $(x, x + h)$ when $h > 0$, and on the interval $(x + h, x)$ when $h < 0$. If we expect anything like the Fundamental Theorem of Calculus to hold, we would want this average value to converge to $f(x)$ as $h \to 0$. That is, we would want

$$\frac{1}{|I|} \int_I f(t) \, \mathrm{d}t \to f(x) \quad \text{as } |I| \to 0,$$

where I is any interval containing x, and $|I|$ denotes its length. We can ask that x be the center of I, as I shrinks to x, or that only $x \in I$ as $|I| \to 0$.

Moving into higher dimensions, we can ask a similar question. If f is an integrable function on an open set $U \subseteq \mathbb{R}^n$, can we hope to have

$$\frac{1}{\operatorname{vol} B} \int_B f(y) \, \mathrm{d}y \to f(x) \quad \text{as } \operatorname{vol} B \to 0,$$

where B is any ball containing x, and $\operatorname{vol} B$ denotes its volume? This is known as the **averaging problem**.

Exercise 4.2.1 (Continuity and the Averaging Problem). Let $f : U \to \mathbb{R}$ be an integrable function on an open set $U \subseteq \mathbb{R}^n$, and let $x \in U$ be such that f is continuous at x. Show that

$$\frac{1}{\operatorname{vol} B} \int_B f(y) \, \mathrm{d}y \to f(x) \quad \text{as } \operatorname{vol} B \to 0,$$

where B is any ball containing x.

Proof. Let $\varepsilon > 0$ be arbitrary. Since f is continuous at x, there exists $\delta > 0$ such that

$$|f(y) - f(x)| < \varepsilon \quad \forall y \in B_n(x, \delta).$$

Now, let B be any ball containing x with volume $\operatorname{vol} B < \operatorname{vol} B_n(0, \delta)$. Then $B \subseteq B_n(x, \delta)$, and we have

$$
\begin{aligned}
\left| \frac{1}{\operatorname{vol} B} \int_B f(y) \, \mathrm{d}y - f(x) \right| &= \left| \frac{1}{\operatorname{vol} B} \int_B (f(y) - f(x)) \, \mathrm{d}y \right| \\
&\leqslant \frac{1}{\operatorname{vol} B} \int_B |f(y) - f(x)| \, \mathrm{d}y \\
&< \varepsilon.
\end{aligned}
$$

Since $\varepsilon > 0$ was arbitrary, the result follows. $\qquad\square$

4.2.2 The Hardy-Littlewood Maximal Function

Definition 4.2.1. Let $U \subseteq \mathbb{R}^n$ be open. A function $f : U \to \mathbb{R}$ is said to be *locally integrable* on U if for every compact set $K \subseteq U$, we have $f \in L^1(K)$. The collection of all locally integrable functions on U is denoted by $L^1_{\mathrm{loc}}(U)$, where we identify functions that are equal almost everywhere.

That is, an element of $L^1_{\mathrm{loc}}(U)$ is formally an equivalence class of functions that are equal almost everywhere, but we will often abuse notation and refer to a particular representative of this equivalence class as the function itself.

Later we will generalize this definition to arbitrary Borel outer measures, but for now this will suffice.

Definition 4.2.2. Let $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$. The *(uncentered) Hardy-Littlewood maximal function* of f is the function $Mf : \mathbb{R}^n \to [0, \infty]$ defined by

$$
Mf(x) := \sup_{B \ni x} \frac{1}{\operatorname{vol} B} \int_B |f(y)| \, \mathrm{d}y,
$$

where the supremum is taken over all balls B containing x.

Similarly the *(centered) Hardy-Littlewood maximal function* of f is the function $M_c f : \mathbb{R}^n \to [0, \infty]$ defined by

$$
M_c f(x) := \sup_{B_n(x,r)} \frac{1}{\operatorname{vol} B_n(x,r)} \int_{B_n(x,r)} |f(y)| \, \mathrm{d}y,
$$

where the supremum is taken over all balls $B_n(x, r)$ centered at x.

For many pruposes, we can work interchangeably with either the centered or uncentered maximal function, as they are pointwise comparable.

Exercise 4.2.2 (Pointwise Comparison of Centered and Uncentered Maximal Functions)**.** Let $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$. Show that for each $x \in \mathbb{R}^n$, we have

$$
M_c f(x) \leqslant Mf(x) \leqslant 2^n M_c f(x).
$$

Proof. Let $x \in \mathbb{R}^n$ be arbitrary. For any $r > 0$, we have

$$\frac{1}{\text{vol } B_n(x, r)} \int_{B_n(x,r)} |f(y)| \, \mathrm{d}y \leqslant Mf(x),$$

since $B_n(x, r)$ is a ball containing x. Taking the supremum over all $r > 0$, we obtain $M_c f(x) \leqslant Mf(x)$.

Conversely, let B be any ball containing x. Then let $r := \text{radius}(B)$, and see that $B \subseteq B_n(x, 2r)$. Thus,

$$\begin{aligned}
\frac{1}{\text{vol } B} \int_B |f(y)| \, \mathrm{d}y &\leqslant \frac{1}{\text{vol } B_n(x, 2r)} \int_{B_n(x,2r)} |f(y)| \, \mathrm{d}y \\
&= \frac{1}{2^n \, \text{vol } B_n(x, r)} \int_{B_n(x,2r)} |f(y)| \, \mathrm{d}y \\
&\leqslant 2^n M_c f(x).
\end{aligned}$$

Taking the supremum over all balls B containing x, we obtain $Mf(x) \leqslant 2^n M_c f(x)$. $\square$

Exercise 4.2.3 (Sublinearity of the Hardy-Littlewood Maximal Function). Let $f, g \in L^1_{\text{loc}}(\mathbb{R}^n)$ and $\alpha \in \mathbb{R}$. Show that

$$M(f + g)(x) \leqslant Mf(x) + Mg(x) \quad \text{and} \quad M(\alpha f)(x) = |\alpha| Mf(x) \quad \forall x \in \mathbb{R}^n.$$

Proof. It's easy. Let $x \in \mathbb{R}^n$ be arbitrary. For each ball B containing x, we have

$$\begin{aligned}
\frac{1}{\text{vol } B} \int_B |f(y) + g(y)| \, \mathrm{d}y &\leqslant \frac{1}{\text{vol } B} \int_B |f(y)| \, \mathrm{d}y + \frac{1}{\text{vol } B} \int_B |g(y)| \, \mathrm{d}y \\
&\leqslant Mf(x) + Mg(x).
\end{aligned}$$

Taking the supremum over all balls B containing x, we obtain $M(f+g)(x) \leqslant Mf(x) + Mg(x)$.

Similarly, for each ball B containing x, we have

$$\frac{1}{\text{vol } B} \int_B |\alpha f(y)| \, \mathrm{d}y = |\alpha| \frac{1}{\text{vol } B} \int_B |f(y)| \, \mathrm{d}y \leqslant |\alpha| Mf(x).$$

Taking the supremum over all balls B containing x, we obtain $M(\alpha f)(x) = |\alpha| Mf(x)$. $\square$

Proposition 4.2.3 (Properties of the Hardy-Littlewood Maximal Function). *Let $f \in L^1_{loc}(\mathbb{R}^n)$. Then the following properties hold:*

(i) Mf *is measurable,*

(ii) if $f \in L^1(\mathbb{R}^n)$, *then* $Mf(x) < \infty$ *for almost every* $x \in \mathbb{R}^n$, *and*

(iii) if $f \in L^1(\mathbb{R}^n)$, *then for each* $\alpha > 0$ *the function* Mf *satisfies*

$$\mathcal{L}^n(\{x \in \mathbb{R}^n : Mf(x) > \alpha\}) \leqslant \frac{3^n}{\alpha} \|f\|_{L^1(\mathbb{R}^n)}$$

The inequality in part (iii) is known as the *Hardy-Littlewood Maximal Inequality*, or the *weak* $(1,1)$ *estimate* for the maximal function. The notion of weak estimates will be made precise in later chapters.

Proof. (i) Let $\alpha > 0$ be arbitrary. Then we see that the set $\{x \in \mathbb{R}^n : Mf(x) > \alpha\}$ is open because if x is in this set, then

$$\sup_{r>0} \frac{1}{\operatorname{vol} B_n(x,r)} \int_{B_n(x,r)} |f(y)| \, dy > \alpha.$$

Thus there exists a ball B containing x such that

$$\frac{1}{\operatorname{vol} B} \int_B |f(y)| \, dy > \alpha.$$

and hence for each $\bar{x} \in B$ we also have $Mf(\bar{x}) > \alpha$. Since $\alpha > 0$ was arbitrary, we conclude that Mf is measurable.

Now we assume that $f \in L^1(\mathbb{R}^n)$, i.e. $\|f\|_{L^1(\mathbb{R}^n)} < \infty$. We will prove parts (iii) and then show that part (ii) follows from part (iii).

(iii) Let $\alpha > 0$ be arbitrary, and let

$$E_\alpha := \{x \in \mathbb{R}^n : Mf(x) > \alpha\}.$$

For each $x \in E_\alpha$, there exists a ball B_x containing x such that

$$\frac{1}{\operatorname{vol} B_x} \int_{B_x} |f(y)| \, dy > \alpha.$$

Thus, we have

$$E_\alpha \subseteq \bigcup_{x \in E_\alpha} B_x.$$

Fix a compact subset $K \subseteq E_\alpha$. Then the collection $\{B_x : x \in K\}$ is an open cover of K, so there exists a finite subcover $\{B_j\}_{j=1}^N$ such that

$$K \subseteq \bigcup_{j=1}^N B_j.$$

By the Vitali Covering Lemma (4.1.2), there exists a disjoint subcollection $\{B_{j_k}\}_{k=1}^M$ such that

$$K \subseteq \bigcup_{k=1}^M 3B_{j_k}.$$

Therefore, we have

$$\mathcal{L}^n(K) \leqslant \sum_{k=1}^{M} \mathcal{L}^n(3B_{j_k})$$

$$= \sum_{k=1}^{M} 3^n \mathcal{L}^n(B_{j_k})$$

$$< \sum_{k=1}^{M} \frac{3^n}{\alpha} \int_{B_{j_k}} |f(y)| \, \mathrm{d}y$$

$$= \frac{3^n}{\alpha} \int_{\bigcup_{k=1}^{M} B_{j_k}} |f(y)| \, \mathrm{d}y$$

$$\leqslant \frac{3^n}{\alpha} \|f\|_{L^1(\mathbb{R}^n)}$$

since the balls B_{j_k} are disjoint. Since $K \subseteq E_\alpha$ was an arbitrary compact subset, we conclude that

$$\mathcal{L}^n(E_\alpha) \leqslant \frac{3^n}{\alpha} \|f\|_{L^1(\mathbb{R}^n)}.$$

(ii) See that for each $\alpha > 0$, we have

$$\{x \in \mathbb{R}^n : Mf(x) = \infty\} \subseteq \{x \in \mathbb{R}^n : Mf(x) > \alpha\}$$

so we must have that

$$\mathcal{L}^n(\{x \in \mathbb{R}^n : Mf(x) = \infty\}) \leqslant \mathcal{L}^n(\{x \in \mathbb{R}^n : Mf(x) > \alpha\}) \leqslant \frac{3^n}{\alpha} \|f\|_{L^1(\mathbb{R}^n)}$$

for all $\alpha > 0$ by part (iii). Letting $\alpha \to \infty$, we conclude that $\mathcal{L}^n(\{x \in \mathbb{R}^n : Mf(x) = \infty\}) = 0$. $\qquad\square$

4.2.3   The Lebesgue Differentiation Theorem

We will now show that the estimate in Proposition 4.2.3(iii) allows us to solve the averaging problem for integrable functions.

Theorem 4.2.4 (Lebesgue Differentiation Theorem). *Let $f \in L^1_{loc}(\mathbb{R}^n)$.   Then for almost every $x \in \mathbb{R}^n$, we have*

$$\lim_{\substack{\mathrm{vol}\, B \to 0 \\ x \in B}} \frac{1}{\mathrm{vol}\, B} \int_B f(y) \, \mathrm{d}y = f(x).$$

Proof. Step 1: First we assume that $f \in L^1(\mathbb{R}^n)$.

It suffices to show that for each $\alpha > 0$ the set

$$E_\alpha := \left\{ x \in \mathbb{R}^n : \limsup_{\substack{\mathrm{vol}\, B \to 0 \\ x \in B}} \left| \frac{1}{\mathrm{vol}\, B} \int_B f(y) \, \mathrm{d}y - f(x) \right| > 2\alpha \right\}$$

has measure zero. Indeed, if this is the case, then we see that $E := \bigcup_{m=1}^{\infty} E_{1/m}$ has measure zero, and for each $x \notin E$ we have

$$\limsup_{\substack{\operatorname{vol} B \to 0 \\ x \in B}} \left| \frac{1}{\operatorname{vol} B} \int_B f(y) \, dy - f(x) \right| < 2 \cdot \frac{1}{m} \quad \forall m \in \mathbb{N},$$

which implies that

$$\lim_{\substack{\operatorname{vol} B \to 0 \\ x \in B}} \frac{1}{\operatorname{vol} B} \int_B f(y) \, dy = f(x).$$

So, let $\alpha > 0$ be arbitrary, and let $\epsilon > 0$. Since $f \in L^1(\mathbb{R}^n)$, there exists $g \in C_c(\mathbb{R}^n)$ such that

$$\|f - g\|_{L^1(\mathbb{R}^n)} < \epsilon$$

since $C_c(\mathbb{R}^n)$ is dense in $L^1(\mathbb{R}^n)$ by Lemma 6.2.5. Since g is continuous almost everywhere, by Exercise 4.2.1 we have

$$\lim_{\substack{\operatorname{vol} B \to 0 \\ x \in B}} \frac{1}{\operatorname{vol} B} \int_B g(y) \, dy = g(x) \quad \text{for each } x \in \mathbb{R}^n.$$

Now

$$\frac{1}{\operatorname{vol} B} \int_B f(y) \, dy - f(x) = \frac{1}{\operatorname{vol} B} \int_B (f(y) - g(y)) \, dy + \frac{1}{\operatorname{vol} B} \int_B g(y) \, dy - g(x) + g(x) - f(x)$$

implies that

$$\limsup_{\substack{\operatorname{vol} B \to 0 \\ x \in B}} \left| \frac{1}{\operatorname{vol} B} \int_B f(y) \, dy - f(x) \right| \leq M(f - g) + |g(x) - f(x)|$$

by taking the limit superior (the middle term goes to zero by continuity of g). See that

$$E_\alpha \subseteq \{x : M(f - g)(x) > \alpha\} \cup \{x : |g(x) - f(x)| > \alpha\}$$

and Chebyshev's inequality 2.3.1 implies that

$$\mathcal{L}^n(\{x : |f(x) - g(x)| > \alpha\}) \leq \frac{1}{\alpha} \|f - g\|_{L^1(\mathbb{R}^n)}$$

and the weak $(1,1)$ estimate from Proposition 4.2.3(iii) implies that

$$\mathcal{L}^n(\{x : M(f - g)(x) > \alpha\}) \leq \frac{3^n}{\alpha} \|f - g\|_{L^1(\mathbb{R}^n)}.$$

Therefore, we have

$$\mathcal{L}^n(E_\alpha) \leq \frac{(3^n + 1)}{\alpha} \|f - g\|_{L^1(\mathbb{R}^n)} < \frac{(3^n + 1)}{\alpha} \epsilon.$$

Since $\epsilon > 0$ was arbitrary, we conclude that $\mathcal{L}^n(E_\alpha) = 0$ and thus the result holds when $f \in L^1(\mathbb{R}^n)$.

Step 2: Now we consider the general case when $f \in L^1_{\text{loc}}(\mathbb{R}^n)$. For each $m \in \mathbb{Z}^+$, let $f_m := f \cdot \chi_{B_n(0,m)}$. Then $f_m \in L^1(\mathbb{R}^n)$ for each $m \in \mathbb{N}$, so by Step 1 we have

$$\lim_{\substack{\text{vol }B \to 0 \\ x \in B}} \frac{1}{\text{vol }B} \int_B f_m(y)\, dy = f_m(x) \quad \text{for almost every } x \in \mathbb{R}^n.$$

Let A_m denote the set of measure zero where this limit fails for f_m. Then for each $x \notin \bigcup_{m=1}^\infty A_m$ and all $m \in \mathbb{Z}^+$ such that $\|x\| < m$, we have $f_m(x) = f(x)$ and

$$\lim_{\substack{\text{vol }B \to 0 \\ x \in B}} \frac{1}{\text{vol }B} \int_B f(y)\, dy = \lim_{\substack{\text{vol }B \to 0 \\ x \in B}} \frac{1}{\text{vol }B} \int_B f_m(y)\, dy = f_m(x) = f(x)$$

That is, the set $\bigcup_{m=1}^\infty A_m$ has measure zero, and for each x not in this set we have

$$\lim_{\substack{\text{vol }B \to 0 \\ x \in B}} \frac{1}{\text{vol }B} \int_B f(y)\, dy = f(x)$$

as desired. $\square$

Corollary 4.2.5 (Maximal Function Bounds Function Almost Everywhere). *Let $f \in L^1_{loc}(\mathbb{R}^n)$. Then for almost every $x \in \mathbb{R}^n$, we have $|f(x)| \leqslant Mf(x)$ for almost every $x \in \mathbb{R}^n$.*

Proof. We have $|f| \in L^1_{\text{loc}}(\mathbb{R}^n)$ since $f \in L^1_{\text{loc}}(\mathbb{R}^n)$. By Theorem 4.2.4, for almost every $x \in \mathbb{R}^n$ we have

$$\lim_{\substack{\text{vol }B \to 0 \\ x \in B}} \frac{1}{\text{vol }B} \int_B |f(y)|\, dy = |f(x)|.$$

Since

$$Mf(x) = \sup_{B \ni x} \frac{1}{\text{vol }B} \int_B |f(y)|\, dy \geqslant \lim_{\substack{\text{vol }B \to 0 \\ x \in B}} \frac{1}{\text{vol }B} \int_B |f(y)|\, dy,$$

we conclude that $|f(x)| \leqslant Mf(x)$ for almost every $x \in \mathbb{R}^n$. $\square$

Corollary 4.2.6 (Measure Theoretic Interior and Exterior). *Let $A \subseteq \mathbb{R}^n$ be a Lebesgue measurable set. Then for almost every $x \in A$, we have*

$$\lim_{r \to 0^+} \frac{\mathcal{L}^n(B_n(x,r) \cap A)}{\mathcal{L}(B_n(x,r))} = 1$$

and for almost every $x \in \mathbb{R}^n \backslash A$, we have

$$\lim_{r \to 0^+} \frac{\mathcal{L}^n(B_n(x,r) \cap A)}{\mathcal{L}(B_n(x,r))} = 0.$$

Traditionally, the sets of points $x \in \mathbb{R}^n$ where the above limit equals 1 are called the *measure theoretic interior* of A, and the set of points where the limit equals 0 are called the *measure theoretic exterior* of A. Later, we will study a kind of measure theoretic boundary as well.

Proof. Since $A \subseteq \mathbb{R}^n$ is Lebesgue measurable, we have $\chi_A \in L^1_{\text{loc}}(\mathbb{R}^n)$. By Theorem 4.2.4, for almost every $x \in \mathbb{R}^n$ we have

$$\lim_{r \to 0^+} \frac{1}{\mathcal{L}(B_n(x,r))} \int_{B_n(x,r)} \chi_A(y) \, \mathrm{d}y = \chi_A(x).$$

If $x \in A$ is such a point, then $\chi_A(x) = 1$ and hence

$$\lim_{r \to 0^+} \frac{\mathcal{L}^n(B_n(x,r) \cap A)}{\mathcal{L}(B_n(x,r))} = 1.$$

If $x \in \mathbb{R}^n \backslash A$ is such a point, then $\chi_A(x) = 0$ and hence

$$\lim_{r \to 0^+} \frac{\mathcal{L}^n(B_n(x,r) \cap A)}{\mathcal{L}(B_n(x,r))} = 0.$$

$\square$

Definition 4.2.7 (Lebesgue Point). Let $f \in L^1_{\text{loc}}(\mathbb{R}^n)$. A point $x \in \mathbb{R}^n$ is called a *Lebesgue point* of f if

$$\lim_{\substack{\text{vol } B \to 0 \\ x \in B}} \frac{1}{\text{vol } B} \int_B |f(y) - f(x)| \, \mathrm{d}y = 0.$$

The set of all Lebesgue points of f is called the *Lebesgue set* of f.

Remark 4.2.8. By Exercise 4.2.1, every point where f is continuous is a Lebesgue point of f. Furthermore, the triangle inequality implies thatif x is a Lebesgue point of f, then

$$\lim_{\substack{\text{vol } B \to 0 \\ x \in B}} \frac{1}{\text{vol } B} \int_B f(y) \, \mathrm{d}y = f(x).$$

Corollary 4.2.9. *Let $f \in L^1_{loc}(\mathbb{R}^n)$. Then almost every point $x \in \mathbb{R}^n$ is a Lebesgue point of f.*

Proof. For each rational number $r \in \mathbb{Q}$, consider the function $g_r := |f - r|$. Then $g_r \in L^1_{\text{loc}}(\mathbb{R}^n)$, so by Theorem 4.2.4, there exists a set $N_r \subseteq \mathbb{R}^n$ of measure zero such that for each $x \in \mathbb{R}^n \backslash N_r$, we have

$$\lim_{\substack{\text{vol } B \to 0 \\ x \in B}} \frac{1}{\text{vol } B} \int_B g_r(y) \, \mathrm{d}y = g_r(x) = |f(x) - r|.$$

Let $N := \bigcup_{r \in \mathbb{Q}} N_r$ which has measure zero since $\mathbb{Q}$ is countable.

Now let $x \in \mathbb{R}^n \backslash N$ be arbitrary and assume that $f(x)$ is finite. (Note that by Proposition 4.2.3(ii) and 4.2.5, $f(x)$ is finite for almost every $x \in \mathbb{R}^n$.) Let $\epsilon > 0$ be arbitrary, and choose $r \in \mathbb{Q}$ such that $|f(x) - r| < \epsilon/2$. Because

$$\frac{1}{\operatorname{vol} B} \int_B |f(y) - f(x)| \, \mathrm{d}y \leqslant \frac{1}{\operatorname{vol} B} \int_B |f(y) - r| \, \mathrm{d}y + |f(x) - r|$$

taking the limit superior as $\operatorname{vol} B \to 0$ and $x \in B$, we have

$$\limsup_{\substack{\operatorname{vol} B \to 0 \\ x \in B}} \frac{1}{\operatorname{vol} B} \int_B |f(y) - f(x)| \, \mathrm{d}y \leqslant \lim_{\substack{\operatorname{vol} B \to 0 \\ x \in B}} \frac{1}{\operatorname{vol} B} \int_B |f(y) - r| \, \mathrm{d}y + |f(x) - r| = 2|f(x) - r| < \epsilon$$

since $x \notin N_r$. Since $\epsilon > 0$ was arbitrary, we conclude that

$$\lim_{\substack{\operatorname{vol} B \to 0 \\ x \in B}} \frac{1}{\operatorname{vol} B} \int_B |f(y) - f(x)| \, \mathrm{d}y = 0,$$

so x is a Lebesgue point of f. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Remark 4.2.10 (A Precise Representative). Let $f : \mathbb{R}^n \to \mathbb{R}$ be a locally integrable function. By Corollary 4.2.9, almost every point $x \in \mathbb{R}^n$ is a Lebesgue point of f. Thus, we can define a new function $f^* : \mathbb{R}^n \to \mathbb{R}$ by

$$f^*(x) := \begin{cases} \lim_{\substack{\operatorname{vol} B \to 0 \\ x \in B}} \frac{1}{\operatorname{vol} B} \int_B f(y) \, \mathrm{d}y & \text{if } x \text{ is a Lebesgue point of } f, \\ 0 & \text{otherwise.} \end{cases}$$

Then f^* is equal to f almost everywhere, and has the property that

$$\lim_{\substack{\operatorname{vol} B \to 0 \\ x \in B}} \frac{1}{\operatorname{vol} B} \int_B f^*(y) \, \mathrm{d}y = f^*(x) \quad \text{for every } x \in \mathbb{R}^n.$$

We call f^* the *precise representative* of f.

Recall that $L^1_{\mathrm{loc}}(\mathbb{R}^n)$ is defined as the set of equivalence classes of functions that are equal almost everywhere. Thus, if f, g are two representatives of the same equivalence class in $L^1_{\mathrm{loc}}(\mathbb{R}^n)$, then their precise representatives f^*, g^* are equal everywhere. In particular, the precise representative for an equivalence class in $L^1_{\mathrm{loc}}(\mathbb{R}^n)$ is unique. This is nice, as there is some ambiguity in working with equivalence classes of functions, and a canonical representative removes this ambiguity. Thus, if we ever need to evaluate things pointwise, we can always use the precise representative.

So far the theory has been developed for averaging over balls. One can generalize this of course, to other families of sets. One way to capture this is the following definition.

Definition 4.2.11. A collection of sets $\{U_i\}_{i\in I}$ in $\mathbb{R}^n$ is said to *shrink regularly* to a point $x \in \mathbb{R}^n$ if there exists a constant $c > 0$ such that for each $i \in I$, there exists a ball B_i containing x with

$$U_i \subseteq B_i \quad \text{and} \quad \operatorname{vol} B_i \leqslant c \cdot \operatorname{vol} U_i.$$

Thus the sets U_i are all contained in balls that are not too much larger in volume. For instance, if $x \in \mathbb{R}^n$ then the collection of cubes centered at x shrinks regularly to x. With this definition, we can generalize the Lebesgue Differentiation Theorem as follows.

Corollary 4.2.12. *Let* $f \in L^1_{loc}(\mathbb{R}^n)$, *and let* $\{U_i\}_{i\in I}$ *be a collection of sets that shrink regularly to* $x \in \mathbb{R}^n$. *If* x *is a Lebesgue point of* f, *then*

$$\lim_{\substack{\mathcal{L}^n(U_i)\to 0 \\ x\in U_i}} \frac{1}{\mathcal{L}^n(U_i)} \int_{U_i} f(y) \ \mathrm{d}y = f(x).$$

Proof. Let $x \in \mathbb{R}^n$ be a Lebesgue point of f. Since $\{U_i\}_{i\in I}$ shrinks regularly to x, there exists a constant $c > 0$ such that for each $i \in I$, there exists a ball B_i containing x with

$$U_i \subseteq B_i \quad \text{and} \quad \operatorname{vol} B_i \leqslant c \cdot \operatorname{vol} U_i.$$

Then

$$\frac{1}{\mathcal{L}^n(U_i)} \int_{U_i} |f(y) - f(x)| \ \mathrm{d}y \leqslant \frac{1}{c \operatorname{vol} B_i} \int_{B_i} |f(y) - f(x)| \ \mathrm{d}y.$$

Taking the limit as $\operatorname{vol} B_i \to 0$, the right-hand side goes to zero since x is a Lebesgue point of f. Thus,

$$\lim_{\substack{\mathcal{L}^n(U_i)\to 0 \\ x\in U_i}} \frac{1}{\mathcal{L}^n(U_i)} \int_{U_i} |f(y) - f(x)| \ \mathrm{d}y = 0,$$

which implies the result. $\qquad\square$

4.3 Monotone Functions

Recall the definition of monotone functions.

Definition 4.3.1 (Increasing, Decreasing, and Monotone Function)**.** Let $I \subseteq \mathbb{R}$ be an interval and consider a function $f : I \to \mathbb{R}$. The function f is *increasing* if for all $x, y \in I$, we have

$$x \leqslant y \implies f(x) \leqslant f(y).$$

The function f is *decreasing* if for all $x, y \in I$, we have

$$x \leqslant y \implies f(x) \geqslant f(y).$$

The function f is *monotone* if it is either increasing or decreasing.

Clearly a function f is increasing if and only if $-f$ is decreasing, and vice versa. This observation allows us to often reduce questions about monotone functions to questions about increasing functions.

Exercise 4.3.1 (Monotone Functions Have Left and Right Limits). If $f : [a,b] \to \mathbb{R}$ is monotone, then for each $x \in (a,b)$, the left and right limits

$$f(x^-) := \lim_{t \to x^-} f(t) \quad \text{and} \quad f(x^+) := \lim_{t \to x^+} f(t)$$

exist, and we have

$$f(x^-) \leqslant f(x) \leqslant f(x^+)$$

if f is increasing, and the reversed inequalities if f is decreasing. Also $f(a^+) := \lim_{t \to a^+} f(t)$ and $f(b^-) := \lim_{t \to b^-} f(t)$ exist.

Proof. Let $f : [a,b] \to \mathbb{R}$ be an increasing function, and fix $x \in [a,b)$. Then the set $\{f(t) : t \in [a,x)\}$ is bounded above by $f(x)$ since f is increasing, so by the completeness axiom of $\mathbb{R}$, the supremum

$$f(x^+) := \sup\{f(t) : t \in [a,x)\}$$

exists. We claim that $f(x^+) = \lim_{t \to x^+} f(t)$. To see this, let $\varepsilon > 0$. By the definition of supremum, there exists $t_0 \in [a,x)$ such that

$$f(x^+) - \varepsilon < f(t_0) \leqslant f(x^+)$$

since $f(x^+)$ is the least upper bound of $\{f(t) : t \in [a,x)\}$. Since f is increasing, for each $t \in (t_0, x)$ we have

$$f(x^+) - \varepsilon < f(t_0) \leqslant f(t) \leqslant f(x^+).$$

This shows that for all t sufficiently close to x from the left, $f(t)$ is within ε of $f(x^+)$, proving the limit.

Similarly, for each $x \in (a,b]$, the set $\{f(t) : t \in (x,b]\}$ is bounded below by $f(x)$ since f is increasing, so by the completeness axiom of $\mathbb{R}$, the infimum

$$f(x^-) := \inf\{f(t) : t \in (x,b]\}$$

exists and is the limit $f(x^-) = \lim_{t \to x^-} f(t)$ by a similar argument.

Finally, since f is increasing, for each $x \in (a,b)$ we have

$$f(x^-) = \inf\{f(t) : t \in (x,b]\} \leqslant f(x) \leqslant \sup\{f(t) : t \in [a,x)\} = f(x^+).$$

In the case that f is decreasing, then $-f$ is increasing, so the left and right limits of f exist by the above argument. Also for each $x \in (a,b)$ we have

$$f(x^-) = -(-f)(x^-) \geqslant -f(x) = f(x) \geqslant -(-f)(x^+) = f(x^+)$$

since $-f$ is increasing. $\qquad\square$

Lemma 4.3.2 (Increasing Functions are Integrable). *If $f : [a, b] \to \mathbb{R}$ is increasing, then f is measurable and bounded, and hence integrable.*

Proof. The boundedness of f follows immediately from the fact that for each $x \in [a, b]$, we have

$$f(a) \leqslant f(x) \leqslant f(b).$$

For each $c \in \mathbb{R}$, the set $E_c := \{x \in [a, b] : f(x) < c\}$ is either empty or its not; if E_c is nonempty, let $x_c := \sup E_c$. Then we see that

$$\{x \in [a, b] : f(x) < c\} = [a, x_c) \text{ if } x_c \notin E_c,$$

and

$$\{x \in [a, b] : f(x) < c\} = [a, x_c] \text{ if } x_c \in E_c.$$

In any case, E_c is either empty, or a closed, or a half-open interval, so E_c is Borel measurable. This shows that f is measurable by Proposition 1.4.10.

Since f is bounded and measurable, we see that f is integrable by Exercise 2.1.2 (Bounding an Integral). □

It is obvious that monotone functions need not be continuous. However, we have the following.

Lemma 4.3.3 (Monotone Functions can only have Jump Discontinuities). *Let I be an interval in $\mathbb{R}$. If $f : I \to \mathbb{R}$ is monotone, then f can only have jump discontinuities. That is, if f is increasing and $x_0 \in I$ is a point of discontinuity of f, then*

$$f(x_0^-) < f(x_0^+).$$

Similarly, if f is decreasing and $x_0 \in I$ is a point of discontinuity of f, then

$$f(x_0^-) > f(x_0^+).$$

Proof. Without loss of generality, suppose that f is increasing. Also without loss of generality, suppose that $I = [a, b]$ is a closed interval (if I is open or half-open, we can restrict f to a closed subinterval of I and apply the argument there; since I can be written as a countable union of closed intervals, the result will hold for I as well).

Let $x_0 \in (a, b)$ be a point of discontinuity of f. Then by Exercise 4.3.1, the left and right limits $f(x_0^-)$ and $f(x_0^+)$ exist, and we have

$$f(x_0^-) \leqslant f(x_0) \leqslant f(x_0^+).$$

Since f is discontinuous at x_0, we must have either $f(x_0^-) < f(x_0)$ or $f(x_0) < f(x_0^+)$. In either case, we see that

$$f(x_0^-) < f(x_0^+).$$

Thus f has a jump discontinuity at x_0.

We also check that

$$f(a) \leqslant f(a^+) \quad \text{and} \quad f(b^-) \leqslant f(b)$$

which shows that f can only have jump discontinuities at the endpoints a and b as well. Therefore f can only have jump discontinuities on $[a, b]$. $\square$

Theorem 4.3.4 (Monotone Functions have at most Countably Many Discontinuities). *Let I be an interval in $\mathbb{R}$. If $f : [a, b] \to \mathbb{R}$ is monotone, then f has at most countably many points of discontinuity.*

Proof. Without loss of generality, we may assume that f is increasing. Also without loss of generality, assume that I is a closed, bounded interval $[a, b]$. (If I is not closed and bounded, we can restrict to a closed, bounded subinterval of I and apply the argument there; since I can be covered by a countable collection of closed, bounded intervals, the result will follow for I as well.)

Let D be the set of discontinuities of f in (a, b). For each $x \in (a, b)$, define

$$f(x^-) = \lim_{t \to x^-} f(t) = \sup_{t < x} f(t),$$

$$f(x^+) = \lim_{t \to x^+} f(t) = \inf_{t > x} f(t).$$

Since f is increasing, we have $f(x^-) \leqslant f(x) \leqslant f(x^+)$ for all $x \in (a, b)$. Moreover, f is continuous at x if and only if $f(x^-) = f(x^+)$.

If $x \in D$, then $f(x^-) < f(x^+)$ and we define the *jump interval* of f at x to be

$$J_x = (f(x^-), f(x^+)).$$

Note that for each $x \in D$, the jump interval J_x is a nonempty open interval contained in the interval $[f(a), f(b)]$, and the collection of jump intervals $\{J_x : x \in D\}$ is pairwise disjoint. Therefore for each $k \in \mathbb{Z}^+$ there are at most finitely many $x \in D$ such that the length of J_x is at least $1/k$. As a result, the set of discontinuities D is a countable union of finite sets, and is therefore countable. $\square$

Since all countable subsets of $\mathbb{R}$ have Lebesgue measure zero, by the Lebesgue Criterion for Measurability (Theorem **??**), we have the following corollary.

Corollary 4.3.5 (Monotone Functions are Riemann Integrable). *If $f : [a, b] \to \mathbb{R}$ is monotone, then f is Riemann integrable.*

From this observation, one can actually show that the Lebesgue integral of a positive measurable function can be computed as an improper Riemann integral!

Remark 4.3.6 (The Lebesgue Integral is actually an Improper Riemann Integral). Let X be a set with a measure μ, such that (X, μ) is σ-finite. Then by Exercise **??**, we know that for each μ-measurable function $f : X \to [0, \infty]$, we have

$$\int_X f \, d\mu = \int_0^\infty \mu(\{x \in X : t < f(x)\}) \, dt.$$

However, for each μ-measurable function $f : X \to [0, \infty]$, the function $t \mapsto \mu(\{x \in X : t < f(x)\})$ is a decreasing function from $[0, \infty)$ to $[0, \infty]$ — see 1.4.6. Thus if $f : X \to [0, \infty]$ is μ-measurable, then the function $t \mapsto \mu(\{x \in X : t < f(x)\})$ is Riemann integrable on each closed interval $[0, R]$ for each $R > 0$ by Corollary 4.3.5. Therefore we can write the Lebesgue integral of f as an improper Riemann integral,

$$\int_X f \, d\mu = \int_0^\infty \mu(\{x \in X : t < f(x)\}) \, dt = \lim_{R \to \infty} \int_0^R \mu(\{x \in X : t < f(x)\}) \, dt$$

by Corollary 2.2.5.

This formula is what some people take as the definition of the Lebesgue integral, so they can rely on a theory of Riemann integrals. For us, it is a theorem.

Exercise 4.3.2 (Jump Function). Let C be a finite or countable subset of $[a, b]$, which we enumerate as $C = \{x_1, x_2, \ldots\}$. For each $x_n \in C$, let $c_n > 0$ be a positive real number in such a way that the series $\sum_{n=1}^\infty c_n$ converges. Define the function $f : [a, b] \to \mathbb{R}$ by

$$f(x) := \sum_{\{n \, : \, x_n < x\}} c_n.$$

Then the function f is increasing, continuous from the left on $[a, b]$, and has a jump discontinuity of size c_n at each point $x_n \in C$, and is continuous at each point in $[a, b] \backslash C$.

The function f constructed in this way is the *jump function* with jumps c_n at the points x_n for each $n \in \mathbb{Z}^+$.

In other words, for any finite or countable set of points in $[a, b]$, we can construct an increasing function that has jump discontinuities at exactly those points.

Proof. First we check that f is well-defined. For each $x \in [a, b]$, the set $\{x_n : x_n < x\}$ is finite or countable, and the series $\sum_{x_n < x} c_n$ converges since it is a subseries of the convergent series $\sum_{n=1}^\infty c_n$. Thus $f(x)$ is a well-defined real number for each $x \in [a, b]$.

Next we check that f is increasing. Let $x, y \in [a, b]$ such that $x < y$. Then the set $\{n : x_n < x\}$ is a subset of the set $\{n : x_n < y\}$, so

$$f(x) = \sum_{\{n \, : \, x_n < x\}} c_n \leqslant \sum_{\{n \, : \, x_n < y\}} c_n = f(y).$$

Thus f is increasing.

Now we check that f is continuous from the left. Let $x \in (a, b]$. Then by definition of the limit from the left we have

$$f(x^-) = \lim_{t \to x^-} f(t) = \lim_{\substack{\epsilon \to 0, \\ \epsilon > 0}} f(x - \epsilon) = \lim_{\epsilon \to 0^+} \sum_{\{n : x_n < x - \epsilon\}} c_n.$$

If $x_n \in C$ is such that $x_n < x$, then there exists $\epsilon > 0$ such that $x_n < x - \epsilon$. Therefore

$$\lim_{\epsilon \to 0^+} \sum_{\{n : x_n < x - \epsilon\}} c_n = \sum_{\{n : x_n < x\}} c_n = f(x)$$

which proves that $f(x^-) = f(x)$. Thus f is continuous from the left at each point in $(a, b]$

We also check that f has a jump discontinuity of size c_n at each point $x_n \in C$. Then we have

$$f(x_n^+) = \lim_{\substack{\epsilon \to 0 \\ \epsilon > 0}} f(x + \epsilon) = \lim_{\epsilon \to 0^+} \sum_{\{k : x_k < x_n + \epsilon\}} c_k = \sum_{\{k : x_k \leqslant x_n\}} c_k = f(x_n) + c_n.$$

Thus we have

$$f(x_n^+) - f(x_n^-) = f(x_n^+) - f(x_n) = c_n$$

so f has a jump discontinuity of size c_n at x_n.

Finally, we check that f is continuous at each point in $[a, b] \backslash C$. Since the sum $\sum_{n=1}^{\infty} c_n$ converges, we have $c_n \to 0$ as $n \to \infty$. Let $\epsilon > 0$ and choose $N \in \mathbb{Z}^+$ such that $c_n < \epsilon$ for all $n > N$. Then for each $x \in [a, b] \backslash C$ and each $n > N$, there is an interval $[x, x + \delta)$ that contains none of the points $x_1, x_2, \dots, x_n$. Now we see that for each $y \in [x, x + \delta)$, we have

$$f(y) - f(x) = \sum_{\{k : x_k < y\}} c_k - \sum_{\{k : x_k < x\}} c_k = \sum_{\{k : x_k \in [x, y)\}} c_k \leqslant \sum_{k = N+1}^{\infty} c_k < \epsilon.$$

Since f is increasing, that is

$$|f(y) - f(x)| < \epsilon$$

for all $y \in [x, x + \delta)$, proving that f is continuous from the right at x. Since we already showed that f is continuous from the left at x, we see that f is continuous at x. Thus f is continuous at each point in $[a, b] \backslash C$. $\square$

Lemma 4.3.7. *Let $f : [a, b] \to \mathbb{R}$ be an increasing function which is continuous from the left. Then f is the sum of a continuous increasing function and a jump function.*

Proof. Since f is increasing, by Theorem 4.3.4, the set of discontinuities of f is at most countable. Enumerate the set of discontinuities as $\{x_1, x_2, \dots\}$ if it is infinite, or $\{x_1, x_2, \dots, x_N\}$ if it is finite. For each $n \in \mathbb{Z}^+$, define the jump size at x_n to be

$$c_n := f(x_n^+) - f(x_n^-) = f(x_n^+) - f(x_n)$$

since f is continuous from the left — note that $c_n > 0$ since monotone functions only have jump discontinuities 4.3.3.

Define the jump function $\psi : [a, b] \to \mathbb{R}$ with jumps c_n at the points x_n for each $n \in \mathbb{Z}^+$ as in Exercise 4.3.2,

$$\psi(x) := \sum_{\{n\,:\,x_n < x\}} c_n.$$

Then ψ is increasing, continuous from the left, and has a jump discontinuity of size c_n at each point x_n for each $n \in \mathbb{Z}^+$, and is continuous at each point in $[a, b] \backslash \{x_1, x_2, \ldots\}$. Let $\varphi : [a, b] \to \mathbb{R}$ be defined by

$$\varphi(x) := f(x) - \psi(x).$$

We claim that φ is continuous and increasing. To see that φ is increasing, let $x, y \in [a, b]$ such that $x < y$. Then we have

$$\varphi(y) - \varphi(x) = (f(y) - \psi(y)) - (f(x) - \psi(x)) = (f(y) - f(x)) - (\psi(y) - \psi(x))$$

which is non-negative — since f is increasing, the quantity $f(y) - f(x)$ is greater that the sum of the jumps of f at the points x_n in the interval (x, y), which is exactly $\psi(y) - \psi(x)$. Thus $\varphi(y) - \varphi(x) \geqslant 0$, proving that φ is increasing.

To see that φ is continuous, let $x \in [a, b]$. Then we have

$$
\begin{aligned}
\varphi(x^-) &= \lim_{\substack{\epsilon \to 0 \\ \epsilon > 0}} \varphi(x - \epsilon) = \lim_{\substack{\epsilon \to 0 \\ \epsilon > 0}} f(x - \epsilon) - \lim_{\substack{\epsilon \to 0 \\ \epsilon > 0}} \psi(x - \epsilon) \\
&= f(x^-) - \psi(x^-) \\
&= f(x) - \psi(x) = \varphi(x)
\end{aligned}
$$

since both f and ψ are continuous from the left. Thus φ is continuous from the left at x. If $x \notin \{x_1, x_2, \ldots\}$, then both f and ψ are continuous at x, so φ is continuous at x. If $x = x_n$ for some $n \in \mathbb{Z}^+$, then we have a jump discontinuity of size c_n at x_n for both f and ψ, so

$$\varphi(x_n^+) = f(x_n^+) - \psi(x_n^+) = (f(x_n) + c_n) - (\psi(x_n) + c_n) = f(x_n) - \psi(x_n) = \varphi(x_n).$$

Thus φ is continuous from the right at x_n. $\qquad \square$

4.3.1 Derivatives of Monotone Functions

In this section we study the differentiability properties of monotone functions. We need to introduce the following generalization of the usual notion of derivative.

Definition 4.3.8 (Upper and Lower Derivative from Left and Right)**.** Let $[a,b] \subseteq \mathbb{R}$ be an interval and let $f : [a,b] \to \mathbb{R}$ be a function. We define the *upper derivative from the right* of f at $x \in [a,b)$ to be

$$D^+ f(x) := \limsup_{h \to 0^+} \frac{f(x+h) - f(x)}{h},$$

and the *lower derivative from the right* of f at $x \in [a,b)$ to be

$$D_+ f(x) := \liminf_{h \to 0^+} \frac{f(x+h) - f(x)}{h}.$$

Similarly, we define the *upper derivative from the left* of f at $x \in (a,b]$ to be

$$D^- f(x) := \limsup_{h \to 0^-} \frac{f(x+h) - f(x)}{h},$$

and the *lower derivative from the left* of f at $x \in (a,b]$ to be

$$D_- f(x) := \liminf_{h \to 0^-} \frac{f(x+h) - f(x)}{h}.$$

If $D^+ f(x) = D_+ f(x)$, we say that f is *differentiable from the right* at x, and we denote the common value by $f'(x^+)$. If $D^- f(x) = D_- f(x)$, we say that f is *differentiable from the left* at x, and we denote the common value by $f'(x^-)$. If f is differentiable from both the left and the right at x, and $f'(x^+) = f'(x^-)$, then we say that f is *differentiable* at x, and we denote the common value by $f'(x)$.

Remark 4.3.9 (Basic Observations)**.** There are a few basic facts about the upper and lower derivatives from the left and right that are worth noting.

- While many functions f fail to be differentiable, the one-sided upper and lower derivatives always exist (though they may be infinite).

- We trivially have $D_- f(x) \leqslant D^- f(x)$ and $D_+ f(x) \leqslant D^+ f(x)$ for each $x \in (a,b)$. Also $f'(x)$ exists if and only if $D_- f(x) = D^- f(x) = D_+ f(x) = D^+ f(x)$.

- If f is increasing, then $D_- f(x), D_+ f(x) \geqslant 0$ for each $x \in (a,b)$, and by the first line of this remark, we also have $D^- f(x), D^+ f(x) \geqslant 0$ for each $x \in (a,b)$.

- Similarly, if f is decreasing, then $D^- f(x), D^+ f(x) \leqslant 0$ for each $x \in (a,b)$, and by the first line of this remark, we also have $D_- f(x), D_+ f(x) \leqslant 0$ for each $x \in (a,b)$.

- Since the one-sided upper and lower derivatives are defined using $\limsup$ and $\liminf$, we see that if f is measurable, then the functions $D^- f, D_+ f, D^- f, D_+ f : (a,b) \to [-\infty, \infty]$ are all measurable. For instance, the upper derivative from the right $D^+ f$ is the pointwise limit of the sequence of measurable functions $\{D_n^+ f\}_{n=1}^\infty$ defined by

$$D_n^+ f(x) := \sup_{0 < h < 1/n} \frac{f(x+h) - f(x)}{h} \qquad \forall x \in (a,b).$$
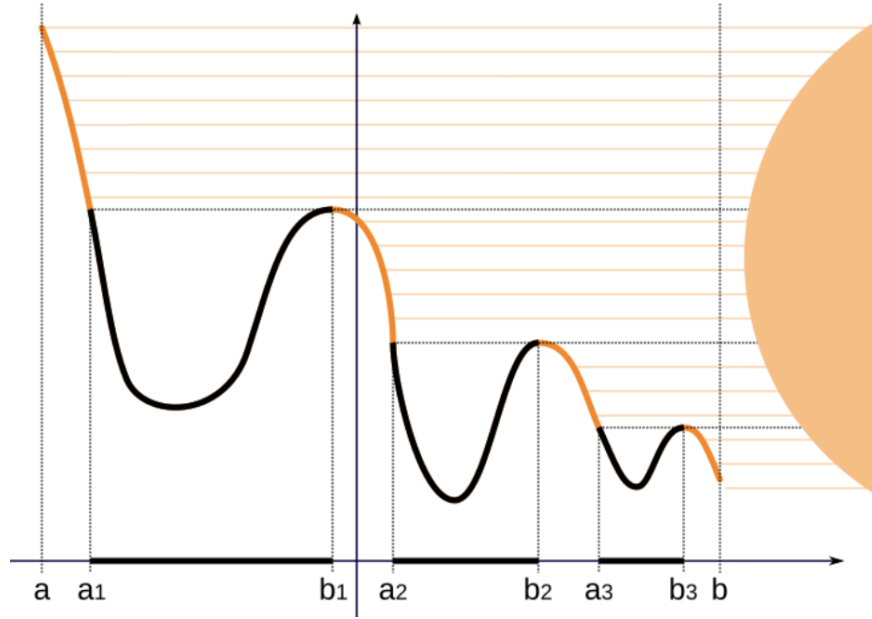
Definition 4.3.10 (Invisible from the Left and Right). Let $f : [a, b] \to \mathbb{R}$ be a continuous function. We say that a point $x \in [a, b]$ is *invisible from the right* if there is a number t such that $x < t \leqslant b$ and $f(x) < f(t)$. Similarly, we say that a point $x \in [a, b]$ is *invisible from the left* if there is a number t such that $a \leqslant t < x$ and $f(x) < f(t)$.

Spivak calls these *shadow points*, which I think fits the picture.

Lemma 4.3.11 (Rising Sun Lemma). *Let $f : [a, b] \to \mathbb{R}$ be a continuous function. Then the set of points in $[a, b]$ that are invisible from the right is the union of at most countably many disjoint open intervals $\{(a_n, b_n)\}_{n=1}^{\infty}$ such that for each $n \in \mathbb{Z}^+$, we have $f(a_n) \leqslant f(b_n)$.*

Similarly, the set of points in $[a, b]$ that are invisible from the left is the union of at most countably many disjoint open intervals $\{(c_n, d_n)\}_{n=1}^{\infty}$ such that for each $n \in \mathbb{Z}^+$, we have $f(c_n) \geqslant f(d_n)$.

It's called the rising sun lemma because if you imagine the graph of f as a landscape, and the sun rising from the right (east), then the points that are invisible from the right are those in shadow.

Proof. We only prove the first assertion; the second follows by a symmetric argument. See that b is not invisible from the right since there is no t such that $b < t \leqslant b$.

First see that if $x \in [a, b)$ is invisible from the right, then by continuity of f, there exists $\delta > 0$ such that $x < y < x + \delta \leqslant b$ and y is also invisible from the right. To see this, let t be such that $x < t \leqslant b$ and $f(x) < f(t)$. By continuity of f at x, there exists $\delta > 0$ such that

for each $y \in (x, x + \delta) \cap [a, b]$, we have

$$f(y) - f(x) \leqslant |f(y) - f(x)| < \frac{f(t) - f(x)}{2}.$$

Thus for each $y \in (x, x + \delta)$, we have

$$f(y) \leqslant \frac{f(t) - f(x)}{2} + f(x) = \frac{f(t) + f(x)}{2} < f(t).$$

Thus each $y \in [x, x + \delta) \cap [a, b]$ is invisible from the right. By taking δ small enough so that $x + \delta \leqslant b$, we that each $y \in [x, x + \delta)$ is invisible from the right.

Since $x \in [a, b)$ was an arbitrary point which is invisible from the right, we see that the set of points in $[a, b]$ that are invisible from the right is open; Hence this set can be written as a countable union of disjoint open intervals $\{(a_n, b_n)\}_{n=1}^{\infty}$.

Notice that for each $n \in \mathbb{Z}^+$, the point b_n is not invisible from the right — if it were, then the above argument shows that b_n belongs to another of the open intervals (a_m, b_m) for some $m \neq n$, and by openness of (a_m, b_m), the intersection $(a_n, b_n) \cap (a_m, b_m)$ is nonempty, contradicting the fact that the intervals are disjoint. Thus for each $n \in \mathbb{Z}^+$, the point b_n is not invisible from the right.

Let $n \in \mathbb{Z}^+$ and suppose towards a contradiction that the interval (a_n, b_n) is nonempty and such that $f(a_n) > f(b_n)$. Then there is a point $x_0 \in (a_n, b_n)$ such that $f(a_n) > f(x_0) > f(b_n)$ by the intermediate value theorem. Define

$$x_* := \sup\{x \in (a_n, b_n) : f(x) = f(x_0)\}$$

to be the least upper bound of the set of points in (a_n, b_n) where f takes the value $f(x_0)$. We claim that $x_* < b_n$ and hence $x_* \in (a_n, b_n)$. To see this, note that since f is continuous, if $x_* = b_n$, then we would have

$$f(x_*) = f(b_n),$$

which contradicts the fact that $f(x_0) > f(b_n)$ and $f(x_0) = f(x_*)$ by definition of x_*. Thus $x_* < b_n$ and hence $x_* \in (a_n, b_n)$. As a result of x_* being in the interval (a_n, b_n), we see that x_* is invisible from the right, which we claim leads to a contradiction.

As x_* is invisible from the right, there is a point t such that $x_* < t \leqslant b$ and $f(x_*) < f(t)$. Clearly we cannot have $t \in (a_n, b_n)$. To see this note that since x_* is the supremum of the set of points in (a_n, b_n) where f takes the value $f(x_0) = f(x_*)$, but $f(b_n) < f(x_0)$, so $t \in (a_n, b_n)$ would imply there is a point in (a_n, b_n) greater than x_* where f takes the value $f(x_0)$, contradicting the definition of x_*. On the other hand, we cannot have $t > b_n$ since this would imply that $f(b_n) < f(x_0) = f(x_*) < f(t)$, contradicting the fact that b_n is not invisible from the right. Thus no such t exists, contradicting the fact that x_* is invisible from the right. Therefore our assumption that $f(a_n) > f(b_n)$ must be false, and we conclude that for each $n \in \mathbb{Z}^+$, we have $f(a_n) \leqslant f(b_n)$. $\square$

Exercise 4.3.3 (Vitali Covering Lemma for Intervals)**.** Let $I_1, I_2, \ldots I_n$ be a finite collection of bounded nonempty open intervals in $\mathbb{R}$. Then there exists a disjoint subcollection $I_{j_1}, I_{j_2}, \ldots, I_{j_k}$ such that

$$\bigcup_{m=1}^{n} I_m \subseteq \bigcup_{m=1}^{k} 3I_{j_m}.$$

Here if I is an interval, we use the notation $3I$ to denote the interval with the same center as I but three times the length of I.

Proof. Use a greedy algorithm.

Select I_{j_1} to be an interval of maximum length among the intervals $I_1, I_2, \ldots, I_n$. (We say "an" instead of "the" because such an interval may not be unique.)

Suppose that the intervals $I_{j_1}, I_{j_2}, \ldots, I_{j_m}$ have been selected. Let $I_{j_{m+1}}$ be an interval of maximum length among the intervals $I_1, I_2, \ldots, I_n$ that are disjoint from $I_{j_1}, I_{j_2}, \ldots, I_{j_m}$. If no such interval exists, then we stop the process. Because we began with a finite collection of intervals, this process must eventually terminate after a finite number of steps, say k steps.

By construction, the intervals $I_{j_1}, I_{j_2}, \ldots, I_{j_k}$ are disjoint. To see the claimed inclusion holds, let $j \in \{1, 2, \ldots, n\}$. If $j = j_m$ for some $m \in \{1, 2, \ldots, k\}$, then clearly $I_j \subseteq 3I_{j_m}$, so we have $I_j \subseteq \bigcup_{m=1}^{k} 3I_{j_m}$.

Thus assume that $j \notin \{j_1, j_2, \ldots, j_k\}$. Then because the process terminated without selecting I_j, we see that I_j is not disjoint from at least one of the intervals $I_{j_1}, I_{j_2}, \ldots, I_{j_k}$. Let $m \in \{1, 2, \ldots, k\}$ be such that I_j is not disjoint from I_{j_m}. Then the length of I_j is at most the length of I_{j_m} by construction, as I_{j_m} was selected to be an interval of maximum length among the intervals that are disjoint from $I_{j_1}, I_{j_2}, \ldots, I_{j_{m-1}}$. Write $I_j = (a_j, b_j)$ and $I_{j_m} = (a_{j_m}, b_{j_m})$. Then we must have

$$b_j - a_j \leqslant b_{j_m} - a_{j_m}$$

since the length of I_j is at most the length of I_{j_m}. Thus we must have $I_j \subseteq 3I_{j_m}$.

[This last sentence requires a moment of thought; draw a picture if necessary. For simplicity, assume that I_{j_m} is centered at the origin, so $I_{j_m} = (-r, r)$ where $r = (b_{j_m} - a_{j_m})/2$ is half the length of I_{j_m}. Then $3I_{j_m} = (-3r, 3r)$. If I_j intersects I_{j_m}, then I_j must contain a point in $(-r, r)$. Since the length of I_j is at most $2r$, it follows that I_j is contained in $(-3r, 3r)$.]

Therefore we have shown that for each $j \in \{1, 2, \ldots, n\}$, we have $I_j \subseteq \bigcup_{m=1}^{k} 3I_{j_m}$. Hence

$$\bigcup_{m=1}^{n} I_m \subseteq \bigcup_{m=1}^{k} 3I_{j_m}$$

as desired. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Lemma 4.3.12 (Derivatives of Jump Functions are a.e. Zero)**.** *Let $f : [a, b] \to \mathbb{R}$ be a jump function. Then the derivative f' exists and vanishes almost everywhere on $[a, b]$.*

Proof. Let $\{c_n\}_{n=1}^{\infty}$ be a sequence of positive real numbers such that the series $\sum_{n=1}^{\infty} c_n$ converges, and let $\{x_n\}_{n=1}^{\infty}$ be a sequence of points in $[a, b]$. Define the jump function $f : [a, b] \to \mathbb{R}$ by

$$f(x) := \sum_{\{n : x_n < x\}} c_n.$$

Then f is increasing, continuous from the left, and has a jump discontinuity of size c_n at each point x_n for each $n \in \mathbb{Z}^+$, and is continuous at each point in $[a, b] \setminus \{x_1, x_2, \ldots\}$ by Exercise 4.3.2.

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

$$j_n(x) := \begin{cases} 1 & \text{if } x_n < x, \\ 0 & \text{if } x \leqslant x_n \end{cases}$$

for each $x \in [a, b]$. Then we have

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

for each $x \in [a, b]$ by definition of f. This expression for f is easier to analyze in what follows.

Since f is increasing, we know it is measurable and bounded by Lemma 4.3.2. Thus $D^+ f$ is measurable, as it is a limit superior of measurable functions. Fix $\epsilon > 0$. Then

$$E_\epsilon := \{x \in [a, b] : D^+ f(x) > \epsilon\}$$

is a measurable set by Proposition 1.4.10. We claim that this set has measure zero.

Let $\eta > 0$. Since the series $\sum_{n=1}^{\infty} c_n$ converges, there exists $N \in \mathbb{Z}^+$ such that $\sum_{n=N+1}^{\infty} c_n < \eta$. Then we let

$$f_0(x) := \sum_{n=N+1}^{\infty} c_n j_n(x) \qquad \forall x \in [a, b].$$

Because of our choice of N, we see see that

$$f_0(b) - f_0(a) = \sum_{n=N+1}^{\infty} c_n < \eta. \tag{$\star$}$$

(We will use this fact later.)

Notice the function $f - f_0$ is a finite sum of the terms $\sum_{n=1}^{N} c_n j_n$. Therefore the set of points

$$E_{\epsilon,0} := \{x \in [a, b] : D^+ f_0(x) > \epsilon\}$$

differs from E_ϵ by at most finitely many points, namely the points $x_1, x_2, \ldots, x_N$. In particular, $E_{\epsilon,0}$ is measurable and $\mathcal{L}^1(E_\epsilon) = \mathcal{L}^1(E_{\epsilon,0})$, so it suffices to show that $\mathcal{L}^1(E_{\epsilon,0}) = 0$.

Since $\mathcal{L}^1$ is a Borel regular outer measure, there exists a compact set $K \subseteq E_{\epsilon,0}$ such that

$$\mathcal{L}^1(K) \geqslant \frac{\mathcal{L}^1(E_{\epsilon,0})}{2}.$$

Then for each $x \in K$, we have $D^+ f_0(x) > \epsilon$. Hence for each $x \in K$, there is an interval $(a_x, b_x) \subset [a, b]$ such that $x \in (a_x, b_x)$ and

$$f_0(b_x) - f_0(a_x) > \epsilon(b_x - a_x).$$

Then the collection of open intervals $\{(a_x, b_x) : x \in K\}$ is an open cover of K, so by by compactness there are finitely many such intervals $(a_1, b_1), \ldots, (a_M, b_M)$ that cover K. By the Vitali Covering Lemma 4.3.3, there is a finite disjoint subcollection of intervals $(a_{i_1}, b_{i_1}), \ldots, (a_{i_L}, b_{i_L})$ such that the intervals $3(a_{i_1}, b_{i_1}), \ldots, 3(a_{i_L}, b_{i_L})$ cover K. Then we have

$$\mathcal{L}^1(K) \leqslant \sum_{n=1}^{M}(b_n - a_n) \leqslant 3\sum_{\ell=1}^{L}(b_{i_\ell} - a_{i_\ell})$$
$$\leqslant \frac{3}{\epsilon}\sum_{\ell=1}^{L}(f_0(b_{i_\ell}) - f_0(a_{i_\ell}))$$
$$\leqslant \frac{3}{\epsilon}(f_0(b) - f_0(a))$$
$$< \frac{3\eta}{\epsilon} \qquad\qquad \text{by } (\star)$$

since the intervals $(a_{i_1}, b_{i_1}), \ldots, (a_{i_L}, b_{i_L})$ are disjoint subintervals of $[a, b]$; hence

$$\mathcal{L}^1(E_{\epsilon,0}) \leqslant 2\mathcal{L}^1(K) < \frac{6\eta}{\epsilon}.$$

Since $\eta > 0$ was arbitrary, we conclude that $\mathcal{L}^1(E_{\epsilon,0}) = 0$. Thus the set $\{x \in [a, b] : D^+ f(x) > \epsilon\}$ has measure zero for each $\epsilon > 0$. Therefore the set

$$\{x \in [a, b] : D^+ f(x) > 0\} = \bigcup_{n=1}^{\infty}\{x \in [a, b] : D^+ f(x) > 1/n\}$$

has measure zero as well. Hence $D^+ f(x) = 0$ for almost every $x \in [a, b]$. Since $D_+ f(x) \leqslant D^+ f(x)$ for each $x \in [a, b)$, we also have $D_+ f(x) = 0$ for almost every $x \in [a, b]$, so f is differentiable from the right at almost every $x \in [a, b)$ with $f'(x^+) = 0$. A similar argument shows that f is differentiable from the left at almost every $x \in (a, b]$ with $f'(x^-) = 0$. Thus f is differentiable at almost every $x \in (a, b)$ with $f'(x) = 0$. $\qquad\square$

Putting everything in this section together, we arrive at the main result.

Theorem 4.3.13 (Monotone Functions are Differentiable a.e.). *Let $f : [a, b] \to \mathbb{R}$ be a monotone function. Then the derivative f' exists and is finite almost everywhere on $[a, b]$.*

Proof. Without loss of generality, we may prove this for increasing functions. We note that because f is increasing, all of the one-sided upper and lower derivatives are non-negative by Remark 4.3.9.

Step 1: We first assume that f is continuous. This assumption will be removed later.

Let $f : [a, b] \to \mathbb{R}$ be a continuous increasing function. We claim that is it enough to show that

$$D^+ f(x) < \infty \quad \text{and} \quad D^+ f(x) \leqslant D_- f(x) \text{ for almost every } x \in [a, b] \qquad (\text{☠})$$

If this is true, then we may apply the same argument to the function $g(y) := -f(-y)$, which is also a continuous increasing function on $[-b, -a]$. Then we would have $D^+ g(y) < \infty$ and $D^+ g(y) \leqslant D_- g(y)$ for almost every $y \in [-b, -a]$. That is, for almost every $x \in [a, b]$, we have

$$\limsup_{h \to 0^+} \frac{g(-x + h) - g(-x)}{h} < \infty \quad \text{and} \quad \limsup_{h \to 0^+} \frac{g(-x + h) - g(-x)}{h} \leqslant \liminf_{h \to 0^-} \frac{g(-x + h) - g(-x)}{h}$$

which is equivalent to

$$\limsup_{h \to 0^+} \frac{f(x - h) - f(x)}{-h} < \infty \quad \text{and} \quad \limsup_{h \to 0^+} \frac{f(x - h) - f(x)}{-h} \leqslant \liminf_{h \to 0^-} \frac{f(x - h) - f(x)}{-h}$$

which is equivalent to

$$D^- f(x) < \infty \quad \text{and} \quad D^- f(x) \leqslant D_+ f(x).$$

Putting this together with the first inequality, we would have

$$D^- f(x), D^+ f(x) < \infty \quad \text{and} \quad D^+ f(x) \leqslant D_- f(x) \quad \text{and} \quad D^- f(x) \leqslant D_+ f(x).$$

That is, for almost every $x \in [a, b]$, we have

$$D^+ f(x) \leqslant D_- f(x) \leqslant D^- f(x) \leqslant D_+ f(x) \leqslant D^+ f(x) < \infty.$$

Hence all of the quantities $D_- f(x), D^- f(x), D_+ f(x), D^+ f(x)$ are finite and equal for almost every $x \in [a, b]$, so f is differentiable at almost every $x \in [a, b]$ with finite derivative. Thus it suffices to prove the two inequalities in (☠).

Step 2: We show that $D^+ f < \infty$ almost everywhere.

For each $M > 0$, define the function $g_M : [a, b] \to \mathbb{R}$ by

$$g_M(x) := f(x) - Mx \qquad \forall x \in [a, b].$$

For each $M > 0$, let E_M be the set of points in $[a, b]$ that are invisible from the right for the function g_M. Then by the Rising Sun Lemma 4.3.11, we can write E_M as the union of at most countably many disjoint open intervals $\{(a_n, b_n)\}_{n=1}^{\infty}$ such that for each $n \in \mathbb{Z}^+$, we have

$$g_M(a_n) \leqslant g_M(b_n)$$

or equivalently

$$f(a_n) - Ma_n \leqslant f(b_n) - Mb_n$$

which rearranges to be

$$M(b_n - a_n) \leqslant f(b_n) - f(a_n).$$

Summing over all $n \in \mathbb{Z}^+$ and dividing by M, we have

$$\sum_{n=1}^{\infty} (b_n - a_n) \leqslant \frac{1}{M} \sum_{n=1}^{\infty} (f(b_n) - f(a_n)) \leqslant \frac{f(b) - f(a)}{M}. \qquad (\text{☺})$$

If $x_0 \in [a, b)$ is such that $D^+ f(x_0) = \infty$, then by definition of $D^+ f(x_0)$, for each $M > 0$, there exists $\delta > 0$ such that for each $x \in [x_0, x_0 + \delta)$, we have

$$\frac{f(x) - f(x_0)}{x - x_0} > M$$

or equivalently

$$f(x) - f(x_0) > M(x - x_0)$$

which rearranges to be

$$f(x) - Mx > f(x_0) - Mx_0.$$

This inequality shows that x_0 is invisible from the right for the function g_M. That is,

$$x_0 \in \{D^+ f = \infty\} \implies (x_0 \in E_M, \ \forall M > 0).$$

Thus for each $M > 0$ the set $\{D^+ f = \infty\}$ is contained in E_M. Since E_M is the union of at most countably many disjoint open intervals $\{(a_n, b_n)\}_{n=1}^{\infty}$ satisfying (☺), we see that

$$\mathcal{L}^1(\{D^+ f = \infty\}) \leqslant \mathcal{L}^1(E_M) = \sum_{n=1}^{\infty} (b_n - a_n) \leqslant \frac{f(b) - f(a)}{M}.$$

Since $M > 0$ was arbitrary, we conclude that $\mathcal{L}^1(\{D^+ f = \infty\}) = 0$, which means that $D^+ f(x) < \infty$ for almost every $x \in [a, b]$.

Step 3: We show that $D^+ f(x) \leqslant D_- f(x)$ for almost every $x \in [a, b]$.

For each pair of positive rational numbers $\alpha, \beta \in \mathbb{Q}^+$ such that $\beta > \alpha$, define the set

$$E_{\alpha,\beta} := \{x \in (\alpha, \beta) : D^+ f(x) > \beta > \alpha > D_- f(x)\}.$$

This set may be empty, but is in any case is measurable since $D^+ f$ and $D_- f$ are measurable. Notice that

$$\{x \in (a, b) : D^+ f(x) > D_- f(x)\} = \bigcup_{\substack{\alpha, \beta \in \mathbb{Q}^+ \\ \beta > \alpha}} E_{\alpha,\beta}$$

by density of the rationals.

We claim that $\mathcal{L}^1(E_{\alpha,\beta}) = 0$ for each pair of positive rational numbers $\alpha, \beta \in \mathbb{Q}^+$ satisfying $\beta > \alpha$. To see this, we suppose that $\mathcal{L}^1(E_{\alpha,\beta}) > 0$ for some pair of positive rational numbers $\alpha, \beta \in \mathbb{Q}^+$ and derive a contradiction. Since $\frac{\beta}{\alpha} > 1$, by Borel regularity of $\mathcal{L}^1$, there exists an open set $U \subset \mathbb{R}$ such that $E_{\alpha,\beta} \subseteq U \subseteq (a, b)$ and

$$\mathcal{L}^1(U) < \frac{\beta}{\alpha} \mathcal{L}^1(E_{\alpha,\beta}).$$

Since U is open, we can write U as a countable union of disjoint open intervals $U = \bigcup_{k=1}^\infty (a_k, b_k)$. By countable disjoint additivity of $\mathcal{L}^1$, we have

$$\mathcal{L}^1(E_{\alpha,\beta}) = \mathcal{L}^1 \left(\bigcup_{k=1}^\infty (E_{\alpha,\beta} \cap (a_k, b_k)) \right) = \sum_{k=1}^\infty \mathcal{L}^1(E_{\alpha,\beta} \cap (a_k, b_k)). \qquad (\star)$$

We claim that for each $k \in \mathbb{Z}^+$, we have

$$\mathcal{L}^1(E_{\alpha,\beta} \cap (a_k, b_k)) \leqslant \frac{\alpha}{\beta}(b_k - a_k). \qquad (\dagger)$$

To see this, let $k \in \mathbb{Z}^+$. If $E_{\alpha,\beta} \cap (a_k, b_k) = \varnothing$, then $(\dagger)$ is trivially true, so assume that $E_{\alpha,\beta} \cap (a_k, b_k) \neq \varnothing$. Then for each $x_0 \in E_{\alpha,\beta} \cap (a_k, b_k)$, we have $D_- f(x_0) < \alpha$, which implies that there exists $\delta > 0$ such that for each $x \in (x_0 - \delta, x_0)$, we have

$$\frac{f(x_0) - f(x)}{x_0 - x} < \alpha$$

or equivalently

$$f(x_0) - f(x) < \alpha(x_0 - x)$$

which rearranges to be

$$f(x_0) - \alpha x_0 < f(x) - \alpha x. \qquad (\heartsuit)$$

This inequality shows that x_0 is invisible from the left for the function $g_\alpha(x) = f(x) - \alpha x$ from step 2.

Thus, by this argument and the Rising Sun Lemma 4.3.11, the set $E_{\alpha,\beta} \cap (a_k, b_k)$ can be written as the union of at most countably many disjoint open intervals $\{(c_n, d_n)\}_{n=1}^{\infty}$ such that for each $n \in \mathbb{Z}^+$, we have

$$f(c_n) - \alpha c_n \geqslant f(d_n) - \alpha d_n$$

or equivalently

$$\alpha(d_n - c_n) \geqslant f(d_n) - f(c_n).$$

For each $n \in \mathbb{Z}^+$, we let $G_n := \{D^+ f > \beta\} \cap (c_n, d_n)$ which is a measurable set. Then for each $n \in \mathbb{Z}^+$, we see that $x_0 \in G_n$ implies that there exists $\delta > 0$ such that for each $x \in (x_0, x_0 + \delta)$, we have

$$\frac{f(x) - f(x_0)}{x - x_0} > \beta$$

or equivalently

$$f(x) - f(x_0) > \beta(x - x_0)$$

which rearranges to be

$$f(x) - \beta x > f(x_0) - \beta x_0.$$

This inequality shows that x_0 is invisible from the right for the function $g_\beta(x) = f(x) - \beta x$ from step 2. Thus, by this argument and the Rising Sun Lemma 4.3.11, the set G_n can be written as the union of at most countably many disjoint open intervals $\{(\tilde{a}_{n,m}, \tilde{b}_{n,m})\}_{m=1}^{\infty}$ such that for each $m \in \mathbb{Z}^+$, we have

$$f(\tilde{a}_{n,m}) - \beta \tilde{a}_{n,m} \leqslant f(\tilde{b}_{n,m}) - \beta \tilde{b}_{n,m}$$

or equivalently

$$\beta(\tilde{b}_{n,m} - \tilde{a}_{n,m}) \leqslant f(\tilde{b}_{n,m}) - f(\tilde{a}_{n,m}).$$

Summing over all $m \in \mathbb{Z}^+$ and dividing by β, we have

$$\sum_{m=1}^{\infty} (\tilde{b}_{n,m} - \tilde{a}_{n,m}) \leqslant \frac{1}{\beta} \sum_{m=1}^{\infty} (f(\tilde{b}_{n,m}) - f(\tilde{a}_{n,m})) \leqslant \frac{f(d_n) - f(c_n)}{\beta}. \qquad (\spadesuit)$$

Putting this all together gives

$$\mathcal{L}^1(E_{\alpha,\beta} \cap (a_k, b_k)) = \mathcal{L}^1\left(\bigcup_{n=1}^{\infty}(E_{\alpha,\beta} \cap (c_n, d_n))\right) \qquad \text{since } E_{\alpha,\beta} \cap (a_k, b_k) = \cup_{n=1}^{\infty}(E_{\alpha,\beta} \cap (c_n, d_n))$$

$$= \sum_{n=1}^{\infty} \mathcal{L}^1(E_{\alpha,\beta} \cap (c_n, d_n)) \qquad \text{by countable disjoint additivity}$$

$$\leqslant \sum_{n=1}^{\infty} \mathcal{L}^1(G_n) \qquad \text{by definition of the sets } G_n$$

$$= \sum_{n=1}^{\infty} \mathcal{L}^1\left(\bigcup_{m=1}^{\infty}(\tilde{a}_{n,m}, \tilde{b}_{n,m})\right) \qquad \text{since } G_n = \bigcup_{m=1}^{\infty}(\tilde{a}_{n,m}, \tilde{b}_{n,m}) \text{ for each } n$$

$$= \sum_{n=1}^{\infty}\sum_{m=1}^{\infty}(\tilde{b}_{n,m} - \tilde{a}_{n,m}) \qquad \text{by countable disjoint additivity}$$

$$\leqslant \frac{1}{\beta}\sum_{n,m=1}^{\infty}(f(\tilde{b}_{n,m}) - f(\tilde{a}_{n,m})) \quad \text{by } (\spadesuit)$$

$$\leqslant \frac{1}{\beta}\sum_{n=1}^{\infty}(f(d_n) - f(c_n)) \qquad \text{since } f \text{ is increasing and } \{(\tilde{a}_{n,m}, \tilde{b}_{n,m})\}_{m=1}^{\infty} \text{ are disjoint}$$

$$\leqslant \frac{\alpha}{\beta}\sum_{n=1}^{\infty}(d_n - c_n) \qquad \text{by } (\heartsuit)$$

$$\leqslant \frac{\alpha}{\beta}(b_k - a_k) \qquad \text{since } \{(c_n, d_n)\}_{n=1}^{\infty} \text{ are disjoint subintervals of } (a_k, b_k)$$

which proves our claim (†).

Combining ($\star$) and (†), using countable disjoint additivity and the definition of $U = \bigcup_{k=1}^{\infty}(a_k, b_k)$ gives

$$\mathcal{L}^1(E_{\alpha,\beta}) \leqslant \frac{\alpha}{\beta}\sum_{k=1}^{\infty}(b_k - a_k) = \frac{\alpha}{\beta}\mathcal{L}^1\left(\bigcup_{k=1}^{\infty}(a_k, b_k)\right) = \frac{\alpha}{\beta}\mathcal{L}^1(U) < \mathcal{L}^1(E_{\alpha,\beta}).$$

But wait, this says that $\mathcal{L}^1(E_{\alpha,\beta}) < \mathcal{L}^1(E_{\alpha,\beta})$, which is a contradiction!

Thus our assumption that $\mathcal{L}^1(E_{\alpha,\beta}) > 0$ for some pair of positive rational numbers $\alpha, \beta \in \mathbb{Q}^+$ satisfying $\beta > \alpha$ must be false. Therefore $\mathcal{L}^1(E_{\alpha,\beta}) = 0$ for each pair of positive rational numbers $\alpha, \beta \in \mathbb{Q}^+$ satisfying $\beta > \alpha$.

After all this work, we now use countable subadditivity to obtain

$$\mathcal{L}^1(\{D^+f > D_-f\}) = \mathcal{L}^1\left(\bigcup_{\substack{\alpha,\beta \in \mathbb{Q}^+ \\ \beta > \alpha}} E_{\alpha,\beta}\right) \leqslant \sum_{\substack{\alpha,\beta \in \mathbb{Q}^+ \\ \beta > \alpha}} \mathcal{L}^1(E_{\alpha,\beta}) = 0.$$

Thus $D^+f(x) \leqslant D_-f(x)$ for almost every $x \in [a, b]$.

By the discussion in step 1 (☺), this completes the proof under the assumption that f is continuous.

Step 4: We now remove the assumption that f is continuous.

Let $f : [a,b] \to \mathbb{R}$ be an arbitrary increasing function. Let J be the set of jump discontinuities of f. Then J is at most countable by Theorem 4.3.4. We define another function $\tilde{f} : [a,b] \to \mathbb{R}$ by

$$\tilde{f}(x) := f(x^-) = \sup_{y < x} f(y) \qquad \forall x \in (a,b],$$

and $\tilde{f}(a) := f(a)$. Then $\tilde{f}$ is an increasing function that is continuous from the left at each point in $[a,b]$ and satisfies $\tilde{f}(x) = f(x)$ for each $x \in [a,b] \backslash J$. That is, if f has a jump discontinuity at $x \in (a,b]$ and for some reason $f(x)$ is strictly between $f(x^-)$ and $f(x^+)$, then we define $\tilde{f}(x)$ to be $f(x^-)$ instead of $f(x)$. Thus f and $\tilde{f}$ differ at only countably many points.

Then since $\tilde{f}$ is an increasing function that is continuous from the left, by Lemma 4.3.7, we can write $\tilde{f}$ as the sum of a continuous increasing function $\varphi : [a,b] \to \mathbb{R}$ and a jump function $\psi : [a,b] \to \mathbb{R}$,

$$\tilde{f}(x) = \varphi(x) + \psi(x) \qquad \forall x \in [a,b].$$

By Step 1, the derivative φ' exists and is finite almost everywhere on $[a,b]$. By Lemma 4.3.12, the derivative ψ' exists and vanishes almost everywhere on $[a,b]$. Therefore the derivative $\tilde{f}' = \varphi' + \psi'$ exists and is finite almost everywhere on $[a,b]$. But since f and $\tilde{f}$ differ at only countably many points, we conclude that the derivative f' exists and is finite almost everywhere on $[a,b]$ as well. $\qquad \square$

Remark 4.3.14 (Monotone Functions on other Intervals). What happens if the domain of a monotone function is not a closed bounded interval $[a,b]$? Say the domain is an open bounded interval (a,b) or a half-open bounded interval $[a,b)$ or $(a,b]$. Or more generally, what if the domain is an unbounded interval such as $(-\infty, b)$, (a, ∞), or $(-\infty, \infty)$, or the half-open unbounded intervals $(-\infty, b]$ or $[a, \infty)$ or even all of $\mathbb{R}$?

It is still true that monotone functions on an arbitrary interval have at most countably many discontinuities. This follows from the fact that any interval - bounded, unbounded, open, half-open, or closed - can be written as a countable union of closed bounded intervals, and that the countable union of countable sets is countable.

The story for derivatives is the same. Suppose that $I \subseteq \mathbb{R}$ is any interval - bounded, unbounded, open, half-open, closed, or all of $\mathbb{R}$ - and let $f : I \to \mathbb{R}$ be a monotone function. Then for each closed interval $[c,d] \subset I$, the restriction of f to $[c,d]$ is a monotone function on a closed bounded interval, so by Theorem 4.3.13, the derivative of f exists and is finite

almost everywhere on $[c, d]$. Since I can be written as a countable union of closed bounded intervals, we conclude that the derivative of f exists and is finite almost everywhere on I.

The following example shows that this is the best we can do; monotone functions can have an uncountable number of points where they fail to be differentiable.

Exercise 4.3.4 (Fundamental Theorem of Calculus Inequality)**.** Let $f : [a, b] \to \mathbb{R}$ be an increasing function. Then the derivative f' is integrable and satisfies

$$\int_a^b f'(x) \, \mathrm{d}x \leqslant f(b) - f(a).$$

Show the Devil's staircase is an increasing continuous function $f : [a, b] \to \mathbb{R}$ such that the inequality is strict.

Proof. (What integral theorem allows us to get an inequality by commuting the integral and limit? Fatou, that's who!) For technical reasons, we first extend the domain of f to $[a, b+1]$ by defining $f(x) := f(b)$ for each $x \in (b, b + 1]$.

For each $k \in \mathbb{Z}^+$ we define a function

$$g_k(x) := \frac{f(x + 1/k) - f(x)}{1/k}, \qquad x \in [a, b].$$

Since f is increasing, we know that f is differentiable almost everywhere on (a, b) by Theorem 4.3.13 and that $g_k(x) \geqslant 0$ for each $x \in [a, b]$. Also, for each $x \in [a, b]$ we have

$$\lim_{k \to \infty} g_k(x) = f'(x)$$

if f is differentiable at x, and $\lim_{k \to \infty} g_k(x)$ does not exist otherwise. Thus $\lim_{k \to \infty} g_k(x) = f'(x)$ for almost every $x \in [a, b]$.

That is $\{g_k\}_{k=1}^\infty$ is a sequence of nonnegative measurable functions that converges pointwise almost everywhere to the nonnegative measurable function f'. Thus by Fatou's Lemma (2.2.1), we have

$$\int_a^b f'(x) \, \mathrm{d}x \leqslant \liminf_{k \to \infty} \int_a^b g_k(x) \, \mathrm{d}x.$$

For each $k \in \mathbb{Z}^+$, we have

$$
\begin{aligned}
\int_a^b g_k(x) \, \mathrm{d}x &= \int_a^b \frac{f(x + 1/k) - f(x)}{1/k} \, \mathrm{d}x \\
&= k \int_a^b f(x + 1/k) - f(x) \, \mathrm{d}x \\
&= k \left(\int_{a+1/k}^{b+1/k} f(t) \, \mathrm{d}t - \int_a^b f(x) \, \mathrm{d}x \right) \qquad \text{by the substitution } t = x + 1/k \\
&= k \left(\int_b^{b+1/k} f(t) \, \mathrm{d}t - \int_a^{a+1/k} f(x) \, \mathrm{d}x \right) \qquad \text{by splitting the integral} \\
&\leqslant k \left(f(b)(1/k) - f(a)(1/k) \right) \qquad\qquad\quad \text{since } f \text{ is increasing} \\
&= f(b) - f(a).
\end{aligned}
$$

Putting these two inequalities together gives

$$
\int_a^b f'(x) \, \mathrm{d}x \leqslant \liminf_{k \to \infty} \int_a^b g_k(x) \, \mathrm{d}x \leqslant f(b) - f(a)
$$

as desired.

For an example, consider the Devil's staircase function $f : [0, 1] \to [0, 1]$ from Example 1.3.2. This function is increasing, but has derivative $f'(x) = 0$ at each point $x \in [0, 1] \backslash C$, where C is the Cantor set, which has Lebesgue measure zero. Thus f' is integrable and satisfies

$$
\int_0^1 f'(x) \, \mathrm{d}x = 0 < 1 = f(1) - f(0).
$$

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

Example 4.3.15 (The Devil's Staircase). Recall that in Example 1.3.2, we constructed the Devil's staircase function $f : [0, 1] \to [0, 1]$, which is an increasing continuous function with $f(0) = 0$ and $f(1) = 1$, and that f is differentiable with at each point in $[0, 1] \backslash C$ with zero derivative, where C is the Cantor set. However, we did not show that the derivative f' fails to exist on a set of points that is uncountable.

It is actually not fully understood exactly where the derivative f' fails to exist on the Cantor set C, however it is known that f' fails to exist at uncountably many points in C.

4.3.2 The Second Fundamental Theorem of Calculus

To finish off this section, we prove a generalization of the Second Fundamental Theorem of Calculus.

Lemma 4.3.16. *Let* $f : [a, b] \to \mathbb{R}$ *be an integrable function, and define the function* $F :$ $[a, b] \to \mathbb{R}$ *by*

$$F(x) := \int_a^x f(t) \, dt \qquad \forall x \in [a, b].$$

Then F *is differentiable and* $F'(x)$ *is finite for almost every* $x \in [a, b]$.

Proof. Write $f = f^+ - f^-$ where $f^+, f^- : [a, b] \to [0, \infty)$ are integrable functions, so that

$$F(x) = \int_a^x f(t) \, dt = \int_a^x f^+(t) \, dt - \int_a^x f^-(t) \, dt.$$

Defining

$$F_1(x) := \int_a^x f^+(t) \, dt \quad \text{and} \quad F_2(x) := \int_a^x f^-(t) \, dt$$

for each $x \in [a, b]$, we see that $F = F_1 - F_2$. Furthermore, the functions F_1 and F_2 are increasing since f^+ and f^- are non-negative.

We check this for F_1; the argument for F_2 is similar. Let $x, y \in [a, b]$ with $x < y$. Then $\mathbf{1}_{[a,x]}(t) \leqslant \mathbf{1}_{[a,y]}(t)$ for each $t \in [a, b]$, so by monotonicity of the Lebesgue integral we Have

$$F_1(x) = \int_a^x f^+(t) \, dt = \int_{[a,x]} f^+(t) \, dt = \int_{\mathbb{R}} \mathbf{1}_{[a,x]}(t) f^+(t) \, dt$$

$$\leqslant \int_{\mathbb{R}} \mathbf{1}_{[a,y]}(t) f^+(t) \, dt = \int_{[a,y]} f^+(t) \, dt = \int_a^y f^+(t) \, dt = F_1(y).$$

Hence $F_1(x) \leqslant F_1(y)$ whenever $x < y$, so F_1 is increasing.

Thus the functions F_1 and F_2 are differentiable and have finite derivatives almost everywhere on $[a, b]$ by Theorem 4.3.13. Therefore the function $F = F_1 - F_2$ is differentiable and has finite derivative $F' = F_1' - F_2'$ almost everywhere on $[a, b]$ as well. $\qquad \square$

With this lemma in hand, we can prove the Second Fundamental Theorem of Calculus.

Theorem 4.3.17 (Second Fundamental Theorem of Calculus). *Let* $f : [a, b] \to \mathbb{R}$ *be an integrable function, and define the function* $F : [a, b] \to \mathbb{R}$ *by*

$$F(x) := \int_a^x f(t) \, dt \qquad \forall x \in [a, b].$$

Then F *is continuous on* $[a, b]$ *and differentiable for almost every* $x \in [a, b]$, *and* $F'(x) = f(x)$ *for almost every* $x \in [a, b]$.

Proof. *Step 0:* We show that F is continuous on $[a, b]$.

Let $\epsilon > 0$. By using the fact from Lemma 2.1.22, there exists $\delta > 0$ such that for each measurable set $E \subseteq [a, b]$ with $\mathcal{L}^1(E) < \delta$, we have

$$\int_E |f(t)| \, dt < \epsilon.$$

The triangle inequality then gives that for each measurable set $E \subseteq [a, b]$ with $\mathcal{L}^1(E) < \delta$, we have

$$\left| \int_E f(t) \, dt \right| \leqslant \int_E |f(t)| \, dt < \epsilon.$$

If $x_0 \in [a, b]$ and $x \in [a, b]$ with $|x - x_0| < \delta$, then letting $E = [x, x_0]$ or $E = [x_0, x]$ as appropriate gives

$$|F(x) - F(x_0)| = \left| \int_{x_0}^x f(t) \, dt \right| = \left| \int_E f(t) \, dt \right| < \epsilon.$$

Since $\epsilon > 0$ was arbitrary, this shows F is continuous at x_0, and since x_0 was arbitrary, F is continuous on $[a, b]$.

Step 1: We claim that it is enough to show $f(x) \geqslant F'(x)$ for almost every $x \in [a, b]$.

If this is true, then we may apply the same argument to the function $-f$, which is also integrable. Then we would have $-f(x) \geqslant (-F)'(x) = -F'(x)$ for almost every $x \in [a, b]$, which is equivalent to $f(x) \leqslant F'(x)$ for almost every $x \in [a, b]$. Together with the first inequality, this would give $f(x) = F'(x)$ for almost every $x \in [a, b]$. Thus it suffices to prove the first inequality.

Step 2: For each pair of positive rational numbers $\alpha, \beta \in \mathbb{Q}^+$ such that $\beta > \alpha$, define the set

$$E_{\alpha, \beta} := \{x \in (\alpha, \beta) : F'(x) > \beta > \alpha > f(x)\}.$$

We claim that $\mathcal{L}^1(E_{\alpha, \beta}) = 0$ for each pair of positive rational numbers $\alpha, \beta \in \mathbb{Q}^+$ satisfying $\beta > \alpha$.

Before proving this claim, observe that $\{x \in (a, b) : F'(x) > f(x)\} = \bigcup_{\substack{\alpha, \beta \in \mathbb{Q}^+ \\ \beta > \alpha}} E_{\alpha, \beta}$ by density of the rationals. Thus by countable subadditivity our claim implies that

$$\mathcal{L}^1(\{F' > f\}) = \mathcal{L}^1 \left(\bigcup_{\substack{\alpha, \beta \in \mathbb{Q}^+ \\ \beta > \alpha}} E_{\alpha, \beta} \right) \leqslant \sum_{\substack{\alpha, \beta \in \mathbb{Q}^+ \\ \beta > \alpha}} \mathcal{L}^1(E_{\alpha, \beta}) = 0.$$

Hence the claim implies $F'(x) \leqslant f(x)$ for almost every $x \in [a, b]$.

To prove the claim, fix $\alpha, \beta \in \mathbb{Q}^+$ with $\beta > \alpha$. Note that $E_{\alpha, \beta}$ is measurable since F' and f are measurable.

Let $\epsilon > 0$, and as in step 0, let $\delta > 0$ be such that for each measurable set $E \subseteq [a, b]$ with $\mathcal{L}^1(E) < \delta$, we have

$$\left| \int_E f(t) \, dt \right| < \epsilon.$$

By Borel regularity of $\mathcal{L}^1$, there exists an open set $U \subset \mathbb{R}$ such that $E_{\alpha,\beta} \subseteq U \subseteq (a, b)$ and

$$\mathcal{L}^1(U) < \mathcal{L}^1(E_{\alpha,\beta}) + \delta.$$

Since U is open, we can write U as a countable union of disjoint open intervals $U = \bigcup_{k=1}^{\infty}(a_k, b_k)$. By countable disjoint additivity of $\mathcal{L}^1$, we have

$$\mathcal{L}^1(E_{\alpha,\beta}) = \mathcal{L}^1\left(\bigcup_{k=1}^{\infty}(E_{\alpha,\beta} \cap (a_k, b_k))\right) = \sum_{k=1}^{\infty}\mathcal{L}^1(E_{\alpha,\beta} \cap (a_k, b_k)). \tag{$\star$}$$

For the moment, fix $k \in \mathbb{Z}^+$. If $x_0 \in E_{\alpha,\beta} \cap (a_k, b_k)$ then by definition of $E_{\alpha,\beta}$ we have $F'(x_0) > \beta$, so there exists $\delta_0 > 0$ such that for each $x \in (x_0 - \delta_0, x_0 + \delta_0)$, we have

$$\frac{F(x) - F(x_0)}{x - x_0} > \beta$$

which rearranges to be

$$F(x) - \beta x > F(x_0) - \beta x_0.$$

This inequality shows that each $x_0 \in E_{\alpha,\beta} \cap (a_k, b_k)$ is invisible from the right with respect to the function $x \mapsto F(x) - \beta x$, which is continuous on $[a, b]$ by step 0. Thus, by this argument and the Rising Sun Lemma 4.3.11, the set $E_{\alpha,\beta} \cap (a_k, b_k)$ can be covered by at most countably many disjoint open intervals $\{(c_{k,j}, d_{k,j})\}_{j=1}^{\infty}$ such that for each $j \in \mathbb{Z}^+$, we have

$$F(c_{k,j}) - \beta c_{k,j} \leqslant F(d_{k,j}) - \beta d_{k,j}$$

or equivalently

$$\beta(d_{k,j} - c_{k,j}) \leqslant F(d_{k,j}) - F(c_{k,j}) = \int_{c_{k,j}}^{d_{k,j}} f(t)\ \mathrm{d}t. \tag{$\star\star$}$$

Then we see that

$$E_{\alpha,\beta} \cap (a_k, b_k) \subseteq \bigcup_{j=1}^{\infty}(c_{k,j}, d_{k,j})$$

which implies

$$\mathcal{L}^1(E_{\alpha,\beta} \cap (a_k, b_k)) \leqslant \sum_{j=1}^{\infty}(d_{k,j} - c_{k,j}) \leqslant \frac{1}{\beta}\sum_{j=1}^{\infty}\int_{c_{k,j}}^{d_{k,j}} f(t)\ \mathrm{d}t = \frac{1}{\beta}\int_{\bigcup_{j=1}^{\infty}(c_{k,j}, d_{k,j})} f(t)\ \mathrm{d}t$$

by monotonicity, using ($\star\star$), and countable disjoint additivity of the Lebesgue integral.

By summing this inequality over all $k \in \mathbb{Z}^+$, using ($\star$), and countable disjoint additivity we obtain

$$\mathcal{L}^1(E_{\alpha,\beta}) \leqslant \frac{1}{\beta}\sum_{k=1}^{\infty}\int_{\bigcup_{j=1}^{\infty}(c_{k,j}, d_{k,j})} f(t)\ \mathrm{d}t = \frac{1}{\beta}\int_{\bigcup_{k,j=1}^{\infty}(c_{k,j}, d_{k,j})} f(t)\ \mathrm{d}t.$$

On the other hand we have

$$\int_{\bigcup_{k,j=1}^{\infty}(c_{k,j},d_{k,j})} f(t) \; \mathrm{d}t = \int_{E_{\alpha,\beta}} f(t) \; \mathrm{d}t + \int_{\left(\bigcup_{k,j=1}^{\infty}(c_{k,j},d_{k,j})\right)\backslash E_{\alpha,\beta}} f(t) \; \mathrm{d}t.$$

$$< \alpha \cdot \mathcal{L}^1(E_{\alpha,\beta}) + \epsilon$$

where the inequality for the first term follows from definition of $E_{\alpha,\beta}$ and Exercise 2.1.2, the inequality for the second term follows from the fact that $\mathcal{L}^1\left(\left(\bigcup_{k,j=1}^{\infty}(c_{k,j},d_{k,j})\right)\backslash E_{\alpha,\beta}\right) \leqslant \mathcal{L}^1(U\backslash E_{\alpha,\beta}) < \delta$ and our choice of δ.

Combinging these last two inequalities gives

$$\mathcal{L}^1(E_{\alpha,\beta}) < \frac{1}{\beta}\left(\alpha \cdot \mathcal{L}^1(E_{\alpha,\beta}) + \epsilon\right) = \frac{\alpha}{\beta}\mathcal{L}^1(E_{\alpha,\beta}) + \frac{\epsilon}{\beta}.$$

Since $\epsilon > 0$ was arbitrary, this implies

$$\mathcal{L}^1(E_{\alpha,\beta}) \leqslant \frac{\alpha}{\beta}\mathcal{L}^1(E_{\alpha,\beta}).$$

But since $\beta > \alpha$, this is only possible if $\mathcal{L}^1(E_{\alpha,\beta}) = 0$, which proves our claim.

As observed in the beginning of step 2, this completes the proof. □

4.4 Functions of Bounded Variation on $\mathbb{R}$

In this section, we study functions of bounded variation on $\mathbb{R}$, and absolutely continuous functions on $\mathbb{R}$.

In this section $[a,b]$ will always denote a closed, bounded interval in $\mathbb{R}$ with $a < b$.

Definition 4.4.1 (Total Variation, Bounded Variation)**.** Let $f : [a,b] \to \mathbb{R}$ be a function. For each partition $P = \{a = x_0 < x_1 < \cdots < x_N = b\}$ of $[a,b]$, we define the *total variation* of f with respect to the partition P to be

$$V(f,P) := \sum_{k=1}^{N} |f(x_k) - f(x_{k-1})|.$$

The *total variation* of f on $[a,b]$ is

$$V_a^b(f) := \sup\{V(f,P) : P \text{ is a partition of } [a,b]\} \in [0,\infty].$$

If $[c,d] \subseteq [a,b]$ is a subinterval, then $V_c^d(f)$ is the total variation of the restriction $f|_{[c,d]}$.

We say that f is of *bounded variation* on $[a,b]$ if $V_a^b(f) < \infty$.

The set of all functions of bounded variation on $[a,b]$ is denoted by $BV([a,b])$.

Remark 4.4.2 (BV Functions on a Closed, Bounded Interval are Bounded). If $f : [a, b] \to \mathbb{R}$ is of bounded variation on $[a, b]$, then f is bounded.

The proof is easy. For each $x \in [a, b]$ we consider the partition $P = \{a, x, b\}$ of $[a, b]$. Then we have

$$V(f, P) = |f(x) - f(a)| + |f(b) - f(x)| \leqslant V_a^b(f) < \infty$$

which implies that

$$|f(x)| \leqslant |f(x) - f(a)| + |f(a)| \leqslant |f(x) - f(a)| + |f(b) - f(x)| + |f(a)| \leqslant V_a^b(f) + |f(a)|.$$

Thus f is bounded above by $V_a^b(f) + |f(a)|$ on $[a, b]$.

Remark 4.4.3 (Functions of Bounded Variation on $\mathbb{R}$). We say that a function $f : \mathbb{R} \to \mathbb{R}$ is of bounded variation on $\mathbb{R}$ if it is of bounded variation on every closed, bounded interval $[a, b] \subseteq \mathbb{R}$ and

$$\sup_{\substack{a, b \in \mathbb{R} \\ a < b}} V_a^b(f) < \infty.$$

This supremum is called the total variation of f on $\mathbb{R}$ and is denoted by $V_{-\infty}^\infty(f)$.

The set of all functions of bounded variation on $\mathbb{R}$ is denoted by $BV(\mathbb{R})$.

Example 4.4.4 (Increasing Functions are BV). If $f : [a, b] \to \mathbb{R}$ is increasing, then for each partition $P = \{a = x_0 < x_1 < \cdots < x_N = b\}$ of $[a, b]$, we have

$$V(f, P) = \sum_{k=1}^N |f(x_k) - f(x_{k-1})| = \sum_{k=1}^N (f(x_k) - f(x_{k-1})) = f(b) - f(a).$$

Thus $V_a^b(f) = f(b) - f(a) < \infty$, so f is of bounded variation on $[a, b]$.

Example 4.4.5 (Lipschitz Functions are BV). If $f : [a, b] \to \mathbb{R}$ is Lipschitz with Lipschitz constant $L > 0$, then for each partition $P = \{a = x_0 < x_1 < \cdots < x_N = b\}$ of $[a, b]$, we have

$$V(f, P) = \sum_{k=1}^N |f(x_k) - f(x_{k-1})| \leqslant \sum_{k=1}^N L|x_k - x_{k-1}| = L(b - a).$$

Thus $V_a^b(f) \leqslant L(b - a) < \infty$, so f is of bounded variation on $[a, b]$.

Example 4.4.6 (Differences of Increasing Functions are BV). If $f, g : [a, b] \to \mathbb{R}$ are increasing functions, then for each partition $P = \{a = x_0 < x_1 < \cdots < x_N = b\}$ of $[a, b]$, we

have

$$
\begin{aligned}
V(f - g, P) &= \sum_{k=1}^{N} |(f - g)(x_k) - (f - g)(x_{k-1})| \\
&= \sum_{k=1}^{N} |(f(x_k) - f(x_{k-1})) - (g(x_k) - g(x_{k-1}))| \\
&\leqslant \sum_{k=1}^{N} |f(x_k) - f(x_{k-1})| + \sum_{k=1}^{N} |g(x_k) - g(x_{k-1})| \\
&= V(f, P) + V(g, P).
\end{aligned}
$$

Taking the supremum over all partitions P of $[a, b]$ gives

$$
V_a^b(f - g) \leqslant V_a^b(f) + V_a^b(g) < \infty,
$$

so $f - g$ is of bounded variation on $[a, b]$.

Example 4.4.7 (A Function of Unbounded Variation)**.** The function ...

Exercise 4.4.1 (*BV* is a Vector Space)**.** Let $[a, b]$ be a closed, bounded interval in $\mathbb{R}$ with $a < b$. Then $BV([a, b])$ is a vector space over $\mathbb{R}$. Also $BV(\mathbb{R})$ is closed under products.

Proof. First it is obvious that the zero function is in $BV([a, b])$ since $V_a^b(0) = 0 < \infty$. Now let $f, g \in BV([a, b])$ and let $\alpha \in \mathbb{R}$. We need to show that $f + g$ and αf are in $BV([a, b])$.

Let $P = \{a = x_0 < x_1 < \cdots < x_N = b\}$ be a partition of $[a, b]$. Then we have

$$
\begin{aligned}
V(f + g, P) &= \sum_{k=1}^{N} |(f + g)(x_k) - (f + g)(x_{k-1})| \\
&= \sum_{k=1}^{N} |(f(x_k) - f(x_{k-1})) + (g(x_k) - g(x_{k-1}))| \\
&\leqslant \sum_{k=1}^{N} |f(x_k) - f(x_{k-1})| + \sum_{k=1}^{N} |g(x_k) - g(x_{k-1})| \\
&= V(f, P) + V(g, P)
\end{aligned}
$$

by the triangle inequality. Taking the supremum over all partitions P of $[a, b]$ gives

$$
V_a^b(f + g) \leqslant V_a^b(f) + V_a^b(g) < \infty,
$$

so $f + g$ is in $BV([a, b])$.

Now let $P = \{a = x_0 < x_1 < \cdots < x_N = b\}$ be a partition of $[a, b]$. Then we have

$$
\begin{aligned}
V(\alpha f, P) &= \sum_{k=1}^{N} |\alpha f(x_k) - \alpha f(x_{k-1})| \\
&= \sum_{k=1}^{N} |\alpha| |f(x_k) - f(x_{k-1})| \\
&= |\alpha| V(f, P).
\end{aligned}
$$

Taking the supremum over all partitions P of $[a, b]$ gives

$$
V_a^b(\alpha f) = |\alpha| V_a^b(f) < \infty,
$$

so αf is in $BV([a, b])$.

This shows that $BV([a, b])$ is closed under addition and scalar multiplication, and so $BV([a, b])$ is a vector space over $\mathbb{R}$.

To see that $BV([a, b])$ is closed under products, let $f, g \in BV([a, b])$. Not that both f and g are bounded on $[a, b]$ since for each $x \in [a, b]$ we have

$$
|f(x)| \leq |f(x) - f(a)| + |f(a)| \leq V_a^b(f) + |f(a)| < \infty
$$

and similarly for g. Thus

$$
\sup_{x \in [a,b]} |f(x)| \leq V_a^b(f) + |f(a)| < \infty \quad \text{and} \quad \sup_{x \in [a,b]} |g(x)| \leq V_a^b(g) + |g(a)| < \infty.
$$

Let $P = \{a = x_0 < x_1 < \cdots < x_N = b\}$ be a partition of $[a, b]$. Then we have

$$
\begin{aligned}
V(fg, P) &= \sum_{k=1}^{N} |(fg)(x_k) - (fg)(x_{k-1})| \\
&= \sum_{k=1}^{N} |f(x_k)g(x_k) - f(x_{k-1})g(x_{k-1})| \\
&= \sum_{k=1}^{N} |f(x_k)g(x_k) - f(x_k)g(x_{k-1}) + f(x_k)g(x_{k-1}) - f(x_{k-1})g(x_{k-1})| \\
&\leq \sum_{k=1}^{N} |f(x_k)||g(x_k) - g(x_{k-1})| + \sum_{k=1}^{N} |g(x_{k-1})||f(x_k) - f(x_{k-1})| \qquad \text{by the triangle inequality} \\
&\leq \left(\sup_{x \in [a,b]} |f(x)| \right) V(g, P) + \left(\sup_{x \in [a,b]} |g(x)| \right) V(f, P).
\end{aligned}
$$

Taking the supremum over all partitions P of $[a, b]$ gives

$$V_a^b(fg) \leqslant \left(\sup_{x \in [a,b]} |f(x)| \right) V_a^b(g) + \left(\sup_{x \in [a,b]} |g(x)| \right) V_a^b(f) < \infty,$$

so fg is in $BV([a, b])$.

This shows that $BV([a, b])$ is closed under products.   □

Proposition 4.4.8 (Additivity of Total Variation). *Let* $f : [a, b] \to \mathbb{R}$ *be a function and let* $c \in (a, b)$. *Then*

$$V_a^b(f) = V_a^c(f) + V_c^b(f).$$

Proof. Step 1: We show that if P is a partition of $[a, b]$ and P' is a refinement of P obtained by adding the point c to P, then

$$V(f, P) \leqslant V(f, P').$$

Let $P = \{a = x_0 < x_1 < \cdots < x_N = b\}$ be a partition of $[a, b]$, and suppose that $x_{j-1} < c < x_j$ for some $j \in \{1, 2, \ldots, N\}$. Then the refinement P' of P obtained by adding the point c is given by the points

$$a = x_0 < x_1 < \cdots < x_{j-1} < c < x_j < \cdots < x_N = b$$

and we have

$$V(f, P') = \sum_{k=1}^{j-1} |f(x_k) - f(x_{k-1})| + |f(c) - f(x_{j-1})| + |f(x_j) - f(c)| + \sum_{k=j+1}^{N} |f(x_k) - f(x_{k-1})|$$

$$\geqslant \sum_{k=1}^{j-1} |f(x_k) - f(x_{k-1})| + |f(x_j) - f(x_{j-1})| + \sum_{k=j+1}^{N} |f(x_k) - f(x_{k-1})|$$

$$= V(f, P)$$

by the triangle inequality.

Thus adding a point to a partition can only increase the total variation with respect to that partition.

Step 2: Let P_1 be a partition of $[a, c]$ and let P_2 be a partition of $[c, b]$. Then the union $P = P_1 \cup P_2$ is a partition of $[a, b]$, and we have

$$V(f, P) = V(f, P_1) + V(f, P_2).$$

Taking the supremum over all partitions P_1 of $[a, c]$ and all partitions P_2 of $[c, b]$ gives

$$V_a^b(f) \geqslant V_a^c(f) + V_c^b(f).$$

To see the reverse inequality, let P be a partition of $[a, b]$. Whether or not c is in P, we can form a refinement $P' := P \cup \{c\}$, and by Step 1 we have

$$V(f, P) \leqslant V(f, P') = V(f, P_1) + V(f, P_2) \leqslant V_a^c(f) + V_c^b(f)$$

where $P_1 := P' \cap [a, c]$ is a partition of $[a, c]$ and $P_2 := P' \cap [c, b]$ is a partition of $[c, b]$. Taking the supremum over all partitions P of $[a, b]$ gives

$$V_a^b(f) \leqslant V_a^c(f) + V_c^b(f).$$

Combining the two inequalities gives the desired result.   □

Corollary 4.4.9 (Total Variation Function is Increasing). *Let $f : [a, b] \to \mathbb{R}$ be a function of bounded variation. Then the function*

$$[a, b] \ni x \longmapsto V_a^x(f) \in [0, V_a^b(f)]$$

is increasing.

Proof. First of all see that by definition, $V_a^x(f) \geqslant 0$ for all $x \in [a, b]$. Thus the additivity of total variation (Proposition 4.4.8) implies that $V_a^x(f) \leqslant V_a^b(f)$ for all $x \in [a, b]$.

Now let $x, y \in [a, b]$ be such that $x < y$. Then by the additivity of total variation, we have

$$V_a^y(f) - V_a^x(f) = V_x^y(f) \geqslant 0.$$

Thus $V_a^x(f) \leqslant V_a^y(f)$, which shows that the function $x \mapsto V_a^x(f)$ is increasing.   □

Theorem 4.4.10 (Jordan Decomposition). *Let $f : [a, b] \to \mathbb{R}$ be a function. Then f is of bounded variation on $[a, b]$ if and only if there exist increasing functions $f_1, f_2 : [a, b] \to \mathbb{R}$ such that*

$$f = f_1 - f_2.$$

In this case, we can take

$$f_1(x) := f(x) + V_a^x(f) \quad and \quad f_2(x) := V_a^x(f)$$

which are both increasing functions.

Proof. ($\Rightarrow$) Suppose that f is of bounded variation on $[a, b]$. Define the functions $f_1, f_2 : [a, b] \to \mathbb{R}$ by

$$f_1(x) := f(x) + V_a^x(f) \quad and \quad f_2(x) := V_a^x(f) \qquad \forall x \in [a, b].$$

By Corollary 4.4.9, we know that f_2 is increasing.

Now we show that f_1 is increasing. See that for each $x, y \in [a, b]$ such that $x < y$, we have

$$f(x) - f(y) \leqslant |f(x) - f(y)| = V(f|_{[x,y]}, \{x, y\}) \leqslant V_x^y(f) = V_a^y(f) - V_a^x(f)$$

which implies that

$$V_a^x(f) + f(x) \leqslant V_a^y(f) + f(y).$$

That is, $a \leqslant x < y \leqslant b$ implies that $f_1(x) \leqslant f_1(y)$, so f_1 is increasing.

For each $x \in [a, b]$ we trivially have $f(x) = f_1(x) - f_2(x)$.

($\Longleftarrow$) Suppose that there exist increasing functions $f_1, f_2 : [a, b] \to \mathbb{R}$ such that $f = f_1 - f_2$. Then by Example 4.4.6, we see that f is of bounded variation on $[a, b]$. $\qquad\square$

Corollary 4.4.11 (BV Functions are Differentiable a.e.)**.** *Let* $f : [a, b] \to \mathbb{R}$ *be a function of bounded variation. Then the derivative* f' *exists and is finite almost everywhere on* (a, b), *and* f' *is integrable.*

Proof. Since f is of bounded variation on $[a, b]$, by the Jordan decomposition (Theorem 4.4.10) there exist increasing functions $f_1, f_2 : [a, b] \to \mathbb{R}$ such that $f = f_1 - f_2$. By Theorem **??**, both f_1 and f_2 are differentiable almost everywhere on (a, b), and their derivatives f_1' and f_2' are integrable. Thus $f' = f_1' - f_2'$ exists and is finite almost everywhere on (a, b), and $f' = f_1' - f_2'$ is integrable. $\qquad\square$

This fact is pretty shocking — much more so than the fact that monotone functions are differentiable almost everywhere. In particular, since Lipschitz functions are of bounded variation (Example 4.4.5), we see that Lipschitz functions are differentiable almost everywhere, which is a very strong regularity property.

Exercise 4.4.2 (Rademacher's Theorem in One Dimension)**.** Let $I \subseteq \mathbb{R}$ be an interval, and let $f : I \to \mathbb{R}$ be a Lipschitz function. Then f is differentiable almost everywhere on I.

Proof. Let $I \subseteq \mathbb{R}$ be an interval, and let $f : I \to \mathbb{R}$ be a Lipschitz function. If I is closed and bounded, then by Example 4.4.5 and Corollary 4.4.11, we know that f is differentiable almost everywhere on I.

Now suppose that I is not closed and bounded, so is one of the forms

$$(-\infty, b), (a, \infty), (-\infty, b], [a, \infty), (a, b), [a, b), (a, b], \text{ or } \mathbb{R}.$$

Then we can write I as a countable union of closed, bounded intervals $\{[a_k, b_k]\}_{k=1}^{\infty}$ such that $[a_k, b_k] \subseteq I$ for all $k \in \mathbb{N}$. For each $k \in \mathbb{N}$, the restriction $f|_{[a_k, b_k]} : [a_k, b_k] \to \mathbb{R}$ is Lipschitz, so by Example 4.4.5 and Corollary 4.4.11, we know that $f|_{[a_k, b_k]}$ is differentiable almost everywhere on $[a_k, b_k]$.

Thus for each $k \in \mathbb{N}$, there exists a set $N_k \subseteq [a_k, b_k]$ with measure zero such that $f|_{[a_k,b_k]}$ is differentiable at every point in $[a_k, b_k] \backslash N_k$. Let

$$N := \bigcup_{k=1}^{\infty} N_k.$$

Then N has measure zero since it is a countable union of measure zero sets. Also if $x \in I \backslash N$, then there exists some $k \in \mathbb{N}$ such that $x \in [a_k, b_k]$, and since $x \notin N$, we have $x \notin N_k$, so $f|_{[a_k,b_k]}$ is differentiable at x. Thus f is differentiable at every point in $I \backslash N$, which shows that f is differentiable almost everywhere on I. $\qquad \square$

Exercise 4.4.3 (BV Functions on $\mathbb{R}$ are Differentiable a.e.)**.** Every function of bounded variatin on $\mathbb{R}$ is differentiable almost everywhere.

Proof. Let $f : \mathbb{R} \to \mathbb{R}$ be a function of bounded variation on $\mathbb{R}$. Then by Remark 4.4.3, we know that f is of bounded variation on every closed, bounded interval $[a, b] \subseteq \mathbb{R}$. Thus by Corollary 4.4.11, we know that f is differentiable almost everywhere on every closed, bounded interval $[a, b] \subseteq \mathbb{R}$.

Now consider the collection $\{[-k, k] : k \in \mathbb{Z}^+\}$ which is a collection of closed, bounded intervals that cover $\mathbb{R}$. For each $k \in \mathbb{Z}^+$, there exists a set $N_k \subseteq [-k, k]$ with measure zero such that f is differentiable at every point in $[-k, k] \backslash N_k$. Let

$$N := \bigcup_{k=1}^{\infty} N_k.$$

Then N has measure zero since it is a countable union of measure zero sets. Also if $x \in \mathbb{R} \backslash N$, then there exists some $k \in \mathbb{Z}^+$ such that $x \in [-k, k]$, and since $x \notin N$, we have $x \notin N_k$, so f is differentiable at x. Thus f is differentiable at every point in $\mathbb{R} \backslash N$, which shows that f is differentiable almost everywhere on $\mathbb{R}$. $\qquad \square$

4.5   Fundamental Theorem of Calculus on $\mathbb{R}$

In this section, we prove both versions of the Fundamental Theorem of Calculus for Lebesgue integrals on the real line. We are not striving for maximum generalizty here, because that will come later.

In short section, we always let $[a, b]$ denoted a closed, bounded interval in $\mathbb{R}$ with $a < b$.

4.5.1   Absolute Continuity

Definition 4.5.1 (Absolute Continuity)**.** A function $f : [a, b] \to \mathbb{R}$ is said to be *absolutely continuous* on $[a, b]$ if for each $\epsilon > 0$, there exists a $\delta > 0$ such that for any finite collection

of pairwise disjoint sub-intervals $\{(a_k, b_k)\}_{k=1}^{N}$ of $[a, b]$ with

$$\sum_{k=1}^{N} (b_k - a_k) < \delta,$$

we have

$$\sum_{k=1}^{N} |f(b_k) - f(a_k)| < \epsilon.$$

We define the set of all absolutely continuous functions on $[a, b]$ to be $AC([a, b])$.

Remark 4.5.2 (Absolute Continuity Implies Uniform Continuity). If f is absolutely continuous on $[a, b]$, then it is uniformly continuous on $[a, b]$.

Let $\epsilon > 0$, and let $\delta > 0$ be such that for any finite collection of pairwise disjoint sub-intervals $\{(a_k, b_k)\}_{k=1}^{N}$ of $[a, b]$ with

$$\sum_{k=1}^{N} (b_k - a_k) < \delta,$$

we have

$$\sum_{k=1}^{N} |f(b_k) - f(a_k)| < \epsilon.$$

Then by choosing $N = 1$, we have that for any sub-interval (a_1, b_1) of $[a, b]$ with

$$b_1 - a_1 < \delta,$$

we have

$$|f(b_1) - f(a_1)| < \epsilon.$$

This is precisely the definition of uniform continuity.

Example 4.5.3 (Lipschitz Functions are Absolutely Continuous). If $f : [a, b] \to \mathbb{R}$ is Lipschitz with Lipschitz constant $L > 0$, then for each $\epsilon > 0$, if we let $\delta := \epsilon/L$, then for any finite collection of pairwise disjoint sub-intervals $\{(a_k, b_k)\}_{k=1}^{N}$ of $[a, b]$ with

$$\sum_{k=1}^{N} (b_k - a_k) < \delta,$$

we have

$$\sum_{k=1}^{N} |f(b_k) - f(a_k)| \leqslant \sum_{k=1}^{N} L|b_k - a_k| = L \sum_{k=1}^{N} (b_k - a_k) < L\delta = \epsilon.$$

Thus f is absolutely continuous on $[a, b]$.

Example 4.5.4 (Not All Absolutely Continuous Functions are Lipschitz)**.** For example, the function $f : [0,1] \to \mathbb{R}$ defined by $f(x) = \sqrt{x}$ is absolutely continuous on $[0,1]$ but not Lipschitz.

   ....

Example 4.5.5 (The Cantor-Lebesgue Function is Not Absolutely Continuous)**.** The Cantor-Lebesgue function $f : [0,1] \to [0,1]$ is continuous and increasing, but not absolutely continuous.

   ....

Exercise 4.5.1 (*AC* is a Vector Space)**.** Show that $AC([a,b])$ is a vector space over $\mathbb{R}$.

Proof. First see that 0, the zero function, is in $AC([a,b])$. Now let $f, g \in AC([a,b])$ and $\alpha \in \mathbb{R}\backslash\{0\}$.

   Let $\epsilon > 0$. Since f and g are absolutely continuous on $[a,b]$, there exist $\delta_1, \delta_2 > 0$ such that for any finite collection of pairwise disjoint sub-intervals $\{(a_k, b_k)\}_{k=1}^{N}$ of $[a,b]$ with

$$\sum_{k=1}^{N}(b_k - a_k) < \delta_1 \quad \text{and} \quad \sum_{k=1}^{N}(b_k - a_k) < \delta_2,$$

we have

$$\sum_{k=1}^{N}|f(b_k) - f(a_k)| < \frac{\epsilon}{2|\alpha|} \quad \text{and} \quad \sum_{k=1}^{N}|g(b_k) - g(a_k)| < \frac{\epsilon}{2}.$$

Then letting $\delta := \min\{\delta_1, \delta_2\}$, we have that for any finite collection of pairwise disjoint sub-intervals $\{(a_k, b_k)\}_{k=1}^{N}$ of $[a,b]$ with

$$\sum_{k=1}^{N}(b_k - a_k) < \delta,$$

we have

$$\sum_{k=1}^{N}|(\alpha f + g)(b_k) - (\alpha f + g)(a_k)| \leqslant \sum_{k=1}^{N}|\alpha||f(b_k) - f(a_k)| + \sum_{k=1}^{N}|g(b_k) - g(a_k)|$$

$$< |\alpha|\frac{\epsilon}{2|\alpha|} + \frac{\epsilon}{2} = \epsilon.$$

Thus $\alpha f + g \in AC([a,b])$. Therefore $AC([a,b])$ is a vector space over $\mathbb{R}$. $\square$

Exercise 4.5.2 (Product of *AC* is *AC*)**.** If $f, g \in AC([a,b])$, then $fg \in AC([a,b])$.

Proof. Let $f, g \in AC([a,b])$. Since f and g are continuous on the compact set $[a,b]$, they are bounded. That is, there exist $M_1, M_2 > 0$ such that

$$|f(x)| \leqslant M_1 \quad \text{and} \quad |g(x)| \leqslant M_2 \quad \forall x \in [a,b].$$

Let $\epsilon > 0$. Since f and g are absolutely continuous on $[a, b]$, there exist $\delta_1, \delta_2 > 0$ such that for any finite collection of pairwise disjoint sub-intervals $\{(a_k, b_k)\}_{k=1}^N$ of $[a, b]$ with

$$\sum_{k=1}^N (b_k - a_k) < \delta_1 \quad \text{and} \quad \sum_{k=1}^N (b_k - a_k) < \delta_2,$$

we have

$$\sum_{k=1}^N |f(b_k) - f(a_k)| < \frac{\epsilon}{2M_2} \quad \text{and} \quad \sum_{k=1}^N |g(b_k) - g(a_k)| < \frac{\epsilon}{2M_1}.$$

Then letting $\delta := \min\{\delta_1, \delta_2\}$, we have that for any finite collection of pairwise disjoint sub-intervals $\{(a_k, b_k)\}_{k=1}^N$ of $[a, b]$ with

$$\sum_{k=1}^N (b_k - a_k) < \delta,$$

we have

$$
\begin{aligned}
\sum_{k=1}^N |(fg)(b_k) - (fg)(a_k)| &= \sum_{k=1}^N |f(b_k)g(b_k) - f(a_k)g(a_k)| \\
&= \sum_{k=1}^N |f(b_k)g(b_k) - f(a_k)g(b_k) + f(a_k)g(b_k) - f(a_k)g(a_k)| \\
&\leqslant \sum_{k=1}^N |g(b_k)||f(b_k) - f(a_k)| + \sum_{k=1}^N |f(a_k)||g(b_k) - g(a_k)| \\
&\leqslant M_2 \sum_{k=1}^N |f(b_k) - f(a_k)| + M_1 \sum_{k=1}^N |g(b_k) - g(a_k)| \\
&< M_2 \frac{\epsilon}{2M_2} + M_1 \frac{\epsilon}{2M_1} = \epsilon.
\end{aligned}
$$

$\square$

Exercise 4.5.3 (Equivalent Definition of *AC* Functions)**.** Let $f : [a, b] \to \mathbb{R}$ be a function. Then f is absolutely continuous on $[a, b]$ if and only if for each $\epsilon > 0$, there exists a $\delta > 0$ such that for any countable collection of pairwise disjoint sub-intervals $\{(a_k, b_k)\}_{k=1}^\infty$ of $[a, b]$ with

$$\sum_{k=1}^\infty (b_k - a_k) < \delta,$$

we have

$$\sum_{k=1}^\infty |f(b_k) - f(a_k)| < \epsilon.$$

Deduce that if f is absolutely continuous on $[a, b]$, then f maps sets of Lebesgue measure zero to sets of Lebesgue measure zero.

Proof. ($\Rightarrow$) Suppose that f is absolutely continuous on $[a, b]$. Let $\epsilon > 0$, and let $\delta > 0$ be such that for any finite collection of pairwise disjoint sub-intervals $\{(a_k, b_k)\}_{k=1}^N$ of $[a, b]$ with

$$\sum_{k=1}^{N} (b_k - a_k) < \delta,$$

we have

$$\sum_{k=1}^{N} |f(b_k) - f(a_k)| < \frac{\epsilon}{2}.$$

Now let $\{(a_k, b_k)\}_{k=1}^{\infty}$ be any countable collection of pairwise disjoint sub-intervals of $[a, b]$ with

$$\sum_{k=1}^{\infty} (b_k - a_k) < \delta.$$

Then for each $N \in \mathbb{Z}^+$, we have

$$\sum_{k=1}^{N} (b_k - a_k) < \delta,$$

so

$$\sum_{k=1}^{N} |f(b_k) - f(a_k)| < \frac{\epsilon}{2}.$$

By taking the limit as $N \to \infty$ we see that

$$\sum_{k=1}^{\infty} |f(b_k) - f(a_k)| \leqslant \frac{\epsilon}{2} < \epsilon.$$

Since the collection $\{(a_k, b_k)\}_{k=1}^{\infty}$ satisfying $\sum_{k=1}^{\infty} (b_k - a_k) < \delta$ was arbitrary, and $\epsilon > 0$ was arbitrary, we have shown that the function f satisfies the condition in the statement of the exercise. Thus the forward direction is proven.

($\Leftarrow$) The reverse direction is trivial since any finite collection of pairwise disjoint sub-intervals is also a countable collection of pairwise disjoint sub-intervals.

Suppose now that f is increasing and absolutely continuous on $[a, b]$, and let $E \subseteq [a, b]$ be a set of Lebesgue measure zero. Let $\epsilon > 0$, and let $\delta > 0$ be such that for any countable collection of pairwise disjoint sub-intervals $\{(a_k, b_k)\}_{k=1}^{\infty}$ of $[a, b]$ with

$$\sum_{k=1}^{\infty} (b_k - a_k) < \delta,$$

we have

$$\sum_{k=1}^{\infty} |f(b_k) - f(a_k)| < \epsilon.$$

Since E has Lebesgue measure zero, there exists a countable collection of pairwise disjoint open intervals $\{(a_k, b_k)\}_{k=1}^{\infty}$ of $[a, b]$ such that

$$E \subseteq \bigcup_{k=1}^{\infty}(a_k, b_k) \quad \text{and} \quad \sum_{k=1}^{\infty}(b_k - a_k) < \delta.$$

Then

$$f(E) \subseteq f\left(\bigcup_{k=1}^{\infty}(a_k, b_k)\right) \subseteq \bigcup_{k=1}^{\infty}[f(a_k), f(b_k)]$$

since f is increasing, so

$$m(f(E)) \leqslant m\left(\bigcup_{k=1}^{\infty}[f(a_k), f(b_k)]\right) \leqslant \sum_{k=1}^{\infty}|f(b_k) - f(a_k)| < \epsilon.$$

Since $\epsilon > 0$ was arbitrary, we have shown that $m(f(E)) = 0$. $\qquad\qquad\square$

Lemma 4.5.6 (*AC* implies *BV*). *If $f : [a, b] \to \mathbb{R}$ is absolutely continuous on $[a, b]$, then it is of bounded variation on $[a, b]$ and its total variation function is absolutely continuous.*

Proof. Let $f : [a, b] \to \mathbb{R}$ be absolutely continuous on $[a, b]$. Then there exists $\delta > 0$ such that for any finite collection of pairwise disjoint sub-intervals $\{(a_k, b_k)\}_{k=1}^{N}$ of $[a, b]$ with

$$\sum_{k=1}^{N}(b_k - a_k) < \delta,$$

we have

$$\sum_{k=1}^{N}|f(b_k) - f(a_k)| < 1.$$

Then let P be a partition

$$a = x_0 < x_1 < \cdots < x_N = b$$

of $[a, b]$ such that $x_k - x_{k-1} < \delta$ for each $k = 1, \ldots, N$. Then

$$V_{x_{k-1}}^{x_k}(f) = \sup\left\{\sum_{j=1}^{M}|f(y_j) - f(y_{j-1})| : P' = \{y_0 < y_1 < \cdots < y_M\} \text{ is a partition of } [x_{k-1}, x_k]\right\} \leqslant 1$$

for each $k = 1, \ldots, N$, so

$$V_a^b(f) = \sum_{k=1}^{N} V_{x_{k-1}}^{x_k}(f) \leqslant N < \infty.$$

Thus f is of bounded variation on $[a, b]$.

Now we will show that the function $x \longmapsto V_a^x(f)$ is absolutely continuous on $[a, b]$. Let $\epsilon > 0$. Since f is absolutely continuous on $[a, b]$, there exists $\delta > 0$ such that for any finite collection of pairwise disjoint sub-intervals $\{(a_k, b_k)\}_{k=1}^N$ of $[a, b]$ with

$$\sum_{k=1}^N (b_k - a_k) < \delta,$$

we have

$$\sum_{k=1}^N |f(b_k) - f(a_k)| < \frac{\epsilon}{2}.$$

Fix a finite collection of pairwise disjoint sub-intervals $\{(c_k, d_k)\}_{k=1}^M$ of $[a, b]$ with

$$\sum_{k=1}^M (d_k - c_k) < \delta,$$

and for each $k = 1, \ldots, M$, let $P_k = \{c_k = y_0 < y_1 < \cdots < y_{N_k} = d_k\}$ be a partition of $[c_k, d_k]$. We have

$$\sum_{k=1}^M V(f|_{[c_k, d_k]}, P_k) < \frac{\epsilon}{2}$$

by choice of δ.

By taking the supremum over all partitions P_k of $[c_k, d_k]$ for each $k = 1, \ldots, M$, we have

$$\sum_{k=1}^M V_{c_k}^{d_k}(f) \leq \frac{\epsilon}{2} < \epsilon.$$

It follows from additivity of the total variation function (see Proposition 4.4.8) that

$$V_{c_k}^{d_k}(f) = V_a^{d_k}(f) - V_a^{c_k}(f) \quad \forall k = 1, \ldots, M.$$

Putting these two facts together, we have

$$\sum_{k=1}^M |V_a^{d_k}(f) - V_a^{c_k}(f)| = \sum_{k=1}^M V_{c_k}^{d_k}(f) < \epsilon.$$

Since the collection $\{(c_k, d_k)\}_{k=1}^M$ satisfying $\sum_{k=1}^M (d_k - c_k) < \delta$ was arbitrary, and $\epsilon > 0$ was arbitrary, we have shown that the function $x \longmapsto V_a^x(f)$ is absolutely continuous on $[a, b]$. □

Corollary 4.5.7 (Jordan Decomposition for *AC* Functions). *If $f : [a, b] \to \mathbb{R}$ is absolutely continuous on $[a, b]$, then there exist increasing, absolutely continuous functions $f_1, f_2 : [a, b] \to \mathbb{R}$ such that*

$$f = f_1 - f_2.$$

Proof. From the previous lemma 4.5.6 and the Jordan decomposition theorem (Theorem **??**), it follows that

$$f(x) = (f(x) + V_a^x(f)) - V_a^x(f) \qquad \forall x \in [a, b]$$

expresses f as the difference of two increasing functions on $[a, b]$. The previous lemma also shows that $x \mapsto V_a^x(f)$ is absolutely continuous on $[a, b]$, and in Exercise 4.5.1 we showed that the difference of two absolutely continuous functions is absolutely continuous. Thus f can be expressed as the difference of two increasing, absolutely continuous functions on $[a, b]$. $\qquad\square$

The following Corollary is frequently used.

Corollary 4.5.8 (*AC* Functions are Differentiable Almost Everywhere). *If $f : [a, b] \to \mathbb{R}$ is absolutely continuous on $[a, b]$, then it is differentiable almost everywhere on $[a, b]$.*

Proof. Since f is absolutely continuous on $[a, b]$, it is of bounded variation on $[a, b]$ by Lemma 4.5.6. Thus by Corollary 4.4.11, f is differentiable almost everywhere on $[a, b]$. $\qquad\square$

4.5.2 Fundamental Theorem of Calculus

Lemma 4.5.9 ((Improved) Second Fundamental Theorem of Calculus). *Let $f : [a, b] \to \mathbb{R}$ be an integrable function. Then the indefinite integral*

$$F(x) = \int_a^x f(t) \, \mathrm{d}t, \qquad x \in [a, b],$$

is absolutely continuous on $[a, b]$ and has derivative $F'(x) = f(x)$ for almost every $x \in [a, b]$.

We already know F is continuous on $[a, b]$, and that $F'(x)$ exists and equals $f(x)$ for almost every $x \in [a, b]$ by Theorem 4.3.17. The new conclusion here is that F is absolutely continuous on $[a, b]$.

Proof. Let $\{(a_k, b_k)\}_{k=1}^N$ be an arbitrary finite collection of pairwise disjoint sub-intervals of $[a, b]$. Then we have

$$
\begin{aligned}
\sum_{k=1}^N |F(b_k) - F(a_k)| &= \sum_{k=1}^N \left| \int_a^{b_k} f(t)\,dt - \int_a^{a_k} f(t)\,dt \right| \\
&= \sum_{k=1}^N \left| \int_{a_k}^{b_k} f(t)\,dt \right| \\
&\leqslant \sum_{k=1}^N \int_{a_k}^{b_k} |f(t)|\,dt = \int_{\bigcup_{k=1}^N (a_k, b_k)} |f(t)|\,dt.
\end{aligned}
$$

Let $\epsilon > 0$, and let $\delta > 0$ be such that if $E \subseteq [a,b]$ is a measurable set with $\mathcal{L}^1(E) < \delta$, then

$$\int_E |f(t)|\, dt < \epsilon.$$

(This is possible by 2.1.22 since f is integrable on $[a,b]$.) Then for each finite collection of pairwise disjoint sub-intervals $\{(a_k, b_k)\}_{k=1}^N$ of $[a,b]$ with

$$\sum_{k=1}^N (b_k - a_k) < \delta,$$

we have

$$\sum_{k=1}^N |F(b_k) - F(a_k)| \leqslant \int_{\bigcup_{k=1}^N (a_k, b_k)} |f(t)|\, dt < \epsilon.$$

Thus F is absolutely continuous on $[a,b]$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Lemma 4.5.10 (Zero Derivative Implies Constant). *Let $f : [a,b] \to \mathbb{R}$ be an increasing absolutely continuous function on $[a,b]$. If $f'(x) = 0$ for almost every $x \in [a,b]$, then f is constant on $[a,b]$.*

Proof. Step 1: We assume that f is increasing on $[a,b]$ and that $f'(x) = 0$ for almost every $x \in [a,b]$.

Since f is increasing and uniformly continuous on the compact set $[a,b]$, its range must be the closed interval $[f(a), f(b)]$. We will show that the length of this interval is zero. Let

$$E := \{x \in [a,b] : f'(x) = 0\}$$

so that $\mathcal{L}^1([a,b]\backslash E) = 0$ by assumption. By 4.5.3, the absolute continuity of f implies that the image of $[a,b]\backslash E$ under f has Lebesgue measure zero. That is,

$$\mathcal{L}^1(f([a,b]\backslash E)) = 0.$$

Step 2: We claim that $f(E)$ also has Lebesgue measure zero.

Let $\epsilon > 0$. If $x_0 \in E$ then $f'(x_0) = 0$, so by definition of the derivative, there exists a $\delta > 0$ such that for each $x \in [x_0, x_0 + \delta)$ we have

$$\frac{f(x) - f(x_0)}{x - x_0} < \epsilon$$

which is equivalent to

$$f(x) - f(x_0) < \epsilon(x - x_0)$$

which rearranges to

$$\epsilon x_0 - f(x_0) < \epsilon x - f(x).$$

Thus x_0 is invisible from the right with respect to the continuous function $x \mapsto \epsilon x - f(x)$. Since $x_0 \in E$ was arbitrary, we have shown that every point in E is invisible from the right with respect to the continuous function $x \mapsto \epsilon x - f(x)$.

By the Rising Sun Lemma (Lemma 4.3.11), there is a countable collection of pairwise disjoint open intervals $\{(a_k, b_k)\}_{k=1}^{\infty}$ such that

$$E \subseteq \bigcup_{k=1}^{\infty}(a_k, b_k) \quad \text{and} \quad f(b_k) - \epsilon b_k \leq f(a_k) - \epsilon a_k \quad \forall k \in \mathbb{Z}^+.$$

That is,

$$f(b_k) - f(a_k) \leq \epsilon(b_k - a_k) \quad \forall k \in \mathbb{Z}^+.$$

But then we see that

$$\sum_{k=1}^{\infty}(f(b_k) - f(a_k)) \leq \epsilon \sum_{k=1}^{\infty}(b_k - a_k) \leq \epsilon(b - a).$$

Hence

$$\mathcal{L}^1(f(E)) \leq \mathcal{L}^1\left(f\left(\bigcup_{k=1}^{\infty}(a_k, b_k)\right)\right) \leq \mathcal{L}^1\left(\bigcup_{k=1}^{\infty}f((a_k, b_k))\right) \leq \sum_{k=1}^{\infty}(f(b_k) - f(a_k)) \leq \epsilon(b - a).$$

Since $\epsilon > 0$ was arbitrary, we have shown that $\mathcal{L}^1(f(E)) = 0$.

But now we see that

$$\mathcal{L}^1(f([a, b])) = \mathcal{L}^1(f(E) \cup f([a, b] \backslash E)) \leq \mathcal{L}^1(f(E)) + \mathcal{L}^1(f([a, b] \backslash E)) = 0 + 0 = 0.$$

Since $f([a, b]) = [f(a), f(b)]$ is a closed interval, it must be that $f(a) = f(b)$ and $f(x) = f(a)$ for all $x \in [a, b]$.

[It is instructive to note where we have used the fact that f is increasing in the above argument — first of all, this allows us to remove the absolute value bars in the first inequality in step 2, and this is crucial for all the following computations. It is also worth noting that we used the absolute continuity of f in step 1 deduce that the image of $\{x : f'(x) = 0\}$ has Lebesgue measure zero, and also in step 2 to apply the Rising Sun Lemma (which requires continuity).] $\square$

Theorem 4.5.11 (First Fundamental Theorem of Calculus for AC Functions)**.** *If* $f : [a, b] \to \mathbb{R}$ *is absolutely continuous on* $[a, b]$*, then*

$$\int_a^b f'(t) \, \mathrm{d}t = f(b) - f(a).$$

Note that we already know that f' exists almost everywhere on $[a, b]$ and is integrable, since *AC* functions are *BV* functions (Lemma 4.5.6) and *BV* functions are differentiable almost everywhere (Corollary 4.4.11) and have integrable derivatives.

Proof. Step 1: First we consider an increasing absolutely continuous function $f : [a, b] \to \mathbb{R}$. Then as we remarked above, f' exists almost everywhere on $[a, b]$ and is integrable. We define a function $\Phi : [a, b] \to \mathbb{R}$ by

$$\Phi(x) := f(x) - \int_a^x f'(t) \, \mathrm{d}t, \qquad \forall x \in [a, b].$$

Then becuase f' is integrable on $[a, b]$, the function $x \mapsto \int_a^x f'(t) \, \mathrm{d}t$ is absolutely continuous on $[a, b]$ by Lemma 4.5.9, and hence Φ is absolutely continuous on $[a, b]$ as well since it is the difference of two absolutely continuous functions.

See that Φ is increasing on $[a, b]$ since for all $x, y \in [a, b]$ with $x < y$, we have

$$\Phi(y) - \Phi(x) = f(y) - f(x) - \int_a^y f'(t) \, \mathrm{d}t + \int_a^x f'(t) \, \mathrm{d}t$$
$$= f(y) - f(x) - \int_x^y f'(t) \, \mathrm{d}t$$
$$\geqslant f(y) - f(x) - (f(y) - f(x)) = 0,$$

where we have used Exercise 4.3.4 in the last line. [This is where we use the fact that f is increasing.]

We also see that Φ is differentiable almost everywhere on $[a, b]$ and that

$$\Phi'(x) = f'(x) - f'(x) = 0$$

for almost every $x \in [a, b]$, where we have used the fact that f' exists almost everywhere on $[a, b]$, and that the derivative of the indefinite integral is the integrand almost everywhere (the Second Fundamental Theorem of Calculus 4.3.17).

Therefore by Lemma 4.5.10, Φ is constant on $[a, b]$, so for each $x \in [a, b]$ we have $\Phi(x) = \Phi(a)$, which is equivalent to

$$f(x) - \int_a^x f'(t) \, \mathrm{d}t = f(a) - \int_a^a f'(t) \, \mathrm{d}t = f(a).$$

Rearranging this gives

$$f(x) = f(a) + \int_a^x f'(t) \, \mathrm{d}t \qquad \forall x \in [a, b].$$

In particular, letting $x = b$ gives

$$f(b) - f(a) = \int_a^b f'(t) \, \mathrm{d}t.$$

Step 2: Now let $f : [a,b] \to \mathbb{R}$ be an arbitrary absolutely continuous function on $[a,b]$. By Corollary 4.5.7, there exist increasing absolutely continuous functions $f_1, f_2 : [a,b] \to \mathbb{R}$ such that

$$f = f_1 - f_2.$$

Then by Step 1, we have

$$f_1(x) - f_1(a) = \int_a^x f_1'(t) \, \mathrm{d}t \quad \text{and} \quad f_2(x) - f_2(a) = \int_a^x f_2'(t) \, \mathrm{d}t \qquad \forall x \in [a,b].$$

Subtracting these two equations gives

$$f(x) - f(a) = (f_1(x) - f_2(x)) - (f_1(a) - f_2(a)) = \int_a^x f_1'(t) \, \mathrm{d}t - \int_a^x f_2'(t) \, \mathrm{d}t = \int_a^x (f_1'(t) - f_2'(t)) \, \mathrm{d}t \qquad \forall x \in [a,b]$$

But since $f = f_1 - f_2$, we have $f' = f_1' - f_2'$ almost everywhere on $[a,b]$, so

$$f(x) - f(a) = \int_a^x f'(t) \, \mathrm{d}t \qquad \forall x \in [a,b].$$

In particular, letting $x = b$ gives

$$f(b) - f(a) = \int_a^b f'(t) \, \mathrm{d}t.$$

$\square$

Corollary 4.5.12 (Integration by Parts for *AC* Functions). *If $f, g : [a,b] \to \mathbb{R}$ are absolutely continuous on $[a,b]$, then*

$$\int_a^b f(t) g'(t) \, \mathrm{d}t = f(b) g(b) - f(a) g(a) - \int_a^b f'(t) g(t) \, \mathrm{d}t.$$

Proof. Let $f, g : [a,b] \to \mathbb{R}$ be absolutely continuous on $[a,b]$. By exercise 4.5.2, the product fg is absolutely continuous on $[a,b]$. Then by Theorem 4.5.11, we have

$$(fg)(b) - (fg)(a) = \int_a^b (fg)'(t) \, \mathrm{d}t.$$

But by the product rule for derivatives, we have

$$(fg)'(t) = f'(t) g(t) + f(t) g'(t)$$

for almost every $t \in [a,b]$, so

$$f(b) g(b) - f(a) g(a) = \int_a^b f'(t) g(t) \, \mathrm{d}t + \int_a^b f(t) g'(t) \, \mathrm{d}t.$$

Rearranging this gives the desired result. $\square$

This finishes our generalization of one-variable calculus to the Lebesgue integral.

everyone has a unique background and focus why do you want emph this program

why are you passionate about polcy and why data change in those areas

lean into what makes you unique

HARRIS BUILD YOU R APPLICATION SESSION

WORDS OF ADVICE: MSCAPP is special believe in yourself

Chapter 5

Differentiation II

5.1 Signed and Vector-Valued Measures, Decomposition Theorems

5.1.1 Motivation

In the previous chapter, we studied generalizations of the fundamental theorem of calculus — in particular, we found that

- (First Fundamental Theorem of Calculus) If $f : [a, b] \to \mathbb{R}$ is absolutely continuous, then f is differentiable almost everywhere, $f' \in L^1([a, b])$, and

$$f(x) - f(a) = \int_a^x f'(t) \, \mathrm{d}t \quad \text{for all } x \in [a, b].$$

- (Second Fundamental Theorem of Calculus) If $f \in L^1([a, b])$, then the function F defined by

$$F(x) = \int_a^x f(t) \, \mathrm{d}t$$

is absolutely continuous, differentiable almost everywhere, and $F' = f$ almost everywhere.

In spirit, the Radon-Nikodym Theorem is a generalization of these results.

5.1.2 Signed and Vector-Valued Measures

Definition 5.1.1 (Signed Measure). A *signed measure* on a measurable space $(X, \mathcal{A})$ is a function $\nu : \mathcal{A} \to \mathbb{R}$ such that for each countable collection $\{A_j\}_{j=1}^\infty$ of pairwise disjoint sets

in $\mathcal{A}$, we have

$$\nu\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \nu(A_j).$$

Note the *by definition* a signed measure is finite on the whole space, i.e., $|\nu|(X) < \infty$. That is, we do not allow any signed measures that take on infinite values.

In particular, many measures are *not* signed measures, since they take on infinite values. A measure μ is a signed measure if and only if it is finite on all measurable sets, i.e., $\mu(A) < \infty$ for all $A \in \mathcal{A}$.

The following is the key example of a signed measure.

Example 5.1.2 (Signed Measure from an L^1 Function). Let $(X, \mathcal{A}, \mu)$ be a measure space, and let $f \in L^1(X, \mu)$. Define $\nu : \mathcal{A} \to \mathbb{R}$ by

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

Then ν is a signed measure on $(X, \mathcal{A})$.

We check this. Let $\{A_j\}_{j=1}^{\infty}$ be a countable collection of pairwise disjoint sets in $\mathcal{A}$. Then

$$
\begin{aligned}
\nu\left(\bigcup_{j=1}^{\infty} A_j\right) &= \int_{\bigcup_{j=1}^{\infty} A_j} f \, \mathrm{d}\mu \\
&= \sum_{j=1}^{\infty} \int_{A_j} f \, \mathrm{d}\mu \qquad \text{by countable additivity of the integral} \\
&= \sum_{j=1}^{\infty} \nu(A_j).
\end{aligned}
$$

Thus, ν is a signed measure.

Remark 5.1.3 (Signed Measures as "Charges"). Some authors, particularly the older Russian literature, refer to signed measures as *charges*. The idea is that in the case of an electrical charge distributed over a surface, each region of the surface can have either positive or negative charge, so the charge distribution is naturally modeled by a signed measure. This is a nice intuition to keep in mind.

Lemma 5.1.4 (Absolute Convergence for Disjoint Unions). *Let ν be a signed measure on a measurable space $(X, \mathcal{A})$. Then for each countable collection $\{A_j\}_{j=1}^{\infty}$ of pairwise disjoint sets in $\mathcal{A}$ we have*

$$\sum_{j=1}^{\infty} |\nu(A_j)| < \infty.$$

Proof. First note that $\nu(\varnothing) = 0$ since

$$\nu(\varnothing) = \nu\left(\bigcup_{j=1}^{\infty}\varnothing\right) = \sum_{j=1}^{\infty}\nu(\varnothing)$$

is a countable disjoint union of sets in $\mathcal{A}$, so

$$\nu(\varnothing) = \sum_{j=1}^{\infty}\nu(\varnothing)$$

and this can only hold if $\nu(\varnothing) = 0$.

Now let $\{A_j\}_{j=1}^{\infty}$ be a countable collection of pairwise disjoint sets in $\mathcal{A}$. Then

$$\nu\left(\bigcup_{\{j:\nu(A_j)>0\}} A_j\right) = \sum_{\{j:\nu(A_j)>0\}}\nu(A_j) = \sum_{\{k:\nu(A_k)>0\}}|\nu(A_k)|$$

and

$$-\nu\left(\bigcup_{\{j:\nu(A_j)<0\}} A_j\right) = \sum_{\{j:\nu(A_j)<0\}}-\nu(A_j) = \sum_{\{k:\nu(A_k)<0\}}|\nu(A_k)|.$$

Since $\nu(A) \in \mathbb{R}$ for each $A \in \mathcal{A}$, the right-hand side of both of these sums must be finite. Therefore the total sum

$$\sum_{j=1}^{\infty}|\nu(A_j)| < \infty$$

converges absolutely as desired. $\qquad\square$

Definition 5.1.5 (Complex Measure, Vector Measure). Let $(X, \mathcal{A})$ be a measurable space. A *complex measure* on $(X, \mathcal{A})$ is a function $\nu : \mathcal{A} \to \mathbb{C}$ such that for each countable collection $\{A_j\}_{j=1}^{\infty}$ of pairwise disjoint sets in $\mathcal{A}$, we have

$$\nu\left(\bigcup_{j=1}^{\infty}A_j\right) = \sum_{j=1}^{\infty}\nu(A_j).$$

A *vector measure* on $(X, \mathcal{A})$ taking values in $\mathbb{R}^m$ is a function $\nu : \mathcal{A} \to \mathbb{R}^m$ such that for each countable collection $\{A_j\}_{j=1}^{\infty}$ of pairwise disjoint sets in $\mathcal{A}$, we have

$$\nu\left(\bigcup_{j=1}^{\infty}A_j\right) = \sum_{j=1}^{\infty}\nu(A_j).$$

Technically we could define vector measures taking values in any Banach space, but we will only consider $\mathbb{R}^m$-valued vector measures.

The following exercise shows that vector and complex measures can be understood component-wise as signed measures.

Exercise 5.1.1 (Components of a Vector Measure)**.** Show that a function $\nu : \mathcal{A} \to \mathbb{R}^m$ is a vector measure if and only if each of its component functions $\nu_k : \mathcal{A} \to \mathbb{R}$ defined by

$$\nu_k(A) = \text{the } k\text{-th component of } \nu(A)$$

is a signed measure.

Similarly, show that a function $\nu : \mathcal{A} \to \mathbb{C}$ is a complex measure if and only if its real and imaginary parts are signed measures.

Proof. Fix $m \geqslant 1$ and for each $1 \leqslant k \leqslant m$, let $\pi_k : \mathbb{R}^m \to \mathbb{R}$ be the projection onto the k-th coordinate,

$$\pi_k(x_1, x_2, \ldots, x_m) = x_k.$$

Suppose that $\nu : \mathcal{A} \to \mathbb{R}^m$ is a vector measure and let $\{A_j\}_{j=1}^\infty$ be a countable collection of pairwise disjoint sets in $\mathcal{A}$. Then for each $1 \leqslant k \leqslant m$, we have

$$\nu_k \left(\bigcup_{j=1}^\infty A_j \right) = \pi_k \circ \nu \left(\bigcup_{j=1}^\infty A_j \right) = \pi_k \left(\sum_{j=1}^\infty \nu(A_j) \right) = \sum_{j=1}^\infty \nu_k(A_j).$$

Thus, each component function ν_k is a signed measure.

Conversely, suppose that each component function ν_k is a signed measure. Then for each countable collection $\{A_j\}_{j=1}^\infty$ of pairwise disjoint sets in $\mathcal{A}$, we have

$$
\begin{aligned}
\nu \left(\bigcup_{j=1}^\infty A_j \right) &= \left(\nu_1 \left(\bigcup_{j=1}^\infty A_j \right), \nu_2 \left(\bigcup_{j=1}^\infty A_j \right), \ldots, \nu_m \left(\bigcup_{j=1}^\infty A_j \right) \right) \\
&= \left(\sum_{j=1}^\infty \nu_1(A_j), \sum_{j=1}^\infty \nu_2(A_j), \ldots, \sum_{j=1}^\infty \nu_m(A_j) \right) \\
&= \sum_{j=1}^\infty \left(\nu_1(A_j), \nu_2(A_j), \ldots, \nu_m(A_j) \right) \\
&= \sum_{j=1}^\infty \nu(A_j).
\end{aligned}
$$

Thus, ν is a vector measure as desired.

The complex measure case is similar, using the projections onto the real and imaginary parts. $\qquad\square$

We generalize Example 5.1.2 as follows.

Example 5.1.6 (Vector Measure from a Vector-Valued L^1 Function). Let $f \in L^1(X, \mu; \mathbb{R}^m)$ be a vector-valued L^1 function. That is, for each $1 \leqslant k \leqslant m$, the k-th component function f_k is in $L^1(X, \mu)$. Define $\nu : \mathcal{A} \to \mathbb{R}^m$ by

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

Then ν is a vector measure on $(X, \mathcal{A})$ by Exercise 5.1.1 and Example 5.1.2.

Proposition 5.1.7 (Properties of Vector Measures). *Let $(X, \mathcal{A})$ be a measurable space and let ν be a vector measure or a complex measure on $(X, \mathcal{A})$. Then*

- $\nu(A \backslash B) = \nu(A) - \nu(B)$ *for all $A, B \in \mathcal{A}$ with $B \subseteq A$,*

- $\nu(A \cup B) = \nu(A) + \nu(B) - \nu(A \cap B)$ *for all $A, B \in \mathcal{A}$,*

- *if $\{A_j\}_{j=1}^{\infty}$ is a countable collection of sets in $\mathcal{A}$ with $A_j \subseteq A_{j+1}$ for all $j \geqslant 1$, then*

$$\nu \left(\bigcup_{j=1}^{\infty} A_j \right) = \lim_{j \to \infty} \nu(A_j),$$

- *if $\{A_j\}_{j=1}^{\infty}$ is a countable collection of sets in $\mathcal{A}$ with $A_{j+1} \subseteq A_j$ for all $j \geqslant 1$, then*

$$\nu \left(\bigcap_{j=1}^{\infty} A_j \right) = \lim_{j \to \infty} \nu(A_j).$$

Proof. By Exercise 5.1.1, it suffices to prove these properties for signed measures.

(i) Let $A, B \in \mathcal{A}$ with $B \subseteq A$. Then

$$\nu(A) = \nu(B \cup (A \backslash B)) = \nu(B) + \nu(A \backslash B)$$

since B and $A \backslash B$ are disjoint, so rearranging gives the desired result.

(ii) Let $A, B \in \mathcal{A}$. Then

$$\nu(A \cup B) = \nu(A \backslash B) + \nu(B)$$
$$= \nu(A) - \nu(A \cap B) + \nu(B)$$

by part (i), giving the desired result.

The proofs of (iii) and (iv) are identical to as in 1.1.6. Note that the assumptions that $\nu(A_1) < \infty$ or $\nu(A_1) > -\infty$ are not needed here since ν is a signed measure and thus finite on all measurable sets. $\quad\square$

5.1.3   Hahn and Jordan Decompositions, Total Variation

In the case of an electrical charge distributed on a surface, we can divide the surface into regions of positive and negative charge. This idea is formalized in the following decomposition theorems for signed measures.

Definition 5.1.8 (Positive and Negative Sets). Let $(X, \mathcal{A})$ be a measurable space and let ν be a signed measure on $(X, \mathcal{A})$. A set $A \in \mathcal{A}$ is called *positive* (with respect to ν) if for every measurable subset $E \subseteq A$, we have $\nu(E) \geqslant 0$. Similarly, a set $A \in \mathcal{A}$ is called *negative* if for every measurable subset $E \subseteq A$, we have $\nu(E) \leqslant 0$.

Theorem 5.1.9 (Hahn Decomposition). *Let $(X, \mathcal{A})$ be a measurable space and let ν be a signed measure on $(X, \mathcal{A})$. Then there exists sets $A^+, A^- \in \mathcal{A}$ such that*

(a) $A^+ \cup A^- = X$ and $A^+ \cap A^- = \varnothing$,

(b) A^+ is positive and A^- is negative (with respect to ν).

Proof. As we remarked after the definition of signed measure, ν is finite on all measurable sets. Therefore we let

$$a := \inf\{\nu(A) : A \in \mathcal{A}, A \text{ is negative }\}$$

and know that $a \in \mathbb{R}$ is well-defined. Let $\{A_j^-\}_{j=1}^\infty$ be a sequence of negative sets such that

$$\lim_{j \to \infty} \nu(A_j^-) = a.$$

Define the set

$$A^- := \bigcup_{j=1}^\infty A_j^-$$

which is measurable since $\mathcal{A}$ is a σ-algebra, and satisfies $\nu(A^-) = a$ by Proposition 5.1.7.(iv). We also define the set

$$A^+ := X \backslash A^-$$

which is measurable since $\mathcal{A}$ is a σ-algebra.

   We see that A^- is negative by construction. It remains to show that A^+ is positive. Suppose for the sake of contradiction that A^+ is not positive. Then there exists a measurable set $E_0 \subseteq A^+$ such that $\nu(E_0) < 0$. But then the set E_0 cannot be negative — otherwise the set $A^\star := A^- \cup E_0$ would be negative with

$$\nu(A^\star) = \nu(A^-) + \nu(E_0) < \nu(A^-) = a$$

contradicting the definition of a.

Therefore there is a smallest positive integer $k_1 \geqslant 1$ such that there exists a measurable set $E_1 \subseteq E_0$ with

$$\nu(E_1) > \frac{1}{k_1}.$$

Clearly we cannot have $E_0 = E_1$ since $\nu(E_0) < 0$. Now by applying the same reasoning to the nonempty set $E_0 \backslash E_1$, we can find a smallest positive integer $k_2 \geqslant 1$ and a measurable set $E_2 \subseteq E_0 \backslash E_1$ such that

$$\nu(E_2) > \frac{1}{k_2}.$$

Continuing in this manner, we obtain a sequence of measurable sets $\{E_j\}_{j=1}^\infty$ that are pairwise disjoint and satisfy

$$\nu(E_j) > \frac{1}{k_j}$$

for some positive integer $k_j \geqslant 1$ for each $j \geqslant 1$. We let

$$F := E_0 \backslash \bigcup_{j=1}^\infty E_j$$

and see that $F \neq \varnothing$ since $\nu(E_0) < 0$ but $\nu(E_j) > 0$ for all $j \geqslant 1$. Also see that F is negative — see that if there were a measurable set $G \subseteq F$ with $\nu(G) > 0$, then G would be a subset of E_0 contradicting the definition of E_1. Hence the set

$$A^\dagger := A^- \cup F$$

is negative with

$$\nu(A^\dagger) = \nu(A^-) + \nu(F) < \nu(A^-) = a$$

contradicting the definition of a. Therefore A^+ is positive as desired. $\qquad\square$

Remark 5.1.10 (Non-uniqueness of the Hahn Decomposition). In general, the Hahn decomposition is not unique. We can always modify a Hahn decomposition (A^+, A^-) by changing A^+ and A^- on a set of ν-measure zero to obtain a different Hahn decomposition. In fact, this is the only way to obtain a different Hahn decomposition as will be shown in the proof of the Jordan decomposition.

Corollary 5.1.11 (Jordan Decomposition). *Let $(X, \mathcal{A})$ be a measurable space and let ν be a signed measure on $(X, \mathcal{A})$. Then there exist unique measures $\nu^+, \nu^- : \mathcal{A} \to [0, \infty)$ such that*

$$\nu = \nu^+ - \nu^-.$$

Maybe this reminds you of the Jordan decomposition 4.4.10 for functions of bounded variation on an interval.

Proof. Step 1: We begin by investigating the uniqueness of the Hahn decomposition.

Assume that

$$X = A_1^+ \cup A_1^- \quad \text{and} \quad X = A_2^+ \cup A_2^-$$

are two Hahn decompositions of X with respect to ν. Then we claim that

$$\nu(E \cap A_1^+) = \nu(E \cap A_2^+) \quad \text{and} \quad \nu(E \cap A_1^-) = \nu(E \cap A_2^-) \qquad (\star)$$

for each set $E \in \mathcal{A}$.

To see this, let $E \in \mathcal{A}$ be arbitrary. Then we have

$$E \cap (A_1^- \backslash A_2^-) \subset E \cap A_1^-$$

and at the same time

$$E \cap (A_1^- \backslash A_2^-) \subset E \cap A_2^+$$

so that

$$\nu(E \cap (A_1^- \backslash A_2^-)) \leqslant 0 \quad \text{and} \quad \nu(E \cap (A_1^- \backslash A_2^-)) \geqslant 0$$

which implies that

$$\nu(E \cap (A_1^- \backslash A_2^-)) = 0.$$

A symmetric argument shows that

$$\nu(E \cap (A_2^- \backslash A_1^-)) = 0.$$

Therefore we have

$$\begin{aligned}
\nu(E \cap A_1^-) &= \nu(E \cap (A_1^- \backslash A_2^-)) + \nu(E \cap (A_1^- \cap A_2^-)) \\
&= 0 + \nu(E \cap (A_2^- \backslash A_1^-)) + \nu(E \cap (A_1^- \cap A_2^-)) \\
&= \nu(E \cap A_2^-)
\end{aligned}$$

as claimed. A similar argument shows that $\nu(E \cap A_1^+) = \nu(E \cap A_2^+)$.

Step 2: Now we prove the existence of the Jordan decomposition.

We define the measures ν^+ and ν^- by

$$\nu^+(A) := \nu(A \cap A^+) \quad \text{and} \quad \nu^-(A) := -\nu(A \cap A^-)$$

for each $A \in \mathcal{A}$, where (A^+, A^-) is a Hahn decomposition of X with respect to ν. It is clear that ν^+ and ν^- are well-defined, and do not depend on the choice of Hahn decomposition by $(\star)$ in Step 1.

Let's check that ν^+ is a measure — the proof for ν^- is similar. Let $\{A_j\}_{j=1}^\infty$ be a countable collection of pairwise disjoint sets in $\mathcal{A}$. Then

$$\nu^+\left(\bigcup_{j=1}^\infty A_j\right) = \nu\left(\left(\bigcup_{j=1}^\infty A_j\right) \cap A^+\right)$$

$$= \nu\left(\bigcup_{j=1}^\infty (A_j \cap A^+)\right)$$

$$= \sum_{j=1}^\infty \nu(A_j \cap A^+)$$

$$= \sum_{j=1}^\infty \nu^+(A_j)$$

as desired.

Finally, we check that $\nu = \nu^+ - \nu^-$. Let $A \in \mathcal{A}$ be arbitrary. Then

$$\nu^+(A) - \nu^-(A) = \nu(A \cap A^+) + \nu(A \cap A^-)$$

$$= \nu(A \cap (A^+ \cup A^-))$$

$$= \nu(A)$$

as desired. $\square$

Definition 5.1.12 (Positive and Negative Variation of a Signed Measure). Let $(X, \mathcal{A})$ be a measurable space and let ν be a signed measure on $(X, \mathcal{A})$. The *positive variation* of ν is the measure ν^+ from the Jordan decomposition, and the *negative variation* of ν is the measure ν^- from the Jordan decomposition.

Exercise 5.1.2 (Positive plus Negative Variation is a Measure). Let $(X, \mathcal{A})$ be a measurable space and let ν be a signed measure on $(X, \mathcal{A})$. Show that $\nu^+ + \nu^-$ is a measure on $(X, \mathcal{A})$.

Proof. Let $\{A_j\}_{j=1}^\infty$ be a countable collection of pairwise disjoint sets in $\mathcal{A}$. Then

$$(\nu^+ + \nu^-)\left(\bigcup_{j=1}^\infty A_j\right) = \nu^+\left(\bigcup_{j=1}^\infty A_j\right) + \nu^-\left(\bigcup_{j=1}^\infty A_j\right)$$

$$= \sum_{j=1}^\infty \nu^+(A_j) + \sum_{j=1}^\infty \nu^-(A_j)$$

$$= \sum_{j=1}^\infty (\nu^+(A_j) + \nu^-(A_j))$$

$$= \sum_{j=1}^\infty (\nu^+ + \nu^-)(A_j)$$

as desired. $\square$

Definition 5.1.13 (Total Variation of a Vector Measure). Let $(X, \mathcal{A})$ be a measurable space and let ν be a vector measure on $(X, \mathcal{A})$ taking values in $\mathbb{R}^m$. The *total variation* of ν is the function $|\nu| : \mathcal{A} \to [0, \infty]$ defined by

$$|\nu|(A) := \sup \left\{ \sum_{j=1}^{n} \|\nu(A_j)\| : \{A_j\}_{j=1}^{n} \text{ is a finite measurable partition of } A \right\}$$

for each $A \in \mathcal{A}$, where $\| \cdot \|$ is the standard Euclidean norm on $\mathbb{R}^m$.

Lemma 5.1.14. *Let $(X, \mathcal{A})$ be a measurable space and let ν be a vector measure on $(X, \mathcal{A})$ taking values in $\mathbb{R}^m$. Then*

(i) $\|\nu(A)\| \leqslant |\nu|(A)$ for all $A \in \mathcal{A}$,

(ii) $|\nu|(A) = \nu(A)$ for all $A \in \mathcal{A}$ if and only if ν is a measure,

(iii) $|\nu|(A) = 0$ if and only if $\nu(C) = 0$ for all $C \subseteq A$, $C \in \mathcal{A}$

(iv) if ν is a signed measure, then

$$|\nu|(A) = \sup\{|\nu(C_1)| + |\nu(C_2)| : C_1, C_2 \text{ are disjoint measurable subsets of } A\}.$$

Proof.   (i) Let $A \in \mathcal{A}$ be arbitrary. Then the partition $\{A\}$ is a finite measurable partition of A, so

$$|\nu|(A) = \sup \left\{ \sum_{j=1}^{n} \|\nu(A_j)\| : \{A_j\}_{j=1}^{n} \text{ is a finite measurable partition of } A \right\} \geqslant \|\nu(A)\|.$$

(ii) Assume that ν is a measure. Then for each finite measurable partition $\{A_j\}_{j=1}^{n}$ of A, we have

$$\sum_{j=1}^{n} \|\nu(A_j)\| = \sum_{j=1}^{n} \nu(A_j) = \nu(A)$$

so taking the supremum over all such partitions gives $|\nu|(A) = \nu(A)$.

Conversely, assume that $|\nu|(A) = \nu(A)$ for all $A \in \mathcal{A}$. Let $\{A_j\}_{j=1}^{\infty}$ be a countable collection of pairwise disjoint sets in $\mathcal{A}$. Then for each $n \geqslant 1$, we have

$$\sum_{j=1}^{n} \nu(A_j) = \nu \left(\bigcup_{j=1}^{n} A_j \right) = |\nu| \left(\bigcup_{j=1}^{n} A_j \right) \leqslant |\nu| \left(\bigcup_{j=1}^{\infty} A_j \right) = \nu \left(\bigcup_{j=1}^{\infty} A_j \right).$$

Taking the limit as $n \to \infty$ gives

$$\sum_{j=1}^{\infty} \nu(A_j) \leqslant \nu\left(\bigcup_{j=1}^{\infty} A_j\right).$$

The reverse inequality follows from Proposition 5.1.7.(ii), so ν is a measure.

(iii) Assume that $|\nu|(A) = 0$. Then for each measurable subset $C \subseteq A$, we have

$$\|\nu(C)\| \leqslant |\nu|(C) \leqslant |\nu|(A) = 0$$

so that $\nu(C) = 0$.

Conversely, assume that $\nu(C) = 0$ for all measurable subsets $C \subseteq A$. Then for each finite measurable partition $\{A_j\}_{j=1}^{n}$ of A, we have

$$\sum_{j=1}^{n} \|\nu(A_j)\| = 0$$

so taking the supremum over all such partitions gives $|\nu|(A) = 0$.

(iv) Assume that ν is a signed measure. Let $n \in \mathbb{Z}^+$ and let $A_1, A_2, \ldots, A_n \in \mathcal{A}$ be disjoint subsets of A. Let

$$C_1 := \bigcup_{\{k \,:\, \nu(A_k) > 0\}} A_k \quad \text{and} \quad C_2 := \bigcup_{\{k \,:\, \nu(A_k) < 0\}} A_k.$$

Then C_1 and C_2 are disjoint measurable subsets of A with

$$|\nu(C_1)| + |\nu(C_2)| = \sum_{k=1}^{n} |\nu(A_k)|.$$

Taking the supremum over all such collections $\{A_j\}_{j=1}^{n}$ gives

$$|\nu|(A) \leqslant \sup\left\{|\nu(C_1)| + |\nu(C_2)| : C_1, C_2 \text{ are disjoint measurable subsets of } A\right\}.$$

The reverse inequality follows trivially. Therefore we have the desired equality. $\qquad\square$

We return to our key example.

Proposition 5.1.15 (Total Variation of Vector Measure from L^1 Function). *Let $(X, \mathcal{A}, \mu)$ be a measure space. Let $f \in L^1(X, \mu; \mathbb{R}^m)$ be a vector-valued L^1 function, and let ν be the vector measure defined by*

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

Then the total variation of ν is given by

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

Proof. Let $A \in \mathcal{A}$ be arbitrary. First, let $A_1, A_2, \ldots, A_n$ be a disjoint collection of measurable subsets of A. Then by the triangle inequality, we have

$$\sum_{k=1}^{n} \|\nu(A_k)\| = \sum_{k=1}^{n} \left\| \int_{A_k} f \, \mathrm{d}\mu \right\| \leqslant \sum_{k=1}^{n} \int_{A_k} \|f\| \, \mathrm{d}\mu = \int_{A} \|f\| \, \mathrm{d}\mu.$$

Hence taking the supremum over all such collections $\{A_k\}_{k=1}^{n}$ gives

$$|\nu|(A) \leqslant \int_{A} \|f\| \, \mathrm{d}\mu.$$

To prove the reverse inequality, suppose first that ν is a signed measure (i.e., $m = 1$). See that $\{x \in A : f(x) > 0\}$ and $\{x \in A : f(x) < 0\}$ are disjoint measurable subsets of A and that

$$|\nu \left(\{x \in A : f(x) > 0\}\right)| + |\nu \left(\{x \in A : f(x) < 0\}\right)| = \int_{\{x : f(x) > 0\}} |f| \, \mathrm{d}\mu + \int_{\{x : f(x) < 0\}} |f| \, \mathrm{d}\mu = \int_{A} |f| \, \mathrm{d}\mu.$$

Therefore we have

$$|\nu|(A) \geqslant \int_{A} |f| \, \mathrm{d}\mu$$

which gives the desired result in this case by Lemma 5.1.14.(iv).

Now suppose that ν is a vector measure taking values in $\mathbb{R}^m$ for some $m \geqslant 1$. Let $\epsilon > 0$ be arbitrary. Since simple functions are dense in $L^1(X, \mu; \mathbb{R}^m)$ (2.3.5), we can find a simple function $s : X \to \mathbb{R}^m$ such That

$$\|f - s\|_{L^1(X, \mu; \mathbb{R}^m)} < \frac{\epsilon}{2}.$$

Write $s|_A$ in the form

$$s|_A = \sum_{j=1}^{N} c_j \chi_{E_j}$$

where $E_1, E_2, \ldots, E_N$ are disjoint measurable subsets of A and $c_1, c_2, \ldots, c_N \in \mathbb{R}^m$. Then we

have

$$\sum_{j=1}^{N} \|\nu(E_j)\| = \sum_{j=1}^{N} \left\| \int_{E_j} f \, d\mu \right\|$$

$$\geqslant \sum_{j=1}^{N} \left\| \int_{E_j} s \, d\mu \right\| - \sum_{j=1}^{N} \left\| \int_{E_j} (f - s) \, d\mu \right\|$$

$$= \sum_{j=1}^{N} \|c_j\| \mu(E_j) - \sum_{j=1}^{N} \left\| \int_{E_j} (f - s) \, d\mu \right\|$$

$$= \int_A \|s\| \, d\mu - \left\| \int_A (f - s) \, d\mu \right\|$$

$$\geqslant \int_A \|s\| \, d\mu - \int_A \|f - s\| \, d\mu$$

$$\geqslant \int_A \|s\| \, d\mu - \frac{\epsilon}{2}$$

$$\geqslant \int_A \|f\| \, d\mu - \int_A \|f - s\| \, d\mu - \frac{\epsilon}{2}$$

$$\geqslant \int_A \|f\| \, d\mu - \epsilon$$

which implies that

$$|\nu|(A) \geqslant \int_A \|f\| \, d\mu - \epsilon.$$

Since $\epsilon > 0$ was arbitrary, we obtain the desired result. $\qquad\square$

Lemma 5.1.16 (Total Variation is a Measure). *Let $(X, \mathcal{A})$ be a measurable space and let ν be a vector measure on $(X, \mathcal{A})$ taking values in $\mathbb{R}^m$. Then the total variation $|\nu|$ is a measure on $(X, \mathcal{A})$.*

Proof. First note that $|\nu|(\varnothing) = 0$ since the only partition of $\varnothing$ is the empty partition and $\nu(\varnothing) = 0$. Now let $\{A_j\}_{j=1}^{\infty}$ be a countable collection of disjoint sets in $\mathcal{A}$. Fix $m \in \mathbb{Z}^+$. Then for each $1 \leqslant j \leqslant m$, let $A_{j,1}, A_{j,2}, \ldots, A_{j,n_j}$ be a finite disjoint collection of measurable subsets of A_j. Then the collection

$$\{A_{j,k} : 1 \leqslant j \leqslant m, 1 \leqslant k \leqslant n_j\}$$

is a finite disjoint collection of measurable subsets of $\bigcup_{j=1}^{\infty} A_j$, so we have

$$\sum_{j=1}^{m} \sum_{k=1}^{n_j} |\nu(A_{j,k})| \leqslant |\nu| \left(\bigcup_{j=1}^{\infty} A_j \right).$$

Taking the supremum over all such collections $\{A_{j,k}\}_{k=1}^{n_j}$ gives

$$\sum_{j=1}^{m} |\nu|(A_j) \leqslant |\nu|\left(\bigcup_{j=1}^{\infty} A_j\right).$$

Finally, taking the limit as $m \to \infty$ gives

$$\sum_{j=1}^{\infty} |\nu|(A_j) \leqslant |\nu|\left(\bigcup_{j=1}^{\infty} A_j\right).$$

To prove the reverse inequality, let $E_1, E_2, \ldots, E_n$ be a finite disjoint collection of measurable subsets of $\bigcup_{j=1}^{\infty} A_j$. Then

$$\begin{aligned}
\sum_{j=1}^{\infty} |\nu|(A_j) &\geqslant \sum_{j=1}^{\infty} \sum_{k=1}^{n} |\nu(E_k \cap A_j)| \\
&= \sum_{k=1}^{n} \sum_{j=1}^{\infty} |\nu(E_k \cap A_j)| \\
&\geqslant \sum_{k=1}^{n} \left|\sum_{j=1}^{\infty} \nu(E_k \cap A_j)\right| \\
&= \sum_{k=1}^{n} |\nu(E_k)|
\end{aligned}$$

where we have used the definition of $|\nu|(A_j)$ in the first inequality, the triangle inequality in the second inequality, and the countable additivity of ν in the last equality. Taking the supremum over all such collections $\{E_k\}_{k=1}^{n}$ gives

$$\sum_{j=1}^{\infty} |\nu|(A_j) \geqslant |\nu|\left(\bigcup_{j=1}^{\infty} A_j\right).$$

Therefore $|\nu|$ is a measure on $(X, \mathcal{A})$ as desired. $\square$

Exercise 5.1.3 (Total Variation of Signed Measure is Positive plus Negative Variation). Let $(X, \mathcal{A})$ be a measurable space and let ν be a signed measure on $(X, \mathcal{A})$. Show that the total variation of ν is given by

$$|\nu| = \nu^+ + \nu^-$$

where ν^+ and ν^- are the positive and negative variations of ν.

Proof. Recall from Lemma 5.1.14.(iv) that

$$|\nu|(A) = \sup\{|\nu(C_1)| + |\nu(C_2)| : C_1, C_2 \text{ are disjoint measurable subsets of } A\}$$

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

Let

$$X = A^+ \cup A^-$$

be a Hahn decomposition of X with respect to ν. Then by definition of ν^+ and ν^-, we have

$$\nu^+(A) = \nu(A \cap A^+) \quad \text{and} \quad \nu^-(A) = -\nu(A \cap A^-)$$

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

Let $A \in \mathcal{A}$ be arbitrary. Then we see that

$$\nu^+(A) + \nu^-(A) = \nu(A \cap A^+) - \nu(A \cap A^-) = |\nu(A \cap A^+)| + |\nu(A \cap A^-)| \leqslant |\nu|(A)$$

since $A \cap A^+$ and $A \cap A^-$ are disjoint measurable subsets of A. To prove the reverse inequality, let $\epsilon > 0$ be arbitrary. Then by the definition of $|\nu|(A)$, there exist disjoint measurable subsets $C_1, C_2 \subseteq A$ such that

$$|\nu|(A) < |\nu(C_1)| + |\nu(C_2)| + \epsilon.$$

Now we have

$$|\nu(C_1)| \leqslant |\nu(C_1 \cap A^+) + \nu(C \cap A^-)|$$
$$\leqslant \nu(C_1 \cap A^+) - \nu(C_1 \cap A^-)$$

and similarly

$$|\nu(C_2)| \leqslant \nu(C_2 \cap A^+) - \nu(C_2 \cap A^-).$$

Therefore we have

$$|\nu|(A) < |\nu(C_1)| + |\nu(C_2)| + \epsilon$$
$$\leqslant \nu(C_1 \cap A^+) - \nu(C_1 \cap A^-) + \nu(C_2 \cap A^+) - \nu(C_2 \cap A^-) + \epsilon$$
$$= \nu((C_1 \cup C_2) \cap A^+) - \nu((C_1 \cup C_2) \cap A^-) + \epsilon$$
$$\leqslant \nu(A \cap A^+) - \nu(A \cap A^-) + \epsilon$$
$$= \nu^+(A) + \nu^-(A) + \epsilon$$

by disjoint additivity of ν and since $C_1 \cup C_2 \subseteq A$. Since $\epsilon > 0$ was arbitrary, obtain

$$|\nu|(A) \leqslant \nu^+(A) + \nu^-(A).$$

Therefore we have shown that

$$|\nu|(A) = \nu^+(A) + \nu^-(A)$$

for each $A \in \mathcal{A}$ as desired. $\qquad \square$

5.1.4   Absolute Continuity and Lebesgue Decomposition Theorem

Definition 5.1.17 (Absolutely Continuous). Let $(X, \mathcal{A}, \mu)$ be a measure space, and let ν be a vector measure on $(X, \mathcal{A})$ taking values in $\mathbb{R}^m$. We say that ν is *absolutely continuous* with respect to μ, denoted $\nu \ll \mu$, if for each $A \in \mathcal{A}$ with $\mu(A) = 0$, we have $\nu(A) = 0$. That is,

$$A \in \mathcal{A} \quad \text{and} \quad \mu(A) = 0 \implies \nu(A) = 0.$$

Similarly, if ν is a measure on $(X, \mathcal{A})$, we say that ν is absolutely continuous with respect to μ if

$$A \in \mathcal{A} \quad \text{and} \quad \mu(A) = 0 \implies \nu(A) = 0.$$

Notice that we really have to include the second definition for measures since only finite measures are vector measures taking values in $\mathbb{R}$.

Example 5.1.18. Let $(X, \mathcal{A}, \mu)$ be a measure space. Let $f \in L^1(X, \mu; \mathbb{R}^m)$ be a vector-valued L^1 function, and let ν be the vector measure defined by

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

Then ν is absolutely continuous with respect to μ.

To see this, let $A \in \mathcal{A}$ be such that $\mu(A) = 0$. Then we have

$$\|\nu(A)\| \leqslant |\nu|(A) = \int_A \|f\| \, \mathrm{d}\mu = 0$$

by Proposition 5.1.15, so that $\nu(A) = 0$ as desired.

Exercise 5.1.4 (Absolute Continuity of Variations). Let $(X, \mathcal{A})$ be a measurable space.

(i) If ν is a signed measure, then $v^+ \ll |\nu|$ and $v^- \ll |\nu|$.

(ii) If ν is a vector measure taking values in $\mathbb{R}^m$, then each component measure ν_k is absolutely continuous with respect to $|\nu|$ and also $\nu \ll |\nu|$.

Proof.   1. Assume ν is a signed measure. Let $A \in \mathcal{A}$ be such that $|\nu|(A) = 0$. Then by Exercise 5.1.3, we have

$$\nu^+(A) + \nu^-(A) = |\nu|(A) = 0$$

so that $\nu^+(A) = 0$ and $\nu^-(A) = 0$. Since A was an arbitrary measurable set with $|\nu|(A) = 0$, we conclude that $\nu^+ \ll |\nu|$ and $\nu^- \ll |\nu|$ as desired.

2. Assume ν is a vector measure taking values in $\mathbb{R}^m$. Let $A \in \mathcal{A}$ be such that $|\nu|(A) = 0$. Then by Lemma 5.1.14.(i) we have

$$|\nu_k(A)| \leqslant \|\nu(A)\| \leqslant |\nu|(A) = 0$$

for each $1 \leqslant k \leqslant m$, so that $\nu_k(A) = 0$ for each $1 \leqslant k \leqslant m$. Since A was an arbitrary measurable set with $|\nu|(A) = 0$, we conclude that $\nu_k \ll |\nu|$ for each $1 \leqslant k \leqslant m$. Finally, since $\nu(A) = (\nu_1(A), \nu_2(A), \dots, \nu_m(A))$, we also have $\nu(A) = 0$. Therefore we have $\nu \ll |\nu|$ as desired.

$\square$

Definition 5.1.19 (Mutually Singular). Let $(X, \mathcal{A})$ be a measurable space, and let ν and μ be two countably additive set functions on $\mathcal{A}$ of the same type, i.e. both measures, both signed measures, or both vector measures taking values in $\mathbb{R}^m$. We say that ν and μ are *mutually singular*, denoted $\nu \perp \mu$, if there exist disjoint measurable sets $A, B \in \mathcal{A}$ such that

$$X = A \cup B, \quad \text{and} \quad \nu(E) = \nu(E \cap A), \quad \mu(E) = \mu(E \cap B)$$

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

That is, the measure ν only "sees" the set A while the measure μ only "sees" the set B, and the two sets do not overlap.

Exercise 5.1.5 (Positive and Negative Variations of Signed Measure are Mutually Singular). Let $(X, \mathcal{A})$ be a measurable space and let ν be a signed measure. Show that ν^+ and ν^- from the Jordan decomposition are mutually singular.

Proof. By the proof of the Jordan decomposition (Theorem 4.4.10), let

$$X = A^+ \cup A^-$$

be a Hahn decomposition of X with respect to ν. Then by definition of ν^+ and ν^-, we have

$$\nu^+(E) = \nu(E \cap A^+) \quad \text{and} \quad \nu^-(E) = -\nu(E \cap A^-)$$

for each $E \in \mathcal{A}$. Therefore ν^+ only "sees" the set A^+ while ν^- only "sees" the set A^-, and since A^+ and A^- are disjoint, we conclude that ν^+ and ν^- are mutually singular as desired.

$\square$

Absolute continuity and mutual singularity are in a sense "opposites" of each other, as shown by the following lemma.

Lemma 5.1.20 (Absolutely Continuous and Mutually Singular implies Zero Measure). *Let $(X, \mathcal{A}, \mu)$ be a measure space and let ν be a vector measure on $(X, \mathcal{A})$. Then $\nu \ll \mu$ and $\nu \perp \mu$ implies that $\nu = 0$.*

Proof. Assume that $\nu \ll \mu$ and $\nu \perp \mu$. Then by definition of mutual singularity, there exist disjoint measurable sets $A, B \in \mathcal{A}$ such that

$$X = A \cup B, \qquad \nu(E) = \nu(E \cap A), \quad \mu(E) = \mu(E \cap B)$$

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

Let $E \in \mathcal{A}$ be arbitrary. Then

$$\mu(E \cap A) = \mu((E \cap A) \cap B) = \mu(\varnothing) = 0.$$

Since $\nu \ll \mu$, we have $\nu(E \cap A) = 0$. Therefore we have

$$\nu(E) = \nu(E \cap A) = 0.$$

Since E was an arbitrary measurable set, we conclude that $\nu = 0$ as desired. $\square$

Theorem 5.1.21 (Lebesgue Decomposition Theorem). *Let $(X, \mathcal{A}, \mu)$ be a measure space, and let ν be a vector measure on $(X, \mathcal{A})$ taking values in $\mathbb{R}^m$. Then there exist unique vector measures ν_{ac} and ν_s on $(X, \mathcal{A})$ taking values in $\mathbb{R}^m$ such that $\nu = \nu_{ac} + \nu_s$ and*

$$\nu_{ac} \ll \mu \quad and \quad \nu_s \perp \mu.$$

The measures ν_{ac} and ν_s are called the absolutely continuous and singular parts of ν with respect to μ, respectively.

Proof. Let

$$b := \sup\{|\nu|(A) : A \in \mathcal{A}, \mu(A) = 0\}.$$

Then for each $j \in \mathbb{Z}^+$, we can find a measurable set $A_j \in \mathcal{A}$ such that $\mu(A_j) = 0$ and

$$|\nu|(A_j) > b - \frac{1}{j}.$$

Let

$$A := \bigcup_{j=1}^{\infty} A_j.$$

Then $\mu(A) = 0$ since countable unions of μ measure zero sets have μ measure zero. Moreover, by countable subadditivity of $|\nu|$ (Lemma 5.1.16), we have

$$|\nu|(A) \leqslant \sum_{j=1}^{\infty} |\nu|(A_j) \leqslant b.$$

On the other hand, since $A_j \subseteq A$ for each $j \in \mathbb{Z}^+$, we have

$$|\nu|(A) \geqslant |\nu|(A_j) > b - \frac{1}{j}$$

for each $j \in \mathbb{Z}^+$, so that taking the limit as $j \to \infty$ gives $|\nu|(A) \geqslant b$. Therefore we have shown that $|\nu|(A) = b$.

Define vector measures ν_{ac} and ν_{s} on $(X, \mathcal{A})$ by

$$\nu_{\mathrm{ac}}(E) := \nu(E \cap A^c), \quad \nu_{\mathrm{s}}(E) := \nu(E \cap A)$$

for each $E \in \mathcal{A}$. Then it is clear that $\nu = \nu_{\mathrm{ac}} + \nu_{\mathrm{s}}$ by disjoint additivity of ν.

If $E \in \mathcal{A}$, then

$$\mu(E) = \mu(E \cap A) + \mu(E \cap A^c) = 0 + \mu(E \cap A^c) = \mu(E \cap A^c)$$

since $\mu(A) = 0$. This shows that $\nu_{\mathrm{s}} \perp \mu$.

To see that $\nu_{\mathrm{ac}} \ll \mu$, let $E \in \mathcal{A}$ be such that $\mu(E) = 0$. Then $\mu(E \cup A) = 0$, so by definition of b we have

$$b \geqslant |\nu|(E \cup A) = |\nu|(E) + |\nu|(A) = |\nu|(E) + b = b$$

which forces $|\nu|(E \backslash A) = 0$. Therefore

$$\nu_{\mathrm{ac}}(E) = \nu(E \cap A^c) = \nu(E \backslash A) = 0.$$

Since E was an arbitrary measurable set with $\mu(E) = 0$, we conclude that $\nu_{\mathrm{ac}} \ll \mu$.

Note: The construction of ν_{ac} and ν_{s} shows that if ν is a (finite) measure, then so are ν_{ac} and ν_{s}. That is, if ν never takes on negative values, then neither do ν_{ac} and ν_{s}.

To prove uniqueness, suppose that there exist vector measures $\tilde{\nu}_{\mathrm{ac}}$ and $\tilde{\nu}_{\mathrm{s}}$ on $(X, \mathcal{A})$ taking values in $\mathbb{R}^m$ such that

$$\nu = \tilde{\nu}_{\mathrm{ac}} + \tilde{\nu}_{\mathrm{s}}$$

and

$$\tilde{\nu}_{\mathrm{ac}} \ll \mu \quad \text{and} \quad \tilde{\nu}_{\mathrm{s}} \perp \mu.$$

Then we have

$$\nu_{\mathrm{ac}} + \nu_{\mathrm{s}} = \tilde{\nu}_{\mathrm{ac}} + \tilde{\nu}_{\mathrm{s}}$$

which implies that

$$\nu_{\mathrm{ac}} - \tilde{\nu}_{\mathrm{ac}} = \tilde{\nu}_{\mathrm{s}} - \nu_{\mathrm{s}}.$$

Since the left-hand side is absolutely continuous with respect to μ and the right-hand side is mutually singular with respect to μ, we conclude from Lemma 5.1.20 that

$$\nu_{\mathrm{ac}} - \tilde{\nu}_{\mathrm{ac}} = 0 \quad \text{and} \quad \tilde{\nu}_{\mathrm{s}} - \nu_{\mathrm{s}} = 0.$$

Therefore we have shown that $\nu_{\mathrm{ac}} = \tilde{\nu}_{\mathrm{ac}}$ and $\nu_{\mathrm{s}} = \tilde{\nu}_{\mathrm{s}}$, completing the proof. $\qquad \square$

5.2   The Radon-Nikodym Theorem

The Radon-Nikodym Theorem is a fundamental result in measure theory that generalizes the concept of derivatives to measures.

First we need a lemma.

Lemma 5.2.1. *Let $(X, \mathcal{A}, \mu)$ be a measure space, and let λ be another measure on $(X, \mathcal{A})$ that is absolutely continuous with respect to μ and is not the zero measure. Then there exists a positive number $\epsilon > 0$ and a measurable set $A_0 \in \mathcal{A}$ such that $\mu(A_0) > 0$ and A_0 is positive with respect to the signed measure $\lambda - \epsilon\mu$, i.e.*

$$\lambda(A) - \epsilon\mu(A) \geqslant 0 \quad \text{for all } A \in \mathcal{A} \text{ such that } A \subseteq A_0.$$

Proof. Let $N \in \mathbb{Z}^+$ be arbitrary; consider the signed measure $\lambda - \frac{1}{N}\mu$. Then let

$$X = A_N^+ \cup A_N^-$$

be a Hahn decomposition (5.1.9) for this signed measure.

Let

$$A^+ = \bigcap_{N=1}^{\infty} A_N^+ \quad \text{and} \quad A^- = \bigcap_{N=1}^{\infty} A_N^-.$$

Then for each $N \in \mathbb{Z}^+$ we have

$$\lambda(A^-) \leqslant \frac{1}{N}\mu(A^-)$$

which implies that $\lambda(A^-) = 0$. Hence we must have $\lambda(A^+) > 0$ since λ is not the zero measure. But then absolute continuity of λ with respect to μ implies that $\mu(A^+) > 0$. By **??**, we see that

$$\lim_{N \to \infty} \mu(A_N^+) = \mu(A^+)$$

because the sets A_N^+ are decreasing and $A^+ = \bigcap_{N=1}^{\infty} A_N^+$. Thus there exists some $N \in \mathbb{Z}^+$ such that $\mu(A_N^+) > 0$, and we may take $A_0 = A_N^+$.

That is, for this choice of N, we have

$$\lambda(A) - \frac{1}{N}\mu(A) \geqslant 0 \quad \text{for all } A \in \mathcal{A} \text{ such that } A \subseteq A_0.$$

The proof is complete. $\qquad\qquad\square$

Theorem 5.2.2 (Radon-Nikodym Theorem for Signed Measures)**.** *Let $(X, \mathcal{A}, \mu)$ be a measure space, and let ν be a signed measure on $(X, \mathcal{A})$ that is absolutely continuous with respect to μ. Then there exists a unique $f \in L^1(X, \mu)$ such that*

$$\nu(A) = \int_A f \, \mathrm{d}\mu \quad \text{for all } A \in \mathcal{A}.$$

Thus our primary example is essentially the only example. After the proof, we will discuss how this is a generalization of the fundamental theorem of calculus.

The more popular proof is by John von Neumann using Hilbert spaces, but we will present the more elementary proof found in Kolmogorov and Fomin.

Proof. If ν is the zero measure, then $f = 0$ works, and uniqueness μ-almost everywhere is clear. Thus we may assume that ν is not the zero measure. We let

$$\mathcal{K} := \left\{ f \in L^1(X, \mu) : \int_A f \, \mathrm{d}\mu \leqslant \nu(A) \quad \text{for all} \quad A \in \mathcal{A} \right\}$$

and let

$$M := \sup \left\{ \int_X f \, \mathrm{d}\mu : f \in \mathcal{K} \right\}.$$

Then there exists a sequence of functions $\{f_k\}_{k=1}^\infty \subseteq \mathcal{K}$ such that

$$\lim_{k \to \infty} \int_X f_k \, \mathrm{d}\mu = M.$$

For each $n \in \mathbb{Z}^+$ we define

$$g_n := \max\{f_1, f_2, \ldots, f_n\}.$$

Then we obtain an increasing sequence $\{g_n\}_{n=1}^\infty$ of measurable functions on X and we claim that

$$g_n \in \mathcal{K} \quad \text{for all } n \in \mathbb{Z}^+.$$

To see this, fix $n \in \mathbb{Z}^+$ and let $A \in \mathcal{A}$ be arbitrary. Then there are disjoint sets $A_{n,1}, A_{n,2}, \ldots, A_{n,n} \in \mathcal{A}$ such that

$$A = \bigcup_{k=1}^n A_{n,k} \quad \text{and} \quad g_n(x) = f_k(x) \text{ for all } x \in A_{n,k}.$$

Thus we have

$$\int_A g_n \, \mathrm{d}\mu = \int_{\bigcup_{k=1}^n A_{n,k}} g_n \, \mathrm{d}\mu = \sum_{k=1}^n \int_{A_{n,k}} f_k \, \mathrm{d}\mu$$

$$\leqslant \sum_{k=1}^n \nu(A_{n,k}) = \nu(A)$$

since $f_k \in \mathcal{K}$ for all $k \in \{1, 2, \ldots, n\}$. Since $A \in \mathcal{A}$ was arbitrary, this shows $g_n \in \mathcal{K}$. Since $n \in \mathbb{Z}^+$ was arbitrary, the claim is proved.

Since $\{g_n\}_{n=1}^\infty$ is an increasing sequence of measurable functions in $\mathcal{K}$ we must have

$$\lim_{n \to \infty} \int_X g_n \, \mathrm{d}\mu = M.$$

We check this — by virtue of $\{g_n\}_{n=1}^{\infty} \subseteq \mathcal{K}$, the inequality $\lim_{n\to\infty} \int_X g_n \, \mathrm{d}\mu \leqslant M$ is clear. If we had a strict inequality $\lim_{n\to\infty} \int_X g_n \, \mathrm{d}\mu < M$, then monotonicity of the integral would imply that

$$\lim_{k\to\infty} \int_X f_k \, \mathrm{d}\mu \leqslant \lim_{n\to\infty} \int_X g_n \, \mathrm{d}\mu < M,$$

contradicting our choice of $\{f_k\}_{k=1}^{\infty}$. Thus we have equality as desired.

We now define

$$f(x) := \lim_{n\to\infty} g_n(x) = \sup_{n\in\mathbb{Z}^+} g_n(x)$$

for all $x \in X$. Then f is a measurable function on X, as a limit of measurable functions, and the Bounded Convergence Theorem **??** implies that $f \in L^1(X, \mu)$ and

$$\int_X f \, \mathrm{d}\mu = \int_X \lim_{n\to\infty} g_n \, \mathrm{d}\mu = \lim_{n\to\infty} \int_X g_n \, \mathrm{d}\mu = M < \infty.$$

We now show that f is the required function.

We define a set function $\lambda : \mathcal{A} \to \mathbb{R}$ by

$$\lambda(A) := \nu(A) - \int_A f \, \mathrm{d}\mu \quad \text{for all } A \in \mathcal{A}.$$

Then λ is a signed measure on $(X, \mathcal{A})$. Since we must have $f \in \mathcal{K}$, it follows that λ is a nonnegative measure, i.e. $\lambda(A) \geqslant 0$ for all $A \in \mathcal{A}$. We claim that λ is the zero measure.

Assume towards a contradiction that this is not the case, and there is a set $E \in \mathcal{A}$ for which $\lambda(E) > 0$. Then by Lemma 5.2.1 there exists $\epsilon > 0$ and a measurable set $A_0 \in \mathcal{A}$ such that $\mu(A_0) > 0$ and

$$\lambda(A) - \epsilon\mu(A) \geqslant 0 \quad \text{for all } A \in \mathcal{A} \text{ such that } A \subseteq A_0.$$

That is,

$$\epsilon\mu(A) \leqslant \lambda(A) \quad \text{for all } A \in \mathcal{A} \text{ such that } A \subseteq A_0.$$

We define

$$h(x) := f(x) + \epsilon\chi_{A_0}(x) \qquad \forall x \in X.$$

Then for each $A \in \mathcal{A}$ we have

$$\int_A h \, \mathrm{d}\mu = \int_A f \, \mathrm{d}\mu + \int_A \epsilon\chi_{A_0} \, \mathrm{d}\mu = \int_A f \, \mathrm{d}\mu + \epsilon\mu(A \cap A_0)$$
$$\leqslant \nu(A) - \lambda(A) + \lambda(A \cap A_0) = \nu(A)$$

which shows that $h \in \mathcal{K}$. However, we also have

$$\int_X h \, \mathrm{d}\mu = \int_X f \, \mathrm{d}\mu + \epsilon\mu(A_0) > M$$

since $\mu(A_0) > 0$. This contradicts the definition of M as a supremum, and implies that λ is the zero measure. As a result, we have

$$\nu(A) = \int_A f \, d\mu \quad \text{for all } A \in \mathcal{A}.$$

It remains to show that the integrable function f is unique in $L^1(X, \mu)$. Recall that $L^1(X, \mu)$ is defined as the set of equivalence classes of functions that are equal μ-almost everywhere, so we need to show that if $f^\star \in L^1(X, \mu)$ is another function such that

$$\nu(A) = \int_A f^\star \, d\mu \quad \text{for all } A \in \mathcal{A},$$

then $f = f^\star$ μ-almost everywhere.

We let

$$E := \{x \in X : f(x) \neq f^\star(x)\}$$

and see that

$$E = \bigcup_{n=1}^{\infty} \left(\left\{ x \in X : f(x) - f^\star(x) > \frac{1}{n} \right\} \cup \left\{ x \in X : f^\star(x) - f(x) > \frac{1}{n} \right\} \right).$$

For each $n \in \mathbb{Z}^+$ Chebyshev's inequality 2.3.1 implies that

$$\mu\left(\left\{ x \in X : f(x) - f^\star(x) > \frac{1}{n} \right\} \right) \leqslant n \cdot \int_{\left\{ x : f(x) - f^\star(x) > \frac{1}{n} \right\}} |f - f^\star| \, d\mu = 0$$

and similarly

$$\mu\left(\left\{ x \in X : f^\star(x) - f(x) > \frac{1}{n} \right\} \right) = 0.$$

Thus we have $\mu(E) = 0$, as the countable union of μ measure zero sets has μ measure zero. This shows that $f = f^\star$ μ-almost everywhere, and the proof is complete. $\qquad\square$

Corollary 5.2.3 (Radon-Nikodym Theorem for Vector Measures)**.** *Let $(X, \mathcal{A}, \mu)$ be a measure space, and let ν be a vector measure on $(X, \mathcal{A})$ taking values in $\mathbb{R}^m$ that is absolutely continuous with respect to μ. Then there exists a unique function $f \in L^1(X, \mu; \mathbb{R}^m)$ such that*

$$\nu(A) = \int_A f \, d\mu \quad \text{for all } A \in \mathcal{A}.$$

Proof. For each $k \in \{1, 2, \ldots, m\}$ we define the signed measure $\nu_k : \mathcal{A} \to \mathbb{R}$ by

$$\nu_k(A) := \langle \nu(A), e_k \rangle \quad \text{for all } A \in \mathcal{A}$$

where $e_1, e_2, \ldots, e_m$ is the standard basis for $\mathbb{R}^m$. Then for each $k \in \{1, 2, \ldots, m\}$ the signed measure ν_k is absolutely continuous with respect to μ since

$$|\nu_k(A)| = |\langle \nu(A), e_k \rangle| \leqslant \|\nu(A)\| \cdot \|e_k\| = \|\nu(A)\|$$

for all $A \in \mathcal{A}$. Thus for each $k \in \{1, 2, \ldots, m\}$ we may apply Theorem 5.2.2 to obtain a unique $f_k \in L^1(X, \mu)$ such that

$$\nu_k(A) = \int_A f_k \, \mathrm{d}\mu \quad \text{for all } A \in \mathcal{A}.$$

We then define the function $f : X \to \mathbb{R}^m$ by

$$f(x) := \sum_{k=1}^m f_k(x) e_k \quad \text{for all } x \in X.$$

Then $f \in L^1(X, \mu; \mathbb{R}^m)$ since each $f_k \in L^1(X, \mu)$, and for each $A \in \mathcal{A}$ we have

$$\int_A f \, \mathrm{d}\mu = \int_A \sum_{k=1}^m f_k(x) e_k \, \mathrm{d}\mu(x) = \sum_{k=1}^m \left(\int_A f_k(x) \, \mathrm{d}\mu(x) \right) e_k$$

$$= \sum_{k=1}^m \nu_k(A) e_k = \sum_{k=1}^m \langle \nu(A), e_k \rangle e_k = \nu(A)$$

which shows existence. Uniqueness follows from the uniqueness in Theorem 5.2.2 applied to each component. □

By combining the Lebesgue Decomposition Theorem **??** and the Radon-Nikodym Theorem for $\mathbb{R}^m$-valued vector measures, we obtain the following fundamental result.

Corollary 5.2.4 (Lebesgue-Radon-Nikodym Theorem)**.** *Let $(X, \mathcal{A}, \mu)$ be a measure space, and let ν be a vector measure on $(X, \mathcal{A})$ taking values in $\mathbb{R}^m$. Then there exist unique vector measures ν_{ac} and ν_s on $(X, \mathcal{A})$ such that*

1. $\nu = \nu_{ac} + \nu_s$,

2. $\nu_{ac} \ll \mu$,

3. $\nu_s \perp \mu$,

and there exists a unique function $f \in L^1(X, \mu; \mathbb{R}^m)$ such that

$$\nu(A) = \int_A f \, \mathrm{d}\mu + \nu_s(A) \quad \text{for all } A \in \mathcal{A}.$$

Proof. The existence and uniqueness of the measures ν_{ac} and ν_{s} follow from the Lebesgue Decomposition Theorem **??**. Since $\nu_{\mathrm{ac}} \ll \mu$, the Radon-Nikodym Theorem for Vector Measures 5.2.3 implies that there exists a unique function $f \in L^1(X, \mu; \mathbb{R}^m)$ such that

$$\nu_{\mathrm{ac}}(A) = \int_A f \, \mathrm{d}\mu \quad \text{for all } A \in \mathcal{A}.$$

Thus we have

$$\nu(A) = \nu_{\mathrm{ac}}(A) + \nu_{\mathrm{s}}(A)$$
$$= \int_A f \, \mathrm{d}\mu + \nu_{\mathrm{s}}(A)$$

for all $A \in \mathcal{A}$, and the proof is complete. $\qquad \square$

Definition 5.2.5. Let $(X, \mathcal{A}, \mu)$ be a measure space, and let ν be a vector measure on $(X, \mathcal{A})$ that is absolutely continuous with respect to μ. The *Radon-Nikodym derivative* of ν with respect to μ is the unique function $f \in L^1(X, \mu; \mathbb{R}^m)$ such that

$$\nu(A) = \int_A f \, \mathrm{d}\mu \quad \text{for all } A \in \mathcal{A}.$$

We denote the Radon-Nikodym derivative by $D_\mu \nu$ or $\frac{\partial \nu}{\partial \mu}$.

Some people also write $\nu = \mu \downharpoonright f$ or $\mathrm{d}\nu = f \, \mathrm{d}\mu$ to denote this relationship, but we will not use this notation.

With our preferred notation, the Radon-Nikodym theorem states that if ν is a signed measure absolutely continuous with respect to μ, then

$$\frac{\partial \nu}{\partial \mu} \in L^1(X, \mu) \quad \text{and} \quad \nu(A) = \int_A \frac{\partial \nu}{\partial \mu} \, \mathrm{d}\mu \text{ for all } A \in \mathcal{A}$$

which shows how it can be viewed as a generalization of the fundamental theorem of calculus.

Corollary 5.2.6 ($\frac{\partial \nu}{\partial |\nu|}$ has Unit Norm). *Let $(X, \mathcal{A}, \mu)$ be a measure space, and let ν be a vector measure on $(X, \mathcal{A})$ taking values in $\mathbb{R}^m$ that is absolutely continuous with respect to μ. Then the Radon-Nikodym derivative $\frac{\partial \nu}{\partial |\nu|} : X \to \mathbb{R}^m$ satisfies $\left\| \frac{\partial \nu}{\partial |\nu|}(x) \right\| = 1$ for $|\nu|$-almost every $x \in X$.*

Proof. Since $\nu \ll |\nu|$, the Radon-Nikodym Theorem for Vector Measures 5.2.3 implies that there exists a unique function $f \in L^1(X, |\nu|; \mathbb{R}^m)$ such that

$$\nu(A) = \int_A f \, \mathrm{d}|\nu| \quad \text{for all } A \in \mathcal{A}.$$

By 5.1.15, we have

$$|\nu|(A) = \int_A \|f(x)\| \, \mathrm{d}|\nu|(x) \quad \text{for all } A \in \mathcal{A}.$$

In particular, taking $A = X$ gives

$$|\nu|(X) = \int_X \|f(x)\| \, \mathrm{d}|\nu|(x).$$

Thus we must have $\|f(x)\| = 1$ for $|\nu|$-almost every $x \in X$. By definition, f is the Radon-Nikodym derivative $\frac{\partial \nu}{\partial |\nu|}$, so we are done. $\square$

In the case that we have two σ-finite measures, we have the following version of the Radon-Nikodym theorem.

Corollary 5.2.7 (Radon-Nikodym Theorem for σ-Finite Measures). *Let $(X, \mathcal{A}, \mu)$ be a σ-finite measure space, and let ν be a σ-finite measure on $(X, \mathcal{A})$ that is absolutely continuous with respect to μ. Then there exists a unique measurable function $f : X \to [0, \infty)$ such that*

$$\nu(A) = \int_A f \, \mathrm{d}\mu \quad \text{for all } A \in \mathcal{A}.$$

Notice that we need *both* measures to be σ-finite, and we lose the conclusion of integrability of f in this case.

Proof. In the case that ν is a finite measure, then ν is also a signed measure, and we may apply Theorem 5.2.2 to obtain the desired result.

Now assume that ν is a σ-finite measure. Then there exists a disjoint sequence of sets $\{X_k\}_{k=1}^{\infty} \subseteq \mathcal{A}$ such that

$$X = \bigcup_{k=1}^{\infty} X_k \quad \text{and} \quad \nu(X_k) < \infty \quad \text{for all } k \in \mathbb{Z}^+.$$

For each $k \in \mathbb{Z}^+$ we define the measure $\nu \restriction X_k$ which is a finite measure on $(X, \mathcal{A})$. Then $\nu \restriction X_k$ is absolutely continuous with respect to μ since ν is absolutely continuous with respect to μ. Thus by the finite measure case, there exists a unique function $f_k \in L^1(X, \mu)$ such that

$$\nu(A \cap X_k) = \int_{A \cap X_k} f_k \, \mathrm{d}\mu \quad \text{for all } A \in \mathcal{A}.$$

We then define the function $f : X \to \mathbb{R}$ by

$$f(x) := f_k(x)$$

if $x \in X$ is such that $x \in X_k$. More formally, we have

$$f(x) := \sum_{k=1}^{\infty} f_k(x) \chi_{X_k}(x) \quad \text{for all } x \in X.$$

Then f is a measurable function on X, as a countable sum of measurable functions.

We compute that for each $A \in \mathcal{A}$ we have

$$\int_A f \, \mathrm{d}\mu = \sum_{k=1}^{\infty} \int_{A \cap \left(X_k \setminus \bigcup_{j=1}^{k-1} X_j\right)} f_k \, \mathrm{d}\mu$$

$$= \sum_{k=1}^{\infty} \nu \left(A \cap \left(X_k \setminus \bigcup_{j=1}^{k-1} X_j \right) \right) = \nu(A)$$

which shows existence.

Now suppose that $f^\star : X \to [0, \infty)$ is another measurable function such that

$$\nu(A) = \int_A f^\star \, \mathrm{d}\mu \quad \text{for all } A \in \mathcal{A}.$$

Then for each $k \in \mathbb{Z}^+$ we have

$$\nu(A \cap X_k) = \int_{A \cap X_k} f^\star \, \mathrm{d}\mu \quad \text{for all } A \in \mathcal{A}.$$

By uniqueness in the finite measure case, it follows that $f_k = f^\star$ μ-almost everywhere on X_k for each $k \in \mathbb{Z}^+$. Thus $f = f^\star$ μ-almost everywhere on X, and the proof is complete. $\qquad \square$

Corollary 5.2.8 (Lebesgue-Radon-Nikodym Theorem for σ-Finite Measures). *Let $(X, \mathcal{A}, \mu)$ be a σ-finite measure space, and let ν be a σ-finite measure on $(X, \mathcal{A})$. Then there exist unique measures ν_{ac} and ν_s on $(X, \mathcal{A})$ such that $\nu = \nu_{ac} + \nu_s$, $\nu_{ac} \ll \mu$, and $\nu_s \perp \mu$, and there exists a unique measurable function $f : X \to [0, \infty)$ such that*

$$\nu(A) = \int_A f \, \mathrm{d}\mu + \nu_s(A) \quad \text{for all } A \in \mathcal{A}.$$

The proof is the same as that of Corollary 5.2.4 using Corollary 5.2.7 in place of Theorem 5.2.2.

Exercise 5.2.1 (Properties of the Radon-Nikodym Derivative). Let $(X, \mathcal{A})$ be a measurable space, and let μ and λ be measures on $(X, \mathcal{A})$, and let ν and σ be vector measures on $(X, \mathcal{A})$ taking values in $\mathbb{R}^m$. The Radon-Nikodym derivatives satisfy the following properties:

1. If $\nu \ll \mu$ and $\sigma \ll \mu$, then

$$\frac{\partial(\nu + \sigma)}{\partial \mu} = \frac{\partial \nu}{\partial \mu} + \frac{\partial \sigma}{\partial \mu}.$$

2. If $\lambda \ll \mu$ and $g \in L^1(X, \lambda)$, then

$$\int_X g \, \mathrm{d}\lambda = \int_X g \frac{\partial \lambda}{\partial \mu} \, \mathrm{d}\mu.$$

3. If $\nu \ll \lambda$ and $\lambda \ll \mu$, then

$$\frac{\partial \nu}{\partial \mu} = \frac{\partial \nu}{\partial \lambda} \frac{\partial \lambda}{\partial \mu}.$$

4. If ν and μ are mutually absolutely continuous, then

$$\frac{\partial \nu}{\partial \mu} = \frac{1}{\left(\frac{\partial \mu}{\partial \nu}\right)}$$

Proof.   1. Assume that $\nu \ll \mu$ and $\sigma \ll \mu$. Then for each $A \in \mathcal{A}$ we have

$$(\nu + \sigma)(A) = \nu(A) + \sigma(A) = \int_A \tfrac{\partial \nu}{\partial \mu} \, \mathrm{d}\mu + \int_A \tfrac{\partial \sigma}{\partial \mu} \, \mathrm{d}\mu$$

$$= \int_A \left(\tfrac{\partial \nu}{\partial \mu} + \tfrac{\partial \sigma}{\partial \mu} \right) \mathrm{d}\mu$$

which shows the desired result by uniqueness of the Radon-Nikodym derivative.

2. Assume that $\lambda \ll \mu$. Then for each $A \in \mathcal{A}$ we have

$$\lambda(A) = \int_A 1 \, \mathrm{d}\lambda = \int_A 1 \cdot \frac{\partial \lambda}{\partial \mu} \, \mathrm{d}\mu = \int_X \mathbf{1}_A \frac{\partial \lambda}{\partial \mu} \, \mathrm{d}\mu$$

and

$$\lambda(A) = \int_X \mathbf{1}_A \, \mathrm{d}\lambda$$

so the result is true for characteristic functions. By the linearity of the integral, we see that if $s = \sum_{j=1}^n a_j \mathbf{1}_{A_j}$ is a simple function on X, then

$$\int_X s \, \mathrm{d}\lambda = \sum_{j=1}^n a_j \lambda(A_j) = \int_X s \frac{\partial \lambda}{\partial \mu} \, \mathrm{d}\mu.$$

Thus the result is true for simple functions.

Now let $g \in L^1(X, \lambda)$ be arbitrary. Then there exists a sequence of simple functions $\{s_k\}_{k=1}^\infty$ on X such that

$$\lim_{k \to \infty} \int_X |g - s_k| \, \mathrm{d}\lambda = 0.$$

By the Dominated Convergence Theorem **??**, we have

$$\int_X g \, \mathrm{d}\lambda = \int_X \lim_{k \to \infty} s_k \, \mathrm{d}\lambda = \lim_{k \to \infty} \int_X s_k \, \mathrm{d}\lambda$$

$$= \lim_{k \to \infty} \int_X s_k \tfrac{\partial \lambda}{\partial \mu} \, \mathrm{d}\mu = \int_X \lim_{k \to \infty} s_k \tfrac{\partial \lambda}{\partial \mu} \, \mathrm{d}\mu = \int_X g \tfrac{\partial \lambda}{\partial \mu} \, \mathrm{d}\mu$$

which shows the desired result.

3. Assume that $\nu \ll \lambda$ and $\lambda \ll \mu$. Then for each $A \in \mathcal{A}$ we have

$$\nu(A) = \int_A \frac{\partial \nu}{\partial \lambda} \, \mathrm{d}\lambda = \int_A \frac{\partial \nu}{\partial \lambda} \frac{\partial \lambda}{\partial \mu} \, \mathrm{d}\mu$$

   where the second equality follows from part (2). By uniqueness of the Radon-Nikodym derivative, we obtain the desired result.

4. Assume that ν and μ are mutually absolutely continuous. Then for each $A \in \mathcal{A}$ we have

$$\mu(A) = \int_A \frac{\partial \mu}{\partial \nu} \, \mathrm{d}\nu = \int_A \frac{\partial \mu}{\partial \nu} \frac{\partial \nu}{\partial \mu} \, \mathrm{d}\mu$$

   where the second equality follows from part (2). Since we also have

$$\mu(A) = \int_A 1 \, \mathrm{d}\mu,$$

   for each $A \in \mathcal{A}$, uniqueness of the Radon-Nikodym derivative implies that

$$1 = \frac{\partial \mu}{\partial \nu} \frac{\partial \nu}{\partial \mu} \quad \mu\text{-almost everywhere,}$$

   which shows the desired result.

$\square$

This finishes our discussion of the abstract Radon-Nikodym theorem.

5.3 Besicovich Covering Theorem

5.3.1 Introduction

In this section, we present the Besicovich Covering Theorem. It is similar in spirit to the Vitali Covering Theorem we proved earlier, but has different hypotheses and conclusions. The Besicovich Covering Theorem is particularly useful in differentiation theory and geometric measure theory, as it allows us to cover sets in $\mathbb{R}^n$ with families of balls in a controlled manner.

The Vitali Covering Theorem says that given an arbitrary collection of balls with uniformly bounded radii, we can extract a disjoint subcollection such that if we scale these balls by a factor of 5, they cover the same set as the original collection. This is a powerful obeservation, but does not really help us unless we have a way to control $\mu(5B)$ in terms of $\mu(B)$ for a measure μ — this is known as a *doubling condition*, and is not satisfied by all measures. The Besicovich Covering Theorem, on the other hand, allows us to cover a set using a bounded number of disjoint subcollections of balls from the original collection,

without needing to scale the balls. This is particularly useful when working with measures that do not satisfy a doubling condition.

It is the Besicovich Covering Theorem that allows us to extend the Lebesgue Differentiation Theorem to Radon measures on $\mathbb{R}^n$.

5.3.2   Besicovich Covering Theorem

Theorem 5.3.1 (Besicovich Covering Theorem). *For each $n \in \mathbb{Z}^+$, there exists an integer constant $M_n \in \mathbb{Z}^+$ such that the following holds.*

Let $A \subseteq \mathbb{R}^n$ be a nonempty set, and let $\mathcal{B}$ be a collection of closed balls in $\mathbb{R}^n$ such that for each $a \in A$, there exists a ball $B \in \mathcal{B}$ centered at a and the diameters are uniformly bounded, i.e.

$$\sup\{\operatorname{diam}(B) : B \in \mathcal{B}\} < \infty.$$

Then there exist subcollections $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_{M_n} \subseteq \mathcal{B}$ such that

(i) for each $1 \leqslant i \leqslant M_n$, the balls in $\mathcal{B}_i$ are disjoint, and

(ii) we have

$$A \subseteq \bigcup_{i=1}^{M_n} \bigcup_{B \in \mathcal{B}_i} B.$$

The conclusion is saying that within each subcollection, the balls do not overlap, but different subcollections may have overlapping balls; moreover, the number of subcollections needed to cover the set of centers is bounded by a constant depending only on the dimension. Our proof will give little indication on what the best possible value of M_n is; we will only show that such a constant exists.

The proof is pretty damn long, and apparently at least one popular account has a major flaw in its proof — beware of *Geometric Integration Theory* by Krantz and Parks; they make a subtle claim (which they take as self-evident, and not even identified as something to be proved) that is actually false. We follow the proof by Evans and Gariepy in *Measure Theory and Fine Properties of Functions*.

Proof. We begin by assuming that the set of centers A is bounded; we will remove this assumption in the very last step of the proof.

Step 1: First Greedy Algorithm.

Start by selecting a ball B_1 from the collection $\mathcal{B}$ with radius satisfying

$$\operatorname{radius}(B_1) > \frac{3}{4} \sup\{\operatorname{radius} B : B \in \mathcal{B}\}.$$

Set $r_1 := \text{radius}(B_1)$ and let a_1 be the center of B_1. Then let $A_2 := A \backslash B_1$, and if A_2 is nonempty, select a ball $B_2 \in \mathcal{B}$ centered at a point $a_2 \in A_2$ with radius r_2 satisfying

$$r_2 > \frac{3}{4} \sup\{r : \overline{B}(a, r) \in \mathcal{B}, a \in A_2\}.$$

Continuing in this manner, for each $j \in \mathbb{Z}^+$, we define

$$A_j := A \backslash \bigcup_{k=1}^{j-1} B_k,$$

and if A_j is nonempty, we select a ball $B_j \in \mathcal{B}$ centered at a point $a_j \in A_j$ with radius r_j satisfying

$$r_j > \frac{3}{4} \sup\{r : \overline{B}(a, r) \in \mathcal{B}, a \in A_j\}$$

If $A_j \neq \varnothing$ for all $j \in \mathbb{Z}^+$, then we obtain an infinite sequence of balls $\{B_j\}_{j=1}^{\infty}$, each selected from $\mathcal{B}$; in this case we set $J = \infty$. If at some step $N \in \mathbb{Z}^+$ we have $A_N = \varnothing$, then the process terminates and we obtain a finite sequence of balls $\{B_j\}_{j=1}^{N-1}$, and in this case we set $J = N - 1$.

Step 2: Properties of the First Greedy Algorithm.

Claim 1: If $j > k$, then $r_j \leqslant \frac{4}{3} r_k$.

Proof of Claim 1. Assume $j > k$. Then $a_j \in A_k$, so we have

$$r_k \geqslant \frac{3}{4} \sup\{r : \overline{B}(a, r) \in \mathcal{B}, a \in A_k\} \geqslant \frac{3}{4} r_j$$

as claimed. $\qquad\square$

Claim 2: The collection of balls $\{\frac{1}{3} B_j\}_{j=1}^{J}$ is disjoint.

Proof of Claim 2. Let $1 \leqslant i < j \leqslant J$. Then $a_j \notin B_i$, so we have

$$\|a_i - a_j\| \geqslant r_i = \frac{r_i}{3} + \frac{2r_i}{3} \geqslant \frac{r_i}{3} + \frac{2}{3}\frac{3}{4} r_j = \frac{r_i}{3} + \frac{1}{2} r_j > \frac{r_i}{3} + \frac{r_j}{3}$$

which shows that the distance between the centers of B_i and B_j is greater than the sum of the radii of the balls $\frac{1}{3} B_i$ and $\frac{1}{3} B_j$, and thus these two balls do not intersect. $\qquad\square$

Claim 3: If $J = \infty$, then $r_j \to 0$ as $j \to \infty$.

Proof of Claim 3. By the previous claim, we know that the balls $\{\frac{1}{3} B_j\}_{j=1}^{\infty}$ are disjoint. Since the set of centers A is bounded, we must have $r_j \to 0$ as $j \to \infty$. $\qquad\square$

Claim 4: We have $A \subseteq \bigcup_{j=1}^{J} B_j$.

Proof of Claim 4. If J is finite, then the greedy algorithm in Step 1 terminated because $A_{J+1} = \varnothing$, which implies that

$$A \subseteq \bigcup_{j=1}^{J} B_j.$$

Thus we assume that $J = \infty$. If $a \in A$ then by assumption there exists $r > 0$ and a closed ball $\overline{B}(a, r) \in \mathcal{B}$. By Claim 3, there exists $j \in \mathbb{Z}^+$ such that $r_j < \frac{3}{4}r$ — in particular, this implies that $a \in \bigcup_{i=1}^{j-1} B_i$; otherwise, we would have selected a ball centered at a with radius at least r at step j of the greedy algorithm, contradicting the choice of r_j. $\square$

Fix an integer $k > 1$ for the remainder of this step and the next step. Define

$$I_k := \{j : 1 \leqslant j < k, B_j \cap B_k \neq \varnothing\} \quad \text{and} \quad K_k := I_k \cap \{j : r_j \leqslant 3r_k\}.$$

Claim 5: We have $\#K_k \leqslant 20^n$.

In words, we are claiming that there are at most 20^n balls from the first greedy algorithm that intersect B_k and have radius at most $3r_k$.

Proof of Claim 5. Let $j \in K_k$. Then $B_j \cap B_k \neq \varnothing$ and $r_j \leqslant 3r_k$. For each point $x \in \overline{B}(a_j, \frac{r_j}{3})$ see that

$$\|x - a_k\| \leqslant \|x - a_j\| + \|a_j - a_k\| \leqslant \frac{r_j}{3} + r_j + r_k$$

$$= \frac{4}{3}r_j + r_k \leqslant 4r_k + r_k = 5r_k$$

which shows that

$$B\left(a_j, \frac{r_j}{3}\right) \subseteq \overline{B}(a_k, 5r_k).$$

By Claim 2, the balls $\{\overline{B}(a_j, \frac{r_j}{3})\}_{j \in K}$ are disjoint, so we have

$$5^n \omega_n r_k^n = \mathcal{L}^n\left(\overline{B}(a_k, 5r_k)\right)$$

$$\geqslant \sum_{j \in K} \mathcal{L}^n\left(B\left(a_j, \frac{r_j}{3}\right)\right) \qquad \text{by disjointness}$$

$$= \sum_{j \in K_k} \omega_n \left(\frac{r_j}{3}\right)^n$$

$$\geqslant \sum_{j \in K_k} \omega_n \left(\frac{r_k}{4}\right)^n \qquad \text{by Claim 1 and } r_j \leqslant 3r_k \text{ for each } j \in K_k$$

$$= \#K_k \cdot \omega_n \left(\frac{r_k}{4}\right)^n.$$

Dividing both sides by $\omega_n \left(\frac{r_k}{4}\right)^n$ gives

$$\#K_k \leqslant 5^n \cdot 4^n = 20^n$$

as claimed. $\square$

Step 3: We now bound the number of elements in $I_k \backslash K_k$.

Let $i, j \in I_k \backslash K_k$ with $i \neq j$. Then $1 \leqslant i, j < k$, both $B_i \cap B_k \neq \varnothing$ and $B_j \cap B_k \neq \varnothing$, and both $r_i > 3r_k$ and $r_j > 3r_k$.

For distinct $i, j \in I_k \backslash K_k$, we let $\theta_{i,j} \in [0, \pi]$ be the angle between the vectors $a_i - a_k$ and $a_j - a_k$. Since we know $i, j < k$, we know that $a_k \notin B_i \cup B_j$ which implies $r_i < \|a_i - a_k\|$ and $r_j < \|a_j - a_k\|$. Since $B_i \cap B_k \neq \varnothing$ and $B_j \cap B_k \neq \varnothing$, we must have $\|a_i - a_k\| < r_i + r_k$ and $\|a_j - a_k\| < r_j + r_k$. Finally without loss of generality we may assume that $\|a_i - a_k\| \leqslant \|a_j - a_k\|$. In summary,

$$\begin{cases} 3r_k < r_i < \|a_i - a_k\| < r_i + r_k, \\ 3r_k < r_j < \|a_j - a_k\| < r_j + r_k, \\ \|a_i - a_k\| \leqslant \|a_j - a_k\|. \end{cases}$$

Claim 6: For distinct $i, j \in I_k \backslash K_k$, if $\cos \theta_{i,j} > \frac{5}{6}$, then $a_i \in B_j$.

Proof of Claim 6. We prove the contrapositive. Assume that $a_i \notin B_j$. Then we consider two possibilities — either $\|a_i - a_j\| \geqslant \|a_j - a_k\|$ or $\|a_i - a_j\| < \|a_j - a_k\|$. If $\|a_i - a_j\| \geqslant \|a_j - a_k\|$, then the Law of Cosines gives

$$\begin{aligned} \cos \theta_{i,j} &= \frac{\|a_i - a_k\|^2 + \|a_j - a_k\|^2 - \|a_i - a_j\|^2}{2\|a_i - a_k\|\|a_j - a_k\|} \\ &\leqslant \frac{\|a_i - a_k\|^2}{2\|a_i - a_k\|\|a_j - a_k\|} \\ &= \frac{\|a_i - a_k\|}{2\|a_j - a_k\|} \leqslant \frac{1}{2} < \frac{5}{6} \end{aligned}$$

as desired. If instead $\|a_i - a_j\| < \|a_j - a_k\|$, then $r_j < \|a_i - a_j\|$ since $a_i \notin B_j$, and thus

$$\begin{aligned} \cos \theta_{i,j} &= \frac{\|a_i - a_k\|^2 + \|a_j - a_k\|^2 - \|a_i - a_j\|^2}{2\|a_i - a_k\|\|a_j - a_k\|} \\ &= \frac{\|a_i - a_k\|^2}{2\|a_i - a_k\|\|a_j - a_k\|} + \frac{(\|a_j - a_k\| - \|a_i - a_j\|)(\|a_j - a_k\| + \|a_i - a_j\|)}{2\|a_i - a_k\|\|a_j - a_k\|} \\ &= \frac{\|a_i - a_k\|}{2\|a_j - a_k\|} + \frac{((\|a_j - a_k\| - \|a_i - a_j\|)(\|a_j - a_k\| + \|a_i - a_j\|))}{2\|a_i - a_k\|\|a_j - a_k\|} \\ &\leqslant \frac{1}{2} + \frac{(\|a_j - a_k\| - \|a_i - a_j\|)(2\|a_j - a_k\|)}{2\|a_i - a_k\|\|a_j - a_k\|} \\ &= \frac{1}{2} + \frac{r_j + r_k - r_j}{r_i} = \frac{1}{2} + \frac{r_k}{r_i} < \frac{1}{2} + \frac{1}{3} = \frac{5}{6} \end{aligned}$$

$\square$

Claim 7: If $a_i \in B_j$, then

$$0 \leqslant \|a_i - a_j\| + \|a_i - a_k\| - \|a_j - a_k\| \leqslant \|a_j - a_k\| \cdot \frac{8}{3} \left(1 - \cos \theta_{i,j}\right)$$

Proof of Claim 7. Assume that $a_i \in B_j$. Then we have

$$0 \leqslant \frac{\|a_i - a_j\| + \|a_i - a_k\| - \|a_j - a_k\|}{\|a_j - a_k\|}$$

$$\leqslant \frac{\|a_i - a_j\| + \|a_i - a_k\| - \|a_j - a_k\|}{\|a_j - a_k\|} \cdot \frac{\|a_i - a_j\| - \|a_i - a_k\| + \|a_j - a_k\|}{\|a_i - a_j\|}$$

$$= \frac{\|a_i - a_j\|^2 - (\|a_i - a_k\| - \|a_j - a_k\|)^2}{\|a_j - a_k\| \|a_i - a_j\|}$$

$$= \frac{\|a_i - a_k\|^2 - 2\|a_i - a_k\| \|a_j - a_k\| \cos\theta_{i,j} + \|a_j - a_k\|^2 - \|a_i - a_k\|^2 + 2\|a_i - a_k\| \|a_j - a_k\| - \|a_j - a_k\|^2}{\|a_j - a_k\| \|a_i - a_j\|}$$

$$= \frac{2\|a_i - a_k\| \|a_j - a_k\| (1 - \cos\theta_{i,j})}{\|a_j - a_k\| \|a_i - a_j\|}$$

$$= \frac{2\|a_i - a_k\| (1 - \cos\theta_{i,j})}{\|a_i - a_j\|}$$

$$\leqslant \frac{2(r_i + r_k)(1 - \cos\theta_{i,j})}{r_i}$$

$$\leqslant \frac{2\left(1 + \frac{1}{3}\right) r_i (1 - \cos\theta_{i,j})}{r_i}$$

$$= \frac{8}{3}(1 - \cos\theta_{i,j}).$$

Hence

$$0 \leqslant \|a_i - a_j\| + \|a_i - a_k\| - \|a_j - a_k\| \leqslant \|a_j - a_k\| \cdot \frac{8}{3}\left(1 - \cos\theta_{i,j}\right)$$

as desired.   □

Claim 8: If $a_i \in B_j$, then $\cos\theta_{i,j} \leqslant \frac{61}{64}$.

Proof of Claim 8. Since $a_i \in B_j$ and $a_j \notin B_i$, we have $r_i < \|a_i - a_j\| < r_j$. Because $i < j$, we have $r_j \leqslant \frac{4}{3} r_i$ by Claim 1. Therefore

$$\|a_i - a_j\| + \|a_i - a_k\| - \|a_j - a_k\| \geqslant r_i + r_i - r_j - r_k$$

$$\geqslant \frac{3}{4}r_j + \frac{3}{4}r_j - r_j - \frac{1}{3}r_j$$

$$= \frac{1}{2}r_j - r_k \geqslant \frac{1}{6}r_j$$

$$= \frac{1}{6}\frac{3}{4}\left(r_j + \frac{1}{3}r_j\right)$$

$$\geqslant \frac{1}{8}(r_j + r_k)$$

$$\geqslant \frac{1}{8}\|a_j - a_k\|$$

so by Claim 7 we have

$$\frac{1}{8}\|a_j - a_k\| \leqslant \|a_i - a_j\| + \|a_i - a_k\| - \|a_j - a_k\| \leqslant \|a_j - a_k\| \cdot \frac{8}{3}\left(1 - \cos\theta_{i,j}\right).$$

Hence

$$\frac{3}{64} \leqslant 1 - \cos\theta_{i,j}$$

which gives the desired result. □

Claim 9: There exists a constant L_n depending only on n such that $\#(I_k \backslash K_k) \leqslant L_n$.

Proof of Claim 9. See that for distinct $i, j \in I_k \backslash K_k$, then either $a_i \notin B_j$ or $a_i \in B_j$, so we have either $\cos\theta_{i,j} \leqslant \frac{5}{6}$ or $\cos\theta_{i,j} \leqslant \frac{61}{64}$ by Claims 6 and 8. In either case, we have $\cos\theta_{i,j} \leqslant \frac{61}{64}$. Thus the angle $\theta_{i,j}$ between the vectors $a_i - a_k$ and $a_j - a_k$ satisfies

$$\theta_{i,j} \geqslant \arccos\left(\frac{61}{64}\right) > 0.$$

Now fix a radius $r_0 > 0$ so small that if $x \in \mathbb{S}^{n-1}$ and $y, z \in B(x, r_0)$ then the angle between y and z is less than $\arccos\left(\frac{61}{64}\right)$. Since $\mathbb{S}^{n-1}$ is compact, we can cover it with finitely many balls of radius r_0 with centers in $\mathbb{S}^{n-1}$; by the well-ordering principle, there exists a smallest number L_n of such balls needed to cover $\mathbb{S}^{n-1}$. Then ∂B_k can be covered by L_n balls of radius $r_0 r_k$ with centers on ∂B_k. By the previous paragraph, if $i, j \in I_k \backslash K_k$ with $i \neq j$, then the angle between $a_i - a_k$ and $a_j - a_k$ is at least $\arccos\left(\frac{61}{64}\right)$, so the points $\frac{a_i - a_k}{\|a_i - a_k\|}$ and $\frac{a_j - a_k}{\|a_j - a_k\|}$ lie in different balls of radius r_0 on $\mathbb{S}^{n-1}$. Therefore, each ball of radius $r_0 r_k$ on ∂B_k contains at most one of the points a_i for $i \in I_k \backslash K_k$. Hence we have $\#(I_k \backslash K_k) \leqslant L_n$ as desired. □

Step 4: We now finish the proof in the case that A is bounded.

We set $C_n := 20^n + L_n + 1$, and see that by claims 5 and 9, we have

$$\#I_k = \#K_k + \#(I_k \backslash K_k) \leqslant 20^n + L_n < C_n$$

for each integer $k > 1$. We define a function $\sigma : \mathbb{Z}^+ \to \{1, 2, \ldots, C_n\}$ as follows —

- if $1 \leqslant j \leqslant C_n$, set $\sigma(j) := j$;

- if $k \geqslant C_n$, we can inductively define $\sigma(k+1)$ by noting that the previous estimate shows that

$$\#\{j : 1 \leqslant j \leqslant k, B_j \cap B_{k+1} \neq \varnothing\} = \#I_k < C_n$$

so there exists $m \in \{1, 2, \ldots, C_n\}$ such that $B_{k+1} \cap B_j = \varnothing$ for all $j \in \{1, 2, \ldots, k\}$ with $\sigma(j) = m$; we set $\sigma(k+1) := m$.

Now for each integer $1 \leqslant m \leqslant C_n$, define

$$\mathcal{B}_m := \{B_j : j \in \mathbb{Z}^+, \sigma(j) = m\}.$$

By definition of the function σ, the balls in each collection $\mathcal{B}_m$ are disjoint — indeed, if $B_i, B_j \in \mathcal{B}_m$ with $i < j$ and $\sigma(i) = \sigma(j) = m$, then we must have $B_i \cap B_j = \varnothing$ by the definition of $\sigma(j)$. Finally, by Claim 4, we have

$$A \subseteq \bigcup_{j=1}^{J} B_j = \bigcup_{m=1}^{C_n} \bigcup_{B \in \mathcal{B}_m} B$$

which completes the proof in the case that A is bounded.

Step 5: Finally we assume that A is unbounded.

Assume that A is unbounded, for each integer $l \in \mathbb{Z}^+$ define

$$A_l := A \cap \left\{ x \in \mathbb{R}^n : 3(l-1) < \frac{\|x\|}{\sup_{B \in \mathcal{B}} \operatorname{diam} B} < 3l \right\}$$

and

$$\mathcal{B}^l := \{\overline{B}(a, r) \in \mathcal{B} : a \in A_l\}.$$

Then for each $l \in \mathbb{Z}^+$, the set A_l is bounded, so we may apply the reasoning from Steps 1 to 4 to the set A_l and the collection $\mathcal{B}^l$ to obtain subcollections $\mathcal{B}_1^l, \mathcal{B}_2^l, \ldots, \mathcal{B}_{M_n}^l \subseteq \mathcal{B}^l$ such that

(i) for each $1 \leqslant m \leqslant C_n$, the balls in $\mathcal{B}_m^l$ are disjoint, and

(ii) we have

$$A_l \subseteq \bigcup_{m=1}^{C_n} \bigcup_{B \in \mathcal{B}_m^l} B.$$

For each integer $1 \leqslant m \leqslant C_n$, define

$$\mathcal{B}_j := \bigcup_{l=1}^{\infty} \mathcal{B}_j^{2l-1}$$

and

$$\mathcal{B}_{j+C_n} := \bigcup_{l=1}^{\infty} \mathcal{B}_j^{2l}.$$

Then for each $1 \leqslant j \leqslant 2C_n$, the balls in $\mathcal{B}_j$ are disjoint — if $B, B' \in \mathcal{B}_j$ with $B \neq B'$, then there exists $l, l' \in \mathbb{Z}^+$ such that $B \in \mathcal{B}_j^l$ and $B' \in \mathcal{B}_j^{l'}$; then if $l = l'$, we have $B \cap B' = \varnothing$ since the balls in $\mathcal{B}_j^l$ are disjoint, and if $l \neq l'$, then by construction of the sets A_l and the collections $\mathcal{B}_j^l$, we have $B \cap B' = \varnothing$ as well.

Finally, by construction of the sets A_l for $l \in \mathbb{Z}^+$, we have

$$A = \bigcup_{l=1}^{\infty} A_l$$

$$\subseteq \bigcup_{l=1}^{\infty} \bigcup_{m=1}^{C_n} \bigcup_{B \in \mathcal{B}_m^l} B$$

$$= \bigcup_{m=1}^{C_n} \bigcup_{l=1}^{\infty} \bigcup_{B \in \mathcal{B}_m^l} B$$

$$= \bigcup_{m=1}^{C_n} \bigcup_{B \in \mathcal{B}_m \cup \mathcal{B}_{m+C_n}} B$$

$$= \bigcup_{j=1}^{2C_n} \bigcup_{B \in \mathcal{B}_j} B$$

The proof is now complete upon setting $M_n := 2C_n$. $\qquad\qquad\square$

5.3.3   Applications

Next we would like to mimic some results that we had for the Lebesgue measure on $\mathbb{R}^n$, but now for Borel regular outer measures on $\mathbb{R}^n$ which are finite on compact subsets.

Lemma 5.3.2 (Filling Finite Measure Sets with Balls). *Let μ be a Borel regular outer measure on $\mathbb{R}^n$ which is finite on compact subsets, and let $A \subset \mathbb{R}^n$ be such that $\mu(A) < \infty$. Let $\mathcal{B}$ be a collection of closed balls such that the center of each ball in $\mathcal{B}$ belongs to A, and*

$$\inf\{r : \overline{B}(a, r) \in \mathcal{B}, a \in A\} = 0.$$

Then there is a countable subcollection of disjoint balls $\{B_j\}_{j=1}^{\infty} \subseteq \mathcal{B}$ such that

$$\mu\left(A \backslash \bigcup_{j=1}^{\infty} B_j\right) = 0.$$

Note that the set A in the statement of the lemma is not assumed to be μ-measurable.

Proof. Step 1: Consider the family of closed balls

$$\mathcal{B}^1 := \{B \in \mathcal{B} : \operatorname{diam} B \leqslant 1\}$$

which covers A because if $a \in A$, then there exists a ball $\overline{B}(a, r) \in \mathcal{B}$ with radius $r \leqslant 1$ by the assumption on $\mathcal{B}$. By the Besicovich Covering Theorem (Theorem 5.3.1), there exists subcollections $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_{M_n} \subseteq \mathcal{B}^1$ such that each collection only contains disjoint balls and

$$A \subseteq \bigcup_{j=1}^{M_n} \bigcup_{B \in \mathcal{B}_j} B.$$

Thus

$$\mu(A) \leqslant \sum_{j=1}^{M_n} \mu\left(A \cap \bigcup_{B \in \mathcal{B}_j} B\right)$$

so there is some $1 \leqslant j \leqslant M_n$ such that

$$\mu\left(A \cap \bigcup_{B \in \mathcal{B}_j} B\right) \geqslant \frac{1}{M_n}\mu(A).$$

Therefore by 1.1.6, there exist finitely many disjoint balls $B_1, B_2, \ldots, B_{N_1} \in \mathcal{B}_j$ such that

$$\mu\left(A \cap \bigcup_{k=1}^{N_1} B_k\right) \geqslant \frac{1}{2M_n}\mu(A).$$

Since $\bigcup_{k=1}^{N_1} B_k$ is a measurable set, we have

$$\mu(A) = \mu\left(A \cap \bigcup_{k=1}^{N_1} B_k\right) + \mu\left(A \backslash \bigcup_{k=1}^{N_1} B_k\right).$$

which implies that

$$\mu\left(A \backslash \bigcup_{k=1}^{N_1} B_k\right) = \mu(A) - \mu\left(A \cap \bigcup_{k=1}^{N_1} B_k\right)$$

$$\leqslant \mu(A) - \frac{1}{2M_n}\mu(A) = \left(1 - \frac{1}{2M_n}\right)\mu(A).$$

Step 2: Now the collection of closed balls

$$\mathcal{B}^2 := \left\{B \in \mathcal{B}^1 : B \cap \left(\bigcup_{k=1}^{N_1} B_k\right) = \varnothing\right\}$$

clearly covers the set $A \backslash \bigcup_{k=1}^{N_1} B_k$, and we also have

$$\inf\left\{r : \overline{B}(a, r) \in \mathcal{B}^2, a \in A \backslash \bigcup_{k=1}^{N_1} B_k\right\} = 0$$

for each $a \in A \backslash \bigcup_{k=1}^{N_1} B_k$ by the assumption on $\mathcal{B}$. Repeating the above argument with the set $A \backslash \bigcup_{k=1}^{N_1} B_k$ and the collection $\mathcal{B}^2$, we obtain finitely many disjoint balls $B_{N_1+1}, B_{N_1+2}, \ldots, B_{N_2} \in \mathcal{B}^2$ such that

$$\mu\left(A \backslash \bigcup_{k=1}^{N_2} B_k\right) \leqslant \left(1 - \frac{1}{2M_n}\right)\mu\left(A \backslash \bigcup_{k=1}^{N_1} B_k\right) \leqslant \left(1 - \frac{1}{2M_n}\right)^2 \mu(A).$$

Continuing in this manner, for each $k > 1$ we let $\mathcal{B}^k$ be the collection of closed balls in $\mathcal{B}^1$ which are disjoint from $\bigcup_{j=1}^{N_{k-1}} B_j$, and we obtain finitely many disjoint balls $B_{N_{k-1}+1}, B_{N_{k-1}+2}, \ldots, B_{N_k} \in \mathcal{B}^k$ such that

$$\mu\left(A \setminus \bigcup_{j=1}^{N_k} B_j\right) \leqslant \left(1 - \frac{1}{2M_n}\right)^k \mu(A).$$

Finally, taking the limit as $k \to \infty$, we have

$$\mu\left(A \setminus \bigcup_{j=1}^{\infty} B_j\right) = \lim_{k \to \infty} \mu\left(A \setminus \bigcup_{j=1}^{N_k} B_j\right) \leqslant \lim_{k \to \infty} \left(1 - \frac{1}{2M_n}\right)^k \mu(A) = 0$$

as desired. $\qquad \square$

Corollary 5.3.3 (More on Filling Open Sets with Balls). *Let μ be a Borel regular outer measure on $\mathbb{R}^n$ which is finite on compact subsets, and let $\mathcal{B}$ be a nonempty collection of closed balls. Let A be the set of centers of balls in $\mathcal{B}$, and assume that $\mu(A) < \infty$ and that*

$$\inf\{r : \overline{B}(a, r) \in \mathcal{B}, a \in A\} = 0$$

for each $a \in A$.

Then for each open set $U \subseteq \mathbb{R}^n$ there exists a countable subcollection of disjoint balls $\{B_j\}_{j=1}^{\infty} \subseteq \mathcal{B}$ such that

$$\bigcup_{j=1}^{\infty} B_j \subset U$$

and

$$\mu\left((A \cap U) \setminus \bigcup_{j=1}^{\infty} B_j\right) = 0.$$

Proof. Define the collection of closed balls

$$\mathcal{B}_U := \{B \in \mathcal{B} : B \subset U\}.$$

Then $\mathcal{B}_U$ covers the set $A \cap U$, and for each $a \in A \cap U$, we have

$$\inf\{r : \overline{B}(a, r) \in \mathcal{B}_U, a \in A \cap U\} = 0$$

by the assumption on $\mathcal{B}$. Since $\mu(A \cap U) < \infty$, we may apply Lemma 5.3.2 to the set $A \cap U$ and the collection $\mathcal{B}_U$ to obtain a countable subcollection of disjoint balls $\{B_j\}_{j=1}^{\infty} \subseteq \mathcal{B}_U$ such that

$$\mu\left((A \cap U) \setminus \bigcup_{j=1}^{\infty} B_j\right) = 0$$

as desired. $\qquad \square$

Definition 5.3.4 (Maximal Function for Borel Regular Measures)**.** Let μ be a Borel regular outer measure on $\mathbb{R}^n$ which is finite on compact subsets. For each locally integrable function $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n, \mu)$, we define the *maximal function* $M_\mu f : \mathbb{R}^n \to [0, \infty]$ by

$$M_\mu f(x) := \sup_{r>0} \frac{1}{\mu(\overline{B}(x,r))} \int_{\overline{B}(x,r)} |f| \; \mathrm{d}\mu$$

for each $x \in \mathbb{R}^n$, where we interpret the expression $\frac{1}{0}$ as ∞ and use the convention that $\infty \cdot 0 = 0$.

Remark 5.3.5 (About the Definition of the Maximal Function)*.* Recall that f is said to be locally integrable with respect to μ if for each compact set $K \subseteq \mathbb{R}^n$, we have

$$\int_K |f(y)| \; \mathrm{d}\mu(y) < \infty.$$

We require that μ is finite on compact subsets to ensure that the fraction $\frac{1}{\mu(B)}$ is not zero. The locally integrability of f with respect to μ ensures that the integral $\int_B |f(y)| \; \mathrm{d}\mu(y)$ is finite for each closed ball $B \subseteq \mathbb{R}^n$, so that the expression defining $M_\mu f(x)$ is well-defined (possibly infinite) for each $x \in \mathbb{R}^n$.

Here we mimic the definition of the centered Hardy-Littlewood maximal function from before, but now using a general Borel regular outer measure μ on $\mathbb{R}^n$ which is finite on compact sets.

Exercise 5.3.1 (Weak $(1, 1)$ Inequality for the Maximal Function)**.** Let μ be a Borel regular outer measure on $\mathbb{R}^n$ which is finite on compact subsets. Then there exists a constant $C_n > 0$ depending only on n such that for each $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n, \mu)$ and each $t > 0$, we have

$$\mu \left(\{ x \in \mathbb{R}^n : M_\mu f(x) > t \} \right) \leqslant \frac{C_n}{t} \int_{\mathbb{R}^n} |f| \; \mathrm{d}\mu.$$

I am not 100% sure if this statement and the following proof is correct, I am maybe 50% sure. I cannot find another reference — the Wikipedia page on Besicovich's Covering Theorem is where I found the statement of the result, but their proof begins with the incorrect statement that Mf is lower semicontinuous. (To see that Mf is not necessarily lower semicontinuous, consider the measure $\mu = \delta_0$ (the Dirac delta at 0) on $\mathbb{R}$, and the function $f = \chi_{\{0\}}$; then $Mf(0) = 1$, but for each $x \neq 0$, we have $Mf(x) = 0$, so Mf is not lower semicontinuous at 0.) Because the first statement is false (and not even acknowledged as something to be proved), I find it hard to trust the rest of their proof; their proof also uses a different version of Besicovich than the one I have stated.

I also had several false starts by trying to invoke Besicovich at the wrong time in this proof. We will not use this result later, so it is not a big deal if the proof is wrong, but I would appreciate any feedback on this proof.

Proof. Let $f \in L^1_{\text{loc}}(\mathbb{R}^n, \mu)$. If $\int_{\mathbb{R}^n} |f| \, d\mu = \infty$, then the inequality is trivial since the right-hand side is infinite. Thus we assume that

$$\int_{\mathbb{R}^n} |f| \, d\mu < \infty$$

i.e. that $\|f\|_{L^1(\mathbb{R}^n, \mu)} < \infty$.

For each $t > 0$, define the set

$$E_t := \{x \in \mathbb{R}^n : M_\mu f(x) > t\}.$$

Let $K \subseteq E_t$ be a compact set. Then for each $x \in K$, by definition of E_t there exists a radius $r_x > 0$ such that

$$\frac{1}{\mu(\overline{B}(x, r_x))} \int_{\overline{B}(x, r_x)} |f| \, d\mu > t.$$

[It is tempting to say two false statments here, based on the following observation — for each $x \in K$, by definition of E_t there exists a radius $r_x > 0$ such that

$$\frac{1}{\mu(\overline{B}(x, r_x))} \int_{\overline{B}(x, r_x)} |f| \, d\mu > t.$$

The *first mistake* is to state that compactness of K implies that $\sup_{x \in K} r_x < \infty$. This is false; take a moment to think of an example.

The *second mistake* is more subtle, and was my attempt at rectifying the first mistake — I was tempted to get an open cover $\{B(x, r_x) : x \in K\}$ of K and then use compactness to extract a finite subcover to get a finite collection of balls covering K ; this finite subcover then *does* have uniformly bounded diameters, since there are only finitely many balls. However, the center of these balls is no longer guaranteed to be all of K, so we cannot apply Besicovich to this finite collection of balls with K as the set of centers. This point is subtle, but important.]

For each integer $N \in \mathbb{Z}^+$, define

$$K_N := \left\{ x \in K : \exists \, 0 < r \leqslant N \text{ such that } \frac{1}{\mu(\overline{B}(x, r))} \int_{\overline{B}(x, r)} |f| \, d\mu > t \right\}$$

so that $K = \bigcup_{N=1}^\infty K_N$. Fix $N \in \mathbb{Z}^+$. Then for each $x \in K_N$, there exists a radius $0 < r_x \leqslant N$ such that

$$\frac{1}{\mu(\overline{B}(x, r_x))} \int_{\overline{B}(x, r_x)} |f| \, d\mu > t.$$

Thus the collection of closed balls

$$\mathcal{B}_N := \{\overline{B}(x, r_x) : x \in K_N\}$$

has K_N as the set of centers, and the set of diameters of balls in $\mathcal{B}_N$ is bounded above by $2N$. Therefore we may apply the Besicovich Covering Theorem (Theorem 5.3.1) to the set K_N and the collection $\mathcal{B}_N$ to see that there exists disjoint subcollections $\mathcal{B}_{N,1}, \mathcal{B}_{N,2}, \ldots, \mathcal{B}_{N,C_n} \subseteq \mathcal{B}_N$ such that

$$K_N \subseteq \bigcup_{j=1}^{C_n} \bigcup_{B \in \mathcal{B}_{N,j}} B$$

where for each $1 \leqslant j \leqslant C_n$, the balls in $\mathcal{B}_{N,j}$ are disjoint. We remark that the constant C_n depends only on n and not on N, K, f or anything else. Therefore, we have

$$
\begin{aligned}
\mu(K_N) &\leqslant \mu\left(\bigcup_{j=1}^{C_n} \bigcup_{B \in \mathcal{B}_{N,j}} B\right) \\
&\leqslant \sum_{j=1}^{C_n} \mu\left(\bigcup_{B \in \mathcal{B}_{N,j}} B\right) \\
&= \sum_{j=1}^{C_n} \sum_{B \in \mathcal{B}_{N,j}} \mu(B) \qquad\qquad \text{by disjointness} \\
&< \sum_{j=1}^{C_n} \sum_{B \in \mathcal{B}_{N,j}} \frac{1}{t} \int_B |f| \, \mathrm{d}\mu \qquad \text{by choice of balls in } \mathcal{B}_N \\
&= \frac{1}{t} \sum_{j=1}^{C_n} \sum_{B \in \mathcal{B}_{N,j}} \int_B |f| \, \mathrm{d}\mu \\
&\leqslant \frac{1}{t} \sum_{j=1}^{C_n} \int_{\mathbb{R}^n} |f| \, \mathrm{d}\mu \qquad\quad \text{by disjointness of balls in each } \mathcal{B}_{N,j} \\
&= \frac{C_n}{t} \int_{\mathbb{R}^n} |f| \, \mathrm{d}\mu.
\end{aligned}
$$

Since $K = \bigcup_{N=1}^{\infty} K_N$ and this is an increasing union, by 1.1.6 we have

$$\mu(K) = \lim_{N \to \infty} \mu(K_N) \leqslant \frac{C_n}{t} \int_{\mathbb{R}^n} |f| \, \mathrm{d}\mu.$$

Since $K \subseteq E_t$ was an arbitrary compact subset, Borel regularity of μ implies that

$$\mu(E_t) = \sup\{\mu(K) : K \subseteq E_t, K \text{ compact}\} \leqslant \frac{C_n}{t} \int_{\mathbb{R}^n} |f| \, \mathrm{d}\mu$$

as desired. $\square$

5.4   Densities

Now that we are armed with the powerful Besicovich Covering Theorem, we can show that the Radon-Nikodym derivative can be expressed as a limit of ratios of measures in many

cases of interest. For this, we will need to define the notion of *density* of a measure at a point.

In this entire section, (X, d) will be a metric space.

5.4.1 Definitions and the Upper Density Theorem

Definition 5.4.1 (*n*-Dimensional Upper and Lower Densities)**.** Let μ be an outer measure on X, and let $n \in \mathbb{Z}^+$. For a subset $A \subseteq X$ and a point $x \in X$, the *n-dimensional upper density* of μ at x with respect to A is defined as

$$\Theta^{*n}(\mu, A, x) := \limsup_{r \to 0} \frac{\mu\big(A \cap \overline{B}(x, r)\big)}{\omega_n r^n},$$

and the *n-dimensional lower density* of μ at x with respect to A is defined as

$$\Theta_*^n(\mu, A, x) := \liminf_{r \to 0} \frac{\mu\big(A \cap \overline{B}(x, r)\big)}{\omega_n r^n},$$

where ω_n is the volume of the unit ball in $\mathbb{R}^n$.

In the case $A = X$, we simply write $\Theta^{*n}(\mu, x)$ and $\Theta_*^n(\mu, x)$ for the upper and lower densities, respectively.

If the upper and lower densities at a point $x \in X$ are equal, we call the common value the *n-dimensional density* of μ at x with respect to A, denoted

$$\Theta^n(\mu, A, x) := \lim_{r \to 0} \frac{\mu\big(A \cap \overline{B}(x, r)\big)}{\omega_n r^n}.$$

Exercise 5.4.1 (Measurability of Densities)**.** Let μ be a Borel outer measure on X such that $\mu(B) < \infty$ for each closed ball $B \subseteq X$. Then for each $r > 0$ and each $x \in X$, we have

$$\mu(A \cap \overline{B}(x, r)) \geqslant \limsup_{y \to x} \mu(A \cap \overline{B}(y, r)).$$

That is, for each $r > 0$ the function

$$X \ni x \longmapsto \frac{\mu(A \cap \overline{B}(x, r))}{\omega_n r^n} \in [0, \infty]$$

is upper semicontinuous.

Conclude that the lower n-dimensional density $\Theta_*^n(\mu, A, \cdot) : X \to [0, \infty]$ is a Borel measurable function.

Also the upper n-dimensional density $\Theta^{*n}(\mu, A, \cdot) : X \to [0, \infty]$ is a Borel measurable function.

Proof. Fix $r > 0$ and let $x \in X$. Let $\{y_k\}_{k=1}^{\infty} \subseteq X$ be a sequence such that $y_k \to x$ as $k \to \infty$. Then for each $\epsilon > 0$, there exists $N \in \mathbb{Z}^+$ such that

$$\overline{B}(y_k, r) \subseteq \overline{B}(x, r + \epsilon) \quad \forall k \geqslant N$$

and hence

$$\mu(A \cap \overline{B}(y_k, r)) \leqslant \mu(A \cap \overline{B}(x, r + \epsilon))$$

by monotonicity of μ. Taking the limit superior as $k \to \infty$ gives

$$\limsup_{k \to \infty} \mu(A \cap \overline{B}(y_k, r)) \leqslant \mu(A \cap \overline{B}(x, r + \epsilon)).$$

Since this holds for each $\epsilon > 0$, we get

$$\limsup_{r \to 0} \frac{\mu(A \cap \overline{B}(x, r))}{\omega_n r^n} \leqslant \mu(A \cap \overline{B}(x, r))$$

by 1.1.6. Since $\{y_k\}_{k=1}^{\infty}$ was an arbitrary sequence which converged to x, we have shown that

$$\mu(A \cap \overline{B}(x, r)) \geqslant \limsup_{y \to x} \mu(A \cap \overline{B}(y, r)).$$

Since $x \in X$ was arbitrary and $r > 0$ is fixed, this shows that the function

$$X \ni x \longmapsto \frac{\mu(A \cap \overline{B}(x, r))}{\omega_n r^n} \in [0, \infty]$$

is upper semicontinuous.

Thus for each $m \in \mathbb{Z}^+$, the function

$$X \ni x \longmapsto \inf_{0 < r < \frac{1}{m}} \frac{\mu(A \cap \overline{B}(x, r))}{\omega_n r^n}$$

is also upper semicontinuous, and hence Borel measurable. As a result, the lower n-dimensional density

$$\Theta_*^n(\mu, A, x) = \lim_{m \to \infty} \inf_{0 < r < \frac{1}{m}} \frac{\mu(A \cap \overline{B}(x, r))}{\omega_n r^n}$$

is a Borel measurable function on X.

We claim that for each $r > 0$ the map

$$X \ni x \longmapsto \frac{\mu(A \cap B(x, r))}{\omega_n r^n} \in [0, \infty]$$

is lower semicontinuous.

To see this, let $x \in X$ and let $\{y_k\}_{k=1}^{\infty} \subseteq X$ be a sequence such that $y_k \to x$ as $k \to \infty$. Then for each $\epsilon > 0$, there exists $N \in \mathbb{Z}^+$ such that

$$B(x, r - \epsilon) \subseteq B(y_k, r) \quad \forall k \geqslant N$$

and hence

$$\mu(A \cap B(y_k, r)) \geqslant \mu(A \cap B(x, r - \epsilon))$$

by monotonicity of μ. Taking the limit infimum as $k \to \infty$ gives

$$\liminf_{k \to \infty} \mu(A \cap B(y_k, r)) \geqslant \mu(A \cap B(x, r - \epsilon)).$$

Since this holds for each $\epsilon > 0$, we get

$$\liminf_{r \to 0} \frac{\mu(A \cap B(x, r))}{\omega_n r^n} \geqslant \mu(A \cap B(x, r))$$

by 1.1.6. Since $\{y_k\}_{k=1}^{\infty}$ was an arbitrary sequence which converged to x, we have shown that

$$\mu(A \cap B(x, r)) \leqslant \liminf_{y \to x} \mu(A \cap B(y, r)).$$

Since $x \in X$ was arbitrary and $r > 0$ is fixed, this shows that the function

$$X \ni x \longmapsto \frac{\mu(A \cap B(x, r))}{\omega_n r^n} \in [0, \infty]$$

is lower semicontinuous.

As a result, for each $m \in \mathbb{Z}^+$, the function

$$X \ni x \longmapsto \sup_{0 < r < \frac{1}{m}} \frac{\mu(A \cap B(x, r))}{\omega_n r^n}$$

is also lower semicontinuous, and hence Borel measurable. Thus the upper n-dimensional density

$$\Theta^{*n}(\mu, A, x) = \lim_{m \to \infty} \sup_{0 < r < \frac{1}{m}} \frac{\mu(A \cap B(x, r))}{\omega_n r^n}$$

is a Borel measurable function on X. $\qquad\square$

If we have a certain inquality for the n-dimensional upper density, then we can relate the measure μ to the n-dimensional Hausdorff measure $\mathcal{H}^n$.

Lemma 5.4.2 (Comparison Lemma). *Let μ be a Borel regular outer measure on X, and let $t \geqslant 0$ and $A_1 \subseteq A_2 \subseteq X$. Then*

$$\Theta^{*n}(\mu, A_2, x) \geqslant t \quad \forall x \in A_1 \implies t\mathcal{H}^n(A_1) \leqslant \mu(A_2).$$

Of course an important special case is when $A_1 = A_2$. Note that we *do not* assume that A_1 and A_2 are Borel measurable.

Proof. If either $t = 0$ or $\mu(A_2) = \infty$, the inequality is trivial, so we may assume that $t > 0$ and $\mu(A_2) < \infty$.

Fix $\tau \in (0, t)$ so that

$$\Theta^{*n}(\mu, A_2, x) \geqslant \tau \quad \forall x \in A_1.$$

Also let $\delta > 0$ be arbitrary and define

$$\mathcal{B}_\delta := \left\{ \overline{B}(x, r) : x \in A_1, 0 < r < \frac{\delta}{2}, \mu(A_2 \cap \overline{B}(x, r)) \geqslant \tau \omega_n r^n \right\}.$$

Then $\mathcal{B}_\delta$ is a collection of closed balls which covers A_1, the diameters of the balls in $\mathcal{B}_\delta$ are uniformly bounded above by δ, and for each $a \in A_1$ we have

$$\inf\{r : \overline{B}(a, r) \in \mathcal{B}_\delta\} = 0.$$

By the Vitali Covering Lemma 4.1.3, there exists a disjoint subcover $\mathcal{B}'_\delta \subseteq \mathcal{B}_\delta$ such that

$$A_1 \subseteq \bigcup_{B \in \mathcal{B}'_\delta} 5B.$$

We claim that $\mathcal{B}'_\delta$ is countable. If X is separable, then this is immediate as no uncountable collection of disjoint balls can exist in a separable metric space. If X is not separable, then we can argue as follows. Since $\mu(A_2 \cap B) > 0$ for each $B \in \mathcal{B}'_\delta$, and

$$B_1, B_2, \ldots, B_N \in \mathcal{B}'_\delta \implies \sum_{j=1}^N \mu(A_2 \cap B_j) = \mu\left(A_2 \cap \bigcup_{j=1}^N B_j\right) \leqslant \mu(A_2) < \infty$$

it follows that $\mathcal{B}'_\delta$ is at most countable. Thus

$$A_1 \backslash \bigcup_{j=1}^N B_j \subseteq \bigcup_{B \in \mathcal{B}'_\delta \backslash \{B_1, \ldots, B_N\}} 5B$$

for each $N \in \mathbb{Z}^+$ because $\inf\{r : \overline{B}(a, r) \in \mathcal{B}_\delta\} = 0$ for each $a \in A_1$. Also

$$\tau \sum_{j=1}^\infty \omega_n r_j^n \leqslant \sum_{j=1}^\infty \mu(A_2 \cap \overline{B}(x_j, r_j))$$

$$= \mu\left(A_2 \cap \bigcup_{j=1}^\infty \overline{B}(x_j, r_j)\right)$$

$$\leqslant \mu(A_2) < \infty.$$

As a result, we see that

$$A_2 \subseteq \left(\bigcup_{j=1}^N B_j\right) \cup \left(\bigcup_{j=N+1}^\infty 5B_j\right)$$

for each $N \in \mathbb{Z}^+$, and hence the definition of $\mathcal{H}^n_{5\delta}$ gives

$$\mathcal{H}^n_{5\delta}(A_1) \leqslant \sum_{j=1}^{N} \omega_n r_j^n + 5^n \sum_{j=N+1}^{\infty} \omega_n r_j^n.$$

Taking the limit as $N \to \infty$, the first term converges to $\sum_{j=1}^{\infty} \omega_n r_j^n$ and the second term converges to 0, so we get

$$\mathcal{H}^n_{5\delta}(A_1) \leqslant \sum_{j=1}^{\infty} \omega_n r_j^n \leqslant \frac{1}{\tau} \mu(A_2).$$

Since $\delta > 0$ was arbitrary, taking the limit as $\delta \to 0^+$ and then letting $\tau \to t^-$ gives

$$\mathcal{H}^n_{5\delta}(A_1) \leqslant \frac{1}{t} \mu(A_2)$$

which completes the proof. $\qquad \square$

Theorem 5.4.3 (Upper Density Theorem). *Let μ be a Borel regular outer measure on X, and let $A \subseteq X$ be a Borel measurable set with $\mu(A) < \infty$. Then for $\mathcal{H}^n$-almost every $x \in X \backslash A$, we have*

$$\Theta^{*n}(\mu, A, x) = 0.$$

Proof. Let $t \geqslant 0$ and define

$$E_t := \{x \in X \backslash A : \Theta^{*n}(\mu, A, x) > t\}$$

and let $C \subseteq A$ be an arbitrary closed subset. Since $X \backslash C$ is an open set and

$$E_t \subseteq X \backslash A \subseteq X \backslash C$$

we see that

$$\Theta^{*n}(\mu, A \cap (X \backslash C), x) = \Theta^{*n}(\mu, A, x) \geqslant t$$

for each $x \in E_t$. Thus by the Comparison Lemma 5.4.2, we have

$$t \mathcal{H}^n(E_t) \leqslant \mu(A \backslash C).$$

Since $C \subseteq A$ was an arbitrary closed subset, we note this holds for each closed subset of A.

By Borel regularity of μ, there exists a sequence of closed sets $\{C_j\}_{j=1}^{\infty}$ such that $C_j \subseteq A$ for each $j \in \mathbb{Z}^+$ and

$$\mu(A) = \lim_{j \to \infty} \mu(C_j).$$

Thus taking the limit as $j \to \infty$ gives

$$t \mathcal{H}^n(E_t) \leqslant \lim_{j \to \infty} \mu(A \backslash C_j) = \mu(A) - \lim_{j \to \infty} \mu(C_j) = 0.$$

Since $t \geqslant 0$ was arbitrary, we conclude that

$$\mathcal{H}^n(E_t) = 0 \qquad \forall\, t > 0$$

which implies that

$$\mathcal{H}^n(\{x \in X \backslash A : \Theta^{*n}(\mu, A, x) > 0\}) = 0.$$

and hence

$$\Theta^{*n}(\mu, A, x) = 0 \quad \text{for } \mathcal{H}^n\text{-a.e. } x \in X \backslash A.$$

$\square$

Exercise 5.4.2 (Upper Density Theorem for σ-Finite Measures)**.** Show that in the previous theorem, you can drop the hypothesis $\mu(A) < \infty$ if you assume that (X, μ) is open σ-finite.

Proof. Assume that (X, μ) is open σ-finite, so there exists a countable collection of open sets $\{U_j\}_{j=1}^{\infty}$ such that $X = \bigcup_{j=1}^{\infty} U_j$ and $\mu(U_j) < \infty$ for each $j \in \mathbb{Z}^+$. Then for each $j \in \mathbb{Z}^+$, we consider the measure $\mu \restriction U_j$ which satisfies $(\mu \restriction U_j)(X) = \mu(U_j) < \infty$; thus we may apply Theorem 5.4.3 to conclude that

$$\Theta^{*n}(\mu \restriction U_j, A, x) = \Theta^{*n}(\mu, A \cap U_j, x) = 0 \quad \text{for } \mathcal{H}^n\text{-a.e. } x \in U_j \backslash A.$$

Then we have $X \backslash A = \bigcup_{j=1}^{\infty} (U_j \backslash A)$, so it follows that

$$\Theta^{*n}(\mu, A, x) = 0 \quad \text{for } \mathcal{H}^n\text{-a.e. } x \in X \backslash A.$$

$\square$

Corollary 5.4.4 (Density of Lebesgue Measure)**.** *If $A \subset \mathcal{L}^n$ is Lebesgue measurable, then the density $\Theta^n(\mathcal{L}^n, A, x)$ exists for $\mathcal{L}^n$-almost every $x \in \mathbb{R}^n$, and*

$$\Theta^n(\mathcal{L}^n, A, x) = 1 \quad \text{for } \mathcal{L}^n\text{-a.e. } x \in A,$$

and

$$\Theta^n(\mathcal{L}^n, A, x) = 0 \quad \text{for } \mathcal{L}^n\text{-a.e. } x \in \mathbb{R}^n \backslash A.$$

Proof. Since A is Lebesgue measurable, we have

$$\mathcal{L}^n(\overline{B}(x, r)) = \mathcal{L}^n(A \cap \overline{B}(x, r)) + \mathcal{L}^n(\overline{B}(x, r) \backslash A) \qquad \forall\, x \in \mathbb{R}^n, r > 0$$

which implies that

$$1 = \frac{\mathcal{L}^n(\overline{B}(x, r))}{\omega_n r^n} = \frac{\mathcal{L}^n(A \cap \overline{B}(x, r))}{\omega_n r^n} + \frac{\mathcal{L}^n(A \backslash \overline{B}(x, r))}{\omega_n r^n} \qquad \forall\, x \in \mathbb{R}^n, r > 0.$$

Taking the limit superior as $r \to 0^+$, the Upper Density Theorem 5.4.3 implies that the first term on the right-hand side converges to $\Theta^{*n}(\mathcal{L}^n, A, x)$ and the second term converges to $\Theta^{*n}(\mathcal{L}^n, \mathbb{R}^n \backslash A, x)$ for $\mathcal{L}^n$-almost every $x \in \mathbb{R}^n$. That is,

$$1 = \Theta^{*n}(\mathcal{L}^n, A, x) + 0 \qquad \text{for } \mathcal{L}^n\text{-a.e. } x \in A$$

and

$$1 = 0 + \Theta^{*n}(\mathcal{L}^n, \mathbb{R}^n \backslash A, x) \qquad \text{for } \mathcal{L}^n\text{-a.e. } x \in \mathbb{R}^n \backslash A.$$

Thus the limit defining the density $\Theta^n(\mathcal{L}^n, A, x)$ exists for $\mathcal{L}^n$-almost every $x \in \mathbb{R}^n$, and the desired equalities hold. $\qquad \square$

5.4.2 The Symmetric Vitali Property

We want to generalize the Comparison Lemma 5.4.2 and the Upper Density Theorem 5.4.3.

Exercise 5.4.3 (The Set where a Measure Vanishes is Open). Let μ be a Borel regular outer measure on X. Then the set

$$U_\mu := \{x \in X : \mu(\overline{B}(x, r)) = 0 \text{ for some } r > 0\}$$

is open. If X is a separable metric space, then $\mu(U_\mu) = 0$.

Proof. Let $x \in U_\mu$ be arbitrary. Then there exists $r > 0$ such that $\mu(\overline{B}(x, r)) = 0$. If $y \in B(x, r)$, then $d(x, y) < r$ and $\overline{B}(y, r - d(x, y)) \subseteq \overline{B}(x, r)$ by the triangle inequality, so by monotonicity of μ we have

$$\mu(\overline{B}(y, r - d(x, y))) \leqslant \mu(\overline{B}(x, r)) = 0$$

which implies that $y \in U_\mu$. Thus $B(x, r) \subseteq U_\mu$, showing that U_μ is open.

Now assume that X is a separable metric space. Then U_μ is also separable, so there exists a countable dense subset $\{x_j\}_{j=1}^\infty \subseteq U_\mu$. For each $j \in \mathbb{Z}^+$, there exists $r_j > 0$ such that $\mu(\overline{B}(x_j, r_j)) = 0$. Thus

$$U_\mu = \bigcup_{j=1}^\infty \overline{B}(x_j, r_j)$$

by density of $\{x_j\}_{j=1}^\infty$ in U_μ, so by countable subadditivity of μ we have

$$\mu(U_\mu) \leqslant \sum_{j=1}^\infty \mu(\overline{B}(x_j, r_j)) = 0.$$

$\qquad \square$

Definition 5.4.5 (Upper Densities with respect to a Measure). Let μ and μ_0 be Borel regular outer measures on X, and assume that μ_0 is locally finite. Let

$$U_{\mu_0} := \{x \in X : \mu_0(\overline{B}(x,r)) = 0 \text{ for some } r > 0\}$$

and

$$U_\mu := \{x \in X : \mu(\overline{B}(x,r)) = 0 \text{ for some } r > 0\}.$$

We define the *upper density* of μ with respect to μ_0 by

$$\Theta^{*\mu_0}(\mu,x) := \begin{cases} \limsup\limits_{r\to 0^+} \dfrac{\mu(\overline{B}(x,r))}{\mu_0(\overline{B}(x,r))}, & \text{if } x \in X \backslash (U_{\mu_0} \cup U_\mu), \\ \infty, & \text{if } x \in U_{\mu_0} \backslash U_\mu, \\ 0, & \text{if } x \in U_\mu. \end{cases}$$

Note that the set $U_{\mu_0} \backslash U_\mu$ in the above definition is not the only way that $\Theta^{*\mu_0}(\mu,x)$ can be infinite; it can also be infinite if the limit superior diverges to infinity.

Remark 5.4.6 (Upper Densities with respect to Lebesgue Measure on $\mathbb{R}^n$). Notice that if $X = \mathbb{R}^n$ and $\mu_0 = \mathcal{L}^n$ is Lebesgue measure on $\mathbb{R}^n$, then the definition of $\Theta^{*\mu_0}(\mu,x)$ agrees with the definition of the n-dimensional upper density $\Theta^{*n}(\mu,x)$ from Definition 5.4.1, i.e., we have

$$\Theta^{*\mathcal{L}^n}(\mu,x) = \Theta^{*n}(\mu,x) \qquad \forall x \in \mathbb{R}^n.$$

Remark 5.4.7 (Conditions to Ensure $\mu_0(U_{\mu_0}) = 0$.). There are various conditions one can impose on μ_0 to ensure that $\mu_0(U_{\mu_0}) = 0$. For example, if X is a seperable metric space then this was shown in Exercise 5.4.3.

Also if (X, μ_0) is open σ-finite and μ_0 has the Symmetric Vitali Property (Definition 5.4.8 below), then $\mu_0(U_{\mu_0}) = 0$. See Exercise 5.4.5 for the proof.

Definition 5.4.8 (Symmetric Vitali Property). Let μ be a Borel outer measure on X. We say that μ has the *symmetric Vitali property* if for each subset $A \subseteq X$ such that $\mu(A) < \infty$, and for each collection $\mathcal{B}$ of closed balls in X such that the center of each ball in $\mathcal{B}$ is in A and

$$\inf\{r : \overline{B}(a,r) \in \mathcal{B}\} = 0$$

for each $a \in A$, there exists a countable disjoint subcollection $\{B_j\}_{j=1}^\infty \subseteq \mathcal{B}$ such that

$$\mu\left(A \backslash \bigcup_{j=1}^\infty B_j\right) = 0.$$

An informal way to think about the symmetric Vitali property is that we can "almost" cover an arbitrary finite measure set A by disjoint balls with centers in A.

Example 5.4.9 (Borel Regular Measures on $\mathbb{R}^n$ have the Symmetric Vitali Property). Each Borel regular outer measure on $\mathbb{R}^n$ which is finite on compact sets has the symmetric Vitali property — this is exactly what was proven in Lemma 5.3.2 by using the Besicovich Covering Theorem **??**.

Exercise 5.4.4 (Finite Borel Regular Measures with a Doubling Condition have the Symmetric Vitali Property). Let μ be a Borel regular outer measure on X such that $\mu(X) < \infty$, and assume that there exists a constant $\mathbf{b} > 0$ such that

$$\mu(2B) \leqslant \mathbf{b}\mu(B) \qquad \text{for each closed ball } B \subseteq X. \tag{$\ddagger$}$$

Then μ has the symmetric Vitali property.

The condition ($\ddagger$) is called a *doubling condition*, and it says that we can uniformly control the measure of a ball by the measure of a ball with half the radius. The constant $\mathbf{b}$ is called a *doubling constant* for μ. Also note that some people prefer to state the doubling condition in terms tripling the radius (which is cleaner for using things like Vitali's Covering Theorem), but the two versions are equivalent up to changing the constant $\mathbf{b}$.

Proof. Let $A \subseteq X$ be such that $\mu(A) < \infty$, and let $\mathcal{B}$ be a collection of closed balls in X such that the center of each ball in $\mathcal{B}$ is in A and

$$\inf\{r : \overline{B}(a, r) \in \mathcal{B}\} = 0$$

for each $a \in A$. By Vitali's Covering Lemma (Infinite Version) 4.1.3, there exists a countable disjoint subcollection $\{B_j\}_{j=1}^{\infty} \subseteq \mathcal{B}$ such that

$$A \subseteq \bigcup_{j=1}^{\infty} 5B_j.$$

Since

$$\mu(5B_j) \leqslant \mu(8B_j) \leqslant \mathbf{b}^3 \mu(B_j)$$

for each $j \in \mathbb{Z}^+$ by applying the doubling condition ($\ddagger$) three times, we have

$$\mu\left(A \backslash \bigcup_{j=1}^{N} B_j\right) \leqslant \mu\left(\bigcup_{j=N+1}^{\infty} 5B_j\right) \leqslant \mathbf{b}^3 \sum_{j=N+1}^{\infty} \mu(B_j)$$

for each $N \in \mathbb{Z}^+$. Since $\sum_{j=1}^{\infty} \mu(B_j) \leqslant \mu(X) < \infty$, we have

$$\lim_{N \to \infty} \mu\left(A \backslash \bigcup_{j=1}^{N} B_j\right) = \lim_{N \to \infty} \sum_{j=N+1}^{\infty} \mu(B_j) = 0,$$

which implies that

$$\mu\left(A\backslash\bigcup_{j=1}^{\infty} B_j\right) = 0.$$

Since $\mathcal{B}$ and A were arbitrary, we conclude that μ has the symmetric Vitali property. $\qquad\square$

Exercise 5.4.5 (Measure of Set Where Measure Vanishes is Zero). Let μ be a Borel regular outer measure on X which is open σ-finite and has the symmetric Vitali property. Show that $\mu(U_\mu) = 0$ where

$$U_\mu := \{x \in X : \mu(\overline{B}(x,r)) = 0 \text{ for some } r > 0\}.$$

Proof. First we assume that $\mu(U_\mu) < \infty$. See that the collection of closed balls

$$\mathcal{B} := \{\overline{B}(x,r) : x \in U_\mu, r > 0, \mu(\overline{B}(x,r)) = 0\}$$

satisfies

$$\inf\{r : \overline{B}(a,r) \in \mathcal{B}\} = 0$$

for each $a \in U_\mu$. Since μ has the symmetric Vitali property, there exists a countable disjoint subcollection $\{B_j\}_{j=1}^{\infty} \subseteq \mathcal{B}$ such that

$$\mu\left(U_\mu\backslash\bigcup_{j=1}^{\infty} B_j\right) = 0.$$

But $\mu(B_j) = 0$ for each $j \in \mathbb{Z}^+$ by construction, so by countable subadditivity of μ we have

$$\mu(U_\mu) \leqslant \mu\left(U_\mu\backslash\bigcup_{j=1}^{\infty} B_j\right) + \sum_{j=1}^{\infty} \mu(B_j) = 0.$$

Now if $\mu(U_\mu) = \infty$, since μ is open σ-finite there exists a countable collection of open sets $\{U_k\}_{k=1}^{\infty}$ such that $X = \bigcup_{k=1}^{\infty} U_k$ and $\mu(U_k) < \infty$ for each $k \in \mathbb{Z}^+$. Then we have

$$U_\mu = \bigcup_{k=1}^{\infty}(U_\mu \cap U_k)$$

and $\mu(U_\mu \cap U_k) \leqslant \mu(U_k) < \infty$ for each $k \in \mathbb{Z}^+$, so by the previous case we have $\mu(U_\mu \cap U_k) = 0$ for each $k \in \mathbb{Z}^+$. Thus by countable subadditivity of μ we have

$$\mu(U_\mu) \leqslant \sum_{k=1}^{\infty} \mu(U_\mu \cap U_k) = 0.$$

$\square$

Exercise 5.4.6 (Symmetric Vitali Property and Open σ-Finiteness). Let μ be a Borel regular outer measure on X which is open σ-finite and has the symmetric Vitali property. Then the symmetric Vitali property holds for all subsets $A \subseteq X$ such that $\mu(A) = \infty$ as well.

Proof. Let $A \subseteq X$ be such that $\mu(A) = \infty$. Since μ is open σ-finite, there exists a countable collection of open sets $\{U_k\}_{k=1}^\infty$ such that $X = \bigcup_{k=1}^\infty U_k$ and $\mu(U_k) < \infty$ for each $k \in \mathbb{Z}^+$. Then we have

$$A = \bigcup_{k=1}^\infty (A \cap U_k)$$

and $\mu(A \cap U_k) \leqslant \mu(U_k) < \infty$ for each $k \in \mathbb{Z}^+$. By the symmetric Vitali property for finite measure sets, for each $k \in \mathbb{Z}^+$ there exists a countable disjoint subcollection $\left\{B_j^{(k)}\right\}_{j=1}^\infty$ of closed balls in $\mathcal{B}$ with centers in $A \cap U_k$ such that

$$\mu\left((A \cap U_k) \backslash \bigcup_{j=1}^\infty B_j^{(k)}\right) = 0.$$

Then the collection

$$\left\{B_j^{(k)} : j, k \in \mathbb{Z}^+\right\}$$

is a countable disjoint subcollection of $\mathcal{B}$ with centers in A such that

$$\mu\left(A \backslash \bigcup_{k=1}^\infty \bigcup_{j=1}^\infty B_j^{(k)}\right) \leqslant \mu\left(\bigcup_{k=1}^\infty \left((A \cap U_k) \backslash \bigcup_{j=1}^\infty B_j^{(k)}\right)\right)$$

$$\leqslant \sum_{k=1}^\infty \mu\left((A \cap U_k) \backslash \bigcup_{j=1}^\infty B_j^{(k)}\right) = 0.$$

$\square$

Lemma 5.4.10 (General Comparison Lemma). *Let μ and μ_0 be Borel regular outer measures on X which are both open σ-finite, and assume that μ_0 has the symmetric Vitali property. Also let $t \geqslant 0$ and $A \subseteq X$. Then*

$$\Theta^{*\mu_0}(\mu, x) \geqslant t \quad \forall\, x \in A \quad \implies \quad \mu(A) \geqslant t \cdot \mu_0(A).$$

Proof. The proof is similar to that of the Comparison Lemma 5.4.2, but we use the Symmetric Vitali Property in place of the Vitali Covering Lemma.

Let U_{μ_0} and U_μ be as in Definition 5.4.5. Because (X, μ_0) is open σ-finite and μ_0 has the symmetric Vitali property, we have $\mu_0(U_{\mu_0}) = 0$ by Exercise 5.4.5. Also the desired inequality is trivial if $t = 0$ so we assume that $t > 0$. Let $A \subseteq X$ be a subset such that

$$\Theta^{*\mu_0}(\mu, x) \geqslant t \quad \forall\, x \in A.$$

Let $U \subseteq X$ be an open set such that $A \subseteq U$; fix $\tau \in (0, t)$.

We consider the collection of closed balls

$$\mathcal{B} := \left\{ \overline{B}(x, r) : x \in X \cap (X \setminus U_{\mu_0}), \overline{B}(x, r) \subset U, \text{ and } \mu(\overline{B}(x, r)) > \tau \mu_0(\overline{B}(x, r)) \right\}.$$

See that for each $a \in A \setminus U_{\mu_0}$, we have

$$\Theta^{*\mu_0}(\mu, a) = \limsup_{r \to 0^+} \frac{\mu(\overline{B}(a, r))}{\mu_0(\overline{B}(a, r))} \geqslant t > \tau$$

which implies that there exists $r_a > 0$ such that for all $0 < r < r_a$ we have

$$\mu(\overline{B}(a, r)) > \tau \mu_0(\overline{B}(a, r)).$$

Thus

$$\inf\{r : \overline{B}(a, r) \in \mathcal{B}\} = 0$$

for each $a \in A \setminus U_{\mu_0}$.

Since μ_0 has the symmetric Vitali property, there exists a countable disjoint subcollection $\{B_j\}_{j=1}^{\infty} \subseteq \mathcal{B}$ such that

$$\mu_0 \left((A \setminus U_{\mu_0}) \setminus \bigcup_{j=1}^{\infty} B_j \right) = 0.$$

(By Exercise 5.4.6, we can apply the symmetric Vitali property even if $\mu_0(A) = \infty$.) Thus

$$\mu_0 \left(A \setminus \bigcup_{j=1}^{\infty} B_j \right) \leqslant \mu_0 \left((A \setminus U_{\mu_0}) \setminus \bigcup_{j=1}^{\infty} B_j \right) + \mu_0(U_{\mu_0}) = 0$$

by using that $\mu_0(U_{\mu_0}) = 0$. By definition of $\mathcal{B}$, we have

$$\mu(B_j) > \tau \mu_0(B_j)$$

for each $j \in \mathbb{Z}^+$, so by adding we obtain

$$\tau \mu_0(A) \leqslant \tau \mu_0 \left(A \setminus \bigcup_{j=1}^{\infty} B_j \right) + \tau \mu_0 \left(\bigcup_{j=1}^{\infty} B_j \right)$$

$$= 0 + \tau \sum_{j=1}^{\infty} \mu_0(B_j) \qquad \text{by disjointness of } \{B_j\}_{j=1}^{\infty}$$

$$< \sum_{j=1}^{\infty} \mu(B_j) \qquad \text{by definition of } \mathcal{B}$$

$$= \mu \left(\bigcup_{j=1}^{\infty} B_j \right) \leqslant \mu(U) \qquad \text{since } B_j \subseteq U \text{ for each } j \in \mathbb{Z}^+.$$

Since $U \supseteq A$ was an arbitrary open set, Borel regularity of μ gives

$$\mu(A) \geqslant \tau \mu_0(A).$$

Taking the limit as $\tau \to t^-$ completes the proof. $\qquad\square$

Corollary 5.4.11 (Upper Density Finite almost everywhere). *Let μ and μ_0 be Borel regular outer measures on X which are both open σ-finite, and assume that μ_0 has the symmetric Vitali property. Then*

$$\Theta^{*\mu_0}(\mu, x) < \infty$$

for μ_0-almost every $x \in X$.

Proof. Before the proof, maybe you are naïve and think that this is obvious because our assumptions guaruntee that $\mu_0(U_{\mu_0}) = 0$ by Exercise 5.4.5, so we are done. A moments thought reveals that this is not the only way that $\Theta^{*\mu_0}(\mu, x)$ can be infinite; it can also be infinite if the limit superior diverges to infinity. Thus we need to do some work.

Since X is open σ-finite with respect to μ_0, there exists a countable collection of open sets $\{U_k\}_{k=1}^{\infty}$ such that $X = \bigcup_{k=1}^{\infty} U_k$ and $\mu_0(U_k) < \infty$ for each $k \in \mathbb{Z}^+$. For each $m \in \mathbb{Z}^+$, we consider $\mu \restriction U_j$ which is a Borel regular outer measure on X satisfying $(\mu \restriction U_j)(X) = \mu(U_j) < \infty$.

Fix $m \in \mathbb{Z}^+$ and $t > 0$ and define

$$A_{m,t} := \{x \in U_m \backslash U_{\mu_0} : \Theta^{*\mu_0}(\mu, x) \geqslant t\}.$$

By the General Comparison Lemma 5.4.10, we have

$$t\mu_0(A_{m,t}) \leqslant \mu_m(A_{m,t}) \leqslant \mu(U_m) < \infty$$

and thus

$$\mu_0(\{x \in U_m : \Theta^{*\mu_0}(\mu, x) = \infty\}) \leqslant \frac{1}{t}\mu(U_m).$$

Since this holds for each $t > 0$, we conclude that

$$\mu_0(\{x \in U_m : \Theta^{*\mu_0}(\mu, x) = \infty\}) = 0.$$

Since $X = \bigcup_{m=1}^{\infty} U_m$, by countable subadditivity of μ_0 we have

$$\mu_0(\{x \in X : \Theta^{*\mu_0}(\mu, x) = \infty\}) \leqslant \sum_{m=1}^{\infty} \mu_0(\{x \in U_m : \Theta^{*\mu_0}(\mu, x) = \infty\}) = 0.$$

$\qquad\square$

Theorem 5.4.12 (General Upper Density Theorem). *Let μ and μ_0 be Borel regular outer measures on X, and assume that (X, μ_0) is open σ-finite and that μ_0 has the symmetric Vitali property. If $A \subseteq X$ is a μ-measurable set with $\mu(A) < \infty$, then*

$$\Theta^{*\mu_0}(\mu \llcorner A, x) = 0$$

for μ_0-almost every $x \in X \backslash A$.

Proof. Assume that $A \subseteq X$ is a μ-measurable set with $\mu(A) < \infty$.

Let $t \geqslant 0$ and define

$$E_t := \{x \in X \backslash A : \Theta^{*\mu_0}(\mu \llcorner A, x) > t\}$$

and let $C \subseteq A$ be an arbitrary closed subset. Since $X \backslash C$ is open and

$$E_t \subset X \backslash A \subset X \backslash C$$

we see that

$$\Theta^{*\mu_0}(\mu \llcorner (A \cap (X \backslash C))) = \Theta^{*\mu_0}(\mu \llcorner A, x) \geqslant t$$

for each $x \in E_t$. By the General Comparison Lemma 5.4.10, we have

$$t\mu_0(E_t) \leqslant \mu \llcorner (A \cap (X \backslash C))(E_t) \leqslant \mu(A \backslash C)$$

and we note that this holds for each closed subset $C \subseteq A$.

Sincce μ is Borel regular, there is a sequence of closed sets $\{C_j\}_{j=1}^{\infty}$ such that $C_j \subseteq A$ for each $j \in \mathbb{Z}^+$ and

$$\lim_{j \to \infty} \mu(C_j) = \mu(A).$$

Thus

$$t\mu_0(E_t) \leqslant \lim_{j \to \infty} \mu(A \backslash C_j) = \mu(A) - \lim_{j \to \infty} \mu(C_j) = 0.$$

Since $t \geqslant 0$ was arbitrary, we conclude that

$$\mu_0(E_t) = 0 \qquad \forall\, t > 0.$$

Thus

$$\mu_0(\{x \in X \backslash A : \Theta^{*\mu_0}(\mu \llcorner A, x) > 0\}) = \mu_0\left(\bigcup_{n=1}^{\infty} E_{1/n}\right) \leqslant \sum_{n=1}^{\infty} \mu_0(E_{1/n}) = 0$$

and hence

$$\Theta^{*\mu_0}(\mu \llcorner A, x) = 0$$

for μ_0-almost every $x \in X \backslash A$. $\square$

We come to the final result in this section.

Theorem 5.4.13 (General Density Theorem)**.** *Let μ be a Borel regular outer measure on X which is open σ-finite and has the symmetric Vitali property. Then for each μ-measurable set $A \subseteq X$, we have*

$$\lim_{r \to 0^+} \frac{\mu(\overline{B}(x,r) \cap A)}{\mu(\overline{B}(x,r))} = \begin{cases} 1, & \text{for } \mu\text{-almost every } x \in A, \\ 0, & \text{for } \mu\text{-almost every } x \in X \backslash A. \end{cases}$$

Proof. Letting U_μ be the set

$$U_\mu = \{x \in X : \mu(\overline{B}(x,r)) = 0 \text{ for some } r > 0\}$$

we see that by Exercise 5.4.5, we have $\mu(U_\mu) = 0$.

We assume first that $\mu(X) < \infty$, and let $A \subseteq X$ be a μ-measurable set. For each $x \in X \backslash U_\mu$, since A is μ-measurable we have

$$\mu(\overline{B}(x,r)) = \mu(\overline{B}(x,r) \cap A) + \mu(\overline{B}(x,r) \backslash A) \qquad \forall\, r > 0$$

which implies that

$$1 = \frac{\mu(\overline{B}(x,r))}{\mu(\overline{B}(x,r))} = \frac{\mu(\overline{B}(x,r) \cap A)}{\mu(\overline{B}(x,r))} + \frac{\mu(\overline{B}(x,r) \backslash A)}{\mu(\overline{B}(x,r))}$$

for all $r > 0$.

Taking the limit superior as $r \to 0^+$, the General Upper Density Theorem 5.4.12 implies that the first term on the right converges to 0 for μ-almost every $x \in X \backslash A$ and the second term converges to 0 for μ-almost every $x \in A$. That is,

$$1 = 0 + \Theta^{*\mu}(\mu \llcorner (X \backslash A), x) \qquad \text{for } \mu\text{-a.e. } x \in X \backslash A$$

and

$$1 = \Theta^{*\mu}(\mu \llcorner A, x) + 0 \qquad \text{for } \mu\text{-a.e. } x \in A.$$

Thus the limit defining the density

$$\Theta^\mu(\mu, A, x) := \lim_{r \to 0^+} \frac{\mu(\overline{B}(x,r) \cap A)}{\mu(\overline{B}(x,r))}$$

exists for μ-almost every $x \in X$, and the desired equalities hold.

Removing the assumption that $\mu(X) < \infty$, we assume instead that μ is open σ-finite. Then there exists a countable collection of open sets $\{U_k\}_{k=1}^\infty$ such that $X = \bigcup_{k=1}^\infty U_k$ and $\mu(U_k) < \infty$ for each $k \in \mathbb{Z}^+$. For each $k \in \mathbb{Z}^+$, by the previous case we have

$$\Theta^\mu(\mu \llcorner U_k, x) = \begin{cases} 1, & \text{for } \mu\text{-almost every } x \in A \cap U_k, \\ 0, & \text{for } \mu\text{-almost every } x \in U_k \backslash A. \end{cases}$$

Since $X = \bigcup_{k=1}^{\infty} U_k$, we know that

$$\Theta^{\mu}(\mu, A, x) = \begin{cases} 1, & \text{for } \mu\text{-almost every } x \in A, \\ 0, & \text{for } \mu\text{-almost every } x \in X \backslash A. \end{cases}$$

$\square$

5.5   Densities and the Radon-Nikodym Theorem

In this section, we will finish the goal stated in the previous section of expressing the Radon-Nikodym derivative as a limit of ratios of measures.

5.5.1   Lebesgue-Besicovich Differentiation Theorem

First we want to show that densities can be used to prove a generalization of the Lebesgue Differentiation Theorem 4.2.4.

Theorem 5.5.1 (Lebesgue-Besicovich Differentiation Theorem). *Let μ be a Borel regular outer measure on X which is open σ-finite and has the symmetric Vitali property. Let $f : X \to \mathbb{R}$ be a locally integrable function with respect to μ, i.e. for each $x \in X$ there exists $r > 0$ such that*

$$\int_{\overline{B}(x,r)} |f| \, \mathrm{d}\mu < \infty.$$

Then for μ-almost every $x \in X$, we have

$$f(x) = \lim_{r \to 0^+} \frac{1}{\mu\big(\overline{B}(x,r)\big)} \int_{\overline{B}(x,r)} f \, \mathrm{d}\mu.$$

Proof. Assume first that $f \geqslant 0$. We define $\nu : 2^X \to [0, \infty]$ by

$$\nu(A) := \int_A f \, \mathrm{d}\mu$$

for each Borel set $A \subseteq X$, and

$$\nu(E) := \inf\{\nu(A) : A \subseteq X \text{ is a Borel set with } E \subseteq A\}.$$

We claim that ν is a well-defined Borel regular outer measure on X.

Proof of Claim. See that ν is well-defined, since if $A \subseteq X$ is a Borel set, then clearly

$$\inf \left\{\nu(\hat{A}) : \hat{A} \subseteq X \text{ is a Borel set with } A \subseteq \hat{A}\right\} \leqslant \nu(A)$$

since A is one such Borel set containing itself; also for each Borel set $\hat{A} \subseteq X$ with $A \subseteq \hat{A}$, we have

$$\nu(\hat{A}) = \int_{\hat{A}} f \, \mathrm{d}\mu \geqslant \int_{A} f \, \mathrm{d}\mu = \nu(A)$$

by monotonicity of the Lebesgue integral (Proposition 2.1.15 (iii)). Therefore if $A \subseteq X$ is a Borel set, then

$$\nu(A) = \inf\{\nu(\hat{A}) : \hat{A} \subseteq X \text{ is a Borel set with } A \subseteq \hat{A}\}$$

which shows that ν is well-defined.

To see that ν is a Borel regular outer measure, first note that $\nu(\varnothing) = 0$ and if $\{A_j\}_{j=1}^{\infty}$ is a countable collection of subsets of X, then

$$\nu\left(\bigcup_{j=1}^{\infty} A_j\right) = \int_{\bigcup_{j=1}^{\infty} A_j} f \, \mathrm{d}\mu \leqslant \sum_{j=1}^{\infty} \int_{A_j} f \, \mathrm{d}\mu = \sum_{j=1}^{\infty} \nu(A_j).$$

Therefore ν is an outer measure on X.

Now let $A \subseteq X$ be a Borel set. For an arbitrary subset $S \subseteq X$, we let $C \subset X$ be a Borel set such that $S \subseteq C$ and $\nu(S) = \nu(C)$. Then

$$\nu(S) = \nu(C) = \int_{C} f \, \mathrm{d}\mu = \int_{C \cap A} f \, \mathrm{d}\mu + \int_{C \cap A^c} f \, \mathrm{d}\mu = \nu(C \cap A) + \nu(C \cap A^c) \geqslant \nu(S \cap A) + \nu(S \cap A^c)$$

which proves that A is ν-measurable as S was an arbitrary subset of X. Thus ν is a Borel outer measure on X. The fact that ν is Borel regular follows from the definition of ν as the infimum over Borel sets. Therefore ν is a Borel regular outer measure on X. $\qquad\square$

For each $j, k \in \mathbb{Z}^+$, we define the sets

$$A_{j,k} := f^{-1}\left(\left[\frac{k-1}{j}, \frac{k}{j}\right]\right)$$

and observe that for each $j, k \in \mathbb{Z}^+$, the set $A_{j,k}$ is μ-measurable since f is measurable; hence, by the Density Theorem 5.4.13, for μ-almost every $x \in A_{j,k}$ we have

$$\lim_{r \to 0^+} \frac{\mu(\overline{B}(x,r) \cap A_{j,k})}{\mu(\overline{B}(x,r))} = 1 \qquad (\Diamond)$$

By the Upper Density Theorem 5.4.3, for μ-almost every $x \in A_{j,k}$ we have

$$\lim_{r \to 0^+} \frac{\nu(\overline{B}(x,r) \setminus A_{j,k})}{\mu(\overline{B}(x,r))} = 0. \qquad (\clubsuit)$$

Fix $j, k \in \mathbb{Z}^+$. Let $x \in A_{j,k}$ be a point where both ($\diamondsuit$) and ($\clubsuit$) hold, and let $r > 0$ be small enough so that $\int_{\overline{B}(x,r)} |f| \, \mathrm{d}\mu < \infty$. Then we have the chain of inequalities

$$\mu(\overline{B}(x,r) \cap A_{j,k}) \cdot \frac{k-1}{j} \leqslant \int_{\overline{B}(x,r) \cap A_{j,k}} f \, \mathrm{d}\mu = \nu(\overline{B}(x,r) \cap A_{j,k})$$

$$= \int_{\overline{B}(x,r)} f \, \mathrm{d}\mu - \nu(\overline{B}(x,r) \backslash A_{j,k})$$

$$\leqslant \mu(\overline{B}(x,r) \cap A_{j,k}) \cdot \frac{k}{j}. \qquad (\heartsuit)$$

Extracting the first and thirs terms of this chain of inequalities, we have

$$\frac{k-1}{j} \leqslant \frac{1}{\mu(\overline{B}(x,r) \cap A_{j,k})} \left(\int_{\overline{B}(x,r)} f \, \mathrm{d}\mu - \nu(\overline{B}(x,r) \backslash A_{j,k}) \right)$$

$$= \frac{\mu(\overline{B}(x,r))}{\mu(\overline{B}(x,r) \cap A_{j,k})} \cdot \left(\frac{1}{\mu(\overline{B}(x,r))} \int_{\overline{B}(x,r)} f \, \mathrm{d}\mu - \frac{\nu(\overline{B}(x,r) \backslash A_{j,k})}{\mu(\overline{B}(x,r))} \right). \qquad (\heartsuit_1)$$

by monotonicity of μ. Taking limits as $r \to 0^+$ in ($\heartsuit_1$) and using ($\diamondsuit$) and ($\clubsuit$), we see that

$$\frac{k-1}{j} \leqslant \liminf_{r \to 0^+} \frac{1}{\mu(\overline{B}(x,r))} \int_{\overline{B}(x,r)} f \, \mathrm{d}\mu. \qquad (\spadesuit_1)$$

Similarly we take the third and fourth terms of the chain of inequalities ($\heartsuit$) to see that

$$\frac{k}{j} \geqslant \frac{1}{\mu(\overline{B}(x,r) \cap A_{j,k})} \left(\int_{\overline{B}(x,r)} f \, \mathrm{d}\mu - \nu(\overline{B}(x,r) \backslash A_{j,k}) \right)$$

$$= \frac{\mu(\overline{B}(x,r))}{\mu(\overline{B}(x,r) \cap A_{j,k})} \cdot \left(\frac{1}{\mu(\overline{B}(x,r))} \int_{\overline{B}(x,r)} f \, \mathrm{d}\mu - \frac{\nu(\overline{B}(x,r) \backslash A_{j,k})}{\mu(\overline{B}(x,r))} \right). \qquad (\heartsuit_2)$$

Taking limits as $r \to 0^+$ in ($\heartsuit_2$) and using ($\diamondsuit$) and ($\clubsuit$), we see that

$$\frac{k}{j} \geqslant \limsup_{r \to 0^+} \frac{1}{\mu(\overline{B}(x,r))} \int_{\overline{B}(x,r)} f \, \mathrm{d}\mu. \qquad (\spadesuit_2)$$

Combining ($\spadesuit_1$) and ($\spadesuit_2$), we have

$$\frac{k-1}{j} \leqslant \liminf_{r \to 0^+} \frac{1}{\mu(\overline{B}(x,r))} \int_{\overline{B}(x,r)} f \, \mathrm{d}\mu \leqslant \limsup_{r \to 0^+} \frac{1}{\mu(\overline{B}(x,r))} \int_{\overline{B}(x,r)} f \, \mathrm{d}\mu \leqslant \frac{k}{j}.$$

Since $\frac{k-1}{j} \leqslant f \leqslant \frac{k}{j}$ on $A_{j,k}$, we finally estimate

$$f(x) - \frac{1}{j} \leqslant \frac{k-1}{j}$$

$$\leqslant \liminf_{r \to 0^+} \frac{1}{\mu(\overline{B}(x,r))} \int_{\overline{B}(x,r)} f \, \mathrm{d}\mu$$

$$\leqslant \limsup_{r \to 0^+} \frac{1}{\mu(\overline{B}(x,r))} \int_{\overline{B}(x,r)} f \, \mathrm{d}\mu$$

$$\leqslant \frac{k}{j} \leqslant f(x) + \frac{1}{j}.$$

Since μ-almost every $x \in A_{j,k}$ satisfies both ($\Diamond$) and ($\clubsuit$), we have shown that

$$f(x) - \frac{1}{j} \leqslant \liminf \frac{1}{\mu(\overline{B}(x,r))} \int_{\overline{B}(x,r)} f \; \mathrm{d}\mu \leqslant \limsup \frac{1}{\mu(\overline{B}(x,r))} \int_{\overline{B}(x,r)} f \; \mathrm{d}\mu \leqslant f(x) + \frac{1}{j}$$

For each $j \in \mathbb{Z}^+$ and for each $x \in X$ there exists $k \in \mathbb{Z}^+$ such that $x \in A_{j,k}$. Thus for each $j \in \mathbb{Z}^+$, for μ-almost every $x \in X$ we have

$$f(x) - \frac{1}{j} \leqslant \liminf_{r \to 0^+} \frac{1}{\mu(\overline{B}(x,r))} \int_{\overline{B}(x,r)} f \; \mathrm{d}\mu \leqslant \limsup_{r \to 0^+} \frac{1}{\mu(\overline{B}(x,r))} \int_{\overline{B}(x,r)} f \; \mathrm{d}\mu \leqslant f(x) + \frac{1}{j}.$$

Taking $j \to \infty$, we conclude that for μ-almost every $x \in X$,

$$f(x) = \lim_{r \to 0^+} \frac{1}{\mu(\overline{B}(x,r))} \int_{\overline{B}(x,r)} f \; \mathrm{d}\mu.$$

$\square$

Corollary 5.5.2 (Lebesgue Points). *Let μ be a Borel regular outer measure on X which is open σ-finite and has the symmetric Vitali property, and let $1 \leqslant p < \infty$. Let $f : X \to \mathbb{R}$ be a locally p-integrable function with respect to μ, i.e. for each $x \in X$ there exists $r > 0$ such that*

$$\int_{\overline{B}(x,r)} |f|^p \; \mathrm{d}\mu < \infty.$$

Then for μ-almost every $x \in X$, we have

$$\lim_{r \to 0^+} \frac{1}{\mu\big(\overline{B}(x,r)\big)} \int_{\overline{B}(x,r)} |f(y) - f(x)|^p \; \mathrm{d}\mu(y) = 0. \tag{\dagger}$$

Points $x \in X$ satisfying ($\dagger$) are called Lebesgue points of f with respect to μ.

Proof. Let $\{q_j\}_{j=1}^{\infty}$ be an enumeration of the rational numbers. Then for each $j \in \mathbb{Z}^+$, the function $|f - q_j|^p$ is a locally integrable function with respect to μ, so by the Lebesgue-Besicovich Differentiation Theorem 5.5.1, for μ-almost every $x \in X$ we have

$$|f(x) - q_j|^p = \lim_{r \to 0^+} \frac{1}{\mu(\overline{B}(x,r))} \int_{\overline{B}(x,r)} |f(y) - q_j|^p \; \mathrm{d}\mu(y).$$

Thus there exists a set $A \subset X$ with $\mu(A) = 0$ such that for each $j \in \mathbb{Z}^+$ and for each $x \in X \backslash A$, we have

$$|f(x) - q_j|^p = \lim_{r \to 0^+} \frac{1}{\mu(\overline{B}(x,r))} \int_{\overline{B}(x,r)} |f(y) - q_j|^p \; \mathrm{d}\mu(y)$$

for each $j \in \mathbb{Z}^+$.

Now let $x \in X \backslash A$ be arbitrary, and let $\varepsilon > 0$. Choose $j \in \mathbb{Z}^+$ such that

$$|f(x) - q_j|^p < \frac{\varepsilon}{2^p}.$$

Then

$$
\begin{aligned}
\limsup_{r \to 0^+} \frac{1}{\mu(\overline{B}(x,r))} \int_{\overline{B}(x,r)} |f(y) - f(x)|^p \, \mathrm{d}\mu(y) &\leqslant \limsup_{r \to 0^+} \frac{2^{p-1}}{\mu(\overline{B}(x,r))} \int_{\overline{B}(x,r)} |f(y) - q_j|^p + |q_j - f(x)|^p \, \mathrm{d}\mu(y) \\
&= 2^{p-1} \bigg(\limsup_{r \to 0^+} \frac{1}{\mu(\overline{B}(x,r))} \int_{\overline{B}(x,r)} |f(y) - q_j|^p \, \mathrm{d}\mu(y) \\
&\quad + \limsup_{r \to 0^+} \frac{1}{\mu(\overline{B}(x,r))} \int_{\overline{B}(x,r)} |q_j - f(x)|^p \, \mathrm{d}\mu(y) \bigg) \\
&= 2^{p-1} \left(|f(x) - q_j|^p + |q_j - f(x)|^p \right) \\
&= 2^p |f(x) - q_j|^p < \varepsilon.
\end{aligned}
$$

Since $\varepsilon > 0$ was arbitrary, we conclude that

$$\lim_{r \to 0^+} \frac{1}{\mu(\overline{B}(x,r))} \int_{\overline{B}(x,r)} |f(y) - f(x)|^p \, \mathrm{d}\mu(y) = 0.$$

Since this holds for each $x \in X \backslash A$ and $\mu(A) = 0$, we have shown that for μ-almost every $x \in X$, x is a Lebesgue point of f with respect to μ. $\qquad\square$

Notation 5.5.3 (Averaged Integral). With μ and f as in Corollary 5.5.2, we will sometimes use the notation

$$\fint_{\overline{B}(x,r)} f \, \mathrm{d}\mu := \frac{1}{\mu(\overline{B}(x,r))} \int_{\overline{B}(x,r)} f \, \mathrm{d}\mu$$

to denote the *averaged integral* of f over the closed ball $\overline{B}(x,r)$ with respect to the outer measure μ.

More generally, if $A \subseteq X$ is a μ-measurable set with $\mu(A) > 0$ then we write

$$\fint_A f \, \mathrm{d}\mu := \frac{1}{\mu(A)} \int_A f \, \mathrm{d}\mu$$

to denote the averaged integral of f over the set A with respect to the outer measure μ.

5.5.2   The Lebesgue-Radon-Nikodym Theorem via Densities

We begin with a result that is related to the Lebesgue Decomposition Theorem 5.1.21.

Lemma 5.5.4 (Absolute Continuity Lemma)**.** *Let μ and μ_0 be Borel regular outer measures on X, and assume that (X, μ) is σ-finite. Then there is a Borel set $Z \subseteq X$ such that $\mu_0(Z) = 0$ and $\mu \restriction (X \backslash Z)$ is absolutely continuous with respect to μ_0. The Borel set Z is*

unique up to μ-measure zero, and as a result the Borel regular outer measure $\mu \lfloor (X \backslash Z)$ is uniquely determined.

The measure $\mu \lfloor (X \backslash Z)$ is called the absolutely continuous part of μ with respect to μ_0.

Remark 5.5.5 (Relation to the Lebesgue Decomposition). Recall that, considering μ and μ_0 as measures on the measurable space $(X, \mathcal{B}(X))$, by the Lebesgue Decomposition Theorem 5.1.21 there exist unique measures μ_{ac} and μ_s on $(X, \mathcal{B}(X))$ such that $\mu = \mu_{ac} + \mu_s$, where μ_{ac} is absolutely continuous with respect to μ_0 and μ_s is singular with respect to μ_0. By uniqueness of the absolutely continuous part of μ with respect to μ_0 in Lemma 5.5.4, we see that the measure $\mu \lfloor (X \backslash Z)$ in Lemma 5.5.4 coincides with the absolutely continuous part μ_{ac} of μ with respect to μ_0 in the Lebesgue Decomposition Theorem 5.1.21.

Proof. *Step 1:* First we assume that $\mu(X) < \infty$.

We define the set

$$\mathcal{A} := \{A \subseteq X : \mu_0(A) = 0\}$$

and let

$$\alpha := \sup\{\mu(A) : A \in \mathcal{A}\}.$$

Choose a sequence $\{A_j\}_{j=1}^{\infty} \subset \mathcal{A}$ such tha $\lim_{j \to \infty} \mu(A_j) = \alpha$. Then we have $\bigcup_{j=1}^{\infty} A_j \in \mathcal{A}$ since μ_0 is countably subadditive. We claim that

$$Z := \bigcup_{j=1}^{\infty} A_j$$

is the desired Borel set.

To see this, assume that $S \subseteq X$ is such that $\mu_0(S) = 0$. Then by Borel regularity of μ_0, there exists a Borel set $A \in \mathcal{A}$ such that $S \subseteq A$ and

$$\mu(S \backslash Z) \leqslant \mu(A \backslash Z) = \mu(Z \cup (A \backslash Z)) - \mu(Z) \leqslant \alpha - \alpha = 0$$

That is,

$$\mu \lfloor (X \backslash Z)(S) = 0.$$

Since S was an arbitrary Borel set with $\mu_0(S) = 0$, we have shown that $\mu \lfloor (X \backslash Z)$ is absolutely continuous with respect to μ_0.

To see the uniqueness of Z up to μ-measure zero, suppose that $\hat{Z} \subseteq X$ is another Borel set such that $\mu_0(\hat{Z}) = 0$ and $\mu \lfloor (X \backslash \hat{Z})$ is absolutely continuous with respect to μ_0. Then we have

$$\mu(Z \backslash \hat{Z}) = \mu \lfloor (X \backslash \hat{Z})(Z \backslash \hat{Z}) = 0$$

since $\mu_0(Z\backslash\hat{Z}) = 0$. Similarly, we have $\mu(\hat{Z}\backslash Z) = 0$. Thus $\mu(Z\triangle\hat{Z}) = 0$, which proves the uniqueness of Z up to μ-measure zero. As a result, the Borel regular outer measure $\mu \downharpoonright (X\backslash Z)$ is uniquely determined.

Step 2: Now we assume that (X, μ) is σ-finite.

Then there is an increasing sequence of Borel sets $\{A_j\}_{j=1}^\infty$ such that $X = \bigcup_{j=1}^\infty A_j$ and $\mu(A_j) < \infty$ for each $j \in \mathbb{Z}^+$. For each $j \in \mathbb{Z}^+$, by Step 1 there exists a Borel set $Z_j \subseteq A_j$ such that $\mu_0(Z_j) = 0$ and $\mu \downharpoonright (A_j\backslash Z_j)$ is absolutely continuous with respect to μ_0. Let

$$Z := \bigcup_{j=1}^\infty Z_j.$$

Then $Z \subseteq X$ is a Borel set with $\mu_0(Z) = 0$. We claim that $\mu \downharpoonright (X\backslash Z)$ is absolutely continuous with respect to μ_0. To see this, let $A \subseteq X\backslash Z$ be a Borel set with $\mu_0(A) = 0$. Then for each $j \in \mathbb{Z}^+$, we have

$$\mu \downharpoonright (X\backslash Z)(A \cap A_j) = \mu \downharpoonright (A_j\backslash Z_j)(A \cap A_j) = 0$$

since $\mu \downharpoonright (A_j\backslash Z_j)$ is absolutely continuous with respect to μ_0. Thus by countable sub-additivity of $\mu \downharpoonright (X\backslash Z)$, we have

$$\mu \downharpoonright (X\backslash Z)(A) = \mu \downharpoonright (X\backslash Z)\left(\bigcup_{j=1}^\infty (A \cap A_j)\right) \leq \sum_{j=1}^\infty \mu \downharpoonright (X\backslash Z)(A \cap A_j) = 0.$$

This proves that $\mu \downharpoonright (X\backslash Z)$ is absolutely continuous with respect to μ_0.

The uniqueness of the Borel set Z up to μ-measure zero follows from the uniqueness of Z_j up to μ-measure zero for each $j \in \mathbb{Z}^+$ established in Step 1. Thus the Borel regular outer measure $\mu \downharpoonright (X\backslash Z)$ is uniquely determined. $\qquad\square$

We can also define lower densities in a similar manner to upper densities. The following two definitions and comparison lemma are the last ingredients to prepare before the proof of the Lebesgue-Radon-Nikodym Theorem via Densities.

Definition 5.5.6 (Lower Densities). Let μ and μ_0 be Borel regular outer measures on X, and assume that μ_0 is locally finite. Let

$$U_{\mu_0} := \{x \in X : \mu_0(\overline{B}(x,r)) = 0 \text{ for some } r > 0\} \quad \text{and} \quad U_\mu := \{x \in X : \mu(\overline{B}(x,r)) = 0 \text{ for some } r > 0\}.$$

We define the *lower density* of μ with respect to μ_0 by

$$\Theta_*^{\mu_0}(\mu, x) := \begin{cases} \liminf_{r\to 0^+} \dfrac{\mu(\overline{B}(x,r))}{\mu_0(\overline{B}(x,r))}, & \text{if } x \in X\backslash(U_{\mu_0} \cup U_\mu), \\ \infty, & \text{if } x \in U_{\mu_0}\backslash U_\mu, \\ 0, & \text{if } x \in U_\mu. \end{cases}$$

Definition 5.5.7 (Density of a Measure with respect to another Measure). Let μ and μ_0 be Borel regular outer measures on X, and assume that μ_0 is locally finite. If $x \in X$ is such that $\Theta^{*\mu_0}(\mu, x) = \Theta_*^{\mu_0}(\mu, x) \in [0, \infty]$, then we define the *density* of μ with respect to μ_0 at x as the common value

$$\Theta^{\mu_0}(\mu, x) := \Theta^{*\mu_0}(\mu, x) = \Theta_*^{\mu_0}(\mu, x).$$

If $x \in X$ is such that $\Theta_*^{\mu_0}(\mu, x) < \Theta^{*\mu_0}(\mu, x)$, then $\Theta^{\mu_0}(\mu, x)$ is undefined.

We can also give an alternate definition of the density of a measure with respect to another measure, which more closely resembles the definitions of upper and lower densities.

Remark 5.5.8 (Alternate Definition of Density of a Measure with respect to another Measure). Letting μ, μ_0 be as in Definition 5.5.7, let

$$U_{\mu_0} := \{x \in X : \mu(\overline{B}(x, r)) = 0 \text{ for some } r > 0\} \quad \text{and} \quad U_\mu := \{x \in X : \mu_0(\overline{B}(x, r)) = 0 \text{ for some } r > 0\}.$$

Then we see that

$$\Theta^{\mu_0}(\mu, x) = \begin{cases} \lim_{r \to 0^+} \dfrac{\mu(\overline{B}(x, r))}{\mu_0(\overline{B}(x, r))}, & \text{if } x \in X \backslash (U_{\mu_0} \cup U_\mu) \text{ and this limit exists,} \\ \infty, & \text{if } x \in U_{\mu_0} \backslash U_\mu, \\ 0, & \text{if } x \in U_\mu, \end{cases}$$

and is undefined otherwise.

Lemma 5.5.9 (Comparison Lemma for Lower Densities). *Let μ and μ_0 be Borel regular outer measures on X, and assume that both μ and μ_0 are open σ-finite. Let $t > 0$ and let $A \subseteq X$ be such that $\Theta_*^{\mu_0}(\mu, x) \leqslant t$ for each $x \in A$.*

(a) If μ has the Symmetric Vitali Property, then $\mu(A) \leqslant \mu_0(A) \cdot t$.

(b) If μ_0 has the Symmetric Vitali Property, then with Z as in Lemma 5.5.4, we have $\mu(A \backslash Z) \leqslant \mu_0(A) \cdot t$.

Proof. Step 1: First we assume that $\mu_0(X) < \infty$ and $\mu(A) < \infty$, and we prove (a).

First we observe that $A \subset X \backslash U_{\mu_0}$ since $\Theta_*^{\mu_0}(\mu, x) = \infty$ for each $x \in U_{\mu_0}$. Fix $\tau > t$. Since μ_0 is Borel regular and (X, μ_0) is open σ-finite, Theorem 1.1.14 implies that there exists an open set $U \supseteq A$ such that

$$\mu_0(U) \leqslant \mu_0(A) + \tau - t.$$

We define

$$\mathcal{B} := \{\overline{B}(x, r) : x \in A \text{ and } \mu(\overline{B}(x, r)) < \tau \mu_0(\overline{B}(x, r))\}$$

and notice that for each $x \in A$, the definition of lower density implies that for each $\varepsilon > 0$ there exists $0 < r_x < \varepsilon$ such that

$$\mu(\overline{B}(x, r_x)) < \tau \mu_0(\overline{B}(x, r_x)).$$

Thus for each $x \in A$, there exists a sequence $\{r_{x,j}\}_{j=1}^{\infty}$ such that $r_{x,j} \to 0$ as $j \to \infty$ and $\overline{B}(x, r_{x,j}) \in \mathcal{B}$ for each $j \in \mathbb{Z}^+$. As a result, we see that

$$\inf\{r : \overline{B}(x, r) \in \mathcal{B}\} = 0$$

for each $x \in A$.

Since μ has the Symmetric Vitali Property, there exists a countable collection of disjoint closed balls $\{B_j\}_{j=1}^{\infty} \subset \mathcal{B}$ such that

$$\mu\left(A \backslash \bigcup_{j=1}^{\infty} B_j\right) = 0$$

and then also $\mu(B_j) < \tau \mu_0(B_j)$ for each $j \in \mathbb{Z}^+$ since $B_j \in \mathcal{B}$. Summing over $j \in \mathbb{Z}^+$ gives

$$
\begin{aligned}
\mu(A) &\leqslant \mu\left(A \backslash \bigcup_{j=1}^{\infty} B_j\right) + \mu\left(\bigcup_{j=1}^{\infty} B_j\right) \\
&= 0 + \sum_{j=1}^{\infty} \mu(B_j) \\
&< \tau \sum_{j=1}^{\infty} \mu_0(B_j) \\
&= \tau \mu_0\left(\bigcup_{j=1}^{\infty} B_j\right) \qquad\qquad \text{by disjointness of } \{B_j\}_{j=1}^{\infty} \\
&\leqslant \tau \mu_0(U) \\
&\leqslant \tau(\mu_0(A) + \tau - t).
\end{aligned}
$$

Letting $\tau \to t^+$ gives the desired result

$$\mu(A) \leqslant \mu_0(A) \cdot t.$$

Step 2: Now we assume that both μ and μ_0 are open σ-finite, and we prove (a).

Then there exist countable collections of open sets $\{U_k\}_{k=1}^{\infty}$ and $\{V_k\}_{k=1}^{\infty}$ such that $X = \bigcup_{k=1}^{\infty} U_k = \bigcup_{k=1}^{\infty} V_k$ and $\mu(U_k) < \infty$ and $\mu_0(V_k) < \infty$ for each $k \in \mathbb{Z}^+$. By considering the collection

$$\{U_j \cap V_k : j, k \in \mathbb{Z}^+\}$$

of open set, see that $X = \bigcup_{j,k=1}^{\infty}(U_j \cap V_k)$ and $\mu(U_j \cap V_k) < \infty$ and $\mu_0(U_j \cap V_k) < \infty$ for each $j, k \in \mathbb{Z}^+$. Thus be re-labeling if necessary, we may assume there is an increasing collection of open sets $\{W_k\}_{k=1}^{\infty}$ such that $X = \bigcup_{k=1}^{\infty} W_k$ and $\mu(W_k) < \infty$ and $\mu_0(W_k) < \infty$ for each $k \in \mathbb{Z}^+$.

For each $k \in \mathbb{Z}^+$ we consider the measures $\mu \downharpoonright W_k$ and $\mu_0 \downharpoonright W_k$ which are Borel regular outer measures on X satisfying $(\mu \downharpoonright W_k)(X) = \mu(W_k) < \infty$ and $(\mu_0 \downharpoonright W_k)(X) = \mu_0(W_k) < \infty$. Then by Step 1, we have

$$(\mu \downharpoonright W_k)(A) \leqslant (\mu_0 \downharpoonright W_k)(A) \cdot t$$

for each $k \in \mathbb{Z}^+$. That is,

$$\mu(A \cap W_k) \leqslant \mu_0(A \cap W_k) \cdot t \qquad \forall\, k \in \mathbb{Z}^+.$$

By using the fact that $\{W_k\}_{k=1}^{\infty}$ is an increasing collection of sets whose union is X, we have

$$\mu(A) = \lim_{k \to \infty} \mu(A \cap W_k) \leqslant \lim_{k \to \infty} \mu_0(A \cap W_k) \cdot t = \mu_0(A) \cdot t$$

by Proposition 1.1.6. This proves (a).

Step 3: Finally, we prove (b).

Let $Z \subseteq X$ be the Borel set given by Lemma 5.5.4, so that $\mu \downharpoonright (X \backslash Z)$ is absolutely continuous with respect to μ_0. Thus the Symmetric Vitali Property for μ_0 implies that $\mu \downharpoonright (X \backslash Z)$ also has the Symmetric Vitali Property, so we apply the result of (a) to the set $A \backslash Z$ and the measure $\tilde{\mu}$, instead of A and μ, to see that

$$(\mu \downharpoonright (X \backslash Z))(A \backslash Z) = \mu(A \backslash Z) \leqslant \mu_0(A \backslash Z) \cdot t.$$

Since $\mu_0(Z) = 0$, we have $\mu_0(A \backslash Z) = \mu_0(A)$, and the above inequality implies

$$\mu(A \backslash Z) \leqslant \mu_0(A \backslash Z) \cdot t = \mu_0(A) \cdot t$$

which completes the proof of (b). □

At this point, we are ready to show how these densities relate to the differentiation theorems in the previous chapter.

Theorem 5.5.10 (Differentiation of Measures). *Let μ and μ_0 be Borel regular outer measures on X, which are both open σ-finite.*

(a) *If μ has the symmetric Vitali property, then there is a Borel set $S \subseteq X$ with $\mu(S) = 0$ such that for every $x \in X \backslash S$, the density $\Theta^{\mu_0}(\mu, x)$ is defined.*

(b) If μ_0 has the symmetric Vitali property, then there is a Borel set $S \subseteq X$ with $\mu_0(S) = 0$ such that for every $x \in X \backslash S$, the density $\Theta^{\mu_0}(\mu, x)$ is defined.

In either case, the function

$$(X \backslash S) \ni x \longmapsto \Theta^{\mu_0}(\mu, x)$$

is Borel measurable.

Proof of (a). We will show (a) first, so assume that μ has the symmetric Vitali property.

Step 1: First we assume that $\mu_0(X) < \infty$ and $\mu(A) < \infty$, and we prove (a).

First see that if $A \subseteq X$ is a subset and $a, b > 0$ are positive real numbers such that

$$\Theta_*^{\mu_0}(\mu, x) < a \quad \text{and} \quad b < \Theta^{*\mu_0}(\mu, x) \qquad \forall\, x \in A$$

then

$$\mu(A) \leqslant a\mu_0(A) \quad \text{and} \quad \mu_0(A) \cdot b \leqslant \mu(A)$$

by the Comparison Lemmas 5.4.10 and 5.5.9.

In particular, if $0 < a < b$ and

$$E_{a,b} := \{x \in X : \Theta_*^{\mu_0}(\mu, x) < a < b < \Theta^{*\mu_0}(\mu, x)\}$$

then we have

$$\frac{\mu(E_{a,b})}{a} \leqslant \mu_0(E_{a,b}) \leqslant \frac{\mu(E_{a,b})}{b}$$

which implies that

$$\mu_0(E_{a,b}) = \mu(E_{a,b}) = 0.$$

Because

$$\{x \in X : \Theta_*^{\mu_0}(\mu, x) < \Theta^{*\mu_0}(\mu, x)\} \subseteq \bigcup_{\substack{a,b \in \mathbb{Q}^+ \\ a < b}} E_{a,b}$$

we deduce that $\Theta_*^{\mu_0}(\mu, x) < \Theta^{*\mu_0}(\mu, x)$ only on a set which has both μ-measure and μ_0-measure zero. As a result the density $\Theta^{\mu_0}(\mu, x)$ is defined for μ-almost every $x \in X \backslash U_{\mu_0}$. Since $\Theta^{\mu_0}(\mu, x)$ is also defined in $[0, \infty]$ for each $x \in U_\mu$ by 5.5.8, we conclude that $\Theta^{\mu_0}(\mu, x)$ is defined for μ-almost every $x \in X$. By using Borel regularity of μ, we may therefore choose a Borel set $S \subseteq X$ with $\mu(S) = 0$ such that for every $x \in X \backslash S$, the density $\Theta^{\mu_0}(\mu, x)$ is defined.

We now show that thedensity function

$$(X \backslash S) \ni x \longmapsto \Theta^{\mu_0}(\mu, x)$$

is Borel measurable.

For each fixed $r > 0$, the functions

$$x \longmapsto \mu(B(x,r)), \qquad x \longmapsto \mu_0(B(x,r))$$

defined on $X \backslash (S \cup U_{\mu_0} \cup U_\mu)$ are upper semi-continuous, and hence Borel measurable. Thus their ratio

$$x \longmapsto \frac{\mu(B(x,r))}{\mu_0(B(x,r))}$$

is also Borel measurable on $X \backslash (S \cup U_{\mu_0} \cup U_\mu)$. Therefore

$$\Theta^{\mu_0}(\mu, x) = \lim_{j \to \infty} \frac{\mu(B(x, \frac{1}{j}))}{\mu_0(B(x, \frac{1}{j}))}$$

is a Borel measurable function on $X \backslash (S \cup U_{\mu_0} \cup U_\mu)$, as the pointwise limit of a sequence of Borel measurable functions.

Since $\Theta^{\mu_0}(\mu, x) = \infty$ on $U_{\mu_0} \backslash U_\mu$ and $\Theta^{\mu_0}(\mu, x) = 0$ on U_μ, and the sets U_{μ_0} and U_μ are both open (by 5.4.3), so we conclude that $\Theta^{\mu_0}(\mu, x)$ is Borel measurable on all of $X \backslash S$.

Step 2: Now we assume that both μ and μ_0 are open σ-finite, and we prove (a).

Then there exist countable collections of open sets $\{U_k\}_{k=1}^\infty$ and $\{V_k\}_{k=1}^\infty$ such that $X = \bigcup_{k=1}^\infty U_k = \bigcup_{k=1}^\infty V_k$ and $\mu(U_k) < \infty$ and $\mu_0(V_k) < \infty$ for each $k \in \mathbb{Z}^+$. By considering the collection

$$\{U_j \cap V_k : j, k \in \mathbb{Z}^+\}$$

of open sets, see that $X = \bigcup_{j,k=1}^\infty (U_j \cap V_k)$ and $\mu(U_j \cap V_k) < \infty$ and $\mu_0(U_j \cap V_k) < \infty$ for each $j, k \in \mathbb{Z}^+$. Thus be re-labeling if necessary, we may assume there is an increasing collection of open sets $\{W_k\}_{k=1}^\infty$ such that $X = \bigcup_{k=1}^\infty W_k$ and $\mu(W_k) < \infty$ and $\mu_0(W_k) < \infty$ for each $k \in \mathbb{Z}^+$.

For each $k \in \mathbb{Z}^+$ we consider the measures $\mu \llcorner W_k$ and $\mu_0 \llcorner W_k$ which are Borel regular outer measures on X satisfying $(\mu \llcorner W_k)(X) = \mu(W_k) < \infty$ and $(\mu_0 \llcorner W_k)(X) = \mu_0(W_k) < \infty$. Then by Step 1, there exists a Borel set $S_k \subseteq W_k$ with $(\mu \llcorner W_k)(S_k) = \mu(S_k) = 0$ such that for every $x \in W_k \backslash S_k$, the density $\Theta^{\mu_0 \llcorner W_k}(\mu \llcorner W_k, x)$ is defined. Let

$$S := \bigcup_{k=1}^\infty S_k.$$

Then $S \subseteq X$ is a Borel set with $\mu(S) = 0$. We claim that for every $x \in X \backslash S$, the density $\Theta^{\mu_0}(\mu, x)$ is defined. To see this, let $x \in X \backslash S$ be arbitrary. Then there exists $k_0 \in \mathbb{Z}^+$ such that $x \in W_{k_0}$. Since $x \notin S$, we have $x \notin S_{k_0}$, so the density $\Theta^{\mu_0 \llcorner W_{k_0}}(\mu \llcorner W_{k_0}, x)$ is defined. By the definitions of densities, we see that

$$\Theta^{\mu_0}(\mu, x) = \Theta^{\mu_0 \llcorner W_{k_0}}(\mu \llcorner W_{k_0}, x)$$

is also defined. Since $x \in X \backslash S$ was arbitrary, we have shown that for every $x \in X \backslash S$, the density $\Theta^{\mu_0}(\mu, x)$ is defined.

Lastly, see that the density function $\Theta^{\mu_0}(\mu, x)$ is Borel measurable on $X \backslash S$ by taking the pointwise limit of the Borel measurable functions $\Theta^{\mu_0 \restriction W_k}(\mu \restriction W_k, x)$ on each $W_k \backslash S_k$ as in Step 1. □

Proof of (b). Now we prove (b), so assume that μ_0 has the symmetric Vitali property. By 5.4.11, we know that

$$\Theta^{* \mu_0}(\mu, x) < \infty \qquad \text{for } \mu_0\text{-almost every } x \in X.$$

By 5.5.4, there is a Borel set $Z \subseteq X$ such that $\mu_0(Z) = 0$ and $\hat{\mu} := \mu \restriction (X \backslash Z)$ is absolutely continuous with respect to μ_0. Since $\hat{\mu}$ is absolutely continuous with respect to μ_0, it is clear that $\hat{\mu}$ also has the Symmetric Vitali Property. Now we use the argument of part (a) applied to the measures $\hat{\mu}$ to see that

$$\mu_0(E_{a,b}) = \mu(E_{a,b} \backslash Z) = 0$$

where $E_{a,b}$ is as defined in the proof of part (a). Thus the density $\Theta^{\mu_0}(\mu, x)$ is defined for μ_0-almost every $x \in X \backslash Z$, and by our first remark in this proof, it is finite for μ_0-almost every $x \in X$. By using Borel regularity of μ_0, we may therefore choose a Borel set $S \subseteq X$ with $\mu_0(S) = 0$ such that for every $x \in X \backslash S$, the density $\Theta^{\mu_0}(\mu, x)$ is defined and finite.

The Borel measurability of the density function

$$(X \backslash S) \ni x \longmapsto \Theta^{\mu_0}(\mu, x)$$

follows from the same argument as in part (a). □

Theorem 5.5.11 (The Lebesgue-Radon-Nikodym Theorem). *Let μ, μ_0 be Borel regular outer measures on X, which are both open σ-finite, and suppose that μ_0 has the symmetric Vitali property.*

(i) *If $\mu \ll \mu_0$, then μ also has the symmetric Vitali property, and*

$$\mu(A) = \int_A \Theta^{\mu_0}(\mu, x) \, \mathrm{d}\mu_0(x)$$

for each Borel set $A \subseteq X$.

(ii) *In general, there exists a Borel set $Z \subseteq X$ such that $\mu_0(Z) = 0$ and*

$$\mu(A) = \int_A \Theta^{\mu_0}(\mu, x) \, \mathrm{d}\mu_0(x) + (\mu \restriction Z)(A)$$

for each Borel set $A \subseteq X$.

(iii) If μ also has the symmetric Vitali property, then the set Z in part (b) can be taken to be

$$Z = \{x \in X : \Theta^{\mu_0}(\mu, x) = \infty\}.$$

Remark 5.5.12 (Interpretation of the Lebesgue-Radon-Nikodym Theorem via Densities). Basically this says that if you have Borel regular outer measures μ and μ_0 on X which are open σ-finite and such that μ_0 has the symmetric Vitali property, then the Radon-Nikodym derivative of μ with respect to μ_0 can be computed as a limit of ratios of measures at μ-almost every point.

More precisely, using our notation for the Radon-Nikodym derivative, (i) says that if $\mu \ll \mu_0$, then

$$D_{\mu_0}\mu(x) = \Theta^{\mu_0}(\mu, x) = \lim_{r \to 0^+} \frac{\mu\big(\overline{B}(x, r)\big)}{\mu_0\big(\overline{B}(x, r)\big)}$$

for μ_0-almost every $x \in X$.

Sanity Check. [Sanity check — note that $\Theta^{\mu_0}(\mu, x)$ exists for μ_0-almost every $x \in X$ by Theorem 5.5.10. To justify the second equality computing $\Theta^{\mu_0}(\mu, x)$ as a limit, we use the alternate definition 5.5.8 of density of a measure with respect to another measure. Since

$$\Theta^{\mu_0}(\mu, x) = \begin{cases} \displaystyle\lim_{r \to 0^+} \frac{\mu(\overline{B}(x, r))}{\mu_0(\overline{B}(x, r))}, & \text{if } x \in X \backslash (U_{\mu_0} \cup U_\mu) \text{ and this limit exists,} \\ \infty, & \text{if } x \in U_{\mu_0} \backslash U_\mu, \\ 0, & \text{if } x \in U_\mu, \end{cases}$$

and $\mu_0(U_{\mu_0}) = 0$, so the second branch only occurs on a μ_0-measure zero set.

If $x \in U_\mu \backslash U_{\mu_0}$, then $\mu_0(\overline{B}(x, r)) > 0$ for each $r > 0$ but $\mu(\overline{B}(x, r)) = 0$ for some $r > 0$, so the limit in the first branch is 0 which agrees with the third branch (for such x). Since the set U_{μ_0} has μ_0-measure zero, we conclude that for every $x \in X \backslash U_{\mu_0}$ where $\Theta^{\mu_0}(\mu, x)$ is defined, we have

$$\Theta^{\mu_0}(\mu, x) = \lim_{r \to 0^+} \frac{\mu\big(\overline{B}(x, r)\big)}{\mu_0\big(\overline{B}(x, r)\big)}.$$

Finally, use Theorem 5.5.10 to see that $\Theta^{\mu_0}(\mu, x)$ is defined for μ_0-almost every $x \in X$, so the above equality holds for μ_0-almost every $x \in X$ as well.] □

Also (ii) says that in general, we can decompose μ into an absolutely continuous part with respect to μ_0 and a singular part with respect to μ_0 using these densities, i.e.

$$\begin{aligned} \mu &= \mu_{ac} + \mu_s \\ &= D_{\mu_0}\mu + (\mu \restriction Z) \end{aligned}$$

and the absolutely continuous part and the singular part are unique, by the Lebesgue Decomposition Theorem 5.1.21. Furthermore the absolutely continuous part μ_{ac} can be computed pointwise using the densities $\Theta^{\mu_0}(\mu, x) = \lim_{r \to 0^+} \frac{\mu(\overline{B}(x,r))}{\mu_0(\overline{B}(x,r))}$ for μ_0-almost every $x \in X$ as in the previous paragraph.

First we prove the statements of the theorem in the case that $\mu(X) < \infty$ and $\mu_0(X) < \infty$. The general case where μ and μ_0 are open σ-finite then follows by a standard exhaustion argument as in previous results.

Proof when $\mu(X) < \infty$ and $\mu_0(X) < \infty$. We assume that both $\mu(X) < \infty$ and $\mu_0(X) < \infty$.

(i). Assume that $\mu \ll \mu_0$. Then it is clear that μ has the Symmetric Vitali Property since μ_0 does.

Since μ_0 has the Symmetric Vitali Property, there is a Borel set $S \subseteq X$ with $\mu_0(S) = 0$ such that for every $x \in X \backslash S$, the density $\Theta^{\mu_0}(\mu, x)$ is defined by Theorem 5.5.10. For each Borel set $A \subset X \backslash S$ we define

$$\nu(A) := \int_A \Theta^{\mu_0}(\mu, x) \, \mathrm{d}\mu_0(x)$$

and for an arbitrary set $E \subseteq X$, we define

$$\nu(E) := \inf\{\nu(A) : A \text{ is a Borel set such that } A \supseteq E\}.$$

The same argument as in the first step of the proof of 5.5.4 shows that ν is a well-defined Borel regular outer measure on X.

Let $A \subseteq X$ be an arbitrary Borel set. If $0 < a < b$ and

$$A_{(a,b)} := \{x \in A : a < \Theta^{\mu_0}(\mu, x) < b\},$$

then we have

$$a\mu_0(A_{a,b}) \leqslant \nu(A_{a,b}) \leqslant b\mu_0(A_{a,b})$$

by definition of ν, while the Comparison Lemmas 5.4.10 and 5.5.9 imply that

$$a\mu_0(A_{a,b}) \leqslant \mu(A_{a,b}) \leqslant b\mu_0(A_{a,b})$$

and as a result we see that

$$\frac{a}{b}\mu_0(A_{a,b}) \leqslant \nu(A_{a,b}) \leqslant \frac{b}{a}\mu_0(A_{a,b}).$$

Since a and b are arbitrary positive real numbers with $a < b$, we conclude that

$$\nu(A) = \mu(A).$$

Since $A \subset X$ was an arbitrary Borel set, we have proven (i).

(ii). Now assume that we are in the general case, where we allow sets $A \subset X$ with $\mu(A) > 0$ and $\mu_0(A) = 0$. By 5.5.4, there exists a Borel set $Z \subseteq X$ such that $\mu_0(Z) = 0$ and $\hat{\mu} := \mu \restriction (X \backslash Z)$ is absolutely continuous with respect to μ_0. Applying part (i) to the measures $\hat{\mu}$ and μ_0, we see that

$$\hat{\mu}(A) = \int_A \Theta^{\mu_0}(\hat{\mu}, x) \, \mathrm{d}\mu_0(x) \qquad \text{for each Borel set } A \subseteq X$$

which means that

$$\mu(A \backslash Z) = \int_{A \backslash Z} \Theta^{\mu_0}(\hat{\mu}, x) \, \mathrm{d}\mu_0(x) \qquad \text{for each Borel set} A \subset X.$$

As a result, for each Borel set $A \subseteq X$, we have

$$\mu(A) = \mu(A \backslash Z) + \mu(A \cap Z) = \int_{A \backslash Z} \Theta^{\mu_0}(\hat{\mu}, x) \, \mathrm{d}\mu_0(x) + \mu(A \cap Z) = \int_A \Theta^{\mu_0}(\hat{\mu}, x) \, \mathrm{d}\mu_0(x) + (\mu \restriction Z)(A)$$

as desired.

(iii). Finally, assume that μ also has the Symmetric Vitali Property. Then the density $\Theta^{\mu_0}(\mu, x)$ is defined for μ-almost every $x \in X$ by Theorem 5.5.10 (and also for μ_0-almost every $x \in X$ by the same theorem). Also see that if

$$A \subseteq \{x \in X \backslash U_{\mu_0} : \Theta^{\mu_0}(\mu, x) < \infty\}$$

then the Comparison Lemma for Lower Densities 5.5.9 shows that

$$\mu_0(A) = 0 \implies \mu(A) = 0.$$

Therefore we can apply part (i) to the measure $\mu \restriction (X \backslash \{x : \Theta^{\mu_0}(\mu, x) = \infty\})$ which is absolutely continuous with respect to μ_0 (as we just showed) to see that

$$\mu(A \backslash \{x : \Theta^{\mu_0}(\mu, x) = \infty\}) = \int_{A \backslash \{x : \Theta^{\mu_0}(\mu, x) = \infty\}} \Theta^{\mu_0}(\mu, x) \, \mathrm{d}\mu_0(x)$$

for each Borel set $A \subseteq X$.

Thus if we let

$$Z := \{x \in X : \Theta^{\mu_0}(\mu, x) = \infty\},$$

then $Z \subseteq X$ is a Borel set with $\mu_0(Z) = 0$ and for each Borel set $A \subseteq X$, we have

$$\mu(A) = \int_A \Theta^{\mu_0}(\mu, x) \, \mathrm{d}\mu_0(x) + (\mu \restriction Z)(A)$$

as desired. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \square$

Proof when both μ and μ_0 are open σ-finite. Assume now that both μ and μ_0 are open σ-finite. Then there exist countable collections of open sets $\{U_k\}_{k=1}^{\infty}$ and $\{V_k\}_{k=1}^{\infty}$ such that $X = \bigcup_{k=1}^{\infty} U_k = \bigcup_{k=1}^{\infty} V_k$ and $\mu(U_k) < \infty$ and $\mu_0(V_k) < \infty$ for each $k \in \mathbb{Z}^+$. By considering the collection

$$\{U_j \cap V_k : j, k \in \mathbb{Z}^+\}$$

of open sets, see that $X = \bigcup_{j,k=1}^{\infty}(U_j \cap V_k)$ and $\mu(U_j \cap V_k) < \infty$ and $\mu_0(U_j \cap V_k) < \infty$ for each $j, k \in \mathbb{Z}^+$. Thus be re-labeling if necessary, we may assume there is an increasing collection of open sets $\{W_k\}_{k=1}^{\infty}$ such that $X = \bigcup_{k=1}^{\infty} W_k$ and $\mu(W_k) < \infty$ and $\mu_0(W_k) < \infty$ for each $k \in \mathbb{Z}^+$.

For each $k \in \mathbb{Z}^+$ we consider the measures $\mu \,\llcorner\, W_k$ and $\mu_0 \,\llcorner\, W_k$ which are Borel regular outer measures on X satisfying $(\mu \,\llcorner\, W_k)(X) = \mu(W_k) < \infty$ and $(\mu_0 \,\llcorner\, W_k)(X) = \mu_0(W_k) < \infty$. Then by the previous proof, there exists a Borel set $Z_k \subseteq W_k$ such that $(\mu_0 \,\llcorner\, W_k)(Z_k) = \mu_0(Z_k) = 0$ and

$$(\mu \,\llcorner\, W_k)(A) = \int_A \Theta^{\mu_0 \llcorner W_k}(\mu \,\llcorner\, W_k, x) \; \mathrm{d}(\mu_0 \,\llcorner\, W_k)(x) + ((\mu \,\llcorner\, W_k) \,\llcorner\, Z_k)(A)$$

for each Borel set $A \subseteq X$. Let

$$Z := \bigcup_{k=1}^{\infty} Z_k.$$

Then $Z \subseteq X$ is a Borel set with $\mu_0(Z) = 0$.

See that for each Borel set $A \subseteq X$, we have

$$\begin{aligned}
\mu(A) &= \lim_{k\to\infty} (\mu \,\llcorner\, W_k)(A) \\
&= \lim_{k\to\infty} \left(\int_A \Theta^{\mu_0 \llcorner W_k}(\mu \,\llcorner\, W_k, x) \; \mathrm{d}(\mu_0 \,\llcorner\, W_k)(x) + ((\mu \,\llcorner\, W_k) \,\llcorner\, Z_k)(A) \right) \\
&= \lim_{k\to\infty} \int_A \Theta^{\mu_0 \llcorner W_k}(\mu \,\llcorner\, W_k, x) \; \mathrm{d}(\mu_0 \,\llcorner\, W_k)(x) + \lim_{k\to\infty} ((\mu \,\llcorner\, W_k) \,\llcorner\, Z_k)(A) \\
&= \int_A \Theta^{\mu_0}(\mu, x) \; \mathrm{d}\mu_0(x) + (\mu \,\llcorner\, Z)(A).
\end{aligned}$$

Thus if $\mu \ll \mu_0$, then $\mu \,\llcorner\, Z = 0$ and part (i) holds, while in general part (ii) holds. Lastly, if μ also has the Symmetric Vitali Property, then by the result in the finite case we may take

$$Z_k := \{x \in W_k : \Theta^{\mu_0 \llcorner W_k}(\mu \,\llcorner\, W_k, x) = \infty\} \qquad \forall \, k \in \mathbb{Z}^+$$

which means that

$$Z = \bigcup_{k=1}^{\infty} Z_k = \{x \in X : \Theta^{\mu_0}(\mu, x) = \infty\}$$

as desired in part (iii). □

Chapter 6

Riesz Representation Theorem

6.1 Locally Compact Hausdorff Spaces

In this section, we follow Simon.

We begin by reviewing some definitions and results from general topology. Recall a topological space X is said to be *Hausdorff* if for every pair of distinct points $x, y \in X$, there exist disjoint open sets $U, V \subseteq X$ such that $x \in U$ and $y \in V$.

Lemma 6.1.1 (Compact Sets in a Hausdorff Space are Closed). *Let X be a Hausdorff space, and let $K \subseteq X$ be a compact set. Then K is closed.*

Proof. Let $x \in X\backslash K$ be arbitrary. For each $y \in K$, there exist disjoint open sets $U_y, V_y \subseteq X$ such that $x \in U_y$ and $y \in V_y$. The collection $\{V_y : y \in K\}$ is an open cover of K, and so there exist $y_1, \ldots, y_n \in K$ such that $K \subseteq \bigcup_{j=1}^n V_{y_j}$. Then $U := \bigcap_{j=1}^n U_{y_j}$ is an open set containing x such that $U \cap K = \varnothing$, which shows that $X\backslash K$ is open. Thus K is closed. $\qquad\square$

Remark 6.1.2 (Disjoint Compact Sets in a Hausdorff Space can be Separated by Open Sets). In fact we have proven a bit more — If X is a Hausdorff space and $K \subseteq X$ is compact, then for each $x \in X\backslash K$, there exist disjoint open sets $U, V \subseteq X$ such that $x \in U$ and $K \subseteq V$. Then if $K_2 \subseteq X$ is another compact set such that $K \cap K_2 = \varnothing$, we can repeat this argument to obtain disjoint open sets $U'_x, V'_x \subseteq X$ such that $x \in U'_x$ and $K_2 \subseteq V'_x$ for each $x \in K_2$; compactness of K_2 then implies that there exist $x_1, \ldots, x_m \in K_2$ such that $K_2 \subseteq \bigcup_{j=1}^m U'_{x_j}$ and then $V' := \bigcap_{j=1}^m V'_{x_j}$ is an open set containing K such that $V' \cap K_2 = \varnothing$. In summary, if $K_1, K_2 \subseteq X$ are disjoint compact sets, then there exist disjoint open sets $U, V \subseteq X$ such that $K_1 \subseteq U$ and $K_2 \subseteq V$.

We have thus proven the base case of the following useful lemma.

Lemma 6.1.3 (Disjoint Compact Sets in a Hausdorff Space can be Separated by Open Sets)**.** *Let X be a Hausdorff space and let $K_1, K_2, \ldots, K_n \subseteq X$ be compact sets such that $K_i \cap K_j = \varnothing$ for all $i \neq j$. Then there exist disjoint open sets $U_1, U_2, \ldots, U_n \subseteq X$ such that $K_j \subseteq U_j$ for each $j = 1, 2, \ldots, n$.*

We also recall a useful corollary in the case when X is compact.

Corollary 6.1.4 (Compact Hausdorff Spaces are Normal)**.** *Let X be a compact Hausdorff space. Then X is* normal, *i.e. for every disjoint pair of closed sets $K_1, K_2 \subseteq X$, there exists a pair of disjoint open sets $U_1, U_2 \subseteq X$ such that $K_1 \subseteq U_1$ and $K_2 \subseteq U_2$.*

Proof. This follows immediately from the fact that if X is compact, then every closed subset of X must be compact. □

Recall that a topological space X is said to be *locally compact* if for each $x \in X$, there exists an open set $U_x \subseteq X$ containing x such that the closure $\overline{U_x}$ is compact.

Lemma 6.1.5. *Let X be a locally compact Hausdorff space, and let $x \in X$. Then for each open set $U \subseteq X$ containing x, there exists an open set $V \subseteq X$ such that $x \in V$, $\overline{V}$ is compact, and $\overline{V} \subseteq U$.*

Proof. With $x \in X$ fixed, let $U \subseteq X$ be an open set which contains x. Since X is locally compact, there exists an open set $W_0 \subseteq X$ containing x such that $\overline{W_0}$ is compact. We define $W := W_0 \cap U$. Since $\overline{W} \subseteq \overline{W_0}$, we see that the $\overline{W}$ is compact (a closed subset of a compact set is compact).

By applying Corollary 6.1.4 to the disjoint closed sets $\overline{W}\backslash W$ and $\{x\}$ we see that there exists disjoint open sets $V_1, V_2 \subseteq \overline{W}$ such that

$$x \in V_1 \quad \text{and} \quad \overline{W}\backslash W \subseteq V_2.$$

By definition of the subspace topology on $\overline{W}$, there exist open sets $U_1, U_2 \subseteq X$ such that $V_1 = \overline{W} \cap U_1$ and $V_2 = \overline{W} \cap U_2$. Thus we have

$$x \in U_1, \ \overline{W}\backslash W \subseteq U_2, \quad \text{and} \ (\overline{W} \cap U_1) \cap (\overline{W} \cap U_2) = \varnothing.$$

In particular, this last equation implies that $\overline{W} \cap U_1 \subseteq \overline{W}\backslash U_2$. Therefore

$$x \in W \cap U_1 \subseteq \overline{W}\backslash U_2 \subseteq W \subseteq U.$$

Since $\overline{W}\backslash U_2$ is a closed set, it follows that

$$x \in W \cap U_1 \subseteq \overline{W \cap U_1} \subseteq \overline{W}\backslash U_2 \subseteq U.$$

Now take $V := W \cap U_1$ and the lemma is proved. □

We will normally abbreviate "locally compact Hausdorff" as "LCH".

Combining Lemmas 6.1.3 and the proof in 6.1.2, we obtain the following strengthened version of Lemma 6.1.3.

Corollary 6.1.6 (Disjoints Compact Sets in a LCH Space can be Separated by Open Sets with Compact Closure). *Let X be a LCH space and let $K_1, K_2, \ldots, K_n \subseteq X$ be compact sets such that $K_i \cap K_j = \varnothing$ for all $i \neq j$. Then there exist open sets $U_1, U_2, \ldots, U_n \subseteq X$ such that for each $j = 1, 2, \ldots, n$, we have $K_j \subseteq U_j$, the closure $\overline{U_j}$ is compact, and $\overline{U_i} \cap \overline{U_j} = \varnothing$ for all $i \neq j$.*

Lemma 6.1.7 (Urysohn's Lemma for LCH Spaces). *Let X be a LCH space, and let $K \subseteq X$ be a compact set. Then for each open set $V \subseteq X$ such that $K \subseteq V$, there exists an open set $U \subseteq X$ such that $K \subseteq U \subseteq \overline{U} \subseteq V$, the closure $\overline{U}$ is compact, and there exists a continuous function*

$$f : X \to [0, 1]$$

such that $f \equiv 1$ on an open set containing K and $f \equiv 0$ on $X \backslash U$.

Proof. By lemma 6.1.5, for each $x \in K$, there exists an open set $U_x \subseteq X$ such that $x \in U_x$, $\overline{U_x}$ is compact, and $\overline{U_x} \subseteq V$. Since K is compact, there exist $x_1, x_2, \ldots, x_n \in K$ such that

$$K \subseteq \bigcup_{j=1}^{n} U_{x_j}.$$

Define $U := \bigcup_{j=1}^{n} U_{x_j}$ so that

$$K \subseteq U \subseteq \overline{U} \subseteq V,$$

and $\overline{U}$ is compact as a finite union of compact sets.

By Corollary 6.1.4, we know that $\overline{U}$ is normal. By the classical Urysohn's lemma, there exists a continuous function $f_0 : \overline{U} \to [0, 1]$ such that $f_0 \equiv 1$ on K and $f_0 \equiv 0$ on $\overline{U} \backslash U$. We can extend f_0 to a continuous function f_1 on X by defining $f \equiv 0$ on $X \backslash \overline{U}$.

(Let us check that f_1 is continuous. Since f_0 is continuous on $\overline{U}$ and $f_1|_{\overline{U}} = f_0$, we see that f_1 is continuous on $\overline{U}$. Since $f_1 \equiv 0$ on the overlap $\overline{U} \backslash U = \overline{U} \cap (X \backslash U)$, we see that f_1 is continuous on $X \backslash U$.)

Finally, we let $f := 2 \min\{f_1, 1/2\}$. Then see that $f \equiv 1$ on the set $\left\{f_1 > \frac{1}{2}\right\}$, which is an open set containing K, and $f \equiv 0$ on $X \backslash U$. $\qquad \square$

Okay, we need one last topological lemma about LCH spaces before we can get back to measure theory, yay.

Theorem 6.1.8 (Partition of Unity for LCH Spaces). *Let X be a locally compact Hausdorff space, and let $K \subseteq X$ be a compact set. Let $\{U_j\}_{j=1}^{n}$ be a finite open cover of K. Then there exist finitely many continuous functions $\psi_1, \psi_2, \ldots, \psi_n : X \to [0, 1]$ such that*

(i) $\text{supp}(\psi_j) := \overline{\{x \in X : \psi_j(x) \neq 0\}}$ *is compact for each* $j = 1, 2, \ldots, n$,

(ii) *for each* $j = 1, 2, \ldots, n$, *we have* $\text{supp}(\psi_j) \subseteq U_j$,

(iii) $\sum_{j=1}^n \psi_j(x) = 1$ *for all* x *in an open set containing* K, *and*

(iv) $0 \leqslant \sum_{j=1}^n \psi_j(x) \leqslant 1$ *for all* $x \in X$.

Proof. By Lemma 6.1.7, for each $x \in K$, there exists $j \in \{1, 2, \ldots, n\}$ and an open set U_x containing x such that $\overline{U_x}$ is a compact subset of U_j. Since K is compact, there exist $x_1, x_2, \ldots, x_m \in K$ such that

$$K \subseteq \bigcup_{i=1}^m U_{x_i}.$$

For each $i = 1, 2, \ldots, m$, we define V_j to be the union of all U_{x_i} such that $\overline{U_{x_i}} \subseteq U_j$. Then $\{V_j\}_{j=1}^n$ is an open cover of K such that $\overline{V_j}$ is a compact subset of U_j for each $j = 1, 2, \ldots, n$. Thus for each $j = 1, 2, \ldots, n$, we can apply Lemma 6.1.7 to obtain a continuous function $\varphi_j : X \to [0, 1]$ such that $\varphi_j \equiv 1$ on $\overline{V_j}$ and $\varphi_j \equiv 0$ on $X \backslash W_j$ for some open set W_j such that $\overline{W_j}$ is a compact subset of U_j and $\overline{V_j} \subseteq W_j$.

We can also use the same lemma on the open set $\bigcup_{j=1}^n V_j$ which contains K to obtain a continuous function $f_0 : X \to [0, 1]$ such that $f_0 \equiv 1$ on $\bigcup_{j=1}^n V_j$ and $f_0 \equiv 0$ on $\{x \in X : \sum_{j=1}^n \varphi_j(x) = 0\}$. (Let us check this — the set $\{x \in X : \sum_{j=1}^n \varphi_j(x) = 0\}$ is closed since it is the intersection of the closed sets $\{x \in X : \varphi_j(x) = 0\}$ for $j = 1, 2, \ldots, n$. Thus its complement $\{x \in X : \sum_{j=1}^n \varphi_j(x) \neq 0\}$ is open and contains the compact set $\bigcup_{j=1}^n \overline{V_j}$. Hence Lemma 6.1.7 applies.)

We now define $\varphi_0 := 1 - f_0$ so that by contruction we have

$$\sum_{j=0}^n \varphi_j(x) > 0, \qquad \forall x \in X.$$

For each $j = 1, 2, \ldots, n$, we define

$$\psi_j := \frac{\varphi_j}{\sum_{k=0}^n \varphi_k}.$$

Then the functions $\psi_1, \psi_2, \ldots, \psi_n$ satisfy properties (i)-(iv).

We check this. Property (i) and (ii) hold since

$$\text{supp}(\psi_j) = \text{supp}(\varphi_j) \subseteq W_j \subseteq \overline{W_j} \subseteq U_j.$$

Property (iv) is evident from the definition of ψ_j. Finally, property (iii) holds since for each $x \in \bigcup_{j=1}^n V_j$, we have $\varphi_0(x) = 0$ and so

$$\sum_{j=1}^n \psi_j(x) = \frac{\sum_{j=1}^n \varphi_j(x)}{\sum_{j=0}^n \varphi_j(x)} = 1.$$

Since $\bigcup_{j=1}^n V_j$ is an open set containing K, property (iii) is proved. $\square$

6.2 Radon Measures on LCH Spaces

We now give the definition of a Radon measure on an LCH space. We remark that the definition and the first two lemmas are valid in an arbitrary Hausdorff space.

Definition 6.2.1 (Radon Measure)**.** Let X be a Hausdorff space. A *Radon measure* on X is an outer measure μ on X satisfying:

  (i) μ is Borel regular and $\mu(K) < \infty$ for each compact set $K \subseteq X$,

  (ii) $\mu(A) = \inf\{\mu(U) : U \supseteq A, U \text{ open}\}$ for each set $A \subseteq X$,

  (iii) $\mu(U) = \sup\{\mu(K) : K \subseteq U, K \text{ compact}\}$ for each open set $U \subseteq X$.

You might wonder why we do not require the third condition to hold for all sets $A \subseteq X$, instead of just open sets, but it turns out that this automatically follows from the definition given.

Lemma 6.2.2. *Let X be a Hausdorff space, and let μ be a Radon measure on X. Then for each set $A \subseteq X$ such that $\mu(A) < \infty$, we have*

$$\mu(A) = \sup\{\mu(K) : K \subseteq A, K \text{ compact}\}.$$

Proof. Let $A \subseteq X$ be such that $\mu(A) < \infty$. Let $\varepsilon > 0$ be arbitrary. By condition (ii) in Definition 6.2.1, there exists an open set $U \subseteq X$ such that $A \subseteq U$ and

$$\mu(U \backslash A) < \varepsilon.$$

By condition (iii) in Definition 6.2.1, there exists a compact set $K \subseteq U$ such that

$$\mu(U \backslash K) < \varepsilon.$$

By using condition (ii) again on the set $U \backslash A$, there exists an open set $V \subseteq X$ such that $U \backslash A \subseteq V$ and

$$\mu(V \backslash (U \backslash A)) < \varepsilon.$$

Therefore

$$\mu(V) \leqslant \epsilon + \mu(U \backslash A) < 2\varepsilon.$$

Now $K \backslash W$ is a compact subset of $U \backslash V$, so is a subset of A; we compute

$$\mu(A \backslash (K \backslash V)) \leqslant \mu(U \backslash (K \backslash V)) \leqslant \mu(U \backslash K) + \mu(W) \leqslant 3\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, we conclude that

$$\mu(A) = \sup\{\mu(K) : K \subseteq A, K \text{ compact}\}$$

as desired. $\qquad\square$

The next lemma shows that if we have an outer measure μ on a LCH space X which is finite on compact sets, and finitely additive on disjoint unions of compact sets, and μ satisfies conditions (ii) and (iii) in Definition 6.2.1, then μ is a Radon measure.

Lemma 6.2.3. *Let X be a LCH space, and let μ be an outer measure on X such that $\mu(K) < \infty$ for each compact set $K \subseteq X$. Suppose that μ is finitely additive on disjoint unions of compact sets, i.e. if $K_1, K_2 \subseteq X$ are disjoint compact sets, then*

$$\mu(K_1 \cup K_2) = \mu(K_1) + \mu(K_2) < \infty.$$

If μ also satisfies conditions (ii) and (iii) in Definition 6.2.1, then μ also satisfies condition (i) in Definition 6.2.1, and hence is a Radon measure on X.

Proof. Step 1: We claim that for each set $A \subseteq X$, there is a countable collection of open sets $\{U_j\}_{j=1}^{\infty}$ such that $A \subseteq \bigcap_{j=1}^{\infty} U_j$ and $\mu(A) = \mu(\bigcap_{j=1}^{\infty} U_j)$.

Let $A \subseteq X$ be arbitrary. By condition (ii) in Definition 6.2.1, for each $j = 1, 2, \ldots$, there exists an open set $U_j \subseteq X$ such that $A \subseteq U_j$ and

$$\mu(U_j) < \mu(A) + \frac{1}{j}.$$

Then we have $A \subseteq \bigcap_{j=1}^{\infty} U_j$ and

$$\mu\left(\bigcap_{j=1}^{\infty} U_j\right) \leqslant \mu(U_j) < \mu(A) + \frac{1}{j}, \qquad \forall j = 1, 2, \ldots.$$

Since $\mu(A) \leqslant \mu(\bigcap_{j=1}^{\infty} U_j)$, we conclude that

$$\mu(A) = \mu\left(\bigcap_{j=1}^{\infty} U_j\right).$$

This proves our claim.

Step 2: We claim that all Borel sets in X are μ-measurable.

Since the Borel σ-algebra is the smallest σ-algebra containing all open subsets of X, and the set of all μ-measurable sets is a σ-algebra, it suffices to show that all open sets in X are μ-measurable.

Let $\varepsilon > 0$ be arbitrary, and let $U \subseteq X$ be an open set. Fix an arbitrary set $A \subseteq X$ such that $\mu(A) < \infty$. By (ii) in the definition of Radon measure, there exists an open set $V \subseteq X$ such that $A \subseteq V$ and

$$\mu(V) < \mu(A) + \varepsilon.$$

By (iii) in the definition of Radon measure, there exists a compact set $K_1 \subseteq V \cap U$ such that

$$\mu(V \cap U) < \mu(K_1) + \varepsilon.$$

By (iii) again, there exists a compact set $K_2 \subseteq V \backslash K_1$ such that

$$\mu(V \backslash K_1) < \mu(K_2) + \varepsilon.$$

We estimate

$$
\begin{aligned}
\mu(V \backslash U) + \mu(V \cap U) &< \mu(V \backslash U) + \mu(K_1) + \varepsilon \\
&\leqslant \mu(K_2) + \mu(K_1) + 2\varepsilon \\
&= \mu(K_1 \cup K_2) + 2\varepsilon \qquad \text{(by finite additivity of } \mu \text{ on disjoint compact sets)} \\
&\leqslant \mu((V \backslash K_1) \cup K_1) + 2\varepsilon \\
&= \mu(V) + 2\varepsilon \\
&< \mu(A) + 3\varepsilon
\end{aligned}
$$

which implies that

$$\mu(A \backslash U) + \mu(A \cap U) \leqslant \mu(V \backslash U) + \mu(V \cap U) \leqslant \mu(A) + 3\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, we conclude that

$$\mu(A \backslash U) + \mu(A \cap U) \leqslant \mu(A)$$

which is the Carathéodory criterion for μ-measurability of U.

Since $U \subseteq X$ was an arbitrary open set, we conclude that all open sets in X are μ-measurable, and hence all Borel sets in X are μ-measurable.

Step 3: We claim that (i) in the Definition of Radon measure holds.

By assumption μ being finitely additive on disjoint unions of compact sets, we have $\mu(K) < \infty$ for each compact set $K \subseteq X$ (Look back!). By Step 2, we know that all Borel sets in X are μ-measurable. Thus μ is a Borel measure, and by Step 1, we see that μ is Borel regular.

Combining Steps 2 and 3, we see that μ satisfies condition (i) in Definition 6.2.1. $\qquad \square$

The next lemma is convenient, and gives us a way to check that a Borel regular outer measure is a Radon measure.

Lemma 6.2.4. *Let X be a LCH space such that each open set in X is the countable union of compact sets. Then each Borel regular outer measure which is finite on compact sets is a Radon measure.*

Proof. First we observe that, since X is a Hausdorff space, the statement "each open set in X is the countable union of compact sets" is equivalent to the statement "X is the countable union of compact sets and every closed set in X is the countable intersection of open sets".

(Let's check this. Assume that each open set in X is the countable union of compact sets. Then since X is open, we see that X is the countable union of compact sets. Now let $F \subseteq X$ be a closed set. Then F^c is open, so is the countable union of compact sets, say $F^c = \bigcup_{j=1}^{\infty} K_j$ where each K_j is compact. Then we have

$$F = (F^c)^c = \left(\bigcup_{j=1}^{\infty} K_j \right)^c = \bigcap_{j=1}^{\infty} K_j^c$$

which is a countable intersection of open sets since each K_j is closed.

Conversely, assume that X is the countable union of compact sets and every closed set in X is the countable intersection of open sets. Let $U \subseteq X$ be an open set. Then U^c is closed, so is the countable intersection of open sets, say $U^c = \bigcap_{j=1}^{\infty} V_j$ where each V_j is open. Then we have

$$U = (U^c)^c = \left(\bigcap_{j=1}^{\infty} V_j \right)^c = \bigcup_{j=1}^{\infty} V_j^c$$

which is a countable union of closed sets since each V_j is closed. Since X is the countable union of compact sets, say $X = \bigcup_{k=1}^{\infty} K_k$ where each K_k is compact, we see that

$$U = \bigcup_{j=1}^{\infty} V_j^c = \bigcup_{j=1}^{\infty} \bigcup_{k=1}^{\infty} (V_j^c \cap K_k)$$

which is a countable union of compact sets since each $V_j^c \cap K_k$ is closed in K_k and hence compact.

We have used two facts about Hausdorff spaces here: (1) compact sets are closed, and (2) closed subsets of compact sets are compact.)

Therefore X is the countable union of compact sets, and each closed set in X is the countable intersection of open sets. Let μ be a Borel regular outer measure on X such that $\mu(K) < \infty$ for each compact set $K \subseteq X$. Then Theorem 1.1.14 about Borel regular outer measures implies that μ satisfies the following two conditions:

(b) if $A \subseteq X$ is such that there is a countable collection of open sets $\{V_j\}_{j=1}^{\infty}$ such that $A \subseteq \bigcap_{j=1}^{\infty} V_j$ and $\mu(V_j) < \infty$ for each $j \in \mathbb{Z}^+$, then

$$\mu(A) = \inf\{\mu(U) : U \supseteq A, U \text{ open}\},$$

(c) if $\{A_j\}_{j=1}^{\infty}$ is a countable collection of μ-measurable sets such that $\mu(A_j) < \infty$ for each $j \in \mathbb{Z}^+$, then $A := \bigcup_{j=1}^{\infty} A_j$ has

$$\mu(A) = \sup\{\mu(C) : C \subseteq A, C \text{ closed}\}$$

Now see that because X is an LCH space, for each compact set $K \subseteq X$, there exists an open set $U \subseteq X$ such that $K \subseteq U$ and $\overline{U}$ is compact. (We check — since each point $x \in K$ has an open set V_x such that $x \in V_x$ and $\overline{V_x}$ is compact, we can cover K by finitely many such neighborhoods $V_{x_1}, V_{x_2}, \ldots, V_{x_n}$; then we can take $U := \bigcup_{j=1}^{n} V_{x_j}$ which has compact closure.)

Our first observation shows that there are countably many compact sets $\{K_j\}_{j=1}^{\infty}$ such that $X = \bigcup_{j=1}^{\infty} K_j$. By the previous paragraph, for each $j \in \mathbb{Z}^+$, there exists an open set $U_j \subseteq X$ such that $K_j \subseteq U_j$ and $\overline{U_j}$ is compact. Since $\mu(\overline{U_j}) < \infty$ for each $j \in \mathbb{Z}^+$ by assumption that μ is finite on compact sets, we see that condition (b) holds for each set $A \subseteq X$. That is, condition (ii) in the definition of Radon measure holds for each set $A \subseteq X$.

Next, see that if $A \subseteq X$ is μ-measurable, then we can write $A = \bigcup_{j=1}^{\infty}(A \cap K_j)$ and each $A \cap K_j$ is μ-measurable and $\mu(A \cap K_j) \leqslant \mu(K_j) < \infty$. Thus condition (c) holds for each μ-measurable set A since X is the countable union of compact sets. Also each closed set $C \subseteq X$ we can write C as an increasing union of compact sets $C = \bigcup_{j=1}^{\infty} C_j$ where

$$C_j := C \cap \left(\bigcup_{i=1}^{j} K_i \right)$$

so that each C_j is compact; also $\lim_{j \to \infty} \mu(C_j) = \mu(C)$ which implies that

$$\mu(C) = \sup\{\mu(K) : K \subseteq C, K \text{ compact}\}.$$

This fact combined with condition (c) implies that for each μ-measurable set $A \subseteq X$ we have

$$\mu(A) = \sup\{\mu(K) : K \subseteq A, K \text{ compact}\}.$$

In particular, this shows that condition (iii) in the definition of Radon measure holds for each open set $U \subseteq X$. $\qquad \square$

Theorem 6.2.5 (Density of $C_c(X)$ in $L^p(X, \mu)$)**.** *Let X be a LCH space, and let μ be a Radon measure on X. Then for each $1 \leqslant p < \infty$, the space $C_c(X)$ of continuous functions with compact support is dense in $L^p(X, \mu)$. That is, for each $f \in L^p(X, \mu)$ and each $\varepsilon > 0$, there exists $g \in C_c(X)$ such that*

$$\|f - g\|_{L^p(X,\mu)} < \varepsilon.$$

Before the proof, we record the following useful corollary.

Corollary 6.2.6. *Let X be a LCH space with the property that each open set in X is the countable union of compact sets. Let μ be a Borel measure on X such that $\mu(K) < \infty$ for each compact set $K \subseteq X$. Then $C_c(X)$ is dense in $L^1(X, \mu)$ and there exists a radon measure $\overline{\mu}$ on X such that $\overline{\mu}(B) = \mu(B)$ for each Borel set $B \subseteq X$.*

That is, every Borel measure on X which is finite on compact sets is the restriction of a Radon measure.

Proof of Density of $C_c(X)$ in $L^p(X, \mu)$. Let $p \in [1, \infty)$ be fixed. Let $f \in L^p(X, \mu)$ such that $\|f\|_{L^p} < \infty$, and let $\varepsilon > 0$ be arbitrary.

First see that the set of simple functions is dense in $L^p(X, \mu)$. (We can decompose f into its positive and negative parts as $f = f^+ - f^-$, and then approximate f^+ and f^- by increasing sequences of simple functions using the Dominated Convergence Theorem.) Thus there exists a simple function

$$\varphi = \sum_{j=1}^n a_j \mathbf{1}_{A_j}$$

where the numbers $a_1, a_2, \ldots, a_n$ are distinct and nonzero, and the sets $A_1, A_2, \ldots, A_n \subseteq X$ are disjoint μ-measurable sets, and

$$\|f - \varphi\|_{L^p} < \varepsilon.$$

Since $\|\varphi\|_{L^p} \leqslant \|\varphi - f\|_{L^p} + \|f\|_{L^p} < \infty$, we see that $\mu(A_j) < \infty$ for each $j = 1, 2, \ldots, n$.

Choose $M > \max\{|a_j| : j = 1, 2, \ldots, n\}$ and for each $j = 1, 2, \ldots, n$ use Lemma 6.2.2 to find a compact set $K_j \subseteq A_j$ such that

$$\mu(A_j \backslash K_j) < \frac{\varepsilon^p}{n 2^p M^p}.$$

Similarly, by definition of a Radon measure, for each $j = 1, 2, \ldots, n$ we can find an open set U_j containing K_j such that

$$\mu(U_j \backslash K_j) < \frac{\varepsilon^p}{n 2^p M^p}.$$

We may assume that the sets $U_1, U_2, \ldots, U_n$ are disjoint. (If not, we can use Corollary 6.1.6 to find disjoint open sets $U_1', U_2', \ldots, U_n'$ such that $K_j \subseteq U_j'$ for each $j = 1, 2, \ldots, n$;

then we can replace U_j by $U_j \cap U_j'$.) For each $j = 1, 2, \ldots, n$, we can use Lemma 6.1.7 to find a continuous function $g_j \in C_c(X)$ such that $g_j \equiv a_j$ on an open subset containing K_j, $\operatorname{supp}(g_j) \subseteq\subseteq U_j$, and $|g_j| \leqslant |a_j|$ on X. We define

$$g := \sum_{j=1}^{n} g_j.$$

Then $g \in C_c(X)$ since the supports of $g_1, g_2, \ldots, g_n$ are compact and disjoint. We compute that for each $j = 1, 2, \ldots, n$ and each $x \in K_j$, we have

$$g(x) = \sum_{i=1}^{n} g_i(x) = g_j(x) = a_j = \sum_{i=1}^{n} a_i \mathbf{1}_{K_i}(x) = \varphi(x).$$

Therefore $g \equiv \varphi$ on the set $\bigcup_{j=1}^{n} K_j$, and we have

$$\sup |g| = \sup |\varphi| < M.$$

Now the function $\varphi - g$ is identically zero on the compliment of the set $\bigcup_{j=1}^{n}((U_j \backslash K_j) \cup (A_j \backslash K_j))$, and we have

$$\int_X |\varphi - g|^p \, \mathrm{d}\mu \leqslant \sum_{j=1}^{n} \int_{(U_j \backslash K_j) \cup (A_j \backslash K_j)} |\varphi - g|^p \, \mathrm{d}\mu$$

$$\leqslant \sup |\varphi - g|^p \sum_{j=1}^{n} \mu((U_j \backslash K_j) \cup (A_j \backslash K_j))$$

$$\leqslant (2M)^p \sum_{j=1}^{n} \mu((U_j \backslash K_j) \cup (A_j \backslash K_j))$$

$$\leqslant (2M)^p \sum_{j=1}^{n} \big(\mu(U_j \backslash K_j) + \mu(A_j \backslash K_j)\big)$$

$$< (2M)^p \sum_{j=1}^{n} \frac{\varepsilon^p}{n 2^p M^p}$$

$$\leqslant \varepsilon^p$$

where we have used subadditivity of μ in the first line, and Lemma 2.1.2 (Bounding an Integral) in the second line. Hence

$$\|f - g\|_{L^p} \leqslant \|f - \varphi\|_{L^p} + \|\varphi - g\|_{L^p} \leqslant 2\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, the theorem is proved. $\qquad\square$

We remark that this result is sometimes used "coordinate-wise", and cited as "the space of continuous functions $C_c(X, \mathbb{F}^m)$ is dense in $L^p(X, \mathbb{F}^m, \mu)$, for $p \in [1, \infty)$. "

6.3 The Riesz Representation Theorem

In this section, we present and prove the Riesz Representation Theorem for linear functionals. There are actually several results which go by this name, and they are all closely related.

Throughout this section, we let X be an arbitrary nonempty LCH space. Recall that $C_c(X)$ is the set of all continuous, real-valued functions on X with compact support. We define $C_c^+(X)$ to be the set of all nonnegative functions in $C_c(X)$.

Theorem 6.3.1 (Riesz Representation Theorem for Positive Linear Functionals)**.** *Let* $L : C_c(X) \to \mathbb{R}$ *be a positive linear functional, i.e.* L *is linear and* $L(f) \geqslant 0$ *for each* $f \in C_c^+(X)$. *Then there exists a unique Radon measure* μ *on* X *such that*

$$L(f) = \int_X f \, \mathrm{d}\mu, \qquad \forall f \in C_c^+(X).$$

Proof. We begin with some remarks. First, note that if $f, g \in C_c^+(X)$ are such that $f \leqslant g$, then $g - f \in C_c^+(X)$ and so

$$L(f) = L(g) - L(g - f) \leqslant L(g). \tag{6.1}$$

Thus L is monotone on $C_c^+(X)$.

Second, if $K \subseteq X$ is compact and $f \in C_c^+(X)$ is such that $\mathrm{supp}(f) \subseteq K$, and if $g \in C_c^+(X)$ is such that $g \equiv 1$ on K, then we have $gf = f \leqslant (\sup f)g$ so we have

$$L(f) \leqslant (\sup f)L(g). \tag{6.2}$$

Notice that if U is an arbitrary open subset of X which contains K, then by Urysohn's lemma there exists an open set V such that $K \subseteq V \subseteq \overline{V} \subseteq U$ and a function $g \in C_c^+(X)$ such that $g \equiv 1$ on an oen set containing $\overline{V}$, $g \leqslant 1$ on X, and $\mathrm{supp}(g) \subseteq U$. Then for each $f \in C_c^+(X)$ such that $\mathrm{supp}(f) \subseteq V$ and $f \leqslant 1$, our previous remark implies

$$L(f) \leqslant (\sup f)L(g) \leqslant L(g).$$

Taking the supremum over all such f and the infimum over all such g, we obtain

$$\sup\{L(f) : f \in C_c^+(X), f \leqslant 1, \mathrm{supp}\, f \subseteq V\} \leqslant \inf\{L(g) : g \in C_c^+(X), g \leqslant 1, g \equiv 1 \text{ on an open set containing } \tag{6.3}$$

This concludes our remarks.

Step 1: Definition of the measure μ.

We begin by defining μ on open sets $U \subseteq X$ by

$$\mu(U) := \sup\{L(f) : f \in C_c^+(X), f \leqslant 1, \mathrm{supp}\, f \subseteq U\}$$

and for an arbitrary set $A \subseteq X$, we define

$$\mu(A) := \inf\{\mu(U) : U \supseteq A, U \text{ open}\}.$$

Note that these two definitions are consistent — if $U \subseteq X$ is open, then for each open set $V \supseteq U$, we have

$$\mu(U) := \sup\{L(f) : f \in C_c^+(X), f \leqslant 1, \text{supp } f \subseteq U\} \leqslant \sup\{L(f) : f \in C_c^+(X), f \leqslant 1, \text{supp } f \subseteq V\} =: \mu(V)$$

and this shows

$$\mu(U) \leqslant \inf\{\mu(V) : V \supseteq U, V \text{ open}\}.$$

For the other direction, note that U is one of the open sets containing itself, so

$$\inf\{\mu(V) : V \supseteq U, V \text{ open}\} \leqslant \mu(U).$$

Thus we conclude that

$$\mu(U) = \inf\{\mu(V) : V \supseteq U, V \text{ open}\}$$

for each open set $U \subseteq X$.

Also note that for each compact set $K \subseteq X$, we have

$$\mu(K) := \inf\{\mu(U) : U \supseteq K, U \text{ open}\} = \inf\{L(g) : g \in C_c^+(X), g \leqslant 1, g \equiv 1 \text{ on an open set containing } K\} <$$

$$(*)$$

so the measure μ is finite on compact sets.

Step 2: Now we claim that μ is an outer measure on X.

To prove the claim, let $U \subseteq X$ be open and let $\{U_j\}_{j=1}^{\infty}$ be a countable collection of open subsets of X such that $U \subseteq \bigcup_{j=1}^{\infty} U_j$. Fix a $g \in C_c^+(X)$ with $g \leqslant 1$ and $\text{supp}(g) \subseteq U$. Since $\text{supp}(g)$ is compact, there exists a $k \in \mathbb{N}$ such that $\text{supp}(g) \subseteq \bigcup_{j=1}^{k} U_j$. Let $\{\psi_j\}_{j=1}^{k} \subseteq C_c^+(X)$ be a partition of unity subordinate to the open cover $\{U_j\}_{j=1}^{k}$ of $\text{supp}(g)$ (Theorem 6.1.8). Then we have

$$g = \sum_{j=1}^{k} \psi_j g$$

which implies that

$$L(g) \leqslant \sum_{j=1}^{k} L(\psi_j g) \leqslant \sum_{j=1}^{k} \mu(U_j) \leqslant \sum_{j=1}^{\infty} \mu(U_j).$$

Taking the supremum over all such g, we obtain

$$\mu(U) \leqslant \sum_{j=1}^{\infty} \mu(U_j).$$

This proves that μ is countably subadditive on open sets.

Now let $\{A_j\}_{j=1}^{\infty}$ be an arbitrary countable collection of subset of X and let $A \subseteq \bigcup_{j=1}^{\infty} A_j$. Fix $\varepsilon > 0$. For each $j \in \mathbb{Z}^+$, the definition of $\mu(A_j)$ implies that there exists an open set U_j such that $A_j \subseteq U_j$ and

$$\mu(U_j) \leqslant \mu(A_j) + \frac{\varepsilon}{2^j}.$$

Then $A \subseteq \bigcup_{j=1}^{\infty} U_j$, and so

$$\mu(A) \leqslant \mu\left(\bigcup_{j=1}^{\infty} U_j\right) \leqslant \sum_{j=1}^{\infty} \mu(U_j) \leqslant \sum_{j=1}^{\infty} \mu(A_j) + \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, we conclude that

$$\mu(A) \leqslant \sum_{j=1}^{\infty} \mu(A_j).$$

This proves that μ is countably subadditive on all subsets of X.

Since $\mu(\varnothing) = 0$ by definition, we conclude that μ is an outer measure on X.

Step 3: We claim that μ is a Radon measure on X.

See that μ is an outer measure on X which is finite on compact sets by Step 1, and conditions (i) and (ii) of the definition of a Radon measure are satisfied. Thus, by Lemma 6.2.3, it suffices to show that μ is finitely additive on compact sets.

Let $K_1, K_2 \subseteq X$ be disjoint compact sets and let $\varepsilon > 0$. By $(*)$ in Step 1, since $K_1 \cup K_2$ is compact, there exists a $g \in C_c^+(X)$ such that $g \leqslant 1$, $g \equiv 1$ on an open set W containing $K_1 \cup K_2$, and

$$L(g) \leqslant \mu(K_1 \cup K_2) + \varepsilon.$$

By Lemma 6.1.3, there exist disjoint open sets $U_1, U_2 \subseteq X$ such that $K_1 \subseteq U_1$ and $K_2 \subseteq U_2$. Then Urysohn's lemma 6.1.7 implies that there exist functions $f_1, f_2 \in C_c^+(X)$ such for $i = 1, 2$, we have $f_i \leqslant 1$, $\mathrm{supp}(f_i) \subseteq U_i$, and $f_i \equiv 1$ on an open set containting K_i. Now $(*)$ implies that

$$\mu(K_1) + \mu(K_2) \leqslant L(f_1 g) + L(f_2 g) = L((f_1 + f_2)g)$$
$$\leqslant L(g) \leqslant \mu(K_1 \cup K_2) + \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, we conclude that

$$\mu(K_1) + \mu(K_2) \leqslant \mu(K_1 \cup K_2).$$

The reverse inequality follows from countable subadditivity of μ. Thus μ is finitely additive on compact sets, and we conclude that μ is a Radon measure on X by Lemma 6.2.3.

Step 4: It remains to show that $L(f) = \int_X f \, d\mu$ for each $f \in C_c^+(X)$.

By (2) it follows that for each $h \in C_c^+(X)$, by taking an arbitrary $g \in C_c^+(X)$ such that $g \leqslant 1$, and $g \equiv 1$ on an open set containing $\mathrm{supp}(h)$, we have

$$L(h) \leqslant (\sup h)L(g)$$

and by taking the infimum over all such g, we obtain

$$L(h) \leqslant (\sup h)\mu(\mathrm{supp}(h)) < \infty.$$

By observing that h is the uniform limit of the functions $\{\max\{h - 1/n, 0\}\}_{n=1}^\infty \subseteq C_c^+(X)$, we see that

$$L(h) \leqslant (\sup h)\mu(\{x \in X : h(x) > 0\}). \tag{$\star$}$$

Note that this inequality holds for each $h \in C_c^+(X)$.

Now we will show that the integral identity holds. Let $f \in C_c^+(X)$ be and let $\varepsilon > 0$. If $f \equiv 0$, then the integral identity holds trivially, so we may assume that $\sup f > 0$. Then choose numbers $t_0, t_1, \ldots, t_N$ such that

$$0 = t_0 < t_1 < \cdots < t_{N-1} < \sup f < t_N$$

and

$$\max_{1 \leqslant k \leqslant N} t_k - t_{k-1} < \varepsilon$$

and

$$\mu(f^{-1}(\{t_k\})) = 0 \qquad \forall k = 1, 2, \ldots, N.$$

(This final requirement is no issue, as $\mu(f^{-1}(\{t\})) = 0$ for all but countably many $t \in \mathbb{R}$, by virtue of the fact that $\{x \in X : f(x) > 0\}$ is contained in the compact set $\mathrm{supp} \, f$, which has finite μ measure.)

For each $k = 1, 2, \ldots, N$, define

$$U_k := f^{-1}(\{(t_{k-1}, t_k)\}).$$

Note that the sets $U_1, U_2, \ldots, U_N$ are open and disjoint, and that each U_k is contained in the compact set $\mathrm{supp}(f)$. We define functions $\phi^-, \phi^+ : X \to [0, \infty)$ by

$$\phi^-(x) := \sum_{k=1}^N t_{k-1}\chi_{U_k}(x), \quad \phi^+(x) := \sum_{k=1}^N t_k\chi_{U_k}(x)$$

for each $x \in X$. Then we have

$$\phi^-(x) \leqslant f(x) \leqslant \phi^+(x) \qquad \forall x \in X$$

because if $x \in U_k$ for some $k = 1, 2, \ldots, N$, then $t_{k-1} < f(x) < t_k$, and if $x \notin \bigcup_{k=1}^{N} U_k$, then $f(x) = 0$ and $\phi^-(x) = \phi^+(x) = 0$. Monotonicity of the integral implies that

$$\int_X \phi^-(x) \, \mathrm{d}\mu \leqslant \int_X f(x) \, \mathrm{d}\mu \leqslant \int_X \phi^+(x) \, \mathrm{d}\mu$$

and the fact that ϕ^- and ϕ^+ are simple functions and the sets $U_1, U_2, \ldots, U_N$ are disjoint implies that

$$\sum_{k=1}^{N} t_{k-1}\mu(U_k) \leqslant \int_X f(x) \, \mathrm{d}\mu \leqslant \sum_{k=1}^{N} t_k\mu(U_k). \tag{‡}$$

By definition of the measure μ on open sets, for each $k = 1, 2, \ldots, N$, there exists a function $g_k \in C_c^+(X)$ such that $g_k \leqslant 1$, $\mathrm{supp}(g_k) \subseteq U_k$, and

$$\mu(U_k) \leqslant L(g_k) + \frac{\varepsilon}{N}.$$

Also for each $k = 1, 2, \ldots, N$ and each compact set $K_k \subseteq U_k$, Urysohn's lemma implies that there exists a function $h_k \in C_c^+(X)$ such that $h_k \leqslant 1$, $h_k \equiv 1$ on an open set containing $K_k \cup \mathrm{supp}\, g_j$, and $\mathrm{supp}(h_k) \subseteq U_k$. Then $g_k \leqslant h_k \leqslant 1$ and $\mathrm{supp}(h_k)$ is a compact subset of U_k for each $k = 1, 2, \ldots, N$, and thus

$$\mu(U_k) - \frac{\varepsilon}{N} \leqslant L(g_k) \leqslant L(h_k) \leqslant \mu(U_k) \quad \forall j = 1, 2, \ldots, N. \tag{$\star\star$}$$

Since μ is a Radon measure, we can choose the compact sets $K_1, K_2, \ldots, K_N$ such that $\mu(U_k \backslash K_k) < \varepsilon/N$ for each $k = 1, 2, \ldots, N$. Then we have

$$\left\{ x \in X : \left(f - f \sum_{k=1}^{N} h_k \right)(x) > 0 \right\} \subseteq \bigcup_{k=1}^{N} (U_k \backslash K_k)$$

so we use ($\star$), subadditivity of μ, and monotonicity of L to deduce that

$$L\left(f - f \sum_{k=1}^{N} h_k \right) \leqslant \sup\left(f - f \sum_{k=1}^{N} h_k \right) \mu\left(\bigcup_{k=1}^{N} (U_k \backslash K_k) \right)$$

$$\leqslant (\sup f) \sum_{k=1}^{N} \mu(U_k \backslash K_k) < (\sup f)\varepsilon. \tag{☺}$$

Using $(\star\star)$ and the fact that $t_{k-1}h_k \leqslant fh_k \leqslant t_k h_k$ for each $k = 1, 2, \ldots, N$, we obtain

$$\sum_{k=1}^{N} t_{k-1}\mu(U_k) - \varepsilon \sup f \leqslant L\left(\sum_{k=1}^{N} t_{k-1}h_k\right) \qquad \text{by } (\star\star)$$

$$\leqslant L\left(f \sum_{k=1}^{N} h_k\right) \qquad \text{since } t_{k-1}h_k \leqslant fh_k \text{ for all } k$$

$$\leqslant L(f) \qquad \text{by monotonicity of } L$$

$$\leqslant L\left(f \sum_{k=1}^{N} h_k\right) + \varepsilon \sup f \qquad \text{by } (\text{☺})$$

$$\leqslant L\left(\sum_{k=1}^{N} t_k h_k\right) + \varepsilon \sup f \qquad \text{since } t_k h_k \leqslant fh_k \text{ for all } k$$

$$\leqslant \sum_{k=1}^{N} t_k\mu(U_k) + \varepsilon \sup f. \qquad \text{by } (\star\star)$$

from which we extract that

$$\sum_{k=1}^{N} t_{k-1}\mu(U_k) - \varepsilon \sup f \leqslant L(f) \leqslant \sum_{k=1}^{N} t_k\mu(U_k) + \varepsilon \sup f.$$

Combining the above with ($\ddagger$), we obtain

$$-\varepsilon(\mu(\operatorname{supp} f) + \sup f) \leqslant -\sum_{k=1}^{N}(t_k - t_{k-1})\mu(U_j) - \varepsilon \sup f$$

$$\leqslant \int_X f \, \mathrm{d}\mu - L(f)$$

$$\leqslant \sum_{k=1}^{N}(t_k - t_{k-1})\mu(U_j) + \varepsilon \sup f \leqslant \varepsilon(\mu(\operatorname{supp} f) + \sup f).$$

That is,

$$\left|\int_X f \, \mathrm{d}\mu - L(f)\right| \leqslant \varepsilon(\mu(\operatorname{supp} f) + \sup f).$$

Since $\varepsilon > 0$ is arbitrary, we conclude that $L(f) = \int_X f \, \mathrm{d}\mu$. This completes the proof of the integral identity for each $f \in C_c^+(X)$.

$\square$

Now we are ready to state and prove the Riesz Representation Theorem.

In the theorem below, we let H denote a finite-dimensional real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\| \cdot \|$. The set $C_c(X, H)$ is defined as the vector space of continuous maps $X \to H$ with compact support.

Theorem 6.3.2 (Riesz Representation Theorem for Bounded Linear Functionals on $C_c(X, H)$).
Let $L : C_c(X, H) \to \mathbb{R}$ be a linear functional such that for each compact set $K \subseteq X$ we have

$$\sup \big\{ L(f) : f \in C_c(X, H), \|f\| \leq 1, \operatorname{supp} f \subseteq K \big\} < \infty.$$

Then there exists a unique finite Radon measure μ on X and a μ-measurable map

$$\sigma : X \to H$$

such that $\|\sigma(x)\| = 1$ for μ-a.e. $x \in X$ and

$$L(f) = \int_X \langle f(x), \sigma(x) \rangle \, \mathrm{d}\mu(x) \qquad \forall f \in C_c(X, H).$$

Proof. By using an orthonormal basis of H, we can identify H with $\mathbb{R}^n$ for some $n \in \mathbb{Z}^+$. Thus it suffices to prove the theorem in the case $H = \mathbb{R}^n$.

Step 1: We begin by defining

$$\lambda : C_c^+(X) \to \mathbb{R}, \quad \lambda(f) := \sup\{L(\omega) : \omega \in C_c(X, \mathbb{R}^n), \|\omega\| \leq f\}, \ \forall f \in C_c^+(X)$$

and we claim that λ is a positive linear functional on $C_c^+(X)$.

It is clear that λ is nonnegative and positively homogeneous, since

$$\lambda(cf) = \sup\{L(\omega) : \omega \in C_c(X, \mathbb{R}^n), \|\omega\| \leq cf\} = c\sup\{L(\omega) : \omega \in C_c(X, \mathbb{R}^n), \|\omega\| \leq f\} = c\lambda(f)$$

for each $c \geq 0$ and each $f \in C_c^+(X)$. To see that λ is additive, let $f_1, f_2 \in C_c^+(X)$ be given. If $\omega_1, \omega_2 \in C_c(X, \mathbb{R}^n)$ are such that $\|\omega_1\| \leq f_1$ and $\|\omega_2\| \leq f_2$, then we have $\|\omega_1 + \omega_2\| \leq f_1 + f_2$ and so

$$\lambda(f_1 + f_2) \geq L(\omega_1 + \omega_2) = L(\omega_1) + L(\omega_2).$$

Taking the supremum over all such ω_1 and ω_2, we obtain

$$\lambda(f_1 + f_2) \geq \lambda(f_1) + \lambda(f_2).$$

For the reverse inequality, let $\omega \in C_c(X, \mathbb{R}^n)$ be such that $\|\omega\| \leq f_1 + f_2$, and define

$$\omega_j := \begin{cases} \dfrac{f_j}{f_1 + f_2}\omega & \text{on } \{f_1 + f_2 > 0\}, \\ 0, & \text{on } \{f_1 + f_2 = 0\} \end{cases}$$

for $j = 1, 2$. It is easy to see that $\omega_1, \omega_2 \in C_c(X, \mathbb{R}^n)$. Then $\omega = \omega_1 + \omega_2$, $\|\omega_1\| \leq f_1$, and $\|\omega_2\| \leq f_2$, so

$$L(\omega) = L(\omega_1 + \omega_2) = L(\omega_1) + L(\omega_2) \leq \lambda(f_1) + \lambda(f_2).$$

By taking the supremum over all such ω, we obtain

$$\lambda(f_1 + f_2) \leq \lambda(f_1) + \lambda(f_2).$$

Therefore

$$\lambda(f_1 + f_2) = \lambda(f_1) + \lambda(f_2).$$

Since $f_1, f_2 \in C_c^+(X)$ were arbitrary, we conclude that λ is additive.

This proves the claim.

Step 2: By Step 1 and the Riesz Representation Theorem for Positive Linear Functionals (Theorem 6.3.1), there exists a unique Radon measure μ on X such that

$$\lambda(f) = \int_X f \, \mathrm{d}\mu \qquad \forall f \in C_c^+(X).$$

That is, we have

$$\sup\{L(\omega) : \omega \in C_c(X, \mathbb{R}^n), \|\omega\| \leq f\} = \int_X f \, \mathrm{d}\mu \qquad \forall f \in C_c^+(X). \qquad (\dagger)$$

Let $e_1, e_2, \ldots, e_n$ be the standard orthonormal basis of $\mathbb{R}^n$. Then for each $j = 1, 2, \ldots, n$ and we have

$$|L(fe_j)| \leq \int_X |f| \, \mathrm{d}\mu =: \|f\|_{L^1}, \qquad \forall f \in C_c(X)$$

since $\|fe_j\| = |f| \in C_c^+(X)$ for each $f \in C_c(X)$.

Thus for each $j = 1, 2, \ldots, n$, the map

$$L_j : C_c(X) \to \mathbb{R}, \quad L_j(f) := L(fe_j) , \forall f \in C_c(X)$$

can be extended to a bounded linear functional on $L^1(X, \mu)$ with operator norm at most 1 by the previous inequality and density of $C_c(X)$ in $L^1(X, \mu)$ (Theorem 6.2.5). By the Riesz Representation Theorem for Bounded Linear Functionals on $L^1(\mu)$, for each $j = 1, 2, \ldots, n$ there exists a bounded μ-measurable function $\sigma_j : X \to \mathbb{R}$ such that

$$L_j(f) = \int_X f\sigma_j \, \mathrm{d}\mu \qquad \forall f \in L^1(X, \mu).$$

Hence for each $j = 1, 2, \ldots, n$ we have

$$L(fe_j) = \int_X f\sigma_j \, \mathrm{d}\mu \qquad \forall f \in C_c(X).$$

Since each vector-valued map $f \in C_c(X, \mathbb{R}^n)$ can be written as

$$f = \sum_{j=1}^{n} f_j e_j$$

for some $f_1, f_2, \ldots, f_n \in C_c(X)$, it follows that

$$L(f) = \sum_{j=1}^{n} L(f_j e_j) = \sum_{j=1}^{n} \int_X f_j \sigma_j \, \mathrm{d}\mu = \int_X \langle f, \sigma \rangle \, \mathrm{d}\mu \qquad \forall f \in C_c(X, \mathbb{R}^n)$$

where

$$\sigma := \sum_{j=1}^{n} \sigma_j e_j : X \to \mathbb{R}^n.$$

Step 3: It remains to show that $\|\sigma(x)\| = 1$ for μ-a.e. $x \in X$.

To see this, note that by using Cauchy-Schwarz it follows that for each $f \in C_c^+(X)$ we have

$$\sup\{|L(g)| : g \in C_c(X, \mathbb{R}^n), \|g\| \leqslant f\} \leqslant \int_X f \|\sigma\| \, \mathrm{d}\mu. \tag{☠}$$

Define

$$\hat{\sigma}(x) := \begin{cases} \frac{\sigma(x)}{\|\sigma(x)\|}, & \text{if } \sigma(x) \neq 0, \\ 0, & \text{if } \sigma(x) = 0. \end{cases}$$

Then $\hat{\sigma} : X \to \mathbb{R}^n$ is a μ-measurable map which belongs to $L^\infty(X, \mathbb{R}^n, \mu)$ with $\|\hat{\sigma}\|_{L^\infty} \leqslant 1$. Since $C_c(X, \mathbb{R}^n)$ is dense in $L^1(X, \mathbb{R}^n, \mu)$ (Theorem 6.2.5), there exists a sequence $\{g_k\}_{k=1}^{\infty} \subseteq C_c(X, \mathbb{R}^n)$ such that $g_k \to \hat{\sigma}$ in $L^1(X, \mathbb{R}^n, \mu)$ as $k \to \infty$. That is,

$$\lim_{k \to \infty} \int_X \|g_k - \hat{\sigma}\| \, \mathrm{d}\mu = 0.$$

Also define

$$\Phi : \mathbb{R}^n \to \mathbb{R}^n, \quad \Phi(y) := \begin{cases} \frac{y}{\|y\|}, & \text{if } \|y\| > 1, \\ y, & \text{if } \|y\| \leqslant 1. \end{cases}$$

and for each $k \in \mathbb{Z}^+$, define

$$\hat{g}_k := \Phi \circ g_k \in C_c(X, \mathbb{R}^n).$$

Then for each $k \in \mathbb{Z}^+$, clearly $\|\hat{g}_k\| \leqslant 1$ by definition of Φ. Also for each $k \in \mathbb{Z}^+$ and each $x \in X$ such that $\hat{\sigma}(x) \neq 0$, we have $\|\hat{\sigma}(x)\| = 1$ which implies that

$$\|\hat{g}_k(x) - \hat{\sigma}(x)\| \leqslant \|g_k(x) - \hat{\sigma}(x)\|.$$

(To see this note that if $v \in \mathbb{R}^n$ is such that $\|v\| = 1$ and y is an arbitrary vector in $\mathbb{R}^n$, then

$$\|\Phi(y) - v\| \leqslant \|y - v\|.$$

This follows from the fact that if $\|y\| \leqslant 1$, then $\Phi(y) = y$ and the inequality is trivial. If $\|y\| > 1$, then write $y = r\frac{y}{\|y\|}$ where $r := \|y\| > 1$, and define a function

$$\varphi : [1, \infty) \to \mathbb{R}, \quad \varphi(t) := \left\| t\frac{y}{\|y\|} - v \right\|^2 = t^2 - 2t \left\langle \frac{y}{\|y\|}, v \right\rangle + 1.$$

Then

$$\varphi'(t) = 2t - 2 \left\langle \frac{y}{\|y\|}, v \right\rangle = 2 \left(t - \left\langle \frac{y}{\|y\|}, v \right\rangle \right).$$

In particular, for $t \geqslant 1$ we have

$$\varphi'(t) = 2 \left(t - \left\langle \frac{y}{\|y\|}, v \right\rangle \right) \geqslant 2 \left(1 - \left\langle \frac{y}{\|y\|}, v \right\rangle \right) \geqslant 2 \left(1 - \left\| \frac{y}{\|y\|} \right\| \|v\| \right) = 0$$

by the Cauchy-Schwarz inequality. Thus φ is increasing on $[1, \infty)$, and so $r > 1$ implies that $\varphi(r) \geqslant \varphi(1)$, which says that

$$\left\| r\frac{y}{\|y\|} - v \right\|^2 \geqslant \left\| \frac{y}{\|y\|} - v \right\|^2$$

which by definition of Φ and $r = \|y\|$ is exactly the desired inequality.)

Therefore

$$\lim_{k \to \infty} \int_X \|\hat{g}_k - \hat{\sigma}\| \, d\mu = \lim_{k \to \infty} \left(\int_{\{x:\hat{\sigma}(x) \neq 0\}} \|\hat{g}_k(x) - \hat{\sigma}(x)\| \, d\mu + \int_{\{x:\hat{\sigma}(x) = 0\}} \|\hat{g}_k(x) - \hat{\sigma}(x)\| \, d\mu \right)$$

$$= \lim_{k \to \infty} \left(\int_{\{x:\hat{\sigma}(x) \neq 0\}} \|\hat{g}_k(x) - \hat{\sigma}(x)\| \, d\mu + \int_{\{x:\hat{\sigma}(x) = 0\}} \|\hat{g}_k(x)\| \, d\mu \right). \quad (\text{☺})$$

We can estimate the first integral as

$$\int_{\{x:\hat{\sigma}(x) \neq 0\}} \|\hat{g}_k(x) - \hat{\sigma}(x)\| \, d\mu \leqslant \int_{\{x:\hat{\sigma}(x) \neq 0\}} \|g_k(x) - \hat{\sigma}(x)\| \, d\mu \leqslant \int_X \|g_k - \hat{\sigma}\| \, d\mu$$

by the previous inequality. For the second integral, note that the fact that $g_k \to \hat{\sigma}$ in $L^1(X, \mathbb{R}^n, \mu)$ implies that

$$\lim_{k \to \infty} \int_{\{x:\hat{\sigma}(x) = 0\}} \|g_k(x)\| \, d\mu = \int_{\{x:\hat{\sigma}(x) = 0\}} \|\hat{\sigma}\| \, d\mu = 0$$

so there exists $M \in \mathbb{Z}^+$ such that $\|g_k\|_{L^1} < 1$ for each $k \geqslant M$. Thus for each $k \geqslant M$ and each $x \in X$ such that $\hat{\sigma}(x) = 0$, we have

$$\hat{g}_k(x) = \Phi(g_k(x)) = g_k(x).$$

It follows that

$$\lim_{k\to\infty} \int_{\{x:\hat\sigma(x)=0\}} \|\hat g_k(x)\| \ \mathrm d\mu = \lim_{k\to\infty} \int_{\{x:\hat\sigma(x)=0\}} \|g_k(x)\| \ \mathrm d\mu = 0.$$

Therefore the second integral in (☺) also converges to 0 as $k \to \infty$. Combining these facts, we conclude that

$$\lim_{k\to\infty} \int_X \|\hat g_k - \hat\sigma\| \ \mathrm d\mu = 0.$$

We claim that we actually have equality in (☻), i.e.,

$$\sup\{|L(g)| : g \in C_c(X,\mathbb R^n), \|g\| \leqslant f\} = \int_X f\|\sigma\| \ \mathrm d\mu, \qquad \forall f \in C_c^+(X). \qquad (☠')$$

To see this, note that for each $f \in C_c^+(X)$ the sequence $\{f\hat g_k\}_{k=1}^\infty \subseteq C_c(X,\mathbb R^n)$ satisfies $\|f\hat g_k\| \leqslant f$ for each $k \in \mathbb Z^+$, and by the previous limit and the Dominated Convergence Theorem we have

$$\lim_{k\to\infty} L(f\hat g_k) = \lim_{k\to\infty} \int_X f\langle\hat g_k,\sigma\rangle \ \mathrm d\mu = \int_X f\langle\hat\sigma,\sigma\rangle \ \mathrm d\mu = \int_X f\|\sigma\| \ \mathrm d\mu.$$

That is, for each $f \in C_c^+(X)$ there is a sequence of vector-valued maps satisfying the constraints in the supremum on the left side of (☻) whose L-values converge to the right side of (☻). Thus we must have equality in (☻).

By comparing the left side of (☠') with (†), we conclude that

$$\int_X f \ \mathrm d\mu = \int_X f\|\sigma\| \ \mathrm d\mu, \qquad \forall f \in C_c^+(X).$$

By decomposing an arbitrary $f \in C_c(X)$ into its positive and negative parts, we see that

$$\int_X f \ \mathrm d\mu = \int_X f\|\sigma\| \ \mathrm d\mu, \qquad \forall f \in C_c(X).$$

Since $C_c(X)$ is dense in $L^1(X,\mu)$ (Theorem 6.2.5), we have

$$\int_X g \ \mathrm d\mu = \int_X g\|\sigma\| \ \mathrm d\mu, \qquad \forall g \in L^1(X,\mu).$$

By taking $g = \mathbf 1_{\{\|\sigma\|\neq 1\}}$ we conclude that

$$\int_X \mathbf 1_{\{\|\sigma\|\neq 1\}} \ \mathrm d\mu = \int_X \mathbf 1_{\{\|\sigma\|\neq 1\}}\|\sigma\| \ \mathrm d\mu$$

which implies that either $\mu(\{\|\sigma\| \neq 1\}) = 0$ or $\|\sigma\| = 1$ μ-a.e. on $\{\|\sigma\| \neq 1\}$. The latter is impossible, so we must have $\mu(\{\|\sigma\| \neq 1\}) = 0$. That is, $\|\sigma(x)\| = 1$ for μ-a.e. $x \in X$. $\square$

Remark 6.3.3. Theorem 6.3.2 says that Radon measures are in one-to-one correspondence with bounded linear functionals on $C_c(X)$. Every Radon measure μ on X gives rise to an integral functional

$$L_\mu : C_c(X) \to \mathbb{R}, \quad f \mapsto \int_X f \, \mathrm{d}\mu$$

and the Riesz Representation Theorem says that every bounded linear functional on $C_c(X)$ arises in this way from a unique Radon measure on X.

The proof of the next theorem uses the Banach-Alaoglu theorem from linear functional analysis to prove a compactness result for Radon measures. There is a functional analysis interpretation of the Riesz Representation Theorem, but it is not needed for the rest of these notes so we omit it.

Theorem 6.3.4 (Compactness Theorem for Radon Measures). *Let X be a LCH space which is the union of countably many compact sets. Let $\{\mu_j\}_{j=1}^\infty$ be a sequence of Radon measures on X such that*

$$\sup_{j \in \mathbb{Z}^+} \mu_j(K) < \infty$$

for each compact set $K \subseteq X$. Then there exists a subsequence $\{\mu_{j_k}\}_{k=1}^\infty$ and a Radon measure μ on X such that

$$\lim_{k \to \infty} L_{\mu_{j_k}}(f) = L_\mu(f), \qquad \forall f \in C_c(X).$$

Here we use the notation $L_\mu(f) := \int_X f \, \mathrm{d}\mu$ for each Radon measure μ on X and each $f \in C_c(X)$.

Proof. Let $\{K_m\}_{m=1}^\infty$ be an increasing sequence of compact sets such that $X = \bigcup_{m=1}^\infty K_m$. For $j, m \in \mathbb{Z}^+$ define

$$L_{j,m}(f) := \int_{K_m} f \, \mathrm{d}\mu_j, \qquad \forall f \in C_c(K_m).$$

For each $m \in \mathbb{Z}^+$, the sequence $\{L_{j,m}\}_{j=1}^\infty$ is a sequence of bounded linear functionals on the Banach space $C(K_m)$ since

$$\sup_{j \in \mathbb{Z}^+} |L_{j,m}(f)| \leqslant \left(\sup_{x \in K_m} |f(x)| \right) \sup_{j \in \mathbb{Z}^+} \mu_j(K_m) < \infty, \qquad \forall f \in C(K_m).$$

Thus for each $m \in \mathbb{Z}^+$, applying the Banach-Alaoglu theorem to the sequence $\{L_{j,m}\}_{j=1}^\infty$ we see that there exists a subsequence $\{L_{j_k,m}\}_{k=1}^\infty$ that converges to a limit functional $L_m \in C(K_m)^*$.

By choosing the subsequences $\{L_{j_k,m}\}_{k=1}^\infty$ successively, so that $\{L_{j_k,m+1}\}_{k=1}^\infty$ is a subsequence of $\{L_{j_k,m}\}_{k=1}^\infty$ for each $m \in \mathbb{Z}^+$, we can use a diagonalization argument to obtain a single subsequence $\{L_{j_k,k}\}_{k=1}^\infty$ such that for each $m \in \mathbb{Z}^+$, the sequence $\{L_{j_k,m}\}_{k=m}^\infty$ converges

to L_m in $C(K_m)^*$ as $k \to \infty$. Since the collection $\{K_m\}_{m=1}^{\infty}$ is increasing, it is easy to see that the limit functionals L_m are compatible in the sense that

$$L_{m+1}(f) = L_m(f), \qquad \forall f \in C(K_m), \ \forall m \in \mathbb{Z}^+.$$

Thus we can define a linear functional $L : C_c(X) \to \mathbb{R}$ by

$$L(f) := L_m(f)$$

where $m \in \mathbb{Z}^+$ is such that $\operatorname{supp}(f) \subseteq K_m$. It is easy to see that L is well-defined, since $X = \bigcup_{m=1}^{\infty} K_m$ and the limit functionals L_m are compatible. Unravelling the definitions, we see that if $f \in C_c(X)$ and $m \in \mathbb{Z}^+$ is such that $\operatorname{supp}(f) \subseteq K_m$, then

$$L(f) = L_m(f) = \lim_{k \to \infty} L_{j_k, m}(f) = \lim_{k \to \infty} L_{\mu_{j_k}}(f)$$

Now use the Riesz Representation Theorem for Bounded Linear Functionals on $C_c(X)$ to see that there exists a unique Radon measure μ on X such that

$$L(f) = \int_X f \, d\mu, \qquad \forall f \in C_c(X).$$

This completes the proof of the theorem. $\square$

Chapter 7

Whitney's Extension Theorem

7.1 Taylor's Theorem with Integral Remainder

In this section we will present Taylor's Theorem with integral remainder for C^m functions, as motivation for the criterion in Whitney's C^m Extension Theorem.

Theorem 7.1.1 (Taylor's Theorem with Integral Remainder for C^m Functions on an Interval). *Let $[a, b] \subset \mathbb{R}$ be a closed interval, and let $f : [a, b] \to \mathbb{R}$ be a C^m function. Explicitly, this means that f is m times differentiable on $[a, b]$, and that the mth derivative of f is continuous on $[a, b]$.*

Then for each $x \in [a, b]$, we have

$$f(x) = f(a) + \sum_{k=1}^{m-1} \frac{f^{(k)}(a)}{k!} h^k + \frac{1}{(m-1)!} \int_a^x f^{(m)}(t)(x - t)^{m-1} \, \mathrm{d}t.$$

Proof. Fix a number $x \in [a, b]$. In what follows, all integrals and derivatives are taken with respect to the variable $t \in [a, x]$.

See that by repeated uses of the Fundamental Theorem of Calculus and Integration by

Parts, we have

$$
\begin{aligned}
f(x) - f(a) &= \int_a^x f'(t)\, \mathrm{d}t \\
&= -\int_a^x f'(t)\, \mathrm{d}(x - t) \\
&= -\left[f'(t)(x - t) \right]\Big|_{t=a}^{t=x} + \int_a^x f''(t)(x - t)\, \mathrm{d}t \\
&= f'(a)(x - a) - \frac{1}{2}\int_a^x f''(t)\, \mathrm{d}(x - t)^2 \\
&= f'(a)(x - a) + \left[f''(t) \cdot \frac{1}{2}(x - t)^2 \right]\Big|_{t=a}^{t=x} - \frac{1}{2}\int_a^x f'''(t) \cdot (x - t)^2\, \mathrm{d}t \\
&= f'(a)(x - a) + \frac{f''(a)}{2}(x - a)^2 - \frac{1}{2}\frac{1}{3}\int_a^x f'''(t)\, \mathrm{d}(x - t)^3 \\
&\;\;\vdots \\
&= f'(a)(x - a) + \frac{f''(a)}{2}(x - a)^2 + \cdots + \frac{f^{(m-1)}(a)}{(m-1)!}(x - a)^{m-1} + \frac{1}{(m-1)!}\int_a^x f^{(m)}(t)(x - t)^{m-1} \\
&= \sum_{k=1}^{m-1} \frac{f^{(k)}(a)}{k!}(x - a)^k + \frac{1}{(m-1)!}\int_a^x f^{(m)}(t)(x - t)^{m-1}\, \mathrm{d}t
\end{aligned}
$$

as desired. $\square$

Theorem 7.1.2 (Taylor's Formula with Integral Remainder)**.** *Let U be a open subset of $\mathbb{R}^n$, and let $f \in C^m(U)$. Fix a point $a \in U$.*

Then for each $x \in U$ where the line segment connecting a and x is contained in U, we have

$$
f(x) = \sum_{|\alpha| < m} \frac{D^\alpha f(a)}{\alpha!}(x - a)^\alpha + m \sum_{|\alpha| = m} \frac{(x - a)^\alpha}{\alpha!} \int_0^1 (1 - t)^{m-1} D^\alpha f(a + t(x - a))\, \mathrm{d}t
$$

and as a consequence, we have

$$
f(x) = \sum_{|\alpha| \leqslant m} \frac{D^\alpha f(a)}{\alpha!}(x-a)^\alpha + m \sum_{|\alpha| = m} \frac{(x - a)^\alpha}{\alpha!} \int_0^1 (1-t)^{m-1} \left(D^\alpha f(a + t(x - a)) - D^\alpha f(a) \right)\, \mathrm{d}t.
$$

Proof. Let $x \in U$ be such that the line segment connecting a and x is contained in U.

Define the function $g : [0, 1] \to \mathbb{R}$ by

$$
g(t) := f(a + t(x - a)) \qquad \forall\, t \in [0, 1].
$$

Notice that g is a C^m function on the interval $[0, 1]$, since f is a C^m function on U and the map $t \mapsto a + t(x - a)$ is smooth. We see that

$$
g(t) = f(a_1 + t(x_1 - a_1), \ldots, a_n + t(x_n - a_n)) \forall\, t \in [0, 1]
$$

so the first derivative of g is given by

$$g'(t) = \sum_{i=1}^{n} (x_i - a_i) \cdot D_i f(a + t(x - a)) \qquad \forall\, t \in [0, 1]$$

and the second derivative of g is given by

$$g''(t) = \sum_{i=1}^{n} \sum_{j=1}^{n} (x_i - a_i)(x_j - a_j) \cdot D_{i,j} f(a + t(x - a)) \qquad \forall\, t \in [0, 1]$$

and by continuing in this way, we see that the mth derivative of g is given by

$$g^{(m)}(t) = \sum_{j_1=1}^{n} \cdots \sum_{j_m=1}^{n} (x_{j_1} - a_{j_1}) \cdots (x_{j_m} - a_{j_m}) \cdot D_{j_1,\ldots,j_m} f(a + t(x - a)) \qquad \forall\, t \in [0, 1].$$

Using the multi-index notation, we see that for each $1 \leqslant k \leqslant m$, the kth derivative of g is given by

$$g^{(k)}(t) = \sum_{|\alpha|=k} \frac{k!}{\alpha!} (x - a)^\alpha D^\alpha f(a + t(x - a)) \qquad \forall\, t \in [0, 1].$$

Therefore, we can apply Theorem 7.1.1 to the function g and get

$$f(x) - f(a) = g(1) - g(0) = \sum_{k=1}^{m-1} \frac{g^{(k)}(0)}{k!} + \frac{1}{(m-1)!} \int_0^1 g^{(m)}(t)(1-t)^{m-1}\, dt$$

$$= \sum_{k=1}^{m-1} \sum_{|\alpha|=k} \frac{(x-a)^\alpha}{\alpha!} D^\alpha f(a) + \frac{1}{(m-1)!} \int_0^1 \sum_{|\alpha|=m} \frac{m!}{\alpha!} (x-a)^\alpha D^\alpha f(a + t(x-a))(1$$

$$= \sum_{|\alpha|<m} \frac{D^\alpha f(a)}{\alpha!} (x-a)^\alpha + m \sum_{|\alpha|=m} \frac{(x-a)^\alpha}{\alpha!} \int_0^1 (1-t)^{m-1} D^\alpha f(a + t(x-a))\, dt$$

as desired. This establishes the first formula in the statement of the theorem.

Since

$$m \int_0^1 (1-t)^{m-1}\, dt = 1,$$

we see that

$$
\begin{aligned}
f(x) - f(a) &= \sum_{|\alpha| < m} \frac{D^\alpha f(a)}{\alpha!} (x - a)^\alpha + m \sum_{|\alpha| = m} \frac{(x - a)^\alpha}{\alpha!} \int_0^1 (1 - t)^{m-1} D^\alpha f(a + t(x - a)) \, \mathrm{d}t \\
&= \sum_{|\alpha| < m} \frac{D^\alpha f(a)}{\alpha!} (x - a)^\alpha + \sum_{|\alpha| = m} \frac{D^\alpha f(a)}{\alpha!} (x - a)^\alpha - \sum_{|\alpha| = m} \frac{D^\alpha f(a)}{\alpha!} (x - a)^\alpha \\
&\quad + m \sum_{|\alpha| = m} \frac{(x - a)^\alpha}{\alpha!} \int_0^1 (1 - t)^{m-1} D^\alpha f(a + t(x - a)) \, \mathrm{d}t \\
&= \sum_{|\alpha| \leqslant m} \frac{D^\alpha f(a)}{\alpha!} (x - a)^\alpha - \left(\sum_{|\alpha| = m} \frac{D^\alpha f(a)}{\alpha!} (x - a)^\alpha \right) \left(m \int_0^1 (1 - t)^{m-1} \, \mathrm{d}t \right) \\
&\quad + m \sum_{|\alpha| = m} \frac{(x - a)^\alpha}{\alpha!} \int_0^1 (1 - t)^{m-1} D^\alpha f(a + t(x - a)) \, \mathrm{d}t \\
&= \sum_{|\alpha| \leqslant m} \frac{D^\alpha f(a)}{\alpha!} (x - a)^\alpha + m \sum_{|\alpha| = m} \frac{(x - a)^\alpha}{\alpha!} \int_0^1 (1 - t)^{m-1} \left(D^\alpha f(a + t(x - a)) - D^\alpha f(a) \right) \, \mathrm{d}t
\end{aligned}
$$

as desired.   □

Corollary 7.1.3 (Local Version of Taylor's Formula with Integral Remainder)**.** *Let $U \subseteq \mathbb{R}^n$ be an open subset, and let $f \in C^m(U)$. Then for each $a \in U$ we have*

$$
f(x) - \sum_{|\alpha| \leqslant m} \frac{D^\alpha f(a)}{\alpha!} (x - a)^\alpha = o(\|x - a\|^m) \quad as \quad x \to a.
$$

This hides the particular remainder functions, and is often more convenient to work with.

Proof. Fix a point $a \in U$. Since U is open, there is some $\delta > 0$ such that $B(a, \delta) \subseteq U$. Then, since $B(a, \delta)$ is a convex open subset of $\mathbb{R}^n$, we can apply Theorem 7.1.2 to the function f and the point a to get

$$
f(x) = \sum_{|\alpha| \leqslant m} \frac{D^\alpha f(a)}{\alpha!} (x - a)^\alpha + m \sum_{|\alpha| = m} \frac{(x - a)^\alpha}{\alpha!} \int_0^1 (1 - t)^{m-1} \left(D^\alpha f(a + t(x - a)) - D^\alpha f(a) \right) \, \mathrm{d}t
$$

for each $x \in B(a, \delta)$.

We estimate

$$\left| m \sum_{|\alpha|=m} \frac{(x-a)^\alpha}{\alpha!} \int_0^1 (1-t)^{m-1} \left(D^\alpha f(a+t(x-a)) - D^\alpha f(a) \right) \, \mathrm{d}t \right| \leqslant$$

$$\leqslant \sum_{|\alpha|=m} \frac{|(x-a)^\alpha|}{\alpha!} \left(m \int_0^1 (1-t)^{m-1} \left| D^\alpha f(a+t(x-a)) - D^\alpha f(a) \right| \, \mathrm{d}t \right)$$

$$\leqslant \sum_{|\alpha|=m} \frac{\|x-a\|^m}{\alpha!} \cdot \sup_{t\in[0,1]} \left| D^\alpha f(a+t(x-a)) - D^\alpha f(a) \right| \cdot \left(m \int_0^1 (1-t)^{m-1} \, \mathrm{d}t \right)$$

$$= \left(\sum_{|\alpha|=m} \frac{1}{\alpha!} \cdot \sup_{t\in[0,1]} \left| D^\alpha f(a+t(x-a)) - D^\alpha f(a) \right| \right) \|x-a\|^m.$$

Let $\varepsilon > 0$ be arbitrary. Then for each multi-index α with $|\alpha| = m$, since $D^\alpha f$ is continuous at a, there is some $\delta_\alpha > 0$ such that for each $t \in [0,1]$ and each $x \in B(a, \delta_\alpha)$, we have

$$\left| D^\alpha f(a+t(x-a)) - D^\alpha f(a) \right| < \varepsilon \cdot \left(\sum_{|\alpha|=m} \frac{1}{\alpha!} \right)^{-1}.$$

Let $\delta_0 := \min\{\delta_\alpha : |\alpha| = m\}$. Then, for each $x \in B(a, \delta_0)$, we have

$$\left| m \sum_{|\alpha|=m} \frac{(x-a)^\alpha}{\alpha!} \int_0^1 (1-t)^{m-1} \left(D^\alpha f(a+t(x-a)) - D^\alpha f(a) \right) \, \mathrm{d}t \right|$$

$$< \left(\sum_{|\alpha|=m} \frac{1}{\alpha!} \right) \cdot \varepsilon \cdot \left(\sum_{|\alpha|=m} \frac{1}{\alpha!} \right)^{-1} \cdot \|x-a\|^m$$

$$= \varepsilon \|x-a\|^m.$$

That is, for each $x \in B(a, \delta_0)$, we have

$$\left| f(x) - \sum_{|\alpha|\leqslant m} \frac{D^\alpha f(a)}{\alpha!} (x-a)^\alpha \right| < \varepsilon \|x-a\|^m.$$

Since $\varepsilon > 0$ was arbitrary, we conclude that

$$f(x) - \sum_{|\alpha|\leqslant m} \frac{D^\alpha f(a)}{\alpha!} (x-a)^\alpha = o(\|x-a\|^m) \quad \text{as} \quad x \to a.$$

$\square$

Definition 7.1.4. Let $U \subseteq \mathbb{R}^n$ be an open subset, and let $f \in C^m(U)$. We define the mth *order Taylor polynomial of f at a point $a \in U$* to be the function $T_a^m f : \mathbb{R}^n \to \mathbb{R}$ defined by

$$T_a^m f(x) := \sum_{|\alpha| \leqslant m} \frac{f^{(\alpha)}(a)}{\alpha!} (x - a)^\alpha \qquad \forall \, x \in \mathbb{R}^n.$$

Exercise 7.1.1. Let $U \subseteq \mathbb{R}^n$ be an open set, and let $f \in C^m(U)$. Fix a point $a \in U$.

Show that for each multi-index β with $|\beta| \leqslant m$, the function $D^\beta T_a^m f$ is the $m - |\beta|$ order Taylor polynomial of $D^\beta f$ at a, i.e. that

$$D^\beta T_a^m f = T_a^{m - |\beta|}(D^\beta f).$$

Then as a result, for each multi-index β with $|\beta| \leqslant m$, we have

$$D^\beta f(x) - D^\beta T_a^m f(x) = o(\|x - a\|^{m - |\beta|}) \quad \text{as} \quad x \to a.$$

Proof. Let β be a multi-index with $|\beta| \leqslant m$. The definition of the Taylor polynomial of $D^\beta f$ gives us

$$T_a^{m - |\beta|}(D^\beta f)(x) = \sum_{|\gamma| \leqslant m - |\beta|} \frac{D^\gamma (D^\beta f)(a)}{\gamma!}(x - a)^\gamma = \sum_{|\gamma| \leqslant m - |\beta|} \frac{D^{\gamma + \beta} f(a)}{\gamma!}(x - a)^\gamma \qquad \forall \, x \in \mathbb{R}^n$$

where we have commuted the order of differentiation in the last step, which is justified since $f \in C^m(U)$.

Now we compute $D^\beta T_a^m f$ as

$$D^\beta T_a^m f = \sum_{|\alpha| \leqslant m} \frac{D^\alpha f(a)}{\alpha!} D^\beta (\,\cdot - a)^\alpha.$$

If α is a multi-index with $|\alpha| < |\beta|$, then $D^\beta(\,\cdot - a)^\alpha = 0$. If α is a multi-index with $|\alpha| \geqslant |\beta|$, then see that

$$D^\beta(\,\cdot - a)^\alpha = \frac{\alpha!}{(\alpha - \beta)!}(\,\cdot - a)^{\alpha - \beta}.$$

Therefore, we have

$$\begin{aligned} D^\beta T_a^m f &= \sum_{|\beta| \leqslant |\alpha| \leqslant m} \frac{D^\alpha f(a)}{\alpha!} \cdot \frac{\alpha!}{(\alpha - \beta)!}(\,\cdot - a)^{\alpha - \beta} \\ &= \sum_{|\beta| \leqslant |\alpha| \leqslant m} \frac{D^\alpha f(a)}{(\alpha - \beta)!}(\,\cdot - a)^{\alpha - \beta} \\ &= \sum_{|\gamma| \leqslant m - |\beta|} \frac{D^{\gamma + \beta} f(a)}{\gamma!}(\,\cdot - a)^\gamma \\ &= T_a^{m - |\beta|}(D^\beta f) \end{aligned}$$

as desired.

As a result of corollary 7.1.3, since $D^\beta f \in C^{m-|\beta|}(U)$, we have

$$D^\beta f(x) - D^\beta T_a^m f(x) = D^\beta f(x) - T_a^{m-|\beta|}(D^\beta f)(x) = o(\|x-a\|^{m-|\beta|}) \quad \text{as} \quad x \to a.$$

Because β was an arbitrary multi-index with $|\beta| \leqslant m$, the result follows. $\square$

Lemma 7.1.5 (Uniform Remainder Estimate on Compact Sets). *Let $U \subseteq \mathbb{R}^n$ be an open set, and let $f \in C^m(U)$. Let $K \subset U$ be a compact set.*

Then for each multi-index β with $|\beta| \leqslant m$, and for each $\varepsilon > 0$, there exists $\delta > 0$ such that for each $a, x \in K$ with $0 < \|x-a\| < \delta$, we have

$$\frac{\left|D^\beta f(x) - D^\beta T_a^m f(x)\right|}{\|x-a\|^{m-|\beta|}} < \varepsilon.$$

We will write this as

$$D^\beta f(x) - D^\beta T_a^m f(x) = o(\|x-a\|^{m-|\beta|}) \quad \text{as} \quad x \to a, \quad \text{uniformly for} \quad a, x \in K.$$

In other words, we have a uniform version of the little-o condition in Exercise 7.1.1 on compact subsets of U.

Proof. For each point $a \in K$, there is a radius $\delta_a > 0$ such that $B(a, \delta_a) \subseteq U$. The collection $\{B(a, \delta_a)\}_{a \in K}$ is an open cover of the compact set K, so there is a finite subcover $\{B(a_j, \delta_{a_j})\}_{j=1}^N$ of K. By Lebesgue's number lemma, there is some $\delta > 0$ such that for each $x \in K$, there is some $j \in \{1, \ldots, N\}$ such that $B(x, \delta) \subseteq B(a_j, \delta_{a_j})$.

Let β be a multi-index with $|\beta| \leqslant m$, and let $\varepsilon > 0$ be arbitrary. See that if $a, x \in K$ are such that $0 < \|x-a\| < \delta$, then there is some $j \in \{1, \ldots, N\}$ such that $B(x, \delta) \subseteq B(a_j, \delta_{a_j})$ and thus $\|x - a_j\| < \delta$ implies that $x, a \in B(a_j, \delta_{a_j}) \subseteq U$. Thus if $a, x \in K$ are such that $0 < \|x-a\| < \delta$, then U contains the line segment connecting a and x, so we can apply Theorem 7.1.2 to obtain the estimate

$$\left|D^\beta f(x) - D^\beta T_a^m f(x)\right| = \left|D^\beta f(x) - T_a^{m-|\beta|}(D^\beta f)(x)\right|$$

$$= \left|m \sum_{|\alpha|=m-|\beta|} \frac{(x-a)^\alpha}{\alpha!} \int_0^1 (1-t)^{m-1} \left(D^{\alpha+\beta} f(a + t(x-a)) - D^{\alpha+\beta} f(a)\right) \, dt\right|$$

$$\leqslant \sum_{|\alpha|=m-|\beta|} \frac{|(x-a)^\alpha|}{\alpha!} m \int_0^1 (1-t)^{m-1} \left|D^{\alpha+\beta} f(a + t(x-a)) - D^{\alpha+\beta} f(a)\right| \, dt$$

$$\leqslant \sum_{|\alpha|=m-|\beta|} \frac{\|x-a\|^{m-|\beta|}}{\alpha!} \cdot \sup_{t \in [0,1]} \left|D^{\alpha+\beta} f(a + t(x-a)) - D^{\alpha+\beta} f(a)\right| \cdot \left(m \int_0^1 (1-t)^{m-}\right.$$

$$\leqslant \left(\sum_{|\alpha|=m-|\beta|} \frac{1}{\alpha!} \cdot \sup_{y,z \in K} |D^{\alpha+\beta} f(y) - D^{\alpha+\beta} f(z)|\right) \|x-a\|^{m-|\beta|}.$$

Then for each multi-index α with order $|\alpha| = m - |\beta|$, the function $D^{\alpha+\beta}f$ is continuous on the compact set K, so it is uniformly continuous on K. Thus there is some $\delta_\varepsilon > 0$ such that for each multi-index α with order $|\alpha| = m - |\beta|$, and for each $y, z \in K$ with $\|y - z\| < \delta_\varepsilon$, we have

$$|D^{\alpha+\beta}f(y) - D^{\alpha+\beta}f(z)| < \varepsilon \cdot \left(\sum_{|\alpha|=m-|\beta|} \frac{1}{\alpha!} \right)^{-1}.$$

Now if $a, x \in K$ are such that $0 < \|x - a\| < \min\{\delta, \delta_\varepsilon\}$, then we have

$$\left| D^\beta f(x) - D^\beta T_a^m f(x) \right| \leqslant \left(\sum_{|\alpha|=m-|\beta|} \frac{1}{\alpha!} \cdot \sup_{y,z \in K} |D^{\alpha+\beta}f(y) - D^{\alpha+\beta}f(z)| \right) \|x - a\|^{m-|\beta|}$$

$$< \left(\sum_{|\alpha|=m-|\beta|} \frac{1}{\alpha!} \right) \cdot \varepsilon \cdot \left(\sum_{|\alpha|=m-|\beta|} \frac{1}{\alpha!} \right)^{-1} \cdot \|x - a\|^{m-|\beta|} = \varepsilon \|x - a\|^{m-|\beta|}$$

which gives the desired estimate. $\qquad\qquad\square$

7.2   Decomposition of Open Sets into Cubes

In this section, we will construct a particularly nice decomposition of an open set into almost disjoint cubes. We will then use this decomposition to construct a nice partition of unity for the open set, which will be a crucial tool for defining the extension operators in the next section.

7.2.1   The Whitney Decomposition

Definition 7.2.1 (Dyadic Cubes)**.** For each $k \in \mathbb{Z}$, we let $\mathcal{D}_k$ be the collection of all dyadic closed cubes of the form

$$Q = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : m_j 2^{-k} \leqslant x_j \leqslant (m_j + 1)2^{-k} \text{ for } j = 1, \ldots, n\}$$

where $m_j \in \mathbb{Z}$ for each $j = 1, \ldots, n$.

The set of all dyadic cubes is defined as $\mathcal{D} = \bigcup_{k \in \mathbb{Z}} \mathcal{D}_k$.

Exercise 7.2.1 (Basic Facts about Dyadic Cubes)**.**   (i) Let $k \in \mathbb{Z}$. For each $Q \in \mathcal{D}_k$ there exists 2^n almost disjoint cubes $Q_1, \ldots, Q_{2^n} \in \mathcal{D}_{k+1}$ such that $Q = \bigcup_{j=1}^{2^n} Q_j$.

(ii) If $Q, Q' \in \mathcal{D}$ are such that the interiors of Q and Q' have nonempty intersection, then either $Q \subseteq Q'$ or $Q' \subseteq Q$. In other words, the dyadic cubes are either almost disjoint or one is contained in the other.

(iii) Let $k \in \mathbb{Z}$. For each $Q \in \mathcal{D}_k$ there are exactly $3^n - 1$ other cubes in $\mathcal{D}_k$ which intersect Q.

Proof. 1. Let $Q \in \mathcal{D}_k$ be given by

$$Q = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : m_j 2^{-k} \leqslant x_j \leqslant (m_j + 1)2^{-k} \text{ for } j = 1, \ldots, n\}.$$

For each $j = 1, \ldots, n$, we define the intervals

$$I_j^0 = [m_j 2^{-k}, (m_j + \tfrac{1}{2})2^{-k}] \quad \text{and} \quad I_j^1 = [(m_j + \tfrac{1}{2})2^{-k}, (m_j + 1)2^{-k}].$$

Then we can define the cube $Q_{(i_1, \ldots, i_n)}$ for each $i_j \in \{0, 1\}$ by

$$Q_{(i_1, \ldots, i_n)} = I_1^{i_1} \times \cdots \times I_n^{i_n}.$$

Then we have

$$Q = \bigcup_{i_1, \ldots, i_n \in \{0,1\}} Q_{(i_1, \ldots, i_n)},$$

and the collection of cubes $\{Q_{(i_1, \ldots, i_n)}\}_{i_j \in \{0,1\}}$ are pairwise almost disjoint and belong to $\mathcal{D}_{k+1}$. Furthermore, there are 2^n such cubes, as claimed.

2. Let $Q, Q' \in \mathcal{D}$ be such that $Q \cap Q' \neq \varnothing$. Then there exist $k, k' \in \mathbb{Z}$ such that $Q \in \mathcal{D}_k$ and $Q' \in \mathcal{D}_{k'}$. Without loss of generality, we can assume that $k \leqslant k'$. We write Q and Q' as

$$Q = \prod_{j=1}^{n}[m_j 2^{-k}, (m_j + 1)2^{-k}] \quad \text{and} \quad Q' = \prod_{j=1}^{n}[m_j' 2^{-k'}, (m_j' + 1)2^{-k'}]$$

for some integers $m_1, \ldots, m_n$ and $m_1', \ldots, m_n'$.

Since the interiors of Q and Q' have nonempty intersection, there exists a point $x = (x_1, \ldots, x_n)$ in their common interior. Then for each $j = 1, \ldots, n$, we have

$$m_j 2^{-k} < x_j < (m_j + 1)2^{-k} \quad \text{and} \quad m_j' 2^{-k'} < x_j < (m_j' + 1)2^{-k'}.$$

Since $k \leqslant k'$, we have $2^{-k} \geqslant 2^{-k'}$, and so for each $j = 1, \ldots, n$ the interval $[m_j 2^{-k}, (m_j + 1)2^{-k}]$ contains the interval $[m_j' 2^{-k'}, (m_j' + 1)2^{-k'}]$. Therefore, we have $Q' \subseteq Q$, as claimed.

3. Write

$$Q = \prod_{i=1}^{n}[m_i 2^{-k}, (m_i + 1)2^{-k}]$$

for some $(m_1, \ldots, m_n) \in \mathbb{Z}^n$. If

$$Q' = \prod_{i=1}^{n}[m_i' 2^{-k}, (m_i' + 1)2^{-k}] \in \mathcal{D}_k$$

is such that $Q' \cap Q \neq \varnothing$, then for each $j = 1, \ldots, n$ we must have $|m_j - m'_j| \leqslant 1$, so that $m'_j \in \{m_j - 1, m_j, m_j + 1\}$.

Thus if $Q' \in \mathcal{D}_k$ is such that $Q' \cap Q \neq \varnothing$, then there are 3^n possible choices for the vector $(m'_1, \ldots, m'_n)$, and hence there are 3^n cubes in $\mathcal{D}_k$ which intersect Q. However, one of these cubes is Q itself, so there are exactly $3^n - 1$ other cubes in $\mathcal{D}_k$ which intersect Q as claimed.

$\square$

Definition 7.2.2 (Adjacent Cubes). Two cubes $Q, Q' \subset \mathbb{R}^n$ are said to be adjacent if Q and Q' are almost disjoint but $\partial Q \cap \partial Q' \neq \varnothing$. In other words, Q and Q' are adjacent if their interiors are disjoint but they share a common boundary point.

Proposition 7.2.3 (Whitney Decomposition). *Let $A \subset \mathbb{R}^n$ be a non-empty proper closed subset. Then there exists a countable collection of almost disjoint dyadic closed cubes $\{Q_j\}_{j=1}^{\infty}$ such that*

(a) $A^c = \displaystyle\bigcup_{j=1}^{\infty} Q_j$, and

(b) for each $j \in \mathbb{Z}^+$, we have

$$\operatorname{diam} Q_j \leqslant \operatorname{dist}(Q_j, A) \leqslant 4 \operatorname{diam} Q_j.$$

As a consequence, note that

(c) if $j, k \in \mathbb{Z}^+$ are such that Q_j and Q_k are adjacent, then

$$\frac{1}{4} \leqslant \frac{\operatorname{diam} Q_j}{\operatorname{diam} Q_k} \leqslant 4.$$

As a further consequence, note that

(d) for each $j \in \mathbb{Z}^+$, there are at most $12^n - 4^n$ cubes in the collection which are adjacent to Q_j.

The collection of dyadic cubes $\{Q_j\}_{j=1}^{\infty}$ is called a Whitney decomposition of the open set A^c.

In other words, the key property (b) says the size of a cube in the Whitney decomposition is comparable to its distance from the set A.

Proof of (a) and (b). For each $k \geqslant 1$, we let

$$\Omega_k := \left\{ x \in A^c : 2\sqrt{n}2^{-k} < \operatorname{dist}(x, A) \leqslant 4\sqrt{n} \cdot 2^{-k} \right\}$$

and see that $A^c = \bigcup_{k=1}^{\infty} \Omega_k$.

We define $\mathcal{F}$ to be the collection of all closed cubes $Q \subset \mathbb{R}^n$ such that there exists $k \geqslant 1$ such that $Q \in \mathcal{D}_k$ and $Q \cap \Omega_k \neq \varnothing$. We claim that property (b) holds for the cubes in $\mathcal{F}$.

Let $Q \in \mathcal{F}$, and let $k \geqslant 1$ be such that $Q \in \mathcal{D}_k$ and $Q \cap \Omega_k \neq \varnothing$. Then choose $x \in Q \cap \Omega_k$, and observe that

$$
\begin{aligned}
\operatorname{diam} Q = \sqrt{n}2^{-k} &\leqslant \operatorname{dist}(x, A) - \sqrt{n}2^{-k} \\
&= \operatorname{dist}(x, A) - \operatorname{diam} Q \\
&\leqslant \operatorname{dist}(Q, A) \\
&\leqslant 4\sqrt{n}2^{-k} \\
&= 4\operatorname{diam} Q
\end{aligned}
$$

which gives

$$
\operatorname{diam} Q \leqslant \operatorname{dist}(Q, A) \leqslant 4\operatorname{diam} Q
$$

as claimed.

We also claim that

$$
A^c = \bigcup_{Q \in \mathcal{F}} Q.
$$

Note that each cube in $\mathcal{F}$ is contained in A^c, since it has distance at least $\operatorname{diam} Q$ from A. Thus we have $\bigcup_{Q \in \mathcal{F}} Q \subseteq A^c$.

Conversely, let $x \in A^c$ be given. Then there exists $k \geqslant 1$ such that $x \in \Omega_k$, and so there exists a cube $Q \in \mathcal{D}_k$ such that $x \in Q$. Since $x \in \Omega_k$, we have $\operatorname{dist}(x, A) > 2\sqrt{n}2^{-k}$, and so $Q \cap \Omega_k \neq \varnothing$. Thus $Q \in \mathcal{F}$, and so $x \in \bigcup_{Q \in \mathcal{F}} Q$. Since x was arbitrary, we have $A^c \subseteq \bigcup_{Q \in \mathcal{F}} Q$, and so $A^c = \bigcup_{Q \in \mathcal{F}} Q$ as claimed.

Thus the collection $\mathcal{F}$ satisfies properties (a) and (b). The problem is that the cubes in $\mathcal{F}$ are not almost disjoint. Recall that by Exercise 7.2.1, if $Q_1, Q_2 \in \mathcal{F}$ and $Q_1^{\circ} \cap Q_2^{\circ} \neq \varnothing$, then either $Q_1 \subseteq Q_2$ or $Q_2 \subseteq Q_1$. That is, two cubes in $\mathcal{F}$ either have disjoint interiors or one is contained in the other.

Therefore for each cube $Q \in \mathcal{F}$, we let Q^{max} be the maximal cube in $\mathcal{F}$ which contains Q; such a cube exists and is unique by the previous observation. By maximality and our previous observation, the maximal cubes have pairwise disjoint interiors, and hence are almost disjoint. Noting that the collection $\mathcal{F}$ is countable, we can enumerate the maximal cubes as $\{Q_j\}_{j=1}^{\infty}$. Then $\{Q_j\}_{j=1}^{\infty}$ is a Whitney cover of A^c because

- the cubes $\{Q_j\}_{j=1}^{\infty}$ are almost disjoint by construction;

- we have $A^c = \bigcup_{Q \in \mathcal{F}} Q = \bigcup_{j=1}^{\infty} Q_j$; and

- each cube Q_j belongs to $\mathcal{F}$, and hence satisfies $\operatorname{diam} Q_j \leqslant \operatorname{dist}(Q_j, A) \leqslant 4 \operatorname{diam} Q_j$.

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Proof of (c) and (d). (c) Suppose $j, k \geqslant 1$ are such that the cubes Q_j and Q_k are adjacent. Then Q_j and Q_k are almost disjoint, but $\partial Q_j \cap \partial Q_k \neq \varnothing$, and so $\operatorname{dist}(Q_j, Q_k) = 0$. Then we see that

$$\operatorname{diam} Q_j \leqslant \operatorname{dist}(Q_j, A) \leqslant \operatorname{dist}(Q_j, Q_k) + \operatorname{dist}(Q_k, A) \leqslant 0 + 4 \operatorname{diam} Q_k$$

which gives $\operatorname{diam} Q_j \leqslant 4 \operatorname{diam} Q_k$.

By symmetry, we also have $\operatorname{diam} Q_k \leqslant 4 \operatorname{diam} Q_j$, and hence

$$\frac{1}{4} \leqslant \frac{\operatorname{diam} Q_j}{\operatorname{diam} Q_k} \leqslant 4$$

as desired.

(d) Let $j \in \mathbb{Z}^+$ be arbitrary. Note that the largest number of dyadic cubes adjacent to Q_j is obtained when all of the adjacent cubes have the smallest possible size. By part (c), the smallest possible size of a cube adjacent to Q_j is $\frac{1}{4} \operatorname{diam} Q_j$.

We claim that for each $Q \in \mathcal{D}_k$ there are at most 4^n cubes in $\{Q_j\}_{j=1}^\infty$ with diameter $\geqslant \frac{1}{4} \operatorname{diam} Q$ which are contained in Q.

Proof of Claim. Let $Q \in \mathcal{D}_k$ be arbitrary. Then each cube Q_j with $\operatorname{diam} Q_j \geqslant \frac{1}{4} \operatorname{diam} Q$ which is contained in Q must belong to $\mathcal{D}_k$, or $\mathcal{D}_{k-1}$, or $\mathcal{D}_{k-2}$. We cannot have $Q_j \in \mathcal{D}_{k-3}$ because then $\operatorname{diam} Q_j < \frac{1}{4} \operatorname{diam} Q$, and we cannot have $Q_j \in \mathcal{D}_{k+1}$ because then $\operatorname{diam} Q_j > 4 \operatorname{diam} Q$ and hence Q_j cannot be contained in Q.

In the first case, there is a unique cube in $\mathcal{D}_k$ which is equal to Q, so there is at most one cube in $\{Q_j\}_{j=1}^\infty$ with diameter $\geqslant \frac{1}{4} \operatorname{diam} Q$ which is contained in Q and belongs to $\mathcal{D}_k$.

In the second case where $Q_j \in \mathcal{D}_{k-1}$, there are at most 2^n cubes in $\mathcal{D}_{k-1}$ which are contained in Q, and hence at most 2^n cubes in $\{Q_j\}_{j=1}^\infty$ with diameter $\geqslant \frac{1}{4} \operatorname{diam} Q$ which are contained in Q and belong to $\mathcal{D}_{k-1}$.

In the third case where $Q_j \in \mathcal{D}_{k-2}$, there are at most 4^n cubes in $\mathcal{D}_{k-2}$ which are contained in Q, and hence at most 4^n cubes in $\{Q_j\}_{j=1}^\infty$ with diameter $\geqslant \frac{1}{4} \operatorname{diam} Q$ which are contained in Q and belong to $\mathcal{D}_{k-2}$. $\qquad\qquad$ $\square$

Combining the previous claim with exercise 7.2.1 (c) shows that there are at most $4^n(3^n - 1) = 12^n - 4^n$ cubes in the collection which are adjacent to Q_j.

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Corollary 7.2.4. *Let $A \subseteq \mathbb{R}^n$ be a non-empty proper closed subset, and let $\{Q_j\}_{j=1}^{\infty}$ be the Whitney decomposition of A^c given by Proposition 7.2.3.*

Fix $0 < \varepsilon < \frac{2}{5}$. For each $j \in \mathbb{Z}^+$, let $Q_j^* = (1 + \varepsilon)Q_j$ be the cube obtained by dilating Q_j by a factor of $1 + \varepsilon$ about its center.

Then for each $j \in \mathbb{Z}^+$, we have $Q_j^* \subseteq A^c$. Also if $j, k \in \mathbb{Z}^+$ are such that $Q_j \cap Q_k = \varnothing$, then $Q_j^* \cap Q_k^* = \varnothing$.

Furthermore,

$$\mathbf{1}_{A^c} \leqslant \sum_{j=1}^{\infty} \mathbf{1}_{Q_j^*} \leqslant 2^n \mathbf{1}_{A^c}.$$

Proof. Let $j \in \mathbb{Z}^+$ be arbitrary. Then we have $Q_j^* = (1 + \varepsilon)Q_j$, and since each cube adjacent to Q_j has diameter at least $\frac{1}{4} \operatorname{diam} Q_j$, we see that Q_j^* is contained in the union of Q_j and the cubes adjacent to Q_j. Since each cube adjacent to Q_j is contained in A^c, we have $Q_j^* \subseteq A^c$ as claimed.

Moreover, see that

$$A^c = \bigcup_{j=1}^{\infty} Q_j \subseteq \bigcup_{j=1}^{\infty} Q_j^*$$

implies that

$$\mathbf{1}_{A^c} \leqslant \mathbf{1}_{\bigcup_{j=1}^{\infty} Q_j^*} = \sum_{j=1}^{\infty} \mathbf{1}_{Q_j^*}$$

which is the first inequality.

Suppose now that $j, k \in \mathbb{Z}^+$ are such that $Q_j \cap Q_k = \varnothing$. Then Q_j and Q_k are not adjacent, and since Q_j^* is contained in the union of Q_j and the cubes adjacent to Q_j, we have $Q_j^* \cap Q_k = \varnothing$. We claim that in fact $Q_j^* \cap Q_k^* = \varnothing$.

Proof of Claim. Since Q_j^* is contained in the union of Q_j and the cubes adjacent to Q_j, and since Q_k^* is contained in the union of Q_k and the cubes adjacent to Q_k, we see that for the intersection $Q_j^* \cap Q_k^*$ to be nonempty, there must be a cube in the Whitney decomposition which is adjacent to both Q_j and Q_k. Let Q_l be a cube in the Whitney decomposition which is adjacent to both Q_j and Q_k. Then by part (c) of Proposition 7.2.3, we have

$$\frac{1}{4} \leqslant \frac{\operatorname{diam} Q_j}{\operatorname{diam} Q_l} \leqslant 4 \qquad \text{and} \qquad \frac{1}{4} \leqslant \frac{\operatorname{diam} Q_k}{\operatorname{diam} Q_l} \leqslant 4.$$

Without loss of generality, we can assume that $\operatorname{diam} Q_j \leqslant \operatorname{diam} Q_k$. Then we must have $\operatorname{diam} Q_k \leqslant 16 \operatorname{diam} Q_j$ by combining the previous inequalities, which implies that

$$\operatorname{diam} Q_k = c \cdot \operatorname{diam} Q_j \qquad \text{for some } c \in \{1, 2, 4, 8, 16\}.$$

DRAW PICTURE

Assume that $\operatorname{diam} Q_k = \operatorname{diam} Q_j$. Then Q_j and Q_k are non-adjacent dyadic cubes in $\mathcal{D}_N$ for some $N \in \mathbb{Z}$, and hence must have $\operatorname{dist}(Q_j, Q_k) \geq 2^{-N} = \ell(Q_j)$. (Why? The dyadic cubes in $\mathcal{D}_N$ are arranged in a grid, and two cubes in $\mathcal{D}_k$ are adjacent or separated by at least one cube in $\mathcal{D}_N$.) Hence, the cubes Q_j^* and Q_k^* have a positive distance

$$
\begin{aligned}
\operatorname{dist}(Q_j^*, Q_k^*) &> \operatorname{dist}\left(\left(\frac{7}{5} Q_j \right), \left(\frac{7}{5} Q_k \right) \right) \\
&\geq \operatorname{dist}(Q_j, Q_k) - \frac{1}{2} \frac{2}{5} \ell(Q_j) - \frac{1}{2} \frac{2}{5} \ell(Q_k) \\
&\geq \ell(Q_j) - \frac{1}{5} \ell(Q_j) - \frac{1}{5} \ell(Q_k) \\
&= \frac{3}{5} \ell(Q_j) > 0
\end{aligned}
$$

where the first inequality follows from the fact that $Q_j^* = (1 + \varepsilon) Q_j$ and $1 + \varepsilon < \frac{7}{5}$, and the second inequality follows from the triangle inequality.

Now assume that $\operatorname{diam} Q_k = 2 \operatorname{diam} Q_j$. Then Q_j and Q_k are non-adjacent dyadic cubes in $\mathcal{D}_{N+1}$ and $\mathcal{D}_N$ respectively for some $N \in \mathbb{Z}$, and hence must have $\operatorname{dist}(Q_j, Q_k) \geq 2^{-N-1} = \frac{1}{2} \ell(Q_k)$. (Why? Since $Q_k \in \mathcal{D}_N$ we see that Q_k is a union of 2^n cubes in $\mathcal{D}_{N+1}$, and hence by identical reasoning as in the previous case with one of these cubes in $\mathcal{D}_{N+1}$ we see that Q_j and Q_k must be separated by at least one cube in $\mathcal{D}_{N+1}$, which has side length $\ell(Q_j) = \frac{1}{2} \ell(Q_k)$.) Hence the cubes Q_j^* and Q_k^* have a positive distance

$$
\begin{aligned}
\operatorname{dist}(Q_j^*, Q_k^*) &> \operatorname{dist}\left(\left(\frac{7}{5} Q_j \right), \left(\frac{7}{5} Q_k \right) \right) \\
&\geq \operatorname{dist}(Q_j, Q_k) - \frac{1}{2} \frac{2}{5} \ell(Q_j) - \frac{1}{2} \frac{2}{5} \ell(Q_k) \\
&\geq \frac{1}{2} \ell(Q_k) - \frac{1}{5} \ell(Q_j) - \frac{1}{5} \ell(Q_k) \\
&= \frac{3}{10} \ell(Q_k) - \frac{1}{10} \ell(Q_k) = \frac{1}{5} \ell(Q_k) > 0.
\end{aligned}
$$

Now assume that $\operatorname{diam} Q_k = 4 \operatorname{diam} Q_j$. Then Q_j and Q_k are non-adjacent dyadic cubes in $\mathcal{D}_{N+2}$ and $\mathcal{D}_N$ respectively for some $N \in \mathbb{Z}$, and hence must have $\operatorname{dist}(Q_j, Q_k) \geq 2^{-N-2} = \frac{1}{4} \ell(Q_k)$. (Why? Since $Q_k \in \mathcal{D}_{N-2}$ we see that Q_k is a union of 2^{2n} cubes in $\mathcal{D}_N$, and hence by identical reasoning as in the previous case with one of these cubes in $\mathcal{D}_N$ we see that Q_j and Q_k must be separated by at least one cube in $\mathcal{D}_N$, which has side length $\ell(Q_j)$.) Hence

the cubes Q_j^* and Q_k^* have a positive distance

$$\mathrm{dist}(Q_j^*, Q_k^*) > \mathrm{dist}\left(\left(\tfrac{7}{5}Q_j\right), \left(\tfrac{7}{5}Q_k\right)\right)$$

$$\geqslant \mathrm{dist}(Q_j, Q_k) - \tfrac{1}{2}\tfrac{2}{5}\ell(Q_j) - \tfrac{1}{2}\tfrac{2}{5}\ell(Q_k)$$

$$\geqslant \tfrac{1}{4}\ell(Q_k) - \tfrac{1}{5}\ell(Q_j) - \tfrac{1}{5}\ell(Q_k)$$

$$= \tfrac{5}{20}\ell(Q_k) - \tfrac{1}{20}\ell(Q_k) - \tfrac{4}{20}\ell(Q_k) = 0.$$

In the cases that $\mathrm{diam}\,Q_k = 8\,\mathrm{diam}\,Q_j$ or $\mathrm{diam}\,Q_k = 16\,\mathrm{diam}\,Q_j$, the cubes Q_j and Q_k must have distance at least $\tfrac{1}{4}\ell(Q_k)$, because the smallest cube adjacent to Q_k has side length at least $\tfrac{1}{4}\ell(Q_k)$. In these cases, the cubes Q_j^* and Q_k^* have a positive distance, as

$$\mathrm{dist}(Q_j^*, Q_k^*) \geqslant \mathrm{dist}\left(\left(\tfrac{7}{5}Q_j\right), \left(\tfrac{7}{5}Q_k\right)\right)$$

$$\geqslant \mathrm{dist}(Q_j, Q_k) - \tfrac{1}{2}\tfrac{2}{5}\ell(Q_j) - \tfrac{1}{2}\tfrac{2}{5}\ell(Q_k)$$

$$\geqslant \tfrac{1}{4}\ell(Q_k) - \tfrac{1}{5}\ell(Q_j) - \tfrac{1}{5}\ell(Q_k)$$

$$= \tfrac{1}{20}\ell(Q_k) - \tfrac{1}{5}\ell(Q_j)$$

and in the case $\mathrm{diam}\,Q_k = 8\,\mathrm{diam}\,Q_j$ we have $\ell(Q_j) = \tfrac{1}{8}\ell(Q_k)$, and in the case $\mathrm{diam}\,Q_k = 16\,\mathrm{diam}\,Q_j$ we have $\ell(Q_j) = \tfrac{1}{16}\ell(Q_k)$, so that in either case we have $\mathrm{dist}(Q_j^*, Q_k^*) > 0$.

This case analysis shows that if $Q_j \cap Q_k = \varnothing$, then Q_j^* and Q_k^* are disjoint, as claimed. □

It remains to prove the final inequality.

We claim that the maximum number of pairwise adjacent cubes in the Whitney decomposition such that their dilations by a factor of $1 + \varepsilon$ contain a common point is at most 2^n.

Proof of Claim. Assume that there are cubes $Q_{j_1}, \ldots, Q_{j_M}$ in the Whitney decomposition, and that the intersection of their dilations is nonempty, i.e. $Q_{j_1}^* \cap \cdots \cap Q_{j_M}^* \neq \varnothing$. By the contrapositive of the previous claim, we see that the intersection of any pair of cubes in the collection $\{Q_{j_1}, \ldots, Q_{j_M}\}$ is nonempty. If $M < 2^n$, then we are done. The only way to have $M \geqslant 2^n$ is if the cubes $Q_{j_1}, \ldots, Q_{j_M}$ all have a common corner vertex, and hence are all adjacent to each other, and this forces $M = 2^n$. □

In particular, for each point $x \in A^c$, there are at most 2^n cubes in the collection $\{Q_j^*\}_{j=1}^{\infty}$ which contain x, and hence

$$\sum_{j=1}^{\infty} \mathbf{1}_{Q_j^*}(x) \leqslant 2^n$$

for each $x \in A^c$ as claimed.

If $x \in A$, then x cannot belong to any of the cubes in the Whitney decomposition (as their union equals A^c), and hence for each $j \in \mathbb{Z}^+$ we have $x \notin Q_j^*$, since Q_j^* is contained in the union of Q_j and the cubes adjacent to Q_j. Thus, for each $x \in A$, we have $\sum_{j=1}^{\infty} \mathbf{1}_{Q_j^*}(x) = 0$, so that

$$\sum_{j=1}^{\infty} \mathbf{1}_{Q_j^*}(x) = 0 = 2^n \mathbf{1}_{A^c}(x)$$

as claimed.

This completes the proof of the final inequality, and hence the proof of the corollary. $\square$

7.2.2   Some Lemmas on Higher Order Derivatives and Bump Functions

Exercise 7.2.2 (Higher Order Quotient Rule). Let $\Omega \subseteq \mathbb{R}^n$ be an open set, and let $f : \Omega \to \mathbb{R}$ be a C^m function which is nonzero on Ω. Then for each multi-index $\beta \in \mathbb{N}^n$ with $0 < |\beta| \leqslant m$, we have

$$D^\beta \left(\frac{1}{f} \right) = -\frac{1}{f} \sum_{0 < \gamma \leqslant \beta} \binom{\beta}{\gamma} D^\gamma f \cdot D^{\beta - \gamma} \left(\frac{1}{f} \right).$$

Proof. Let $\beta \in \mathbb{N}^n$ be a multi-index with $0 < |\beta| \leqslant m$. Since f is nonzero on Ω, we can write

$$f(x) \frac{1}{f(x)} = 1 \qquad \forall x \in \Omega$$

so differentiating both sides gives

$$D^\beta \left(f \cdot \frac{1}{f} \right) = D^\beta(1) = 0.$$

Hence the Leibniz rule says that

$$\sum_{\gamma \leqslant \beta} \binom{\beta}{\gamma} D^\gamma f \cdot D^{\beta - \gamma} \left(\frac{1}{f} \right) = 0$$

which implies that

$$f \cdot D^\beta \left(\frac{1}{f} \right) + \sum_{0 < \gamma \leqslant \beta} \binom{\beta}{\gamma} D^\gamma f \cdot D^{\beta - \gamma} \left(\frac{1}{f} \right) = 0.$$

Rearranging this gives

$$D^\beta \left(\frac{1}{f} \right) = -\frac{1}{f} \sum_{0 < \gamma \leqslant \beta} \binom{\beta}{\gamma} D^\gamma f \cdot D^{\beta - \gamma} \left(\frac{1}{f} \right)$$

as desired. $\square$

Definition 7.2.5 (Bump Function Adapted to an Interval or Cube). Fix $0 < \varepsilon < \frac{1}{4}$. Let $\eta \in C_c^\infty(\mathbb{R})$ be a non-negative even function such that

$$\eta(t) = \begin{cases} 0 & \text{if } |t| \geqslant \frac{1}{2} + \frac{\varepsilon}{2} \\ 1 & \text{if } |t| \leqslant \frac{1}{2} \end{cases}$$

and η is strictly decreasing on $(-\frac{1}{2} - \frac{\varepsilon}{2}, -\frac{1}{2})$ and $(\frac{1}{2}, \frac{1}{2} + \frac{\varepsilon}{2})$.

For each dyadic interval $I \subset \mathbb{R}$, we define the function $\eta_I : \mathbb{R} \to \mathbb{R}$ by

$$\eta_I(t) := \eta\left(\frac{t - c_I}{\ell(I)}\right),$$

where c_I is the center of the interval I and $\ell(I)$ is the length of the interval I.

For each dyadic cube $Q \subset \mathbb{R}^n$, write $Q = I_1 \times \cdots \times I_n$ where $I_1, \ldots, I_n$ are dyadic intervals, and define the function $\eta_Q : \mathbb{R}^n \to \mathbb{R}$ by

$$\eta_Q(x) := \eta_{I_1}(x_1) \cdots \eta_{I_n}(x_n).$$

Exercise 7.2.3 (Basic Facts). Fix $0 < \varepsilon < \frac{1}{4}$ and let η be a bump function as in Definition 7.2.5. Then for each dyadic interval $I \subset \mathbb{R}$ we let $I^* := (1 + \varepsilon)I$ to be the interval with the same center as I but with length $(1 + \varepsilon)\ell(I)$, and for each dyadic cube $Q = I_1 \times \cdots \times I_n$ in $\mathbb{R}^n$ we let $Q^* := I_1^* \times \cdots \times I_n^*$.

Then for each dyadic interval $I \subset \mathbb{R}$, we have $\eta_I \in C_c^\infty(\mathbb{R})$ and

$$\eta_I(t) = \begin{cases} 0 & \text{if } t \notin I^* \\ 1 & \text{if } t \in I \end{cases}$$

so that $\operatorname{supp} \eta_I \subseteq I^*$ and $\eta_I|_I = 1$.

Similarly for each dyadic cube $Q = I_1 \times \cdots \times I_n$ in $\mathbb{R}^n$ we have $\eta_Q \in C_c^\infty(\mathbb{R}^n)$ and

$$\eta_Q(x) = \begin{cases} 0 & \text{if } x \notin Q^* := I_1^* \times \cdots \times I_n^* \\ 1 & \text{if } x \in Q \end{cases}$$

so that $\operatorname{supp} \eta_Q \subseteq Q^*$ and $\eta_Q|_Q = 1$.

Proof. Let $I \subset \mathbb{R}$ be a dyadic interval. The definition of η_I shows that $\eta_I \in C_c^\infty(\mathbb{R})$, and that $\eta_I(t) = 0$ if $|t - c_I| \geqslant \ell(I)\left(\frac{1}{2} + \frac{\varepsilon}{2}\right)$, which is equivalent to $t \notin I^*$. Similarly, $\eta_I(t) = 1$ if $|t - c_I| \leqslant \frac{1}{2}\ell(I)$, which is equivalent to $t \in I$.

This shows that $\operatorname{supp} \eta_I \subseteq I^*$ and $\eta_I|_I = 1$.

Let $Q = I_1 \times \cdots \times I_n$ be a dyadic cube in $\mathbb{R}^n$. The definition of η_Q shows that $\eta_Q \in C_c^\infty(\mathbb{R}^n)$, and that $\eta_Q(x) = 0$ if $|x_j - c_{I_j}| \geqslant \ell(I_j)\left(\frac{1}{2} + \frac{\varepsilon}{2}\right)$ for some $1 \leqslant j \leqslant n$, which is equivalent to $x \notin Q^*$. Similarly, $\eta_Q(x) = 1$ if $|x_j - c_{I_j}| \leqslant \ell(I_j)/2$ for each $1 \leqslant j \leqslant n$, which is equivalent to $x \in Q$.

This shows that $\operatorname{supp} \eta_Q \subseteq Q^*$ and $\eta_Q|_Q = 1$. $\qquad\square$

Lemma 7.2.6 (Derivative Bounds). *Fix $0 < \varepsilon < \frac{1}{4}$, and let $\eta \in C_c^\infty(\mathbb{R})$ be an even non-negative bump function as in Definition 7.2.5.*

1. *For each dyadic interval $I \subset \mathbb{R}$, and each $m \in \mathbb{Z}^+$, there is a constant $C_{m,\eta} > 0$ which depends only on m and η such that*

$$\left|\eta_I^{(m)}(t)\right| \leqslant C_{m,\eta}\ell(I)^{-m}$$

   for all $t \in \mathbb{R}$.

2. *For each dyadic cube $Q \subset \mathbb{R}^n$, and each multi-index $\alpha \in \mathbb{N}^n$, there is a constant $C_{n,\alpha,\eta} > 0$ which depends only on n, α and η such that*

$$|\partial^\alpha \eta_Q(x)| \leqslant C_{n,\alpha,\eta}(\operatorname{diam} Q)^{-|\alpha|}$$

   for all $x \in \mathbb{R}^n$.

Proof.   1. See that for each dyadic interval I, and each $m \in \mathbb{Z}^+$, we have the m-th derivative of η_I given by

$$\eta_I^{(m)}(t) = \frac{1}{\ell(I)^m}\eta^{(m)}\left(\frac{t - c_I}{\ell(I)}\right)$$

by iterating the chain rule, which implies that

$$\left|\eta_I^{(m)}(t)\right| \leqslant \|\eta^{(m)}\|_{L^\infty(\mathbb{R})}\ell(I)^{-m} \tag{$\star$}$$

for all $t \in \mathbb{R}$.

2. Let $Q = I_1 \times \cdots \times I_n$ be a dyadic cube in $\mathbb{R}^n$, where $I_1, \ldots, I_n$ are dyadic intervals. Note that for each multi-index $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$, we have

$$\partial^\alpha \eta_Q(x) = \partial^\alpha(\eta_{I_1}(x_1)\cdots\eta_{I_n}(x_n)) = \eta_{I_1}^{(\alpha_1)}(x_1)\cdots\eta_{I_n}^{(\alpha_n)}(x_n)$$

and so by $(\star)$ we have

$$|\partial^\alpha \eta_Q(x)| \leqslant \|\eta^{(\alpha_1)}\|_{L^\infty(\mathbb{R})}\cdots\|\eta^{(\alpha_n)}\|_{L^\infty(\mathbb{R})}\ell(I_1)^{-\alpha_1}\cdots\ell(I_n)^{-\alpha_n}$$

for all $x \in \mathbb{R}^n$. By defining

$$C_{n,\alpha,\eta} := \|\eta^{(\alpha_1)}\|_{L^\infty(\mathbb{R})}\cdots\|\eta^{(\alpha_n)}\|_{L^\infty(\mathbb{R})}n^{\frac{|\alpha|}{2}}$$

and noting that $\ell(I_1) = \cdots = \ell(I_n) = \frac{\operatorname{diam} Q}{\sqrt{n}}$, we have

$$|\partial^\alpha \eta_Q(x)| \leqslant C_{n,\alpha,\eta}(\operatorname{diam} Q)^{-|\alpha|}$$

for all $x \in \mathbb{R}^n$ as desired.

$\square$

Lemma 7.2.7 (Integral Bounds)**.** *Fix* $0 < \varepsilon < \frac{1}{4}$ *and let* η *be a bump function as in Definition 7.2.5.*

1. *There is a constant* $B_\eta > 0$ *such that if* $I, J \subset \mathbb{R}$ *are dyadic intervals such that* $\ell(I) \leqslant 4\ell(J)$ *and* $\eta_I \eta_J$ *is not identically zero, then*

$$\frac{1}{\ell(I)} \int_{\mathbb{R}} \eta_I(y) \eta_J(y) \, \mathrm{d}y \geqslant B_\eta.$$

2. *There is a constant* $B_{n,\eta} > 0$ *such that if* $Q, R \subset \mathbb{R}^n$ *are dyadic cubes such that* $\ell(Q) \leqslant 4\ell(R)$ *and* $\eta_Q \eta_R$ *is not identically zero, then*

$$\frac{1}{\mathrm{vol}(Q)} \int_{\mathbb{R}^n} \eta_Q(y) \eta_R(y) \, \mathrm{d}y \geqslant B_{n,\eta}.$$

Proof. 1. By assumption that $\eta_I \eta_J$ is not identically zero, we must have $J^* \cap I^* \neq \varnothing$, and hence $J \cap I$ is nonempty by the proof of Corollary 7.2.4. Since I and J are dyadic intervals, they must be adjacent or one must be contained in the other.

By translating both I and J by a fixed number, we can assume without loss of generality that $I = [0, 2^k]$ for some $k \in \mathbb{Z}$. (As the expression we want to lower bound is invariant under translations, this does not change the value of the integral.)

By a dilation of both I and J by a factor of $\frac{1}{\ell(I)}$, we can assume without loss of generality that $I = [0, 1]$. (Since the expression we want to lower bound is invariant under dilations, this does not change the value of the integral.) Then J is a dyadic interval such that $\ell(J) \geqslant \frac{1}{4}$.

Since either J is adjacent to $[0, 1]$, or J is contained in $[0, 1]$, or $[0, 1]$ is contained in J, the only possibilities for J are

$$\left[-\frac{1}{4}, 0\right], \left[0, \frac{1}{4}\right], \left[\frac{1}{4}, \frac{1}{2}\right], \left[\frac{1}{2}, \frac{3}{4}\right], \left[\frac{3}{4}, 1\right], \left[1, \frac{5}{4}\right],$$

$$\left[-\frac{1}{2}, \frac{1}{2}\right], \left[0, \frac{1}{2}\right], \left[\frac{1}{2}, 1\right], \left[1, \frac{3}{2}\right]$$

$$[-1, 0], [0, 1], [1, 2]$$

or

$$\left[-2^k, 0\right], \left[0, 2^k\right] \qquad \text{for some } k \geqslant 1.$$

Now see that

$$\frac{1}{\ell(I)} \int_{\mathbb{R}} \eta_I(y) \eta_J(t) \, \mathrm{d}t = \int_{\mathbb{R}} \eta\left(\frac{t - c_J}{\ell(J)}\right) \eta\left(t - \frac{1}{2}\right) \, \mathrm{d}t.$$

Since the function $\eta\left(\cdot - \frac{1}{2}\right)$ is strictly decreasing and never vanishes on $\left[1, 1 + \frac{\varepsilon}{2}\right]$, it follows that for each of the above choices of J, there is a constant $B_\eta > 0$ such that

$$\int_\mathbb{R} \eta\left(\frac{t - c_J}{\ell(J)}\right) \eta\left(t - \frac{1}{2}\right) \, \mathrm{d}t \geqslant B_\eta,$$

and choosing the minimum of these constants over all the above choices of J (this occurs for the *first* listed choice of J), we get a constant $B_\eta > 0$ which satisfies the desired bound.

2. Let $Q = I_1 \times \cdots \times I_n$ and $R = J_1 \times \cdots \times J_n$ be dyadic cubes in $\mathbb{R}^n$, where

$$\ell(I_1) = \ell(I_2) = \cdots = \ell(I_n) = \ell(Q), \qquad \ell(J_1) = \ell(J_2) = \cdots = \ell(J_n) = \ell(R).$$

By the assumption that $\eta_Q \eta_R$ is not identically zero, we must have $R^* \cap Q^* \neq \varnothing$, and hence $R \cap Q$ is nonempty by the proof of Corollary 7.2.4. Therefore, for each $1 \leqslant i \leqslant n$, we have $J_i \cap I_i \neq \varnothing$.

Also $\ell(Q) \leqslant 4\ell(R)$, implies that $\ell(I_i) \leqslant 4\ell(J_i)$ for each $1 \leqslant i \leqslant n$. Then we see that

$$\begin{aligned}
\frac{1}{\mathrm{vol}(Q)} \int_{\mathbb{R}^n} \eta_Q(y) \eta_R(y) \, \mathrm{d}y &= \prod_{i=1}^{n} \frac{1}{\ell(I_i)} \int_\mathbb{R} \eta_{I_i}(y_i) \eta_{J_i}(y_i) \, \mathrm{d}y_i \\
&\geqslant \prod_{i=1}^{n} B_\eta = B_\eta^n =: B_{n,\eta}
\end{aligned}$$

by the one-dimensional case, which gives the desired bound.

$\square$

7.2.3 Whitney Partition of Unity

Proposition 7.2.8 (Whitney Partition of Unity). *Let $A \subset \mathbb{R}^n$ be a non-empty proper closed subset, and let $\{Q_j\}_{j=1}^\infty$ be a Whitney decomposition of A^c. Fix $0 < \varepsilon < \frac{1}{4}$, and for each $j \in \mathbb{Z}^+$ let $Q_j^* = (1 + \varepsilon)Q_j$ be the cube obtained by dilating Q_j by a factor of $1 + \varepsilon$ about its center. Then there exists a collection of bump functions $\{\psi_j\}_{j=1}^\infty \subset C_c^\infty(\mathbb{R}^n)$ such that*

(i) $0 \leqslant \psi_j \leqslant 1$ for each $j \in \mathbb{Z}^+$;

(ii) $\mathrm{supp}\, \psi_j \subseteq Q_j^*$ for each $j \in \mathbb{Z}^+$;

(iii) the collection $\{\psi_j\}_{j=1}^\infty$ forms a partition of unity for A^c, i.e.

$$\sum_{j=1}^{\infty} \psi_j = \mathbf{1}_{A^c},$$

(iv) *for each multi-index α, there exists a constant $C_\alpha > 0$ such that for each $j \in \mathbb{Z}^+$ we have*

$$|\partial^\alpha \psi_j(x)| \leqslant C_\alpha (\operatorname{diam} Q_j)^{-|\alpha|}$$

for all $x \in A^c$;

(v) *for each $j \geqslant 1$, we have*

$$\frac{1}{2^n} \leqslant \frac{1}{\operatorname{vol}(Q_j)} \int_{\mathbb{R}^n} \psi_j(y) \, \mathrm{d}y \leqslant (1 + \varepsilon)^n,$$

(vi) *for each multi-index α, there exists a constant $B_\alpha > 0$ such that for all $j, l \in \mathbb{Z}^+$ we have*

$$|\partial^\alpha (\psi_j \psi_l)(x)| \leqslant \frac{B_\alpha}{(\operatorname{diam} Q_l)^{|\alpha|+n}} \int_{\mathbb{R}^n} \psi_j(y) \psi_l(y) \, \mathrm{d}y$$

for all $x \in A^c$.

The collection $\{\psi_j\}_{j=1}^\infty$ is called a Whitney partition of unity of the open set A^c, which is adapted to the Whitney decomposition $\{Q_j\}_{j=1}^\infty$.

The most important properties of the Whitney partition of unity are (i), (ii), and (iii), which say that the functions $\{\psi_j\}_{j=1}^\infty$ form a smooth partition of unity for the open set A^c subordinated to the Whitney decomposition $\{Q_j\}_{j=1}^\infty$.

Remark 7.2.9. The proof will begin with a non-negative even bump function $\eta \in C_c^\infty(\mathbb{R})$, and all functions in the partition of unity will be built from this. Technically all constants will depend on the choice of η, but we will suppress this dependence in the notation.

Proof. Let $\eta \in C_c^\infty(\mathbb{R})$ be a non-negative even bump function such that $\eta(t) = 1$ for $|t| \leqslant \frac{1}{2}$ and $\eta(t) = 0$ for $|t| \geqslant \frac{1}{2} + \frac{\varepsilon}{2}$, as in Definition 7.2.5.

Let $\{Q_j\}_{j=1}^\infty$ be a Whitney decomposition of the open set A^c. For each $j \in \mathbb{Z}^+$, we define the function $\phi_j := \eta_{Q_j} \in C_c^\infty(\mathbb{R}^n)$, and notice that $\phi_j \leqslant \mathbf{1}_{Q_j^*}$, since $\eta_{Q_j}(x) = 1$ for all $x \in Q_j$ and $\eta_{Q_j}(x) = 0$ for all $x \notin (1 + \varepsilon)Q_j$. Also see that if $x \in A^c$, then there is at least one $j \in \mathbb{Z}^+$ such that $x \in Q_j$, and so $\sum_{j=1}^\infty \phi_j(x) \geqslant 1$. As a result, we have

$$\mathbf{1}_{A^c} \leqslant \sum_{j=1}^\infty \phi_j \leqslant \sum_{j=1}^\infty \mathbf{1}_{Q_j^*} \leqslant 2^n \mathbf{1}_{A^c} \tag{☺}$$

by Corollary 7.2.4.

To obtain a partition of unity, we define the functions

$$\psi_j : \mathbb{R}^n \to \mathbb{R}, \qquad \psi_j(x) := \begin{cases} \frac{\phi_j(x)}{\sum_{k=1}^\infty \phi_k(x)} & \text{if } x \in A^c \\ 0 & \text{if } x \in A \end{cases}$$

for $j \in \mathbb{Z}^+$.

Notice that if $x \in A^c$, then $\sum_{k=1}^{\infty} \phi_k(x) \geqslant 1$, but at most $12^n - 4^n$ of the terms in the sum are nonzero by Proposition 7.2.3 (d), so this sum is also finite. Thus $\psi_j(x)$ is well-defined for each $j \in \mathbb{Z}^+$ and $x \in A^c$.

See that for each $j \in \mathbb{Z}^+$, we have $\psi_j \geqslant 0$ and $\psi_j \leqslant 1$, so that (i) holds. Also see that $\operatorname{supp} \psi_j \subseteq Q_j^*$, since ϕ_j has support in Q_j^*, so (ii) holds. Furthermore, since $\sum_{k=1}^{\infty} \phi_k(x) \geqslant 1$ for each $x \in A^c$, we have

$$\sum_{j=1}^{\infty} \psi_j(x) = \sum_{j=1}^{\infty} \frac{\phi_j(x)}{\sum_{k=1}^{\infty} \phi_k(x)} = 1$$

for each $x \in A^c$, and this sum consists of at most 12^n nonzero terms. Since $\sum_{j=1}^{\infty} \psi_j = 0$ on A, we have

$$\sum_{j=1}^{\infty} \psi_j = \mathbf{1}_{A^c},$$

so (iii) holds.

(iv). Fix a multi-index $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$, and let $j \in \mathbb{Z}^+$ be arbitrary. Let $\ell(Q_j)$ be the length of the cube Q_j, so that $\sqrt{n}\ell(Q_j) = \operatorname{diam} Q_j$. Then the Leibniz rule gives

$$D^\alpha \psi_j = D^\alpha \left(\frac{\varphi_j}{\sum_{k=1}^{\infty} \varphi_k} \right) = \sum_{\beta \leqslant \alpha} \binom{\alpha}{\beta} D^\beta \left(\frac{1}{\sum_{k=1}^{\infty} \varphi_k} \right) D^{\alpha-\beta} \varphi_j$$

on $\mathbb{R}^n$. Then using the Derivative Bounds from Lemma 7.2.6, we have

$$\left| D^\alpha \varphi_j(x) \right| \leqslant \sum_{\beta \leqslant \alpha} \binom{\alpha}{\beta} C_{n,\alpha-\beta} \left| D^\beta \left(\frac{1}{S} \right)(x) \right| \cdot \ell(Q_j)^{-|\alpha|+|\beta|} \tag{$\heartsuit$}$$

for all $x \in \mathbb{R}^n$.

For brevity, we let $S := \sum_{k=1}^{\infty} \varphi_k$. We claim that for each $\beta \leqslant \alpha$ there is a constant $C_{n,\beta}^\bullet > 0$ such that if $l \in \mathbb{Z}^+$ and $x \in Q_l^*$, then

$$\left| D^\beta \left(\frac{1}{S} \right)(x) \right| \leqslant C_{n,\beta}^\bullet \ell(Q_l)^{-|\beta|}. \tag{$\Diamond$}$$

Proof of Claim. Base case: If $\beta = 0$, then the left-hand side of ($\Diamond$) is just $\frac{1}{S(x)}$, and the right-hand side is $C_{n,0}^\bullet$, so the claim holds because $S(x) \geqslant 1$ for each $x \in A^c$.

Base case: If β has $|\beta| = 1$, then $\beta = e_i$ for some $i \in \{1, \ldots, n\}$, and the left-hand side of ($\Diamond$) is

$$\left| D^\beta \left(\frac{1}{S} \right)(x) \right| = \left| -\frac{\partial_i S(x)}{(S(x))^2} \right| \leqslant \frac{|\partial_i S(x)|}{S(x)^2} \leqslant |\partial_i S(x)|$$

since $S(x) \geqslant 1$ for each $x \in A^c$. Now if $l \in \mathbb{Z}^+$ and $x \in Q_l^*$, then there are at most 12^n nonzero terms in the sum $S(x) = \sum_{k=1}^{\infty} \varphi_k(x)$, and each of these nonzero terms corresponds to a cube Q_k which is adjacent to Q_l; for each such cube Q_k, we have $\frac{1}{4} \leqslant \frac{\operatorname{diam} Q_k}{\operatorname{diam} Q_l} \leqslant 4$ by 7.2.3, so that $\ell(Q_k) \leqslant 4\ell(Q_l)$, and hence $\ell(Q_k)^{-1} \leqslant 4\ell(Q_l)^{-1}$. Therefore

$$\left| D^\beta \left(\frac{1}{S} \right)(x) \right| \leqslant |\partial_i S(x)| \leqslant 12^n \max_{k \geqslant 1, x \in Q_k^*} |\partial_i \varphi_k(x)| \leqslant 12^n C_{n,e_i} \max_{k \geqslant 1, x \in Q_k^*} \ell(Q_k)^{-1} \leqslant 12^n C_{n,e_i} 4\ell(Q_l)^{-1}$$

for all $x \in Q_l^*$.

Since β was an arbitrary multi-index with $|\beta| = 1$, this proves the base case.

Inductive step: Suppose that there is an integer N such that for each multi-index with order $< N$, the claim holds.

Let $\beta \in \mathbb{N}^n$ be a multi-index with $|\beta| = N$, and apply the Higher Order Quotient Rule to obtain

$$D^\beta \left(\frac{1}{S} \right) = -\frac{1}{S} \sum_{0 < \gamma \leqslant \beta} \binom{\beta}{\gamma} D^\gamma S \cdot D^{\beta-\gamma} \left(\frac{1}{S} \right).$$

Then we estimate the right-hand side using the triangle inequality and the induction hypothesis — if $l \in \mathbb{Z}^+$ and $x \in Q_l^*$, then

$$\left| D^\beta \left(\frac{1}{S} \right)(x) \right| \leqslant \left| \frac{1}{S(x)} \right| \sum_{0 < \gamma \leqslant \beta} \binom{\beta}{\gamma} |D^\gamma S(x)| \cdot \left| D^{\beta-\gamma} \left(\frac{1}{S} \right)(x) \right|$$

$$\leqslant \sum_{0 < \gamma \leqslant \beta} \binom{\beta}{\gamma} |D^\gamma S(x)| C^\bullet_{n,\beta-\gamma} \ell(Q_l)^{-|\beta|+|\gamma|} \qquad \text{by the induction hypothesis}$$

and for each $0 < \gamma \leqslant \beta$, we have

$$\left| D^\gamma S(x) \right| \leqslant 12^n \max_{k \geqslant 1, x \in Q_k^*} |D^\gamma \varphi_k(x)| \leqslant 12^n C_{n,\gamma} \max_{k \geqslant 1, x \in Q_k^*} \ell(Q_k)^{-|\gamma|} \leqslant 12^n C_{n,\gamma} \left(4^{|\gamma|} \ell(Q_l)^{-|\gamma|} \right),$$

similarly to the base case.

Putting these estimates together, we have

$$\left| D^\beta \left(\frac{1}{S} \right)(x) \right| \leqslant \sum_{0 < \gamma \leqslant \beta} \binom{\beta}{\gamma} \left[12^n C_{n,\gamma} \left(4^{|\gamma|} \ell(Q_l)^{-|\gamma|} \right) \right] C^\bullet_{n,\beta-\gamma} \ell(Q_l)^{-|\beta|+|\gamma|}$$

$$= 12^n \sum_{0 < \gamma \leqslant \beta} \binom{\beta}{\gamma} C_{n,\gamma} C^\bullet_{n,\beta-\gamma} 4^{|\gamma|} \ell(Q_l)^{-|\beta|}$$

$$= C^\bullet_{n,\beta} \ell(Q_l)^{-|\beta|}$$

where we have defined the constant

$$C^\bullet_{n,\beta} := 12^n \sum_{0 < \gamma \leqslant \beta} \binom{\beta}{\gamma} C_{n,\gamma} C^\bullet_{n,\beta-\gamma} 4^{|\gamma|}.$$

This completes the inductive step, and hence the claim holds for all multi-indices $\beta \leqslant \alpha$. $\qquad\square$

Now let $x \in A^c$ be arbitrary, and let $l \in \mathbb{Z}^+$ be such that $x \in Q_l^*$. Then there are at most 12^n nonzero terms in the sum $S(x) = \sum_{k=1}^\infty \varphi_k(x)$, and each of these nonzero terms corresponds to a cube Q_k which is adjacent to Q_l.

To finish the proof of (iv), we split into two cases, depending on if Q_j and Q_l are equal or adjacent, or otherwise if Q_j and Q_l are not adjacent.

If the cubes Q_j and Q_l are not adjacent, then $x \in Q_l^*$ but $x \notin Q_j^*$, so $\varphi_j = 0$ on an open set containing x, and hence $\psi_j(x) = 0$; in this case the desired estimate holds trivially as the left-hand side is zero.

If the cubes Q_j and Q_l are adjacent, then by part (c) of 7.2.3 we have $\frac{1}{4} \leqslant \frac{\text{diam}\, Q_j}{\text{diam}\, Q_l} \leqslant 4$, so that

$$\ell(Q_j) \leqslant 4\ell(Q_l)$$

and hence

$$\ell(Q_l)^{-|\beta|} \leqslant 4^{|\beta|} \ell(Q_j)^{-|\beta|}$$

for each $\beta \leqslant \alpha$, which implies that

$$\left| D^\beta \left(\frac{1}{S} \right)(x) \right| \leqslant C^\bullet_{n,\beta} \ell(Q_l)^{-|\beta|} \leqslant 4^{|\beta|} C^\bullet_{n,\beta} \ell(Q_j)^{-|\beta|} \tag{$\diamondsuit_2$}$$

by using ($\diamondsuit$).

Putting ($\heartsuit$) and ($\diamondsuit_2$) together and using the triangle inequality, we have

$$|D^\alpha \psi_j(x)| \leqslant \sum_{\beta \leqslant \alpha} \binom{\alpha}{\beta} C_{n,\alpha-\beta} \left(4^{|\beta|} C^\bullet_{n,\beta} \ell(Q_l)^{-|\beta|} \right) \ell(Q_j)^{-|\alpha|+|\beta|} \leqslant C^\diamond_{n,\alpha} \ell(Q_j)^{-|\alpha|}$$

where we have defined the constant

$$C^\diamond_{n,\alpha} := \sum_{\beta \leqslant \alpha} \binom{\alpha}{\beta} 4^{|\beta|} C_{n,\alpha-\beta} C^\bullet_{n,\beta}$$

which depends only on n and α.

Since

$$\text{diam}\, Q_j = \sqrt{n}\ell(Q_j)$$

we have

$$|D^\alpha \psi_j(x)| \leqslant C^\diamond_{n,\alpha} \ell(Q_j)^{-|\alpha|} = C^\diamond_{n,\alpha} n^{\frac{|\alpha|}{2}} (\text{diam}\, Q_j)^{-|\alpha|}$$

for all $x \in A^c$, which completes the proof of (iv) with the constant $C_\alpha := C_{n,\alpha}^\diamond n^{\frac{|\alpha|}{2}}$.

(v). For each $j \in \mathbb{Z}^+$, we have

$$
\begin{aligned}
1 \leqslant \frac{1}{\mathrm{vol}(Q_j)} \int_{\mathbb{R}^n} \eta_{Q_j}(y) \, \mathrm{d}y &= \prod_{i=1}^{n} \frac{1}{\ell(I_i)} \int_{\mathbb{R}} \eta_{I_i}(y_i) \, \mathrm{d}y_i \\
&= \prod_{i=1}^{n} \frac{1}{\ell(I_i)} \int_{\mathbb{R}} \eta\left(\frac{y_i - c_{I_i}}{\ell(I_i)}\right) \, \mathrm{d}y_i \\
&\leqslant (1+\varepsilon)^n \qquad\qquad (🧹)
\end{aligned}
$$

where the first equality holds by Fubini's theorem 3.1.19, and the last inequality holds for the following reason — since η is supported in the interval $[-\frac{1}{2} - \frac{\varepsilon}{2}, \frac{1}{2} + \frac{\varepsilon}{2}]$, it follows that for each $i \in \{1, \ldots, n\}$ we have

$$
\int_{\mathbb{R}} \eta\left(\frac{y_i - c_{I_i}}{\ell(I_i)}\right) \, \mathrm{d}y_i = \ell(I_i) \int_{-\frac{1}{2} - \frac{\varepsilon}{2}}^{\frac{1}{2} + \frac{\varepsilon}{2}} \eta(t) \, \mathrm{d}t \leqslant \ell(I_i)\,(1+\varepsilon)
$$

since $\eta(t) = 1$ for $|t| \leqslant \frac{1}{2}$.

Thus for each $j \in \mathbb{Z}^+$ we have

$$
\begin{aligned}
\frac{1}{\mathrm{vol}(Q_j)} \int_{\mathbb{R}^n} \psi_j(y) \, \mathrm{d}y &= \frac{1}{\mathrm{vol}(Q_j)} \int_{A^c} \frac{\phi_j(y)}{\sum_{k=1}^{\infty} \phi_k(y)} \, \mathrm{d}y \qquad \text{by definition of } \psi_j \\
&\leqslant \frac{1}{\mathrm{vol}(Q_j)} \int_{A^c} \phi_j(y) \, \mathrm{d}y \\
&= \frac{1}{\mathrm{vol}(Q_j)} \int_{A^c} \eta_{Q_j}(y) \, \mathrm{d}y \\
&= \frac{1}{\mathrm{vol}(Q_j)} \int_{\mathbb{R}^n} \eta_{Q_j}(y) \, \mathrm{d}y \qquad \text{since } \mathrm{supp}\left(\eta_{Q_j}\right) \subseteq Q_j^* \subset A^c \\
&\leqslant (1+\varepsilon)^n \qquad\qquad \text{by the upper bound in } (🧹).
\end{aligned}
$$

This establishes the upper bound in (v). To establish the lower bound in (v), we recall ((🙂)), which says that

$$
\mathbf{1}_{A^c} \leqslant \sum_{j=1}^{\infty} \phi_j \leqslant 2^n \mathbf{1}_{A^c}.
$$

Thus for each $j \in \mathbb{Z}^+$ we have

$$
\begin{aligned}
\frac{1}{\mathrm{vol}(Q_j)} \int_{\mathbb{R}^n} \psi_j(y) \, \mathrm{d}y &= \frac{1}{\mathrm{vol}(Q_j)} \int_{A^c} \frac{\phi_j(y)}{\sum_{k=1}^{\infty} \phi_k(y)} \, \mathrm{d}y \\
&\geqslant \frac{1}{2^n} \frac{1}{\mathrm{vol}(Q_j)} \int_{A^c} \phi_j(y) \, \mathrm{d}y \\
&= \frac{1}{2^n} \frac{1}{\mathrm{vol}(Q_j)} \int_{\mathbb{R}^n} \eta_{Q_j}(y) \, \mathrm{d}y \\
&\geqslant \frac{1}{2^n} \cdot 1
\end{aligned}
$$

by the lower bound in ($\overset{\text{🏃}}{}$), which completes the proof of (v).

(vi). Fix a multi-index $\alpha \in \mathbb{N}^n$ and let $j, l \in \mathbb{Z}^+$ be arbitrary. If $\psi_j \psi_l \equiv 0$, then the desired estimate holds trivially, so we may assume that $\psi_j \psi_l$ is not identically zero. Then the cubes Q_j and Q_l are either equal or adjacent, so therefore

$$\frac{1}{4} \leqslant \frac{\operatorname{diam} Q_j}{\operatorname{diam} Q_l} \leqslant 4.$$

Thus we have $\ell(Q_j) \geqslant \frac{1}{4}\ell(Q_l)$ which implies

$$\ell(Q_j)^{-|\beta|} \leqslant 4^{|\beta|}\ell(Q_l)^{-|\beta|} \qquad\qquad \forall\, \beta \leqslant \alpha$$

so we estimate

$$
\begin{aligned}
|\partial^\alpha(\psi_j \psi_l)(x)| &\leqslant \sum_{\beta \leqslant \alpha} \binom{\alpha}{\beta} |\partial^\beta \psi_j(x)||\partial^{\alpha-\beta}\psi_l(x)| && \text{by the Leibniz rule and triangle inequality}\\
&\leqslant \sum_{\beta \leqslant \alpha} \binom{\alpha}{\beta} C_{\alpha-\beta}\ell(Q_j)^{-|\beta|} C_\beta \ell(Q_l)^{-|\alpha|+|\beta|} && \text{by part (iv)}\\
&\leqslant \sum_{\beta \leqslant \alpha} \binom{\alpha}{\beta} C_{\alpha-\beta} C_\beta 4^{|\beta|}\ell(Q_l)^{-|\alpha|} && \text{since } \ell(Q_j)^{-|\beta|} \leqslant 4^{|\beta|}\ell(Q_l)^{-|\beta|}\\
&= C_{n,\alpha}^\ddagger (\operatorname{diam} Q_l)^{-|\alpha|}
\end{aligned}
$$

where we have defined the constant

$$C_{n,\alpha}^\ddagger := n^{|\alpha|/2} \sum_{\beta \leqslant \alpha} \binom{\alpha}{\beta} C_{\alpha-\beta} C_\beta 4^{|\beta|}$$

which depends only on n and α.

That is, we have shown that for each $j, l \in \mathbb{Z}^+$ and $x \in A^c$ we have

$$|\partial^\alpha(\psi_j \psi_l)(x)| \leqslant C_{n,\alpha}^\ddagger (\operatorname{diam} Q_l)^{-|\alpha|}. \tag{♣}$$

(Keep this fact in mind for a moment, as we will use it to finish the proof of (vi).)

Also, we see that $\psi_j \geqslant 2^{-n}\eta_{Q_j}$ and $\psi_l \geqslant 2^{-n}\eta_{Q_l}$ by ($\overset{\text{☺}}{}$), so that

$$
\begin{aligned}
\int_{\mathbb{R}^n} \psi_j(y)\psi_l(y)\,\mathrm{d}y &\geqslant 2^{-2n} \int_{\mathbb{R}^n} \eta_{Q_j}(y)\eta_{Q_l}(y)\,\mathrm{d}y\\
&\geqslant 2^{-2n} B_{n,\eta}\operatorname{vol}(Q_l) && \text{by Lemma \textbf{??}}\\
&= B_n^\star(\operatorname{diam} Q_l)^n
\end{aligned}
$$

where we have defined the constant

$$B_n^\star := 2^{-2n} B_{n,\eta}\, n^{-n/2}$$

which depends only on n and η. That is, we have shown that for each $j, l \in \mathbb{Z}^+$ we have

$$\int_{\mathbb{R}^n} \psi_j(y)\psi_l(y) \, \mathrm{d}y \geqslant B_n^\star(\operatorname{diam} Q_l)^n. \tag{$\spadesuit$}$$

Putting ($\clubsuit$) and ($\spadesuit$) together, we have

$$|\partial^\alpha(\psi_j\psi_l)(x)| \leqslant C_{n,\alpha}^\ddagger(\operatorname{diam} Q_l)^{-|\alpha|} \qquad\qquad \text{by}(\clubsuit)$$

$$= \frac{C_{n,\alpha}^\ddagger}{(\operatorname{diam} Q_l)^{|\alpha|}} \cdot \frac{1}{B_n^\star(\operatorname{diam} Q_l)^n} \cdot B_n^\star(\operatorname{diam} Q_l)^n$$

$$= \frac{C_{n,\alpha}^\ddagger}{B_n^\star} \cdot \frac{1}{(\operatorname{diam} Q_l)^{|\alpha|+n}} \cdot B_n^\star(\operatorname{diam} Q_l)^n$$

$$\leqslant \frac{C_{n,\alpha}^\ddagger}{B_n^\star} \cdot \frac{1}{(\operatorname{diam} Q_l)^{|\alpha|+n}} \cdot \int_{\mathbb{R}^n} \psi_j(y)\psi_l(y) \, \mathrm{d}y \qquad \text{by}(\spadesuit)$$

which completes the proof of (vi) with the constant $B_\alpha := \frac{C_{n,\alpha}^\ddagger}{B_n^\star}$. $\qquad\qquad\square$

7.3   Whitney's Extension Theorem

In this section we present the full version of Whitney's Extension Theorem, which we state somewhat imprecisely as follows.

Let $A \subset \mathbb{R}^n$ be a proper nonempty closed subset, and let $f : A \to \mathbb{R}$ be a continuous function. Then Whitney's Extension Theorem states that f is the restriction of a C^m function $F : \mathbb{R}^n \to \mathbb{R}$ if there is a collection of continuous functions $\{f^{(\alpha)} : \alpha \in \mathbb{N}^n, |\alpha| \leqslant m\}$ defined on A which satisfy certain compatibility conditions which arise naturally from Taylor's theorem.

7.3.1   The Language of Jets

We have to introduce some technical language in order to get some more properties out of the extension operator.

Definition 7.3.1 (*m*-jets)**.** Let $A \subseteq \mathbb{R}^n$ be a nonempty closed subset, and let $m \in \mathbb{Z}^+$. An *m-jet* on A is a collection of continuous functions $\{f^{(\alpha)} : \alpha \in \mathbb{N}^n \text{ and } |\alpha| \leqslant m\} \subset C^0(A)$ which are indexed by multi-indices of order at most m.

We will commonly denote an *m*-jet on A by $f^\bullet = \left(f^{(\alpha)}\right)_{|\alpha| \leqslant m}$.

The following is the key example of an *m*-jet.

Example 7.3.2 (Jets from C^m functions)**.** Let $U \subseteq \mathbb{R}^n$ be an open set. Then for each $f \in C^m(U)$ and each compact subset $K \subseteq U$, the collection of functions $\left(D^\alpha f|_K\right)_{|\alpha| \leqslant m}$ is an *m*-jet on K.

Remark 7.3.3 (How to think about m-jets in this context). The previous example also shows us how to think about m-jets in this context.

Let $A \subseteq \mathbb{R}^n$ be a nonempty closed subset, and let $\{f^{(\alpha)} : |\alpha| \leqslant m\}$ be an m-jet on A. Following the example above, we can think of $f^{(0)}$ as the "original" function which we are tying to extend, and for each multi-index α with $1 \leqslant |\alpha| \leqslant m$, we can think of $f^{(\alpha)}$ as the candidate for the αth derivative of the extension F on the set A.

Exercise 7.3.1 (Jets are Banach Spaces). Let $A \subseteq \mathbb{R}^n$ be a nonempty closed subset, and let $m \in \mathbb{Z}^+$. Let $J^m(A)$ be the set of all m-jets on A. Show that $J^m(A)$ is a vector space with the obvious addition and scalar multiplication operations defined by

$$\left(f^{(\alpha)} + g^{(\alpha)}\right)_{|\alpha| \leqslant m} := \left(f^{(\alpha)}\right)_{|\alpha| \leqslant m} + \left(g^{(\alpha)}\right)_{|\alpha| \leqslant m}$$

and

$$\left(cf^{(\alpha)}\right)_{|\alpha| \leqslant m} := c\left(f^{(\alpha)}\right)_{|\alpha| \leqslant m}$$

for all $c \in \mathbb{R}$ and all m-jets $\left(f^{(\alpha)}\right)_{|\alpha| \leqslant m}$ and $\left(g^{(\alpha)}\right)_{|\alpha| \leqslant m}$ on A.

Furthermore, for each compact set $K \subset \mathbb{R}^n$, the function

$$\|\cdot\|_{J^m(K)} : J^m(K) \to [0, \infty), \quad \left\|\left(f^{(\alpha)}\right)_{|\alpha| \leqslant m}\right\|_{J^m(K)} := \max_{|\alpha| \leqslant m} \left\|f^{(\alpha)}\right\|_{C^0(K)}$$

is a norm on $J^m(K)$, and with this norm $J^m(K)$ is a Banach space.

Proof. The fact that $J^m(A)$ is a vector space with these operations follows immediately from the definition of m-jets and the fact that $C^0(A)$ is a vector space.

Now we will show that $\|\cdot\|_{J^m(K)}$ is a norm on $J^m(K)$. See that if $\left\|\left(f^{(\alpha)}\right)_{|\alpha| \leqslant m}\right\|_{J^m(K)} = 0$, then $f^{(\alpha)} = 0$ for each $|\alpha| \leqslant m$, so $\left(f^{(\alpha)}\right)_{|\alpha| \leqslant m}$ is the zero jet; hence $\|\cdot\|_{J^m(K)}$ is positive definite.

Now if $\left(f^{(\alpha)}\right)_{|\alpha| \leqslant m}$ and $c \in \mathbb{R}$, then

$$\left\|c\left(f^{(\alpha)}\right)_{|\alpha| \leqslant m}\right\|_{J^m(K)} = \max_{|\alpha| \leqslant m} \left\|cf^{(\alpha)}\right\|_{C^0(K)} = |c| \max_{|\alpha| \leqslant m} \left\|f^{(\alpha)}\right\|_{C^0(K)} = |c| \left\|\left(f^{(\alpha)}\right)_{|\alpha| \leqslant m}\right\|_{J^m(K)}$$

so $\|\cdot\|_{J^m(K)}$ is absolutely homogeneous.

Finally, if $\left(f^{(\alpha)}\right)_{|\alpha|\leqslant m}$ and $\left(g^{(\alpha)}\right)_{|\alpha|\leqslant m}$ are two m-jets on K, then

$$
\begin{aligned}
\left\|\left(f^{(\alpha)} + g^{(\alpha)}\right)_{|\alpha|\leqslant m}\right\|_{J^m(K)} &= \max_{|\alpha|\leqslant m}\left\|f^{(\alpha)} + g^{(\alpha)}\right\|_{C^0(K)} \\
&\leqslant \max_{|\alpha|\leqslant m}\left(\left\|f^{(\alpha)}\right\|_{C^0(K)} + \left\|g^{(\alpha)}\right\|_{C^0(K)}\right) \\
&\leqslant \max_{|\alpha|\leqslant m}\left\|f^{(\alpha)}\right\|_{C^0(K)} + \max_{|\alpha|\leqslant m}\left\|g^{(\alpha)}\right\|_{C^0(K)} \\
&= \left\|\left(f^{(\alpha)}\right)_{|\alpha|\leqslant m}\right\|_{J^m(K)} + \left\|\left(g^{(\alpha)}\right)_{|\alpha|\leqslant m}\right\|_{J^m(K)}
\end{aligned}
$$

where we have used the triangle inequality for the C^0 norm. Therefore $\|\cdot\|_{J^m(K)}$ satisfies the triangle inequality, and hence is a norm on $J^m(K)$.

Finally, we will show that $J^m(K)$ is complete with respect to this norm. If $\left\{\left(f_j^{(\alpha)}\right)_{|\alpha|\leqslant m}\right\}_{j=1}^{\infty}$ is a Cauchy sequence in $J^m(K)$, then for each multi-index α with $|\alpha| \leqslant m$, the sequence $\left\{f_j^{(\alpha)}\right\}_{j=1}^{\infty}$ is a Cauchy sequence in $C^0(K)$ and hence converges to some function $f^{(\alpha)} \in C^0(K)$. The resulting collection of functions $\left(f^{(\alpha)}\right)_{|\alpha|\leqslant m}$ is an m-jet on K, and we see that $\left\{\left(f_j^{(\alpha)}\right)_{|\alpha|\leqslant m}\right\}_{j=1}^{\infty}$ converges to $\left(f^{(\alpha)}\right)_{|\alpha|\leqslant m}$ in $J^m(K)$. This shows that $J^m(K)$ is complete with respect to the norm $\|\cdot\|_{J^m(K)}$, and hence is a Banach space. $\qquad\square$

Definition 7.3.4 (Formal Taylor Polynomial, Formal Derivative). Let $A \subseteq \mathbb{R}^n$ be a nonempty closed subset, and let $f^{\bullet} = \left(f^{(\alpha)}\right)_{|\alpha|\leqslant m}$ be an m-jet on A.

For each $a \in A$, we define the *formal Taylor polynomial of $f^{\bullet}$ at a* to be the function $T_a^m f^{\bullet} : \mathbb{R}^n \to \mathbb{R}$ defined by

$$
T_a^m f^{\bullet}(x) := \sum_{|\alpha|\leqslant m} \frac{f^{(\alpha)}(a)}{\alpha!}(x-a)^{\alpha} \qquad \forall\, x \in \mathbb{R}^n.
$$

For each multi-index β with $|\beta| \leqslant m$, we define the *formal β^{th} derivative of $f^{\bullet}$* to be the $m - |\beta|$-jet on A defined by

$$
D^{\beta} f^{\bullet} := \left(f^{(\alpha+\beta)}\right)_{|\alpha|\leqslant m-|\beta|}.
$$

Remark 7.3.5 (Justification for the name "formal Taylor polynomial" and "formal derivative"). Let $U \subseteq \mathbb{R}^n$ be an open set, and let $K \subset U$ be a compact set. Let $f \in C^m(U)$, and let $f^{\bullet} = \left(D^{\alpha} f|_K\right)_{|\alpha|\leqslant m}$ be the m-jet on K obtained from f as in Example 7.3.2.

Then for each $a \in K$, the formal Taylor polynomial of $f^{\bullet}$ at a is exactly the mth order Taylor polynomial of f at a, i.e.

$$
T_a^m f^{\bullet} = T_a^m f
$$

because for each multi-index α with $|\alpha| \leqslant m$, we have $f^{(\alpha)}(a) = D^\alpha f(a)$.

Similarly, for each multi-index β with $|\beta| \leqslant m$, the formal β^{th} derivative of $f^\bullet$ is exactly the $m - |\beta|$-jet obtained from the βth derivative of f, i.e.

$$D^\beta f^\bullet = \left(D^{\alpha+\beta} f|_K \right)_{|\alpha| \leqslant m - |\beta|} = \left(D^\alpha \left(D^\beta f \right) \right)_{|\alpha| \leqslant m - |\beta|}$$

by using the fact that partial derivatives commute since f is C^m.

Definition 7.3.6. Let $U \subseteq \mathbb{R}^n$ be an open set, and let $A \subseteq U$ be a closed set. We define the *jet map* $J_A^m : C^m(U) \to J^m(A)$ by

$$J_A^m(f) := \left(D^\alpha f|_A \right)_{|\alpha| \leqslant m} \qquad \forall f \in C^m(U).$$

Definition 7.3.7. Let $A \subseteq \mathbb{R}^n$ be a nonempty closed subset, and let $f^\bullet = \left(f^{(\alpha)} \right)_{|\alpha| \leqslant m}$ be an m-jet on A. For each $a \in A$, we define the *remainder of $f^\bullet$ at a* to be the jet defined by

$$R_a^m f^\bullet = f^\bullet - J_A^m \left(T_a^m f^\bullet \right).$$

Notice that while $f^\bullet$ is an m-jet on A, the Taylor polynomial $T_a^m f^\bullet$ is a polynomial function defined on $\mathbb{R}^n$, and so we must apply the jet map J^m to $T_a^m f^\bullet$ in order to get an m-jet on A which we can subtract from $f^\bullet$. This shows the remainder is well-defined as an m-jet on A.

The next three exrecises may seem like you are just pushing symbols around, but all of these properties will be used in the proof of the Whitney's Extension Theorem.

Exercise 7.3.2 (Remainder Properties)**.** Let $A \subseteq \mathbb{R}^n$ be a nonempty closed subset, and let $f^\bullet = \left(f^{(\alpha)} \right)_{|\alpha| \leqslant m}$ be an m-jet on A. Then for each $a \in A$ and for each each multi-index α with $|\alpha| \leqslant m$, we have

$$(R_a^m f)^{(\alpha)}(x) = f^{(\alpha)}(x) - \sum_{|\beta| \leqslant m - |\alpha|} \frac{f^{(\alpha+\beta)}(a)}{\beta!} (x - a)^\beta \qquad \forall x \in \mathbb{R}^n.$$

Proof. Fix $a \in A$ and a multi-index α with $|\alpha| \leqslant m$. Let $x \in \mathbb{R}^n$ be arbitrary. By the definition of the remainder jet, we have

$$(R_a^m f^\bullet)^{(\alpha)}(x) = f^{(\alpha)}(x) - (J_A^m(T_a^m f^\bullet))^{(\alpha)}(x).$$

Now by the definition of the jet map and the formal Taylor polynomial, we have

$$\left(J_A^m(T_a^m f^\bullet))\right)^{(\alpha)}(x) = D^\alpha(T_a^m f^\bullet)(x)$$

$$= D^\alpha \left(\sum_{|\beta| \leqslant m} \frac{f^{(\beta)}(a)}{\beta!}(\cdot - a)^\beta \right)(x)$$

$$= \sum_{|\beta| \leqslant m} \frac{f^{(\beta)}(a)}{\beta!} D^\alpha \left((\cdot - a)^\beta \right)(x)$$

$$= \sum_{|\beta| \leqslant m} \frac{f^{(\beta)}(a)}{\beta!} \cdot \left(\begin{cases} 0 & \text{if } |\beta| < |\alpha|, \\ \frac{\beta!}{(\beta-\alpha)!}(x-a)^{\beta-\alpha} & \text{if } |\beta| \geqslant |\alpha| \end{cases} \right)$$

$$= \sum_{|\alpha| \leqslant |\beta| \leqslant m} \frac{f^{(\beta)}(a)}{\beta!} \cdot \frac{\beta!}{(\beta-\alpha)!}(x-a)^{\beta-\alpha}$$

$$= \sum_{|\alpha| \leqslant |\beta| \leqslant m} \frac{f^{(\beta)}(a)}{(\beta-\alpha)!}(x-a)^{\beta-\alpha}$$

$$= \sum_{|\gamma| \leqslant m-|\alpha|} \frac{f^{(\alpha+\gamma)}(a)}{\gamma!}(x-a)^\gamma$$

which we summarize and re-index as

$$(J_A^m(T_a^m f^\bullet))^{(\alpha)}(x) = \sum_{|\beta| \leqslant m-|\alpha|} \frac{f^{(\alpha+\beta)}(a)}{\beta!}(x-a)^\beta.$$

Putting these this last equation together with the first gives

$$(R_a^m f^\bullet)^{(\alpha)}(x) = f^{(\alpha)}(x) - \sum_{|\beta| \leqslant m-|\alpha|} \frac{f^{(\alpha+\beta)}(a)}{\beta!}(x-a)^\beta.$$

$\square$

Exercise 7.3.3 (Formal Derivative of Formal Taylor Polynomial)**.** Let $A \subseteq \mathbb{R}^n$ be a nonempty closed subset, and let $f^\bullet = \left(f^{(\alpha)} \right)_{|\alpha| \leqslant m}$ be an m-jet on A. Then for each $a \in A$ and for each each multi-index α with $|\alpha| \leqslant m$, we have

$$D^\beta T_a^m f^\bullet = T_a^{m-|\beta|}(D^\beta f^\bullet)$$

Proof. Let α be a multi-index with $|\alpha| \leqslant m$, and let $a \in A$ be arbitrary. The second computation in the previous exercise shows

$$D^\alpha(T_a^m f^\bullet)(x) = \sum_{|\beta| \leqslant m-|\alpha|} \frac{f^{(\alpha+\beta)}}{\beta!}(x-a)^\beta \qquad \forall\, x \in \mathbb{R}^n$$

and the right hand side is by definition the formal $m - |\alpha|$ order Taylor polynomial at $a \in A$ of the formal derivative $D^\alpha f^\bullet$.

$\square$

Exercise 7.3.4 (More Formal Properties)**.** Let $A \subseteq \mathbb{R}^n$ be a nonempty closed subset, and let $f^{\bullet} = \left(f^{(\alpha)} \right)_{|\alpha| \leqslant m}$ be an m-jet on A. Then for each $a \in A$ and for each each multi-index α with $|\alpha| \leqslant m$, we have

1. $f^{(\alpha)}(a) = D^{\alpha}(T_a^m f^{\bullet})(a)$, and

2. if $b \in A$, then

$$D^{\alpha}(T_a^m f^{\bullet} - T_b^m f^{\bullet})(a) = (R_b^m f^{\bullet})^{(\alpha)}(a)$$

3. and if $b \in A$, then

$$f^{(\alpha)}(b)(x - b)^{\alpha} = D^{\alpha}(T_b^m f^{\bullet})(a) \cdot (x - a)^{\alpha} \qquad \forall\, x \in \mathbb{R}^n.$$

Proof. Fix $a \in A$ and a multi-index α with $|\alpha| \leqslant m$. By definition of the formal Taylor polynomial, we have

$$T_a^m f^{\bullet}(x) = \sum_{|\beta| \leqslant m} \frac{f^{(\beta)}(a)}{\beta!}(x - a)^{\beta} \qquad \forall\, x \in \mathbb{R}^n$$

so taking derivatives gives

$$D^{\alpha}(T_a^m f^{\bullet})(x) = \sum_{|\beta| \leqslant m} \frac{f^{(\beta)}(a)}{\beta!} \cdot \left(\begin{cases} 0 & \text{if } |\beta| < |\alpha|, \\ \frac{\beta!}{(\beta - \alpha)!}(x - a)^{\beta - \alpha} & \text{if } |\beta| \geqslant |\alpha| \end{cases} \right)$$

$$= \sum_{|\alpha| \leqslant |\beta| \leqslant m} \frac{f^{(\beta)}(a)}{(\beta - \alpha)!}(x - a)^{\beta - \alpha}$$

similarly to in the proof of 7.3.2.

Evaluating this at $x = a$ gives

$$D^{\alpha}(T_a^m f^{\bullet})(a) = \frac{f^{(\alpha)}(a)}{0!} = f^{(\alpha)}(a),$$

because all other terms vanish; this proves the first assertion.

Now let $b \in A$ be arbitrary. By exercise 7.3.2 and 7.3.3, we have

$$(R_b^m f^{\bullet})^{(\alpha)}(a) = f^{(\alpha)}(a) - \sum_{|\beta| \leqslant m - |\alpha|} \frac{f^{(\alpha + \beta)}(b)}{\beta!}(a - b)^{\beta}$$

$$= f^{(\alpha)}(a) - T_b^{m - |\alpha|}(D^{\alpha} f^{\bullet})(a)$$

$$= f^{(\alpha)}(a) - D^{\alpha}(T_b^m f^{\bullet})(a).$$

From the first assertion, we have $f^{(\alpha)}(a) = D^{\alpha}(T_a^m f^{\bullet})(a)$, so we have

$$(R_b^m f^{\bullet})^{(\alpha)}(a) = D^{\alpha}(T_a^m f^{\bullet})(a) - D^{\alpha}(T_b^m f^{\bullet})(a) = D^{\alpha}(T_a^m f^{\bullet} - T_b^m f^{\bullet})(a)$$

and the second assertion is proved.

To prove the third assertion, notice that both

$$f^{(\alpha)}(b)(x-b)^\alpha \quad \text{and} \quad D^\alpha(T_b^m f^\bullet)(a)(x-a)^\alpha$$

are homogeneous polynomials of degree $|\alpha|$ in the variable x, so it suffices to check that the coefficients of these polynomials are equal. The Taylor polynomial at b is

$$T_b^m f^\bullet(x) = \sum_{|\beta| \leqslant m} \frac{f^{(\beta)}(b)}{\beta!}(x-b)^\beta \qquad \forall\, x \in \mathbb{R}^n$$

and the degree-$|\alpha|$ homogeneous part of this polynomial is precisely $\frac{f^{(\alpha)}(b)}{\alpha!}(x-b)^\alpha$.

But now expanding the same polynomial around a gives

$$T_b^m f^\bullet(x) = \sum_{|\beta| \leqslant m} \frac{D^\beta(T_b^m f^\bullet)(a)}{\beta!}(x-a)^\beta \qquad \forall\, x \in \mathbb{R}^n$$

and the degree-$|\alpha|$ homogeneous part of this polynomial is precisely $\frac{D^\alpha(T_b^m f^\bullet)(a)}{\alpha!}(x-a)^\alpha$. Equating the homogeneous parts (which is valid since the degree-$|\alpha|$ homogeneous part of a polynomial is unique) gives

$$\frac{f^{(\alpha)}(b)}{\alpha!}(x-b)^\alpha = \frac{D^\alpha(T_b^m f^\bullet)(a)}{\alpha!}(x-a)^\alpha \qquad \forall\, x \in \mathbb{R}^n$$

and the third assertion is proved. $\qquad\square$

Proposition 7.3.8 (Remainder jet of a C^m function). *Let $U \subseteq \mathbb{R}^n$ be an open set, and let $K \subseteq U$ be a compact set. We consider $f \in C^m(U)$, and let $f^\bullet = J_K^m(f) = \big(D^\alpha f|_K\big)_{|\alpha| \leqslant m}$ be the m-jet on K obtained from f as in Example 7.3.2.*

Then for each multi-index α with $|\alpha| \leqslant m$, we have

$$(R_a^m f^\bullet)^\alpha(x) = o(\|x-a\|^{m-|\alpha|}) \quad \text{as} \ \ x \to a, \ \ \text{uniformly for} \ \ x, a \in K.$$

Proof. Because of exercise 7.3.2, we have

$$(R_a^m f^\bullet)^{(\alpha)}(x) = D^\alpha f(x) - \sum_{|\beta| \leqslant m-|\alpha|} \frac{D^{\alpha+\beta}f(a)}{\beta!}(x-a)^\beta$$

for each multi-index α with $|\alpha| \leqslant m$ and all $x, a \in K$.

Thus the conclusion is precisely the content of Lemma 7.1.5, just restated in the language of jets. $\qquad\square$

Definition 7.3.9 (Whitney Jets)**.** Let $A \subseteq \mathbb{R}^n$ be a nonempty closed subset, and let $m \in \mathbb{Z}^+$. An m-jet $f^\bullet = \left(f^{(\alpha)} \right)_{|\alpha| \leqslant m}$ on A is called a *Whitney m-jet* if for each compact set $K \subseteq A$ and each multi-index α with $|\alpha| \leqslant m$, we have

$$(R_a^m f^\bullet)^{(\alpha)}(x) = o(\|x - a\|^{m - |\alpha|}) \quad \text{as} \ \ x \to a, \ \ \text{uniformly for} \ \ x, a \in K. \tag{W}$$

We denote the set of all Whitney m-jets on A by $\mathcal{W}^m(A)$.

While the notation $J^m(A)$ is standard, there is no standard notation for the set of Whitney m-jets on A, so we have chosen to denote it by $\mathcal{W}^m(A)$.

Example 7.3.10 (Every m-jet on a finite set is a Whitney m-jet)**.** Let $A = \{a_1, a_2, \ldots, a_k\} \subseteq \mathbb{R}^n$ be a finite set, and let $m \in \mathbb{Z}^+$. Then every m-jet on A is a Whitney m-jet on A, because each point in A is isolated and hence the remainder condition is trivially satisfied.

Exercise 7.3.5 (Whitney Jets on a compact set are a Banach Space)**.** Let $K \subseteq \mathbb{R}^n$ be a compact set, and let $m \in \mathbb{Z}^+$. Show that $\mathcal{W}^m(K)$ is a vector subspace of $J^m(K)$.

Also we define

$$\| \cdot \|_{\mathcal{W}^m(K)} : \mathcal{W}^m(K) \to [0, \infty), \quad \|f^\bullet\|_{\mathcal{W}^m(K)} := \|f^\bullet\|_{J^m(K)} + \sup \left\{ \frac{|(R_a^m f^\bullet)^{(\alpha)}(x)|}{\|x - a\|^{m - |\alpha|}} : x, a \in K, x \neq a, |\alpha| \leqslant m \right\}.$$

Show that $\| \cdot \|_{\mathcal{W}^m(K)}$ is a norm on $\mathcal{W}^m(K)$, and with this norm $\mathcal{W}^m(K)$ is a Banach space.

Proof that $\| \cdot \|_{\mathcal{W}^m(K)}$ is a norm on $\mathcal{W}^m(K)$. See that by exercise 7.3.2, we have

$$R_a^m(f^\bullet + g^\bullet) = R_a^m f^\bullet + R_a^m g^\bullet$$

for each $a \in K$ and each $f^\bullet, g^\bullet \in \mathcal{W}^m(K)$.

Thus if $f^\bullet, g^\bullet \in \mathcal{W}^m(K)$, then for each multi-index α with $|\alpha| \leqslant m$ we have

$$\left(R_a^m(f^\bullet + g^\bullet) \right)^{(\alpha)}(x) = (R_a^m f^\bullet)^{(\alpha)}(x) + (R_a^m g^\bullet)^{(\alpha)}(x) = o(\|x - a\|) \quad \text{as } x \to a, \ \text{uniformly for } x, a \in K.$$

This shows that $f^\bullet + g^\bullet$ is a Whitney m-jet on K, and hence $\mathcal{W}^m(K)$ is closed under addition.

Similarly $\mathcal{W}^m(K)$ is closed under scalar multiplication, and hence is a vector subspace of $J^m(K)$.

To show that $\| \cdot \|_{\mathcal{W}^m(K)}$ is a norm on $\mathcal{W}^m(K)$, it suffices to show that the quantity on the right-hand side is a semi-norm on $\mathcal{W}^m(K)$, since we already know that $\| \cdot \|_{J^m(K)}$ is a norm on $J^m(K)$ and hence is a norm on $\mathcal{W}^m(K)$.

Clearly

$$\sup \left\{ \frac{|(R_a^m f^\bullet)^{(\alpha)}(x)|}{\|x - a\|^{m - |\alpha|}} : x, a \in K, x \neq a, |\alpha| \leqslant m \right\} \geqslant 0$$

for all $f^\bullet \in \mathcal{W}^m(K)$, and if $f^\bullet = 0 \in \mathcal{W}^m(K)$, then this quantity is zero; hence it is positive semi-definite.

Now if $f^\bullet \in \mathcal{W}^m(K)$ and $c \in \mathbb{R}$, then

$$\sup\left\{\frac{|(R_a^m(cf^\bullet))^{(\alpha)}(x)|}{\|x-a\|^{m-|\alpha|}} : x,a \in K, x \neq a, |\alpha| \leqslant m\right\} = \sup\left\{\frac{|c||(R_a^m f^\bullet)^{(\alpha)}(x)|}{\|x-a\|^{m-|\alpha|}} : x,a \in K, x \neq a, |\alpha| \leqslant m\right\}$$

$$= |c|\sup\left\{\frac{|(R_a^m f^\bullet)^{(\alpha)}(x)|}{\|x-a\|^{m-|\alpha|}} : x,a \in K, x \neq a, |\alpha| \leqslant m\right\}$$

which shows that this quantity is absolutely homogeneous.

Finally, if $f^\bullet, g^\bullet \in \mathcal{W}^m(K)$, then

$$\sup\left\{\frac{|(R_a^m(f^\bullet+g^\bullet))^{(\alpha)}(x)|}{\|x-a\|^{m-|\alpha|}} : x,a \in K, x \neq a, |\alpha| \leqslant m\right\} \leqslant \sup\left\{\frac{|(R_a^m f^\bullet)^{(\alpha)}(x)|}{\|x-a\|^{m-|\alpha|}} : x,a \in K, x \neq a, |\alpha| \leqslant m\right\}$$

$$+ \sup\left\{\frac{|(R_a^m g^\bullet)^{(\alpha)}(x)|}{\|x-a\|^{m-|\alpha|}} : x,a \in K, x \neq a, |\alpha| \leqslant m\right\}$$

which shows that this quantity satisfies the triangle inequality, and hence is a semi-norm on $\mathcal{W}^m(K)$. $\qquad\square$

Proof that $\mathcal{W}^m(K)$ is a Banach space. Now we will show that $\mathcal{W}^m(K)$ is complete with respect to the norm $\|\cdot\|_{\mathcal{W}^m(K)}$. For brevity, we let

$$S(g^\bullet) := \sup\left\{\frac{|(R_a^m g^\bullet)^{(\alpha)}(x)|}{\|x-a\|^{m-|\alpha|}} : x,a \in K, x \neq a, |\alpha| \leqslant m\right\}$$

for each $g^\bullet \in \mathcal{W}^m(K)$.

Let $\{f_j^\bullet\}_{j=1}^\infty \subset \mathcal{W}^m(K)$ be a Cauchy sequence with respect to the norm $\|\cdot\|_{\mathcal{W}^m(K)}$. Then $\{f_j^\bullet\}_{j=1}^\infty$ is also a Cauchy sequence with respect to the norm $\|\cdot\|_{J^m(K)}$, and hence converges to some $f^\bullet \in J^m(K)$ since $J^m(K)$ is a Banach space by Exercise **??**. We claim that

$$\lim_{j\to\infty} \|f_j^\bullet - f^\bullet\|_{\mathcal{W}^m(K)} = 0$$

and that $f^\bullet \in \mathcal{W}^m(K)$.

Since $\{f_j^\bullet\}_{j=1}^\infty$ converges to $f^\bullet$ with respect to the norm $\|\cdot\|_{J^m(K)}$ and $\|\cdot\|_{\mathcal{W}^m(K)} = \|\cdot\|_{J^m(K)} + S(\cdot)$, it suffices to show that

$$\lim_{j\to\infty} S(f_j^\bullet - f^\bullet) = 0. \qquad\qquad (\star)$$

Let $\epsilon > 0$ be arbitrary. Since $\{f_j^\bullet\}_{j=1}^\infty$ is a Cauchy sequence with respect to the norm $\|\cdot\|_{\mathcal{W}^m(K)}$, there exists $N \in \mathbb{Z}^+$ such that for all $j,k \geqslant N$ we have

$$\|f_j^\bullet - f_k^\bullet\|_{\mathcal{W}^m(K)} < \epsilon.$$

Expending this gives

$$\|f_j^\bullet - f_k^\bullet\|_{J^m(K)} + S(f_j^\bullet - f_k^\bullet) < \epsilon \qquad \forall\, j, k \geqslant N.$$

Since both quantities on the left-hand side are non-negative, we have

$$S(f_j^\bullet - f_k^\bullet) < \epsilon \qquad \forall\, j, k \geqslant N.$$

By using the definition of $S(\cdot)$ we see that

$$S(f_j^\bullet - f_k^\bullet) = \sup\left\{ \frac{|(R_a^m(f_j^\bullet - f_k^\bullet))^{(\alpha)}(x)|}{\|x - a\|^{m-|\alpha|}} : x, a \in K, x \neq a, |\alpha| \leqslant m \right\}$$

for all $j \in \mathbb{Z}^+$. Thus for all $j, k \geqslant N$, and for all $x, a \in K$ with $x \neq a$, and each multi-index α with $|\alpha| \leqslant m$, we have

$$\frac{|(R_a^m(f_j^\bullet - f_k^\bullet))^{(\alpha)}(x)|}{\|x - a\|^{m-|\alpha|}} < \epsilon$$

or equivalently

$$\left|(R_a^m(f_j^\bullet - f_k^\bullet))^{(\alpha)}(x)\right| < \epsilon\|x - a\|^{m-|\alpha|}. \tag{$\dagger$}$$

Now let $x, a \in K$ be such that $x \neq a$, and let α be a multi-index with $|\alpha| \leqslant m$. Then for all $j, k \geqslant N$ we have

$$\left|(R_a^m(f_j^\bullet - f_k^\bullet))^{(\alpha)}(x)\right| = \left|(R_a^m f_j^\bullet)^{(\alpha)}(x) - (R_a^m f_k^\bullet)^{(\alpha)}(x)\right|$$

$$= \left| f_j^{(\alpha)}(x) - f_k^{(\alpha)}(x) + \sum_{|\beta| \leqslant m-|\alpha|} \frac{f_j^{(\alpha+\beta)}(a) - f_k^{(\alpha+\beta)}(a)}{\beta!}(x - a)^\beta \right|$$

by exercise 7.3.2 applied to the jet $f_j^\bullet - f_k^\bullet$; for each $j \geqslant N$ we take the limit of the above as $k \to \infty$ to get

$$\lim_{k\to\infty}\left|(R_a^m(f_j^\bullet - f_k^\bullet))^{(\alpha)}(x)\right| = \left| f_j^{(\alpha)}(x) - f^{(\alpha)}(x) + \sum_{|\beta| \leqslant m-|\alpha|} \frac{f_j^{(\alpha+\beta)}(a) - f^{(\alpha+\beta)}(a)}{\beta!}(x - a)^\beta \right|$$

$$= \left|(R_a^m(f_j^\bullet - f^\bullet))^{(\alpha)}(x)\right|$$

by using exercise 7.3.2 applied to the jet $f_j^\bullet - f^\bullet$. Thus for each $j \geqslant N$ we have

$$\left|(R_a^m(f_j^\bullet - f^\bullet))^{(\alpha)}(x)\right| = \lim_{k\to\infty}\left|(R_a^m(f_j^\bullet - f_k^\bullet))^{(\alpha)}(x)\right| \leqslant \epsilon\|x - a\|^{m-|\alpha|}$$

by using the inequality ($\dagger$). Rearranging this gives

$$\frac{\left|(R_a^m(f_j^\bullet - f^\bullet))^{(\alpha)}(x)\right|}{\|x - a\|^{m-|\alpha|}} \leqslant \epsilon \qquad \forall\, j \geqslant N$$

Since $x, a \in K$ such that $x \neq a$ and α with $|\alpha| \leqslant m$ were arbitrary, we have

$$\sup \left\{ \frac{\left|(R_a^m(f_j^\bullet - f^\bullet))^{(\alpha)}(x)\right|}{\|x - a\|^{m-|\alpha|}} : x, a \in K, x \neq a, |\alpha| \leqslant m \right\} \leqslant \epsilon \qquad \forall j \geqslant N.$$

That is, for all $j \geqslant N$ we have

$$S(f_j^\bullet - f^\bullet) \leqslant \epsilon.$$

Since $\epsilon > 0$ was arbitrary, we have shown that $\lim_{j \to \infty} S(f_j^\bullet - f^\bullet) = 0$, which proves the claim $(\star)$.

As a result, we have shown that $\{f_j^\bullet\}_{j=1}^\infty$ converges to $f^\bullet$ with respect to the norm $\|\cdot\|_{\mathcal{W}^m(K)}$. The reverse triangle inequality gives

$$S(f^\bullet) \leqslant S(f^\bullet - f_j^\bullet) + S(f_j^\bullet) \qquad \forall j \in \mathbb{Z}^+$$

and hence $S(f^\bullet) < \infty$, since both terms on the right are uniformly bounded for all $j \geqslant N$. Thus $f^\bullet \in \mathcal{W}^m(K)$, and we have shown that $\{f_j^\bullet\}_{j=1}^\infty$ converges to $f^\bullet$ in $\mathcal{W}^m(K)$.

Since $\{f_j^\bullet\}_{j=1}^\infty$ was an arbitrary Cauchy sequence in $\mathcal{W}^m(K)$, we have shown that $\mathcal{W}^m(K)$ is complete with respect to the norm $\|\cdot\|_{\mathcal{W}^m(K)}$, and hence is a Banach space. $\qquad \square$

Exercise 7.3.6 (Whitney Jets on a non-compact set are a Fréchet Space)**.** Let $A \subseteq \mathbb{R}^n$ be a nonempty closed subset, and let $m \in \mathbb{Z}^+$. Then for each compact set $K \subseteq A$, we consider the semi-norm $\|\cdot\|_{\mathcal{W}^m(K)}$ on $\mathcal{W}^m(A)$.

Let $\{K_j\}_{j=1}^\infty$ be an increasing sequence of compact sets such that $\bigcup_{j=1}^\infty K_j = A$. Show that the family of semi-norms $\{\|\cdot\|_{\mathcal{W}^m(K_j)}\}_{j=1}^\infty$ defines a Fréchet space structure on $\mathcal{W}^m(A)$, and that this structure does not depend on the choice of the sequence $\{K_j\}_{j=1}^\infty$.

Proof. See that for each compact set $K \subseteq A$, the function $\|\cdot\|_{\mathcal{W}^m(K)}$ is a semi-norm on $\mathcal{W}^m(A)$ because it is a norm on $\mathcal{W}^m(K)$.

The fact that $\{\|\cdot\|_{\mathcal{W}^m(K_j)}\}_{j=1}^\infty$ defines a Fréchet space structure on $\mathcal{W}^m(A)$ is shown in exactly the same way that we proved $C^0(A)$ is a Fréchet space with respect to the family of semi-norms $\{\|\cdot\|_{C^0(K_j)}\}_{j=1}^\infty$.

The fact that this structure does not depend on the choice of the sequence $\{K_j\}_{j=1}^\infty$ is shown in exactly the same way that we showed that the Fréchet space structure on $C^0(A)$ does not depend on the choice of the sequence $\{K_j\}_{j=1}^\infty$. $\qquad \square$

7.3.2 Whitney's C^m Extension Theorem

Theorem 7.3.11 (Whitney's C^m Extension Theorem, 1934)**.** *Let $A \subseteq \mathbb{R}^n$ be a nonempty closed subset, and let $m \in \mathbb{Z}^+$. Then there exists a continuous linear map*

$$W : \mathcal{W}^m(A) \to C^m(\mathbb{R}^n)$$

such that for each $f^\bullet \in \mathcal{W}^m(A)$, and each multi-index α with $|\alpha| \leqslant m$, we have

$$D^\alpha(W(f^\bullet))(x) = f^{(\alpha)}(x) \qquad \forall\, x \in A.$$

Moreover, for each $f^\bullet \in \mathcal{W}^m(A)$, the function $W(f^\bullet)$ is C^∞ on $\mathbb{R}^n \backslash A$.

Remark 7.3.12 (Continuity of the Extension Operator). The continuity of the extension operator W means that if $K \subseteq A$ is a compact set, and $Q \subset \mathbb{R}^n$ is an open cube containing K and we set

$$\lambda := \sup_{x \in Q} \operatorname{dist}(x, K)$$

then there is a constant $C_{m,n,\lambda} > 0$ such that for each $f^\bullet \in \mathcal{W}^m(A)$, we have

$$\|W(f^\bullet)\|_{C^m(Q)} \leqslant C_{m,n,\lambda}\|f^\bullet\|_{\mathcal{W}^m(K)}.$$

Example 7.3.13 (The Condition (W) cannot be weakened). The property (W) in the definition of Whitney m-jets cannot be weakened to the condition

$$\lim_{x \to a} \frac{(R_a^m f^\bullet)^{(\alpha)}(x)}{\|x - a\|^{m-|\alpha|}} = 0 \qquad\qquad (*)$$

for each $a \in A$ and each multi-index α with $|\alpha| \leqslant m$. (This is condition (W) without the requirement that this limit must be uniform on compact subsets of A.)

We will show that there exist a m-jet satisfying this weaker condition but which do not admit a C^m extension to $\mathbb{R}^n$.

In this example, we take $n = m = 1$. Choose sequences of positive real numbers $\{x_j\}_{j=1}^\infty$ and $\{y_j\}_{j=1}^\infty$ such that $x_j \to 0$ and $y_j \to 0$ as $j \to \infty$, and such that the line segment joining (x_j, y_j) and (x_{j+1}, y_{j+1}) has slope $(-1)^j$ for each $j \in \mathbb{Z}^+$.

Let $A := \{0\} \cup \{x_j : j \in \mathbb{Z}^+\}$. Define a jet $f^\bullet \in J^1(A)$ by defining $f^{(0)}(0) := 0$ and $f^{(0)}(x_j) = y_j$ for each $j \in \mathbb{Z}^+$, and also $f^{(1)} \equiv 0$. Since each point in A is isolated, the limit condition $(*)$ is trivially satisfied if $x = x_k$ for some $k \in \mathbb{Z}^+$. Also see that

$$(R_0^1 f^\bullet)^0(x_j) = y_j \quad \text{and} \quad (R_0^1 f^\bullet)^1(x_j) = 0 \qquad \forall\, j \in \mathbb{Z}^+,$$

so the limit condition $(*)$ is satisfied at 0 as well, since $y_j \to 0$ as $j \to \infty$.

However, we claim there is no C^1 extension of $f^{(0)}$ to $\mathbb{R}$. See that

$$\frac{y_{k+1} - y_k}{x_{k+1} - x_k} = (-1)^k \qquad \forall\, k \in \mathbb{Z}^+$$

and this does not approach a limit as $k \to \infty$. Thus there can be no extension f with f' continuous at 0, and hence there cannot be a C^1 extension of $f^{(0)}$ to $\mathbb{R}$.
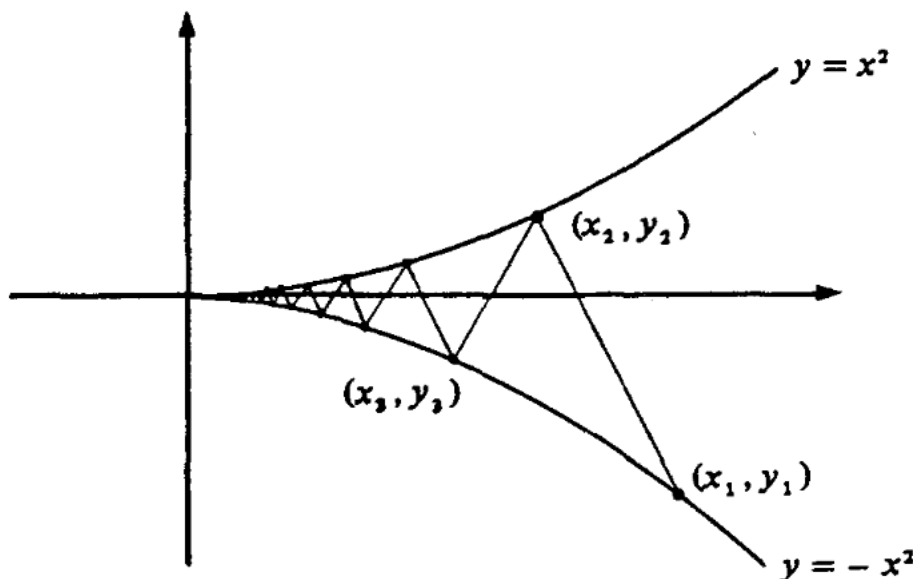
Theorem 7.3.14 (Simplified Version of Whitney's C^m Extension Theorem). *Let $A \subseteq \mathbb{R}^n$ be a nonempty proper closed subset, and let $m \in \mathbb{Z}^+$. Let $f^{(0)} : A \to \mathbb{R}$ be a continuous function, and assume that there exist continuous functions $(f^{(\alpha)})_{1 \leqslant |\alpha| \leqslant m}$ defined on A such that the collection $(f^{(\alpha)})_{|\alpha| \leqslant m}$ is a Whitney m-jet on A.*

Then there exists a function $f \in C^m(\mathbb{R}^n)$ such that for each multi-index α with $|\alpha| \leqslant m$, we have

$$D^\alpha f(x) = f^{(\alpha)}(x) \qquad \forall\, x \in A.$$

Moreover, the function f can be chosen so that it is C^∞ on $\mathbb{R}^n \backslash A$. We say that f is a C^m extension of $f^{(0)}$ to $\mathbb{R}^n$.

Of course, the converse of this theorem is also true — see that if $f \in C^m(\mathbb{R}^n)$ is an extension of $f^{(0)}$ to $\mathbb{R}^n$, i.e. $f|_A = f^{(0)}$, then the collection $\{D^\alpha f|_A : |\alpha| \leqslant m\}$ is a Whitney m-jet on A by Proposition 7.3.8.

7.4 Proof of Whitney's C^m Extension Theorem

In this section we present the proof of Whitney's C^m Extension Theorem.

Lemma 7.4.1 (Modulus of Continuity for Whitney Jets). *A modulus of continuity is a function $\omega : [0, \infty) \to [0, \infty)$ which is increasing, concave, and satisfies $\omega(0) = 0$.*

Let $A \subseteq \mathbb{R}^n$ be a nonempty compact set, and let $f^\bullet \in \mathcal{W}^m(A)$ be a Whitney m-jet on A. Then there exists a modulus of continuity ω such that for each multi-index α with $|\alpha| \leqslant m$, we have

$$\left\| (R_a^m f^\bullet)^{(\alpha)}(x) \right\| \leqslant \omega\left(\|x - a\| \right) \cdot \|x - a\|^{m - |\alpha|} \quad \forall\, x, a \in A$$

and also satisfying

$$\omega(t) = \omega(\operatorname{diam} A) \qquad \forall\, t \geqslant \operatorname{diam} A,$$

and

$$\|f^\bullet\|_{\mathcal{W}^m(A)} = \|f^\bullet\|_{J^m(A)} + \omega(\operatorname{diam} A).$$

Proof. Define $\rho : [0, \infty) \to [0, \infty)$ by

$$\rho(t) := \sup \left\{ \frac{\|(R_a^m f^\bullet)^{(\alpha)}(x)\|}{\|x - a\|^{m-|\alpha|}} : a, x \in A, 0 < \|x - a\| \leqslant t, |\alpha| \leqslant m \right\}$$

if $t > 0$, and $\rho(0) := 0$. Note that ρ is finite-valued since $f^\bullet \in \mathcal{W}^m(A)$. Then ρ is clearly increasing and

$$\lim_{t \to 0^+} \rho(t) = 0 = \rho(0)$$

so ρ is continuous at 0. Also if $t \geqslant \operatorname{diam} A$, then $\rho(t) = \rho(\operatorname{diam} A)$, so ρ is constant on $[\operatorname{diam} A, \infty)$. (We note that ρ is not necessarily concave or continuous on $(0, \infty)$.)

Now define

$$\operatorname{conv}(\rho) := \operatorname{conv}\left([0, \infty) \cup \operatorname{graph}(\rho)\right)$$

to be the closed convex hull of the graph of ρ together with the nonnegative t-axis. Since $\operatorname{conv}(\rho)$ is convex and ρ is non-negative, see that for each $t \geqslant 0$, the set $\{s : (t, s) \in \operatorname{conv}(\rho)\}$ is a closed (possibly degenerate) interval of the form $[0, \omega(t)]$ for some $\omega(t) \geqslant \rho(t) \geqslant 0$.

This defines a function ω on $[0, \infty)$ by $\omega(t) := \sup\{s : (t, s) \in \operatorname{conv}(\rho)\}$.

Let us check the required properties of ω. By construction, $\operatorname{conv}(\rho)$ is a closed subset of the upper half-plane, so $\omega(t) \geqslant 0$ for each $t \geqslant 0$. Thus $\omega : [0, \infty) \to [0, \infty)$ is well-defined.

Now the only point in $[0, \infty) \cup \operatorname{graph}(\rho)$ with $t = 0$ is $(0, 0)$, so the only point in $\operatorname{conv}(\rho)$ with $t = 0$ is $(0, 0)$, which implies $\omega(0) = 0$.

We also see that ω is concave. Let $t_1, t_2 \geqslant 0$ and $\lambda \in [0, 1]$ be arbitrary. Then the points $(t_1, \omega(t_1))$ and $(t_2, \omega(t_2))$ are in the convex set $\operatorname{conv}(\rho)$, so we have

$$(\lambda t_1 + (1 - \lambda)t_2, \lambda\omega(t_1) + (1 - \lambda)\omega(t_2)) \in \operatorname{conv}(\rho).$$

But then the definition of $\omega(\lambda t_1 + (1 - \lambda)t_2)$ implies that

$$\lambda\omega(t_1) + (1 - \lambda)\omega(t_2) \leqslant \omega(\lambda t_1 + (1 - \lambda)t_2).$$

Since $t_1, t_2 \in [0, \infty)$ and $\lambda \in [0, 1]$ were arbitrary, we conclude that ω is concave.

Now $\omega : [0, \infty) \to [0, \infty)$ is a concave function with $\omega(0) = 0$, so we know from Analysis 1 that ω is increasing and continuous on $[0, \infty)$.

Suppose towards a contradiction that there exists $t_0 > \operatorname{diam} A$ such that $\omega(t_0) > \rho(\operatorname{diam} A)$. Then we define the function

$$\hat{\omega} : [0, \infty) \to [0, \infty), \quad \hat{\omega} := \begin{cases} \omega(t) & \text{if } 0 \leqslant t \leqslant \operatorname{diam} A, \\ \omega(\operatorname{diam} A) & \text{if } t \geqslant \operatorname{diam} A. \end{cases}$$

Note that $\hat{\omega}$ is a continuous function since $\hat{\omega}|_{[0,\operatorname{diam} A]} = \omega|_{[0,\operatorname{diam} A]}$ is continuous and $\hat{\omega}(\operatorname{diam} A) = \omega(\operatorname{diam} A)$. Also note that $\hat{\omega}$ is concave because it is the pointwise minimum of the concave function ω and the constant (concave) function $\omega(\operatorname{diam} A)$.

That is, $\hat{\omega}$ is a continuous increasing concave function with $\hat{\omega}(0) = 0$ and $\hat{\omega}(t) = \rho(\operatorname{diam} A)$ for each $t \geqslant \operatorname{diam} A$. (In reality, $\hat{\omega}$ is literally equal to ω, but it is easier to define a new function $\hat{\omega}$ than to prove directly that ω has this last property.)

The final thing to prove is the remainder estimate and the norm identity.

Let α be a multi-index with $|\alpha| \leqslant m$. Then by definition of ρ we see that

$$\|(R_a^m f^\bullet)^{(\alpha)}(x)\| \leqslant \rho(\|x - a\|) \cdot \|x - a\|^{m-|\alpha|} \leqslant \hat{\omega}(\|x - a\|) \cdot \|x - a\|^{m-|\alpha|} \qquad \forall\, x, a \in A$$

as desired. Also

$$\|f^\bullet\|_{\mathcal{W}^m(A)} = \|f^\bullet\|_{J^m(A)} + \sup\left\{ \frac{\|(R_a^m f^\bullet)^{(\alpha)}(x)\|}{\|x - a\|^{m-|\alpha|}} : x, a \in A, x \neq a, |\alpha| \leqslant m \right\}$$

$$= \|f^\bullet\|_{J^m(A)} + \sup\left\{ \frac{\|(R_a^m f^\bullet)^{(\alpha)}(x)\|}{\|x - a\|^{m-|\alpha|}} : x, a \in A, 0 < \|x - a\| \leqslant \operatorname{diam} A, |\alpha| \leqslant m \right\}$$

$$= \|f^\bullet\|_{J^m(A)} + \sup_{t \in (0,\operatorname{diam} A]} \rho(t)$$

$$= \|f^\bullet\|_{J^m(A)} + \rho(\operatorname{diam} A)$$

$$= \|f^\bullet\|_{J^m(A)} + \hat{\omega}(\operatorname{diam} A)$$

where we have used that A is compact in the second equality, and that ρ is increasing in the third equality, and that $\hat{\omega}(\operatorname{diam} A) = \rho(\operatorname{diam} A)$ in the last equality. This finishes the proof of the claim. $\qquad\square$

We are now ready to prove Whitney's C^m Extension Theorem. We warn you that this is one of the *longest* proofs in the book, and it is quite technical.

Proof of Whitney's C^m Extension Theorem in the case $A \subset \mathbb{R}^n$ is compact. Assume that $A \subset \mathbb{R}^n$ is a nonempty compact set, and let $\{Q_j\}_{j=1}^\infty$ be the Whitney decomposition of A^c given by Lemma **??**. Fix $0 < \varepsilon < \frac{1}{4}$ and for each $j \in \mathbb{Z}^+$, let $Q_j^* := (1 + \varepsilon)Q_j$ be the cube with the same center as Q_j but with side length $(1 + \varepsilon)$ times that of Q_j. Let $\{\psi_j\}_{j=1}^\infty$ be the Whitney partition of unity subordinate to the open cover $\{Q_j^*\}_{j=1}^\infty$ of A^c given by Lemma **??**.

For each $j \in \mathbb{Z}^+$, choose a point $a_j \in A$ such that

$$\mathrm{dist}(Q_j, A) = \mathrm{dist}(Q_j, a_j).$$

For each jet $f^\bullet \in \mathcal{J}^m(A)$, we define the function $\mathcal{E}_m(f^\bullet) : \mathbb{R}^n \to \mathbb{R}$ by

$$\mathcal{E}_m(f^\bullet)(x) = \begin{cases} f^{(0)}(x) & \text{if } x \in A, \\ \displaystyle\sum_{j=1}^{\infty} \psi_j(x) \cdot T_{a_j}^m f^\bullet(x) & \text{if } x \in A^c. \end{cases}$$

This defines a linear map $\mathcal{E}_m$ from $\mathcal{J}^m(A)$ to functions on $\mathbb{R}^n$.

Claim 1: We claim that if $f^\bullet \in J^m(A)$, then $\mathcal{E}_m(f^\bullet)$ is C^∞ on A^c.

Proof of Claim 1. See that for each $j \in \mathbb{Z}^+$, at most 12^n other cubes from the Whitney decomposition can intersect Q_j^* by Lemma **??** and 7.2.4. Hence there are only finitely many terms in the sum defining $\mathcal{E}_m(f^\bullet)$ on Q_j^* which are nonzero, and each of these terms is a smooth function on Q_j^*, so $\mathcal{E}_m(f^\bullet)$ is smooth on Q_j^*.

Since $A^c = \bigcup_{j=1}^\infty Q_j^*$ by Lemma **??**, we conclude that $\mathcal{E}_m(f^\bullet)$ is smooth on A^c. $\square$

Now we assume that $f^\bullet \in \mathcal{W}^m(A)$, and let $f := \mathcal{E}_m(f^\bullet)$. We want to show that $f \in C^m(\mathbb{R}^n)$.

By Lemma 7.4.1, there exists a modulus of continuity ω such that for each multi-index α with $|\alpha| \leqslant m$, we have

$$\left\| (R_a^m f^\bullet)^{(\alpha)}(x) \right\| \leqslant \omega\left(\|x - a\| \right) \cdot \|x - a\|^{m - |\alpha|} \quad \forall\, x, a \in A$$

and also satisfying

$$\omega(t) = \omega(\mathrm{diam}\, A) \qquad \forall\, t \geqslant \mathrm{diam}\, A,$$

and

$$\|f^\bullet\|_{\mathcal{W}^m(A)} = \|f^\bullet\|_{J^m(A)} + \omega(\mathrm{diam}\, A).$$

Now we fix an cube $Q \subset \mathbb{R}^n$ such that $A \subset Q$ for the rest of the proof.

Claim 2: For each multi-index α with $|\alpha| \leqslant m$, we define

$$\underline{\partial^\alpha} f(x) := \begin{cases} D^\alpha f(x) & \text{if } x \in A^c, \\ f^{(\alpha)}(x) & \text{if } x \in A \end{cases}$$

which is the candidate for the α partial derivative of f on $\mathbb{R}^n$.

We claim that there exists a constant $C > 0$ which only depends on m, n, and $\sup_{y \in Q} \mathrm{dist}(y, A)$ such that if α is a multi-index with $|\alpha| \leqslant m$, and if $x \in Q$ and $a \in A$, then we have

$$\left| \underline{\partial^\alpha} f(x) - D^\alpha T_a^m f^\bullet(x) \right| \leqslant C \cdot \omega\left(\|x - a\| \right) \cdot \|x - a\|^{m - |\alpha|}. \tag{$\maltese$}$$

For the moment, we will assume Claim 2 and use it to finish the proof of the theorem. We claim that (🐎) implies that $f \in C^m(\mathbb{R}^n)$ and that there is a constant $c > 0$ depending only on m, n, and $\sup_{y \in Q} \mathrm{dist}(y, A)$ such that

$$\|f\|_{C^m(Q)} \leqslant c \, \|f^\bullet\|_{\mathcal{W}^m(A)}. \tag{🎃}$$

Recall that the estimate (🎃) is the required continuity estimate for the extension operator $\mathcal{E}_m$.

Proof that (🐎) implies $f \in C^m(\mathbb{R}^n)$. Fix a multi-index α with $|\alpha| < m$. For each $j \in \{1, 2, \ldots, n\}$, we let e_j be the j-th standard basis vector in $\mathbb{R}^n$. If $a \in A$ and $x \in Q \backslash A$, then

$$\underline{\partial}^\alpha f(a) = f^{(\alpha)}(a) = D^\alpha T_a^m f^\bullet(a)$$

so that

$$\left| \underline{\partial}^\alpha f(x) - \underline{\partial}^\alpha f(a) - \sum_{j=1}^n \underline{\partial}^{\alpha + e_j} f(a)(x_j - a_j) \right| = \left| \underline{\partial}^\alpha f(x) - D^\alpha T_a^m f^\bullet(a) - \sum_{j=1}^n \underline{\partial}^{\alpha + e_j} f(a)(x_j - a_j) \right|$$

$$\leqslant \left| \underline{\partial}^\alpha f(x) - D^\alpha T_a^m f^\bullet(x) \right| +$$

$$+ \left| D^\alpha T_a^m f^\bullet(x) - D^\alpha T_a^m f^\bullet(a) - \sum_{j=1}^n \underline{\partial}^{\alpha + e_j} f(a)(x_j - a_j) \right|$$

$$\leqslant C \cdot \omega \left(\|x - a\| \right) \cdot \|x - a\|^{m - |\alpha|} +$$

$$+ \left| D^\alpha T_a^m f^\bullet(x) - D^\alpha T_a^m f^\bullet(a) - \sum_{j=1}^n D^{\alpha + e_j} T_a^m f^\bullet(a)(x_j - \right.$$

$$= o(\|x - a\|^{m - |\alpha|}) + o(\|x - a\|)$$

$$= o(\|x - a\|)$$

where we have used the estimate (🐎) in the second inequality, and the fact that $T_a^m f^\bullet$ is a polynomial of degree at most m in the second to last equality, and finally the fact that $|\alpha| < m$ in the last equality.

If $a \in A$ and $x \in A$, then we have

$$\left| \underline{\partial}^\alpha f(x) - \underline{\partial}^\alpha f(a) - \sum_{j=1}^n \underline{\partial}^{\alpha + e_j} f(a)(x_j - a_j) \right| = \left| D^\alpha T_a^m f^\bullet(x) - D^\alpha T_a^m f^\bullet(a) - \sum_{j=1}^n \underline{\partial}^{\alpha + e_j} f(a)(x_j - a_j) \right|$$

$$= \left| D^\alpha T_a^m f^\bullet(x) - D^\alpha T_a^m f^\bullet(a) - \sum_{j=1}^n D^{\alpha + e_j} T_a^m f^\bullet(a)(x_j - a_j) \right|$$

$$= o(\|x - a\|).$$

where we have used the fact that $T_a^m f^\bullet$ is a polynomial of degree at most m.

In either case, we have shown that for each multi-index α with $|\alpha| < m$ and each $a \in A$ we have

$$\left| \underline{\partial^\alpha} f(x) - \underline{\partial^\alpha} f(a) - \sum_{j=1}^n \underline{\partial^{\alpha+e_j}} f(a)(x_j - a_j) \right| = o(\|x - a\|) \quad \text{as } x \to a,$$

and as a result, we conclude that $\underline{\partial^\alpha} f$ is differentiable at a with derivative $\underline{\partial^{\alpha+e_j}} f(a) = f^{(\alpha+e_j)}(a)$ for each $j \in \{1, 2, \ldots, n\}$. Since $a \in A$ was arbitrary, we conclude that $\underline{\partial^\alpha} f$ is differentiable at each point of A with derivative $\underline{\partial^{\alpha+e_j}} f(a) = f^{(\alpha+e_j)}(a)$ for each $j \in \{1, 2, \ldots, n\}$; hence $\underline{\partial^\alpha} f$ is C^1 on $\mathbb{R}^n$.

Now taking $\alpha = 0$ shows that for each $j \in \{1, 2, \ldots, n\}$ the function $\underline{\partial_j} f$ is continuous , and the above argument combined with Claim 1 shows that $\underline{\partial_j} f$ is the jth partial derivative of f on $\mathbb{R}^n$. Therefore $f \in C^1(\mathbb{R}^n)$ and $\underline{\partial_j} f = D_j f$ for each $j \in \{1, 2, \ldots, n\}$.

In a similar way, we then see that for each multi-index α with $|\alpha| = 2$, the function $\underline{\partial^\alpha} f$ is continuous; writing $\alpha = e_i + e_j$ for some $i, j \in \{1, 2, \ldots, n\}$, we see that $\underline{\partial^\alpha} f$ is the ith partial derivative of $\underline{\partial_j} f$, and hence $\underline{\partial^\alpha} f$ is the $\alpha = e_i + e_j$ partial derivative of f on $\mathbb{R}^n$. Since α was an arbitrary multi-index with $|\alpha| = 2$, we conclude that $f \in C^2(\mathbb{R}^n)$ and $\underline{\partial^\alpha} f = D^\alpha f$ for each multi-index α with $|\alpha| = 2$.

Bootstrapping this argument, we conclude that for each multi-index α with $|\alpha| \leq m$ the function $\underline{\partial^\alpha} f$ is continuous and is the α partial derivative of f on $\mathbb{R}^n$; because this is true for each multi-index α with $|\alpha| < m$, we conclude that $f \in C^m(\mathbb{R}^n)$. $\qquad \square$

Proof that (🏃) *and* $f \in C^m(\mathbb{R}^n)$ *implies* (☺). Let α be a multi-index with $|\alpha| \leq m$, and let $x \in Q$ be arbitrary. Since A is compact there is a point $a \in A$ such that $\text{dist}(x, A) = \|x - a\|$. Set

$$\lambda := \sup_{y \in Q} \text{dist}(y, A).$$

Then, as shown above we have $D^\alpha f(x) = \underline{\partial^\alpha} f(x)$, and for each $a \in A$ we have

$$D^\alpha(T_a^m f^\bullet)(x) = \sum_{|\beta| \leq m - |\alpha|} \frac{f^{(\alpha+\beta)}(a)}{\beta!}(x - a)^\beta = \sum_{|\beta| \leq m - |\alpha|} \frac{D^{\alpha+\beta} f(a)}{\beta!}(x - a)^\beta \qquad (\dagger)$$

where the first equality follows from the second computation in the proof of 7.3.2, and the second equality follows from the fact that $f^{(\alpha+\beta)}(a) = D^{\alpha+\beta} f(a)$ for all multi-indices β with $|\beta| \leq m - |\alpha|$, which was shown in the previous step.

Then we estimate

$$|D^\alpha f(x)| \leqslant \left|D^\alpha f(x) - D^\alpha T_a^m f^\bullet(x)\right| + \left|D^\alpha T_a^m f^\bullet(x)\right| \qquad \text{by triangle inequality}$$

$$= \left|\underline{\partial^\alpha f(x)} - D^\alpha T_a^m f^\bullet(x)\right| + \left|\sum_{|\beta| \leqslant m - |\alpha|} \frac{D^{\alpha+\beta} f(a)}{\beta!}(x-a)^\beta\right| \qquad \text{by (†)}$$

$$\leqslant C \cdot \omega\left(\|x-a\|\right) \cdot \|x-a\|^{m-|\alpha|} + \sum_{|\beta| \leqslant m - |\alpha|} \frac{\left|D^{\alpha+\beta} f(a)\right|}{\beta!}\left|(x-a)^\beta\right| \qquad \text{by (🧹) and the trian}$$

$$\leqslant C \cdot \omega\left(\lambda\right) \cdot \lambda^{m-|\alpha|} + \sum_{|\beta| \leqslant m - |\alpha|} \frac{\left|D^{\alpha+\beta} f(a)\right|}{\beta!}\|x-a\|^{|\beta|} \qquad \text{since } \omega \text{ is increasing and }|$$

$$\leqslant C \cdot \omega\left(\operatorname{diam} A\right) \cdot \lambda^{m-|\alpha|} + \sum_{|\beta| \leqslant m - |\alpha|} \frac{\|f^\bullet\|_{J^m(A)}}{\beta!}\lambda^{|\beta|}$$

$$= C\left(\|f^\bullet\|_{\mathcal{W}^m(A)} - \|f^\bullet\|_{J^m(A)}\right) \cdot \lambda^{m-|\alpha|} + \left(\sum_{|\beta| \leqslant m - |\alpha|} \frac{1}{\beta!}\lambda^{|\beta|}\right) \cdot \|f^\bullet\|_{J^m(A)}$$

$$=: C\left(\|f^\bullet\|_{\mathcal{W}^m(A)} - \|f^\bullet\|_{J^m(A)}\right) \cdot \lambda^{m-|\alpha|} + c_{m,\lambda} \cdot \|f^\bullet\|_{J^m(A)}$$

$$= C\lambda^{m-|\alpha|} \cdot \|f^\bullet\|_{\mathcal{W}^m(A)} + \left(c_{m,\lambda} - C\lambda^{m-|\alpha|}\right) \cdot \|f^\bullet\|_{J^m(A)}$$

where C is the constant from (🏇) and $c_{m,\lambda} := \sum_{|\beta| \leqslant m - |\alpha|} \frac{1}{\beta!}\lambda^{|\beta|}$. Now let

$$C' := \max\left\{C, \frac{c_{m,\lambda}}{\lambda^{m-|\alpha|}}\right\}$$

so that

$$C \leqslant C' \quad \text{and} \quad \frac{c_{m,\lambda}}{\lambda^{m-|\alpha|}} \leqslant C'$$

which implies

$$c_{m,\lambda} - C'\lambda^{m-|\alpha|} \leqslant 0.$$

Then using the exact same argument as before but with C' in place of C, we see that

$$|D^\alpha f(x)| \leqslant C'\lambda^{m-|\alpha|} \cdot \|f^\bullet\|_{\mathcal{W}^m(A)} + \left(c_{m,\lambda} - C'\lambda^{m-|\alpha|}\right) \cdot \|f^\bullet\|_{J^m(A)}$$
$$\leqslant C'\lambda^{m-|\alpha|} \cdot \|f^\bullet\|_{\mathcal{W}^m(A)}.$$

Setting

$$c := C'\lambda^{m-|\alpha|} = \max\left\{C\lambda^{m-|\alpha|}, c_{m,\lambda}\right\}$$

we conclude that

$$|D^\alpha f(x)| \leqslant c\|f^\bullet\|_{\mathcal{W}^m(A)}.$$

Since α was an arbitrary multi-index with $|\alpha| \leqslant m$ and $x \in Q$ was arbitrary, we conclude that

$$\max_{|\alpha| \leqslant m} \sup_{x \in Q} |D^\alpha f(x)| \leqslant c\|f^\bullet\|_{\mathcal{W}^m(A)}$$

which says that

$$\|f\|_{C^m(Q)} \leqslant c\,\|f^\bullet\|_{\mathcal{W}^m(A)}$$

as desired in (☺).   □

It remains to prove Claim 2. For this, we make a subclaim which will be used in the proof of Claim 2.

Subclaim 2a. We claim that if $a, b \in A$ and α is a multi-index with $|\alpha| \leqslant m$, then

$$\left|D^\alpha T_a^m f^\bullet(x) - D^\alpha T_b^m f^\bullet(x)\right| \leqslant 4^{m-|\alpha|} e^n \cdot \omega(\|a-b\|) \cdot \left(\|x - a\|^{m-|\alpha|} + \|x - b\|^{m-|\alpha|}\right) \qquad \forall\, x \in \mathbb{R}^n.$$

Proof of Subclaim 2a. Let $a, b \in A$ and α be a multi-index with $|\alpha| \leqslant m$.

We will use three properties about jets and their formal derivatives and Taylor polynomials. Namely, for each multi-index β with $|\beta| \leqslant m$, we have

$$f^{(\beta)}(a) = D^\beta(T_a^m f^\bullet)(a) \tag{i}$$

and

$$D^\beta(T_a^m f^\bullet - T_b^m f^\bullet)(a) = (R_b^m f^\bullet)^{(\beta)}(a) \tag{ii}$$

and

$$f^{(\beta)}(b)(x - b)^\beta = D^\beta(T_b^m f^\bullet)(a) \cdot (x - a)^\beta, \tag{iii}$$

which were proven in Exercise 7.3.4.

Then as a result of these properties, for each $x \in \mathbb{R}^n$ we have

$$
\begin{aligned}
T_a^m f^\bullet(x) - T_b^m f^\bullet(x) &= \sum_{|\beta| \leqslant m} \frac{f^{(\beta)}(a)}{\beta!}(x - a)^\beta - \sum_{|\beta| \leqslant m} \frac{f^{(\beta)}(b)}{\beta!}(x - b)^\beta && \text{by definition of formal} \\
&= \sum_{|\beta| \leqslant m} \frac{(x - a)^\beta}{\beta!} D^\beta(T_a^m f^\bullet)(a) - \sum_{|\beta| \leqslant m} \frac{(x - a)^\beta}{\beta!} D^\beta(T_b^m f^\bullet)(a) && \text{by properties (i) and (} \\
&= \sum_{|\beta| \leqslant m} \frac{(x - a)^\beta}{\beta!} \left(D^\beta(T_a^m f^\bullet)(a) - D^\beta(T_b^m f^\bullet)(a)\right) \\
&= \sum_{|\beta| \leqslant m} \frac{(x - a)^\beta}{\beta!} D^\beta(T_a^m f^\bullet - T_b^m f^\bullet)(a) \\
&= \sum_{|\beta| \leqslant m} \frac{(x - a)^\beta}{\beta!} (R_b^m f^\bullet)^{(\beta)}(a) && \text{by property (ii).}
\end{aligned}
$$

As a result, we see that

$$D^\alpha(T_a^m f^\bullet)(x) - D^\alpha(T_b^m f^\bullet)(x) = \sum_{|\beta| \leqslant m - |\alpha|} \frac{(x-a)^\beta}{\beta!} (R_a^m f^\bullet)^{(\beta+\alpha)}(a) \qquad \forall\, x \in \mathbb{R}^n.$$

Hence if $x \in \mathbb{R}^n$ is such that $\|x - a\| \leqslant \|x - b\|$, then we can estimate

$$\left| D^\alpha T_a^m f^\bullet(x) - D^\alpha T_b^m f^\bullet(x) \right| \leqslant \sum_{|\beta| \leqslant m - |\alpha|} \frac{\|x - a\|^{|\beta|}}{\beta!} \cdot \left\| (R_a^m f^\bullet)^{(\beta+\alpha)}(a) \right\|$$

$$\leqslant \sum_{|\beta| \leqslant m - |\alpha|} \frac{\|x - a\|^{|\beta|}}{\beta!} \cdot \|a - b\|^{m - |\alpha| - |\beta|} \cdot \omega(\|a - b\|)$$

$$= \sum_{|\beta| \leqslant m - |\alpha|} \frac{1}{\beta!} \cdot \|x - a\|^{|\beta|} \cdot \|a - b\|^{m - |\alpha| - |\beta|} \cdot \omega(\|a - b\|). \quad (\star)$$

Then for each $x \in \mathbb{R}^n$ such that $\|x - a\| \leqslant \|x - b\|$, and each multi-index β with $|\beta| \leqslant m - |\alpha|$, we have

$$\|x - a\|^{|\beta|} \|a - b\|^{m - |\alpha| - |\beta|} \leqslant \|x - a\|^{|\beta|} \left(\|x - a\| + \|a - b\| \right)^{m - |\alpha| - |\beta|}$$

$$\leqslant \left(\|x - a\| + \|a - b\| \right)^{m - |\alpha|}$$

$$\leqslant 2^{m - |\alpha|} \cdot \left(\|x - a\|^{m - |\alpha|} + \|a - b\|^{m - |\alpha|} \right)$$

$$\leqslant 2^{m - |\alpha|} \cdot \left(\|x - a\|^{m - |\alpha|} + \left(\|a - x\| + \|x - b\| \right)^{m - |\alpha|} \right)$$

$$\leqslant 2^{m - |\alpha|} \cdot \left(\|x - a\|^{m - |\alpha|} + (2\|x - b\|)^{m - |\alpha|} \right)$$

$$\leqslant 4^{m - |\alpha|} \cdot \left(\|x - a\|^{m - |\alpha|} + \|x - b\|^{m - |\alpha|} \right)$$

Returning to the estimate $(\star)$, we conclude that

$$\left| D^\alpha T_a^m f^\bullet(x) - D^\alpha T_b^m f^\bullet(x) \right| \leqslant 4^{m - |\alpha|} \cdot \sum_{|\beta| \leqslant m - |\alpha|} \frac{1}{\beta!} \cdot \omega(\|a - b\|) \cdot \left(\|x - a\|^{m - |\alpha|} + \|x - b\|^{m - |\alpha|} \right)$$

$$\leqslant 4^{m - |\alpha|} \left(\sum_{\beta \in \mathbb{N}^n} \frac{1}{\beta!} \right) \cdot \omega(\|a - b\|) \cdot \left(\|x - a\|^{m - |\alpha|} + \|x - b\|^{m - |\alpha|} \right)$$

$$\leqslant 4^{m - |\alpha|} \left(\prod_{j=1}^n \sum_{k=0}^\infty \frac{1}{k!} \right) \cdot \omega(\|a - b\|) \cdot \left(\|x - a\|^{m - |\alpha|} + \|x - b\|^{m - |\alpha|} \right)$$

$$\leqslant 4^{m - |\alpha|} e^n \cdot \omega(\|a - b\|) \cdot \left(\|x - a\|^{m - |\alpha|} + \|x - b\|^{m - |\alpha|} \right).$$

Similarly, if $x \in \mathbb{R}^n$ is such that $\|x - b\| \leqslant \|x - a\|$, then we can estimate

$$\left| D^\alpha T_a^m f^\bullet(x) - D^\alpha T_b^m f^\bullet(x) \right| \leqslant 4^{m - |\alpha|} e^n \cdot \omega(\|a - b\|) \cdot \|x - b\|^{m - |\alpha|}.$$

Putting these two estimates together, we conclude that for each $x \in \mathbb{R}^n$ we have

$$\left| D^\alpha T_a^m f^\bullet(x) - D^\alpha T_b^m f^\bullet(x) \right| \leqslant 4^{m-|\alpha|} e^n \cdot \omega(\|a - b\|) \cdot \left(\|x - a\|^{m-|\alpha|} + \|x - b\|^{m-|\alpha|} \right)$$

as stated in the subclaim.

$\square$

Now we prove (🏃) in Claim 2 using two cases.

Let α be a multi-index with $|\alpha| \leqslant m$, and let $x \in Q$ and $a \in A$ be arbitrary. First consider the case where $x \in A$.

Proof of (🏃) *when* $x \in A$. If $x \in A$, then $\underline{\partial}^\alpha f(x) = f^{(\alpha)}(x)$ and we can estimate

$$\begin{aligned}
\left| \underline{\partial}^\alpha f(x) - D^\alpha T_a^m f^\bullet(x) \right| &= \left| f^{(\alpha)}(x) - T_a^{m-|\alpha|} D^\alpha f^\bullet(x) \right| \\
&= \left| (R_a^{m-|\alpha|} f^{\bullet+(\alpha)})(x) \right| \\
&\leqslant \omega \left(\|x - a\| \right) \cdot \|x - a\|^{m-|\alpha|}
\end{aligned}$$

where the first equality follows from Exercise 7.3.3, and the final inequality follows from Lemma 7.4.1 applied to the $m-|\alpha|$-jet $D^\alpha f^\bullet$. This shows that (🏃) holds for each $x \in A$. $\square$

Proof of (🏃) *when* $x \in A^c$. (NOTE: We have *not* yet used the definition of the extension operator $\mathcal{E}_m$ in this proof, and the consequences of the Whitney partition of unity and the Whitney decomposition, so this is where we will use these tools.)

Assume that $x \in Q \backslash A$. Then by definition of $f = \mathcal{E}_m(f^\bullet)$ and the fact that $\{\psi_j\}_{j=1}^\infty$ is a partition of unity on A^c, we have

$$f(x) - T_a^m f^\bullet(x) = \sum_{j=1}^\infty \psi_j(x) T_{a_j}^m f^\bullet(x) - \sum_{j=1}^\infty \psi_j(x) T_a^m f^\bullet(x) = \sum_{j=1}^\infty \psi_j(x) \left(T_{a_j}^m f^\bullet(x) - T_a^m f^\bullet(x) \right).$$

Hence we have

$$D^\alpha f(x) - D^\alpha T_a^m f^\bullet(x) = \sum_{\beta \leqslant \alpha} \binom{\alpha}{\beta} \sum_{j=1}^\infty D^\beta \psi_j(x) \cdot D^{\alpha - \beta} \left(T_{a_j}^m f^\bullet(x) - T_a^m f^\bullet(x) \right)$$

by the Leibniz rule, which implies that

$$\left| D^\alpha f(x) - D^\alpha T_a^m f^\bullet(x) \right| \leqslant \sum_{\beta \leqslant \alpha} \binom{\alpha}{\beta} \left| \sum_{j=1}^\infty D^\beta \psi_j(x) \cdot D^{\alpha - \beta} \left(T_{a_j}^m f^\bullet(x) - T_a^m f^\bullet(x) \right) \right|.$$

For each multi-index β with $\beta \leqslant \alpha$, we let

$$S_\beta(x) := \sum_{j=1}^\infty D^\beta \psi_j(x) \cdot D^{\alpha - \beta} \left(T_{a_j}^m f^\bullet(x) - T_a^m f^\bullet(x) \right)$$

so that the previous equation reads

$$\left|D^\alpha f(x) - D^\alpha T_a^m f^\bullet(x)\right| \leqslant \sum_{\beta \leqslant \alpha} \binom{\alpha}{\beta} |S_\beta(x)|. \tag{$\heartsuit$}$$

First we estimate $|S_0(x)|$. If $j \geqslant 1$ is such that $x \in Q_j^*$, then we have

$$\begin{aligned}
\|x - a_j\| &\leqslant \operatorname{diam} Q_j^* + \operatorname{dist}(Q_j^*, A) \\
&\leqslant \frac{5}{4} \operatorname{diam} Q_j + \frac{3}{4} \operatorname{dist}(Q_j, A) \quad \text{since } Q_j^* \text{ is the dilation of } Q_j \text{ by a factor of } 1 + \varepsilon < \frac{5}{4} \\
&\leqslant 2 \operatorname{diam} Q_j + \operatorname{dist}(Q_j, A) \\
&\leqslant 8 \operatorname{dist}(Q_j, A) + \operatorname{dist}(Q_j, A) \quad \text{by Lemma ??} \\
&\leqslant 9\|x - a\|
\end{aligned}$$

where the first inequality follows from the triangle inequality, and the final inequality follows from the fact that $x \in Q_j$ and $a \in A$.

Hence if $j \geqslant 1$ is such that $x \in Q_j^*$, then we have

$$\|a - a_j\| \leqslant \|a - x\| + \|x - a_j\| \leqslant \|a - x\| + 9\|x - a\| = 10\|x - a\|$$

which implies

$$\omega(\|a - a_j\|) \leqslant 10\omega(\|x - a\|)$$

since ω is concave and increasing.

If $j \geqslant 1$ is such that $x \notin Q_j^*$, then $\psi_j(x) = 0$ and so this term does not contribute to the sum defining $S_0(x)$.

Therefore we have

$$\begin{aligned}
|S_0(x)| &\leqslant \sum_{\{j\,:\,x \in Q_j^*\}} 1 \cdot \left|D^\alpha(T_{a_j}^m f^\bullet - T_a^m f^\bullet)(x)\right| \\
&\leqslant C_{m,n} \cdot \sum_{\{j\,:\,x \in Q_j^*\}} \omega(\|a - a_j\|)\left(\|x - a\|^{m-|\alpha|} + \|x - a_j\|^{m-|\alpha|}\right) \qquad \text{by Subclaim 2a} \\
&\leqslant C_{m,n} \cdot \sum_{\{j\,:\,x \in Q_j^*\}} 10\omega(\|x - a\|)\left(\|x - a\|^{m-|\alpha|} + (9\|x - a\|)^{m-|\alpha|}\right) \\
&\leqslant C_{m,n} \cdot 2^n \cdot 10(9^{m-|\alpha|} + 1) \cdot \omega(\|x - a\|) \cdot \|x - a\|^{m-|\alpha|} \qquad \text{by Corollary 7.2.4}
\end{aligned}$$

where $C_{m,n} = 4^m e^n$ is the constant from Subclaim 2a for the case of the zero multi-index. Thus there is a constant C_0 which only depends on m and n such that

$$|S_0(x)| \leqslant C_0 \cdot \omega(\|x - a\|) \cdot \|x - a\|^{m-|\alpha|}. \tag{$\clubsuit$}$$

Next let β be a multi-index with $\beta \leqslant \alpha$ and $|\beta| \geqslant 1$, and consider $|S_\beta(x)|$. Then see that for each $b \in A$ we have

$$S_\beta(x) = \sum_{j=1}^{\infty} D^\beta \psi_j(x) \cdot D^{\alpha-\beta} \left(T_{a_j}^m f^\bullet(x) - T_a^m f^\bullet(x) \right)$$

$$= \sum_{j=1}^{\infty} D^\beta \psi_j(x) \cdot \left(D^{\alpha-\beta} T_{a_j}^m f^\bullet(x) - D^{\alpha-\beta} T_b^m f^\bullet(x) + D^{\alpha-\beta} T_b^m f^\bullet(x) - D^{\alpha-\beta} T_a^m f^\bullet(x) \right)$$

$$= \sum_{j=1}^{\infty} D^\beta \psi_j(x) \cdot \left(D^{\alpha-\beta} T_{a_j}^m f^\bullet(x) - D^{\alpha-\beta} T_b^m f^\bullet(x) \right) + \sum_{j=1}^{\infty} D^\beta \psi_j(x) \cdot \left(D^{\alpha-\beta} T_b^m f^\bullet(x) - D^{\alpha-\beta} T_a^m f^\bullet(x) \right)$$

$$= \sum_{j=1}^{\infty} D^\beta \psi_j(x) \cdot \left(D^{\alpha-\beta} T_{a_j}^m f^\bullet(x) - D^{\alpha-\beta} T_b^m f^\bullet(x) \right) + 0$$

$$= \sum_{j=1}^{\infty} D^\beta \psi_j(x) \cdot \left(D^{\alpha-\beta} T_{a_j}^m f^\bullet(x) - D^{\alpha-\beta} T_b^m f^\bullet(x) \right)$$

since $\sum_j \psi_j \equiv 1$ on A^c and hence $\sum_j D^\beta \psi_j \equiv 0$ on A^c for each multi-index β with $|\beta| \geqslant 1$.

Now choose a point $b \in A$ such that $\text{dist}(x, A) = \|x - b\|$. Then if $j \geqslant 1$ is such that $x \in Q_j^*$, then as before we have $\|x - a_j\| \leqslant 9\|x - b\|$ so that

$$\|a_j - b\| \leqslant \|a_j - x\| + \|x - b\| \leqslant 10\|x - b\|$$

and

$$\omega(\|a_j - b\|) \leqslant 10\omega(\|x - b\|).$$

Hence we can estimate

$$|S_\beta(x)| \leqslant \sum_{\{j : x \in Q_j^*\}} \left| D^\beta \psi_j(x) \cdot D^{\alpha-\beta} \left(T_{a_j}^m f^\bullet - T_b^m f^\bullet \right)(x) \right|$$

$$\leqslant \sum_{\{j : x \in Q_j^*\}} \left| D^\beta \psi_j(x) \right| \cdot \left(4^{m-|\alpha-\beta|} e^n \omega(\|a_j - b\|) \left(\|x - a_j\|^{m-|\alpha-\beta|} + \|x - b\|^{m-|\alpha-\beta|} \right) \right) \qquad \text{by Subclai}$$

$$\leqslant C_{m,n,\beta} \cdot \sum_{\{j : x \in Q_j^*\}} \left| D^\beta \psi_j(x) \right| \cdot 10\omega(\|x - b\|) \left((9\|x - b\|)^{m-|\alpha-\beta|} + \|x - b\|^{m-|\alpha-\beta|} \right) \qquad \text{where } C_{m,n}$$

$$\leqslant C_{m,n,\beta} \cdot 10(9^{m-|\alpha-\beta|} + 1) \cdot \omega(\|x - b\|)\|x - b\|^{m-|\alpha-\beta|} \sum_{\{j : x \in Q_j^*\}} \left| D^\beta \psi_j(x) \right|$$

$$\leqslant C_{m,n,\beta} \cdot 10(9^{m-|\alpha-\beta|} + 1) \cdot \omega(\|x - b\|)\|x - b\|^{m-|\alpha-\beta|} \cdot C_\beta \sum_{\{j : x \in Q_j^*\}} (\text{diam } Q_j)^{-|\beta|}$$

where the last inequality holds by the property (iv) of the Whitney partition of unity 7.2.8, and the constant C_β only depends only on β and n.

Letting

$$\hat{C}_\beta := C_{m,n,\beta} \cdot 10(9^{m-|\alpha-\beta|} + 1) \cdot C_\beta$$

to make things more managable, we have shown that

$$|S_\beta(x)| \leqslant \hat{C}_\beta \cdot \omega(\|x - b\|)\|x - b\|^{m-|\alpha-\beta|} \sum_{\{j\,:\,x\in Q_j^*\}} (\operatorname{diam} Q_j)^{-|\beta|}.$$

Now fix some $l \in \mathbb{Z}^+$ such that $x \in Q_l$. Then for each $j \geqslant 1$ such that $x \in Q_j^*$, the cubes Q_j and Q_l must be adjacent, as shown in the proof of Corollary 7.2.4. Hence there are at most 2^n terms in the sum $\sum_{\{j\,:\,x\in Q_j^*\}} (\operatorname{diam} Q_j)^{-|\beta|}$, as shown in 7.2.4, and each term is at most $4^{|\beta|}(\operatorname{dist}(Q_l, A))^{-|\beta|}$ by Lemma **??** (b) and (c).

Hence we have

$$|S_\beta(x)| \leqslant 2^n \cdot 4^{|\beta|} \cdot \hat{C}_\beta \cdot \omega(\|x - b\|)\|x - b\|^{m-|\alpha-\beta|} \cdot (\operatorname{dist}(Q_l, A))^{-|\beta|}$$
$$= 2^n \cdot 4^{|\beta|} \cdot \hat{C}_\beta \cdot \omega\big(\operatorname{dist}(x, A)\big) \big(\operatorname{dist}(x, A)\big)^{m-|\alpha-\beta|} \cdot \big(\operatorname{dist}(x, A)\big)^{-|\beta|}$$

because $\|x - b\| = \operatorname{dist}(x, A)$ and $\operatorname{dist}(Q_l, A) \geqslant \operatorname{dist}(x, A)$ by virtue of $x \in Q_l$. Since $\beta \leqslant \alpha$, we see that $m - |\alpha - \beta| = m - |\alpha| + |\beta|$ and thus we have

$$|S_\beta(x)| \leqslant 2^n \cdot 4^{|\beta|} \cdot \hat{C}_\beta \cdot \omega\big(\operatorname{dist}(x, A)\big) \big(\operatorname{dist}(x, A)\big)^{m-|\alpha|}.$$

Letting $C'_\beta := 2^n \cdot 4^{|\beta|} \cdot \hat{C}_\beta$, we have shown that

$$|S_\beta(x)| \leqslant C'_\beta \cdot \omega\big(\operatorname{dist}(x, A)\big) \big(\operatorname{dist}(x, A)\big)^{m-|\alpha|}. \tag{$\spadesuit$}$$

Putting this all together, we have

$$|D^\alpha f(x) - D^\alpha T_a^m f^\bullet(x)| \leqslant \sum_{\beta \leqslant \alpha} \binom{\alpha}{\beta} |S_\beta(x)|$$

$$= |S_0(x)| + \sum_{\substack{\beta \leqslant \alpha, \\ |\beta| \geqslant 1}} \binom{\alpha}{\beta} |S_\beta(x)|$$

$$\leqslant C_0 \cdot \omega(\|x - a\|) \cdot \|x - a\|^{m-|\alpha|} + \sum_{\substack{\beta \leqslant \alpha, \\ |\beta| \geqslant 1}} \binom{\alpha}{\beta} C'_\beta \cdot \omega\big(\operatorname{dist}(x, A)\big) \big(\operatorname{dist}(x, A)\big)^{m-|\alpha|}$$

$$= C \cdot \omega\big(\operatorname{dist}(x, A)\big) \big(\operatorname{dist}(x, A)\big)^{m-|\alpha|}$$

where we have set

$$C := C_0 + \sum_{\substack{\beta \leqslant \alpha, \\ |\beta| \geqslant 1}} \binom{\alpha}{\beta} C'_\beta.$$

Because $x \in A^c$, we have $\underline{\partial^\alpha} f(x) = D^\alpha f(x)$, and thus we have shown ($\mathbf{\Large\cdot}$) in the case where $x \in A^c$. $\qquad\square$

$\square$

Proof of Whitney's Extension Theorem in the general case. Let $A \subseteq \mathbb{R}^n$ be a closed set, and let $f^\bullet \in \mathcal{W}^m(A)$ be a Whitney m-jet on A.

partition of unity argument

$\square$

We did it. That is like, *the*, coolest theorem. I have been wanting to go through that full proof since I learned about the result as an undergrad.

It turns out that even more can be said about the extension operator in terms of continuity. If things are properly defined, it can be shown that a small modification of the operator $\mathcal{E}_m$ we constructed is a continuous map from the Banach space of s-Hölder continuous Whitney jets on a compact subset to the Banach space $C^{m,s}(\mathbb{R}^n)$ of C^m functions on $\mathbb{R}^n$ with s-Hölder continuous m^{th} derivatives. The details are in Stein's book *Singular Integrals and Differentiability Properties of Functions*.

The natural follow up question is the following: Is there a continuous extension operator for jets of infinite order, satisfying a Whitney condition?

Grothendieck showed that such an extension operator, with the natural definitions in place, cannot exist on a general closed set — his example shows that the singleton $\{0\}$ is a closed set for which no such extension operator can exist. See Bierstone's paper *Differentiable Functions* in the Bulletin of the Brazilian Mathematical Society for details.

7.5   Applications of Whitney's Extension Theorem

We flush out all the details of Whitney's paper *Functions Differentiable on the Boundaries of Regions*, which is a follow up to his extension theorem paper *Analytic Extensions of Differentiable Functions Defined in Closed Sets*.

7.5.1   Two different definitions of $C^m(\overline{U})$

In this section we want to define and compare two different C^m function spaces on the closure of an open set $U \subseteq \mathbb{R}^n$.

Definition 7.5.1. Let $U \subseteq \mathbb{R}^n$ be an open set. First we recall the definition of the space

$$C_{\mathrm{b}}^m(U) := \{f \in C^m(U) : \text{ for each } \alpha \in \mathbb{N}^n \text{ with } |\alpha| \leqslant m, D^\alpha f \text{ is bounded on } U\}$$

of C^m functions on U with bounded derivatives up to order m.

Then the "geometeric" definition of $C^m(\overline{U})$ is the set

$$C_{\mathrm{geo}}^m(\overline{U}) := \Big\{ f \in C^0(\overline{U}) : \text{ there is an open set } V \subseteq \mathbb{R}^n \text{ such that } \overline{U} \subset V \text{ and there is a function } \tilde{f} \in C^m(V$$

while the "analytic" definition of $C^m(\overline{U})$ is the set

$$C_{\mathrm{ana}}^m(\overline{U}) := \left\{ f \in C^m(U) : \text{ for each } \alpha \in \mathbb{N}^n \text{ with } |\alpha| \leqslant m, D^\alpha f \text{ is uniformly continuous on each bounded su} \right.$$

The definition of $C_{\mathrm{ana}}^m(\overline{U})$ makes it possible to extend a function $f \in C_{\mathrm{ana}}^m(\overline{U})$ and its derivatives to be defined on the closure $\overline{U}$. The definition of $C_{\mathrm{geo}}^m(\overline{U})$ makes it possible to extend a function $f \in C_{\mathrm{geo}}^m(\overline{U})$ and its derivatives to be defined on an open set containing the closure $\overline{U}$. That's the key difference.

Remark 7.5.2 (Natural inclusion $C_{\mathrm{geo}}^m(\overline{U}) \hookrightarrow C_{\mathrm{ana}}^m(\overline{U})$). Assume that $U \subseteq \mathbb{R}^n$ is an open set. Then it is easy to see that the restriction map

$$C_{\mathrm{geo}}^m(\overline{U}) \hookrightarrow C_{\mathrm{ana}}^m(\overline{U}), \qquad f \longmapsto f|_U$$

is a well-defined linear injection.

To see this, let $f \in C_{\mathrm{geo}}^m(\overline{U})$. Then $f \in C^0(\overline{U})$ and there is an open set $V \subseteq \mathbb{R}^n$ such that $\overline{U} \subset V$ and a function $\tilde{f} \in C^m(V)$ such that $\tilde{f}|_{\overline{U}} = f$. But then

$$f|_U = \tilde{f}|_U \in C^m(U)$$

and for each $\alpha \in \mathbb{N}^n$ with $|\alpha| \leqslant m$, the function $D^\alpha \tilde{f}$ is continuous on $\overline{U}$, so the function $D^\alpha(f|_U) = \left(D^\alpha \tilde{f} \right)|_U$ is continuous on each compact subset of U and is hence uniformly continuous on each bounded subset of U.

This shows $f|_U \in C_{\mathrm{ana}}^m(\overline{U})$, so the restriction map is well-defined. Also this map is injective because if $f, g \in C_{\mathrm{geo}}^m(\overline{U})$ and $f|_U = g|_U$, then $f = g$ on $\overline{U}$ by continuity of f and g. Lastly it is clear the restriction map is linear because if $f, g \in C_{\mathrm{geo}}^m(\overline{U})$ and $\lambda \in \mathbb{R}$, then

$$(f + \lambda g)|_U = f|_U + (\lambda g)|_U = f|_U + \lambda \cdot g|_U.$$

Note that we use the language "inclusion map" because we cannot say that $C_{\mathrm{geo}}^m(\overline{U})$ is a subset of $C_{\mathrm{ana}}^m(\overline{U})$ because (by definition) they contain different types of objects — the space $C_{\mathrm{geo}}^m(\overline{U})$ contains functions defined on $\overline{U}$, while $C_{\mathrm{ana}}^m(\overline{U})$ contains functions defined on U.

7.5.2 Properties of the analysts' definition of $C^m(\overline{U})$

Lemma 7.5.3 (Extension Lemma). *Let (X, d_X) be a metric space, let $U \subseteq X$ be an open subset, and let (Y, d_Y) be a complete metric space.*

Then each continuous function $g : U \to Y$ that is uniformly continuous on each bounded subset of U extends to a unique continuous function $\tilde{g} : \overline{U} \to Y$ such that $\tilde{g}|_U = g$.

Proof. Let $g : U \to Y$ be a function that is uniformly continuous on each bounded subset of U. Then we claim that there is a unique continuous function $\tilde{g} : \overline{U} \to \mathbb{R}$ such that $\tilde{g}|_U = g$.

First we will show that there is at most one such function $\tilde{g}$. Let $\tilde{g}_1, \tilde{g}_2 : \overline{U} \to Y$ be two continuous functions such that $\tilde{g}_1|_U = g$ and $\tilde{g}_2|_U = g$. Then for each $x \in \overline{U}$, we have

$$d_Y(\tilde{g}_1(x), \tilde{g}_2(x)) = \lim_{U \ni y \to x} d_Y(\tilde{g}_1(y), \tilde{g}_2(y)) = \lim_{U \ni y \to x} d_Y(g(y), g(y)) = 0,$$

so $\tilde{g}_1(x) = \tilde{g}_2(x)$. This shows that $\tilde{g}_1 = \tilde{g}_2$, and hence there is at most one such function $\tilde{g}$.

Now we will define the desired extension. For each $x \in \overline{U}$, choose a sequence $\{x_k\}_{k=1}^{\infty}$ in U such that $\lim_{k \to \infty} x_k = x$. Then the sequence $\{x_k\}_{k=1}^{\infty}$ is a Cauchy sequence in U, and is contained in a bounded subset V of U; by uniform continuity of g on V, the sequence $\{g(x_k)\}_{k=1}^{\infty}$ is a Cauchy sequence in the complete metric space Y, and hence converges to some point $\tilde{g}(x) \in Y$.

We claim that this function $\tilde{g} : \overline{U} \to Y$ is well-defined. Let $x \in U$ and let $\{x_k\}_{k=1}^{\infty}$ and $\{y_k\}_{k=1}^{\infty}$ be two sequences in U such that $\lim_{k \to \infty} x_k = x$ and $\lim_{k \to \infty} x'_k = x$. Then the sequences $\{x_k\}_{k=1}^{\infty}$ and $\{x'_k\}_{k=1}^{\infty}$ are Cauchy sequences in U, and are contained in a bounded subset V of U; since g is uniformly continuous on V, the sequences $\{g(x_k)\}_{k=1}^{\infty}$ and $\{g(x'_k)\}_{k=1}^{\infty}$ are Cauchy sequences in Y, and hence converge. We need to show that

$$\lim_{k \to \infty} g(x_k) = \lim_{k \to \infty} g(x'_k). \tag{U}$$

Let $\{z_n\}_{n=1}^{\infty}$ be the sequence defined by $z_{2k-1} := x_k$ and $z_{2k} := x'_k$ for each $k \in \mathbb{Z}^+$. (That is the sequence $\{z_n\}_{n=1}^{\infty}$ alternates between the two sequences $\{x_k\}_{k=1}^{\infty}$ and $\{x'_k\}_{k=1}^{\infty}$.) Let $\epsilon > 0$, and let $N \in \mathbb{Z}^+$ be such that

$$d_X(x_k, x) < \epsilon \quad \text{and} \quad d_X(x'_k, x) < \epsilon \qquad \forall\, k \geqslant N.$$

Then if $k \geqslant 2N$ see that $\frac{n}{2} \geqslant N$ and $\frac{n+1}{2} \geqslant N$, so that

$$d_X(z_k, x) = d_X(x_{\frac{k}{2}}, x) < \epsilon$$

if k is even, and

$$d_X(z_k, x) = d_X(x'_{\frac{k+1}{2}}, x) < \epsilon$$

if k is odd. In either case, we have $d(z_k, x) < \epsilon$ for each $k \geqslant 2N$, so $\{z_n\}_{n=1}^{\infty}$ is a sequence in U that converges to x. As we argued above, the sequence $\{g(z_n)\}_{n=1}^{\infty}$ is a Cauchy sequence in Y, and hence converges to some point $y \in Y$. Since the sequences $\{x_k\}_{k=1}^{\infty}$ and $\{x'_k\}_{k=1}^{\infty}$ are subsequences of $\{z_n\}_{n=1}^{\infty}$, we see that the sequences $\{g(x_k)\}_{k=1}^{\infty}$ and $\{g(x'_k)\}_{k=1}^{\infty}$ are subsequences of $\{g(z_n)\}_{n=1}^{\infty}$, and hence converge to the same limit y. This proves (U), and hence $\tilde{g}$ is well-defined.

Finally we will show that $\tilde{g}$ is continuous and that $\tilde{g}|_U = g$.

Let $x \in U$ and see that the constant sequence with each term equal to x converges to x, so $\tilde{g}(x) = \lim_{k \to \infty} g(x) = g(x)$, and hence $\tilde{g}|_U = g$.

To see that $\tilde{g}$ is continuous, we will prove that $\tilde{g}$ is uniformly continuous on each bounded subset of $\overline{U}$. In particular, from this it follows that $\tilde{g}$ is uniformly continuous on each closed ball in $\overline{U}$, and hence that $\tilde{g}$ is continuous on $\overline{U}$.

Fix a bounded subset K of $\overline{U}$, and let $\epsilon > 0$. Since g is uniformly continuous on K, there exists $\delta > 0$ such that

$$x_1, x_2 \in K, d_X(x_1, x_2) < \delta \implies d_Y(g(x_1), g(x_2)) < \frac{\epsilon}{3}.$$

Let $x_1, x_2 \in K$ be points such that $d_X(x_1, x_2) < \frac{\delta}{3}$. Then there are sequences $\{x_{1,k}\}_{k=1}^{\infty}$ and $\{x_{2,k}\}_{k=1}^{\infty}$ in U such that $\lim_{k\to\infty} x_{1,k} = x_1$ and $\lim_{k\to\infty} x_{2,k} = x_2$. Hence there exists $N_1 \in \mathbb{Z}^+$ such that for each $k \geqslant N_1$, we have

$$d_X(x_{1,k}, x_1) < \frac{\delta}{3} \quad \text{and} \quad d_X(x_{2,k}, x_2) < \frac{\delta}{3}.$$

Then for each $k \geqslant N_1$, we have

$$d_X(x_{1,k}, x_{2,k}) \leqslant d_X(x_{1,k}, x_1) + d_X(x_1, x_2) + d_X(x_2, x_{2,k}) < \frac{\delta}{3} + \frac{\delta}{3} + \frac{\delta}{3} = \delta$$

and hence $d_Y(g(x_{1,k}), g(x_{2,k})) < \epsilon$.

Also since $\tilde{g}(x_1) = \lim_{k\to\infty} g(x_{1,k})$ and $\tilde{g}(x_2) = \lim_{k\to\infty} g(x_{2,k})$, there exists $N_2 \in \mathbb{Z}^+$ such that for each $k \geqslant N_2$, we have

$$d_Y(g(x_{1,k}), \tilde{g}(x_1)) < \frac{\epsilon}{3} \quad \text{and} \quad d_Y(g(x_{2,k}), \tilde{g}(x_2)) < \frac{\epsilon}{3}.$$

Then for each $k \geqslant \max\{N_1, N_2\}$, we have

$$d_Y(\tilde{g}(x_1), \tilde{g}(x_2)) \leqslant d_Y(\tilde{g}(x_1), g(x_{1,k})) + d_Y(g(x_{1,k}), g(x_{2,k})) + d_Y(g(x_{2,k}), \tilde{g}(x_2)) < \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon.$$

Since $x_1, x_2 \in K$ were arbitrary points such that $d_X(x_1, x_2) < \frac{\delta}{3}$, this shows that

$$x_1, x_2 \in K, d_X(x_1, x_2) < \frac{\delta}{3} \implies d_Y(\tilde{g}(x_1), \tilde{g}(x_2)) < \epsilon.$$

Since $\epsilon > 0$ was arbitrary, this shows that $\tilde{g}$ is uniformly continuous on K.

Since K was an arbitrary bounded subset of $\overline{U}$, this shows that $\tilde{g}$ is uniformly continuous on each bounded subset of $\overline{U}$, and hence that $\tilde{g}$ is continuous on $\overline{U}$. $\square$

Corollary 7.5.4 (Derivatives of $C_{\text{ana}}^m(\overline{U})$ functions extend to continuous functions on the closure). *Let $U \subseteq \mathbb{R}^n$ be an open set and let $m \in \mathbb{Z}^+$. If $f \in C_{ana}^m(\overline{U})$ then for each $\alpha \in \mathbb{N}^n$ with $|\alpha| \leqslant m$, the function $D^\alpha f : U \to \mathbb{R}$ extends to a continuous function on $\overline{U}$, which we also denote by $D^\alpha f$.*

Proof. Let $f \in C^m_{\mathrm{ana}}(\overline{U})$ and let $\alpha \in \mathbb{N}^n$ with $|\alpha| \leqslant m$. Then the function $D^\alpha f : U \to \mathbb{R}$ is uniformly continuous on each bounded subset of U, so by Lemma 7.5.3, there is a unique continuous function $\tilde{g}_\alpha : \overline{U} \to \mathbb{R}$ such that $\tilde{g}_\alpha|_U = D^\alpha f$. For each $\alpha \in \mathbb{N}^n$ with $|\alpha| \leqslant m$, we will slightly abuse notation and denote $\tilde{g}_\alpha$ by $D^\alpha f$. This shows that for each $\alpha \in \mathbb{N}^n$ with $|\alpha| \leqslant m$, the function $D^\alpha f : U \to \mathbb{R}$ extends to a continuous function on $\overline{U}$, also denoted by $D^\alpha f$. $\qquad\square$

Remark 7.5.5 (Be careful with the abuse of notation at boundary points). In general, you should be careful with the notation $D^\alpha f(x)$ when x is a boundary point of U — of course it is true that

$$\lim_{U \ni y \to x} D^\alpha f(y) = D^\alpha f(x),$$

but the classical definition (as the best linear approximation) of the derivative $D^\alpha f(x)$ at a boundary point x of U *is not guaranteed to hold* with the abuse of notation above.

   It turns out that this is only an issue for some pathological open sets U, which we are not really interested in. Still, be careful.

Exercise 7.5.1 (Norm on C^m spaces). Let $U \subseteq \mathbb{R}^n$ be an open set and let $m \in \mathbb{Z}^+$. For $f \in C^m(U)$ we recall its C^m-norm is defined by

$$\|f\|_{C^m(U)} := \max_{|\alpha| \leqslant m} \|D^\alpha f\|_{C^0(U)},$$

which may be infinite.

   Show that $\|\cdot\|_{C^m(U)}$ is a norm on $C^m_{\mathrm{b}}(U)$ and on $C^m_{\mathrm{ana}}(\overline{U})$.

Also show that $[\![\cdot]\!]_{C^m(U)}$ defined by

$$[\![f]\!]_{C^m(U)} := \sum_{|\alpha| \leqslant m} \|D^\alpha f\|_{C^0(U)}$$

is a norm on $C^m_{\mathrm{b}}(U)$ and on $C^m_{\mathrm{ana}}(\overline{U})$, and that the two norms $\|\cdot\|_{C^m(U)}$ and $[\![\cdot]\!]_{C^m(U)}$ are equivalent on these spaces.

   Because these norms are equivalent, we will use the notation $\|\cdot\|_{C^m(U)}$ for both of them, and will only distinguish them when necessary.

Proof. First we will show that $\|\cdot\|_{C^m(U)}$ is positive definite, homogeneous, and satisfies the triangle inequality for C^m functions on U. Let $f, g \in C^m(U)$ and let $\lambda \in \mathbb{R}$.

   Then clearly $\|f\|_{C^m(U)} \geqslant 0$ and

$$\|f\|_{C^m(U)} = 0 \implies \|D^\alpha f\|_{C^0(U)} = 0 \text{ for each } \alpha \in \mathbb{N}^n \text{ with } |\alpha| \leqslant m$$
$$\implies \|f\|_{C^0(U)} = 0$$
$$\implies f = 0 \text{ on } U$$

by using positive definiteness of the C^0-norm $\|\cdot\|_{C^0(U)}$. This shows that $\|\cdot\|_{C^m(U)}$ is positive definite.

Also we have

$$
\begin{aligned}
\|\lambda f\|_{C^m(U)} &= \max_{|\alpha| \leqslant m} \|D^\alpha(\lambda f)\|_{C^0(U)} \\
&= \max_{|\alpha| \leqslant m} |\lambda| \cdot \|D^\alpha f\|_{C^0(U)} \\
&= |\lambda| \cdot \max_{|\alpha| \leqslant m} \|D^\alpha f\|_{C^0(U)} = |\lambda| \cdot \|f\|_{C^m(U)}
\end{aligned}
$$

by using homogeneity of the C^0-norm $\|\cdot\|_{C^0(U)}$. This shows that $\|\cdot\|_{C^m(U)}$ is homogeneous.

Finally we have

$$
\begin{aligned}
\|f + g\|_{C^m(U)} &= \max_{|\alpha| \leqslant m} \|D^\alpha(f + g)\|_{C^0(U)} \\
&\leqslant \max_{|\alpha| \leqslant m} \left(\|D^\alpha f\|_{C^0(U)} + \|D^\alpha g\|_{C^0(U)} \right) \\
&\leqslant \max_{|\alpha| \leqslant m} \|D^\alpha f\|_{C^0(U)} + \max_{|\alpha| \leqslant m} \|D^\alpha g\|_{C^0(U)} = \|f\|_{C^m(U)} + \|g\|_{C^m(U)}
\end{aligned}
$$

by using the triangle inequality for the C^0-norm $\|\cdot\|_{C^0(U)}$. This shows that $\|\cdot\|_{C^m(U)}$ satisfies the triangle inequality.

In particular, $\|\cdot\|_{C^m(U)}$ is a norm on $C_{\mathrm{b}}^m(U)$ and on $C_{\mathrm{ana}}^m(\overline{U})$ because it is finite-valued on these subsets of $C^m(U)$. (Notice this function is not finite-valued on $C^m(U)$ and hence not a norm on $C^m(U)$; that is our reason for considering these subspaces).

A nearly identical proof shows that $[\![\cdot]\!]_{C^m(U)}$ is a norm on $C_{\mathrm{b}}^m(U)$ and on $C_{\mathrm{ana}}^m(\overline{U})$.

Finally see that if $f \in C^m(U)$, then

$$
\max_{|\alpha| \leqslant m} \|D^\alpha f\|_{C^0(U)} \leqslant \sum_{|\alpha| \leqslant m} \|D^\alpha f\|_{C^0(U)} \leqslant \left(\sum_{|\alpha| \leqslant m} 1 \right) \max_{|\alpha| \leqslant m} \|D^\alpha f\|_{C^0(U)}
$$

which says that

$$
\|f\|_{C^m(U)} \leqslant [\![f]\!]_{C^m(U)} \leqslant \left(\sum_{|\alpha| \leqslant m} 1 \right) \|f\|_{C^m(U)}
$$

so the two norms $\|\cdot\|_{C^m(U)}$ and $[\![\cdot]\!]_{C^m(U)}$ are equivalent on the spaces $C_{\mathrm{b}}^m(U)$ and on $C_{\mathrm{ana}}^m(\overline{U})$.

$\square$

Proposition 7.5.6 ($C_{\mathrm{b}}^1(U)$ is a Banach space)**.** *Let $U \subseteq \mathbb{R}^n$ be an open set. Then the space $C_b^1(U)$ is a Banach space with respect to the C^1-norm $\|\cdot\|_{C^1(U)}$.*

Proof. Let $\{f_k\}_{k=1}^{\infty}$ be a Cauchy sequence in $C_b^1(U)$ with respect to the C^1-norm $\|\cdot\|_{C^1(U)}$. Then $\{f_k\}_{k=1}^{\infty}$ is a Cauchy sequence in $C_b^0(U)$ with respect to the C^0-norm $\|\cdot\|_{C^0(U)}$, and for each $j \in \{1, 2, \ldots, n\}$, the sequence of partial derivatives $\{D_j f_k\}_{k=1}^{\infty}$ is a Cauchy sequence in $C_b^0(U)$. Since $C_b^0(U)$ is a Banach space with respect to the C^0-norm $\|\cdot\|_{C^0(U)}$, there are functions $f \in C_b^0(U)$ and $g_1, g_2, \ldots, g_n \in C_b^0(U)$ such that

$$\lim_{k \to \infty} \|f_k - f\|_{C^0(U)} = 0$$

and

$$\lim_{k \to \infty} \|D_j f_k - g_j\|_{C^0(U)} = 0 \quad \text{for each } j \in \{1, 2, \ldots, n\}.$$

The main difficulty is to show that $f \in C_b^1(U)$ and that $D_j f = g_j$ for each $j \in \{1, 2, \ldots, n\}$.

$$\left\| f(x + h) - f(x) - \sum_{j=1}^{n} g_j(x) h_j \right\| =$$

$\square$

Corollary 7.5.7 ($C_b^m(U)$ is a Banach space)**.** *Let $U \subseteq \mathbb{R}^n$ be an open set and let $m \in \mathbb{Z}^+$. Then the space $C_b^m(U)$ is a Banach space with respect to the C^m-norm $\|\cdot\|_{C^m(U)}$.*

Proof. $\square$

Proposition 7.5.8 ($C_{ana}^1(\overline{U})$ is a Banach space when U is bounded)**.** *Let $U \subseteq \mathbb{R}^n$ be a bounded open set. Then the space $C_{ana}^1(\overline{U})$ is also a Banach space with respect to the C^1-norm $\|\cdot\|_{C^1(U)}$.*

Proof. $\square$

Corollary 7.5.9 ($C_{ana}^m(\overline{U})$ is a Banach space when U is bounded)**.** *Let $U \subseteq \mathbb{R}^n$ be a bounded open set and let $m \in \mathbb{Z}^+$. Then the space $C_{ana}^m(\overline{U})$ is also a Banach space with respect to the C^m-norm $\|\cdot\|_{C^m(U)}$.*

Proof. $\square$

Exercise 7.5.2 (Fréchet space structure on $C_{ana}^m(\overline{U})$ when U is unbounded)**.** Let $U \subseteq \mathbb{R}^n$ be an open set and let $m \in \mathbb{Z}^+$. For each bounded open subset $V \subset U$ we define $\|\cdot\|_{C^m(V)}$ by

$$\|f\|_{C^m(V)} := \max_{|\alpha| \leqslant m} \|D^\alpha f\|_{C^0(V)},$$

and show that $\|\cdot\|_{C^m(V)}$ is a semi-norm on $C_{ana}^m(\overline{U})$.

Now let $\{V_k\}_{k=1}^{\infty}$ be an increasing sequence of bounded open subsets of U such that $\bigcup_{k=1}^{\infty} V_k = U$. Then show that the family of semi-norms $\{\|\cdot\|_{C^m(V_k)}\}_{k=1}^{\infty}$ gives $C_{ana}^m(\overline{U})$ the

structure of a Fréchet space, and that the resulting topology on $C_{\mathrm{ana}}^m(\overline{U})$ is independent of the choice of the sequence $\{V_k\}_{k=1}^\infty$.

In this topology, a sequence $\{f_j\}_{j=1}^\infty$ in $C_{\mathrm{ana}}^m(\overline{U})$ converges to $f \in C_{\mathrm{ana}}^m(\overline{U})$ if and only if for each bounded open subset $V \subset U$, we have $\|f_j - f\|_{C^m(V)} \to 0$ as $j \to \infty$.

Proof. The fact that $\|\cdot\|_{C^m(V)}$ is a semi-norm on $C_{\mathrm{ana}}^m(\overline{U})$ follows from the fact that $\|\cdot\|_{C^m(V)}$ is a norm on $C^m(V)$. Recall that this family of seminorms also defines the Fréchet space structure on $C^m(U)$, and that the resulting topology is independent of the choice of the sequence $\{V_k\}_{k=1}^\infty$.

See that $C_{\mathrm{ana}}^m(\overline{U})$ is a linear subspace of $C^m(U)$, and we claim that $C_{\mathrm{ana}}^m(\overline{U})$ is closed with respect to the topology defined by the family of seminorms $\{\|\cdot\|_{C^m(V_k)}\}_{k=1}^\infty$.

Let $\{f_j\}_{j=1}^\infty$ be a sequence in $C_{\mathrm{ana}}^m(\overline{U})$ that converges to some $f \in C^m(U)$ with respect to the family of seminorms $\{\|\cdot\|_{C^m(V_k)}\}_{k=1}^\infty$. We need to show that $f \in C_{\mathrm{ana}}^m(\overline{U})$, i.e. that f and all of its derivatives of order $\leqslant m$ are uniformly continuous on each bounded subset of U.

Let $V \subset U$ be a bounded open subset. Then there exists $N \in \mathbb{Z}^+$ such that $\overline{V} \subset V_N$, and hence $\|f_j - f\|_{C^m(V_N)} \to 0$ as $j \to \infty$. Since $C^m(\overline{V_N})$ is a Banach space with respect to the C^m-norm we just have $f|_{V_N} \in C^m(\overline{V_N})$. Then V is a bounded subset of V_N, so $f|_V$ and all of its derivatives of order $\leqslant m$ are uniformly continuous on V. Thus f and all of its derivatives of order $\leqslant m$ are uniformly continuous on each bounded subset of U, and hence $f \in C_{\mathrm{ana}}^m(\overline{U})$.

Since $\{f_j\}_{j=1}^\infty$ was an arbitrary convergent sequence in $C_{\mathrm{ana}}^m(\overline{U})$, this shows that $C_{\mathrm{ana}}^m(\overline{U})$ is closed with respect to the topology defined by the family of seminorms $\{\|\cdot\|_{C^m(V_k)}\}_{k=1}^\infty$. Since $C^m(U)$ is a Fréchet space with respect to this topology, it follows that $C_{\mathrm{ana}}^m(\overline{U})$ is also a Fréchet space with respect to the induced topology.

Since the Frechet space topology on $C^m(U)$ does not depend on the choice of the sequence $\{V_k\}_{k=1}^\infty$, it follows that the induced topology on $C_{\mathrm{ana}}^m(\overline{U})$ also does not depend on the choice of the sequence $\{V_k\}_{k=1}^\infty$.

It remains to prove the final remark about convergence of sequences in $C_{\mathrm{ana}}^m(\overline{U})$.

Let $\{f_j\}_{j=1}^\infty$ be a sequence in $C_{\mathrm{ana}}^m(\overline{U})$ and let $f \in C_{\mathrm{ana}}^m(\overline{U})$. if and only if for each $k \in \mathbb{Z}^+$, we have $\|f_j - f\|_{C^m(V_k)} \to 0$ as $j \to \infty$.

Suppose that $\{f_j\}_{j=1}^\infty$ converges to f in the topology defined by the family of seminorms $\{\|\cdot\|_{C^m(V_k)}\}_{k=1}^\infty$. For each bounded open subset $V \subset U$, there exists $N \in \mathbb{Z}^+$ such that $\overline{V} \subset V_N$, and hence

$$\lim_{j \to \infty} \|f_j - f\|_{C^m(V_N)} = 0$$

implies

$$\lim_{j \to \infty} \|f_j - f\|_{C^m(V)} = 0.$$

Since V was an arbitrary bounded open subset of U, this shows that for each bounded open subset $V \subset U$, we have $\|f_j - f\|_{C^m(V)} \to 0$ as $j \to \infty$.

Conversely, if $\{f_j\}_{j=1}^\infty$ converges to f in the $\|\cdot\|_{C^m(V)}$-norm for each bounded open subset $V \subset U$, then clearly $\{f_j\}_{j=1}^\infty$ converges to f in the topology defined by the family of seminorms $\{\|\cdot\|_{C^m(V_k)}\}_{k=1}^\infty$. Since $\{f_j\}_{j=1}^\infty$ was an arbitrary sequence in $C_{\mathrm{ana}}^m(\overline{U})$ and $f \in C_{\mathrm{ana}}^m(\overline{U})$ was arbitrary, this proves the final remark about convergence of sequences in $C_{\mathrm{ana}}^m(\overline{U})$. $\qquad\square$

7.5.3 When the two definitions of $C^m(\overline{U})$ agree

In this section, we will show that under a simple geometric condition on the open set U, the two definitions of $C^m(\overline{U})$ coincide in the sense that the natural inclusion map

$$C_{\mathrm{geo}}^m(\overline{U}) \hookrightarrow C_{\mathrm{ana}}^m(\overline{U}), \qquad f \longmapsto f|_U$$

from **??** is a bijection and the Whitney extension map is the inverse of the restriction map.

Definition 7.5.10 (Quasiconvex). Let $S \subseteq \mathbb{R}^n$ be an arbitrary set. We say that S is *quasiconvex* if there exists a constant $C \geqslant 1$ such that for every pair of points $x, y \in S$, there is a curve $\sigma : [0,1] \to U$ such that $\sigma(0) = x$, $\sigma(1) = y$, and

$$\mathrm{length}(\sigma) \leqslant C\|x - y\|.$$

If the constant C is important, we say that the set S is *C-quasiconvex*.

Note that the curve σ supposed to exist in the definition must be rectifiable, so that its length is well-defined and finite.

Of course, if S is convex, then S is 1-quasiconvex because every pair of points in S can be connected by a line segment, which has length equal to the distance between the two points. The definition of quasiconvexity is a weakening of convexity, in the sense that S may not be convex, but still every pair of points in S can be connected by a curve whose length is at most a constant multiple of the distance between the two points.

Lemma 7.5.11 (The closure of a quasiconvex open set is quasiconvex). *Let $U \subseteq \mathbb{R}^n$ be an open quasiconvex set. Then $\overline{U}$ is also quasiconvex.*

Proof. In fact if U is C-quasiconvex, then $\overline{U}$ is C_*-quasiconvex for each $C_* > C$.

$\qquad\square$

Proposition 7.5.12 (Remainder Formula). *Let $\sigma : [0, L] \to \mathbb{R}^n$ be a rectifiable curve which is parametrized by arc length, and has endpoints $\sigma(0) = a$ and $\sigma(L) = x$. Let $A = \sigma([0, L])$ be the image of σ. If $f^\bullet \in \mathcal{W}^m(A)$, then for each $\alpha \in \mathbb{N}^n$ with $|\alpha| \leqslant m$, the remainder term $R_a^m f^\bullet$ satisfies*

$$(R_a^m f^\bullet)^{(\alpha)}(x) = - \sum_{|\beta|=m-|\alpha|} \frac{1}{\beta!} \int_0^L \left[f^{(\alpha+\beta)}(\sigma(t)) - f^{(\alpha+\beta)}(x) \right] \mathrm{d}\!\left((a - \sigma(t))^\beta \right).$$

Here the integral is a Riemann-Stieltjes integral, as introduced in our Complex Analysis text.

Proof. Fix a Whitney jet $f^\bullet \in \mathcal{W}^m(A)$ and let α be a multi-index with $|\alpha| \leqslant m$.

Step 1: Consider an arbitrary partition P of the interval $[0, L]$ given by

$$0 = t_0 < t_1 < \cdots < t_N = L.$$

$$(R_a^m f^\bullet)^{(\alpha)}(\sigma(t_{j-1})) - (R_a^m f^\bullet)^{(\alpha)}(\sigma(t_j)) = \sum_{|\beta| \leqslant m - |\alpha|} \frac{\left(R_{\sigma(t_j)}^m f^\bullet\right)^{(\alpha+\beta)}(\sigma(t_{j-1}))}{\beta!}(a - \sigma(t_j))^\beta$$

SIMPLEST VERSION

$$(R_a^m f^\bullet)^{(\alpha)}(z) - (R_a^m f^\bullet)^{(\alpha)}(y) = \sum_{|\beta| \leqslant m - |\alpha|} \frac{\left(R_y^m f^\bullet\right)^{(\alpha+\beta)}(z)}{\beta!}(a - y)^\beta \qquad (H)$$

Proof. □

Hence **SOMETHING I DONT UNDERSTAND WHY IS TRUE, BUT IF IT IS THEN**

summing over j gives

$$(R_a^m f^\bullet)^{(\alpha)}(x) = \sum_{|\beta| \leqslant m - |\alpha|} \frac{1}{\beta!} \sum_{j=1}^N \left(R_{\sigma(t_j)}^m f^\bullet\right)^{(\alpha+\beta)}(\sigma(t_{j-1}))(a - \sigma(t_j))^\beta \qquad (9)$$

Step 2: We claim that if β is a multi-index with $|\beta| < m - |\alpha|$ then for each $\epsilon > 0$ there is a $\delta > 0$ such that if P is a partition of $[0, L]$ given by

$$0 = t_0 < t_1 < \cdots < t_N = L$$

and satisfies $\max_{1 \leqslant j \leqslant N}(t_j - t_{j-1}) < \delta$, then

$$\left| \sum_{j=1}^N \left(R_{\sigma(t_j)}^m f^\bullet\right)^{(\alpha+\beta)}(\sigma(t_{j-1}))(a - \sigma(t_j))^\beta \right| < \epsilon.$$

Proof. □

Now for each multi-index β with $|\beta| = m - |\alpha|$, we have

$$\sum_{j=1}^{N} \left(R_{\sigma(t_j)}^m f^\bullet \right)^{(\alpha+\beta)} (\sigma(t_{j-1}))(a - \sigma(t_j))^\beta = \sum_{j=1}^{N} \cdots$$

Thus plugging this into (9) and taking the limit as the mesh of the partition P tends to zero gives

$\square$

Lemma 7.5.13. *Let $\sigma : [0, L] \to \mathbb{R}^n$ be a rectifiable curve which is parametrized by arc length, and has endpoints $\sigma(0) = a$ and $\sigma(L) = x$. Let $A = \sigma([0, L])$ be the image of σ. If $f^\bullet \in \mathcal{W}^m(A)$, then for each $\alpha \in \mathbb{N}^n$ with $|\alpha| = m$, and $\delta > 0$ is such that*

$$\left| f^{(\alpha)}(\sigma(t)) - f^{(\alpha)}(x) \right| < \delta$$

then

$$\left| (R_a^m f^\bullet)^{(\alpha)}(x) \right| < n(m+1)^n L^{m-|\alpha|} \delta^5$$

Remark 7.5.14. The factor $n(m+1)^n$ is the number of multi-indices β with $|\beta| = m - |\alpha|$ — **CHECK THIS** and thus may be replaced by the factor

$$\sum \sum \frac{1}{!} = \frac{n^{m-|\alpha|}}{(m - |\alpha| - 1)!}$$

Proof. $\square$

Lemma 7.5.15.

Proof. $\square$

Theorem 7.5.16.