

525a

DUTCH HANSEN

Fall 2024

# Contents

<b>1</b>	<b>Preliminaries</b>	<b>4</b>
1.1	Motivation . . . . .	4
1.2	Sets . . . . .	5
1.3	Orderings . . . . .	5
1.4	More About Well-Ordered Sets . . . . .	7
1.5	The Extended Reals . . . . .	7
1.5.1	The Standard Topology . . . . .	7
1.5.2	Infinite Limits and Limits at Infinity . . . . .	8
1.5.3	One-Sided Limits and Types of Discontinuities . . . . .	8
1.5.4	Monotonic Functions and Sequences . . . . .	9
1.5.5	Upper and Lower Limits . . . . .	9
1.5.6	Limits of Some Special Sequences . . . . .	11
1.5.7	Additional points from Folland . . . . .	11
<b>2</b>	<b>Measures</b>	<b>12</b>
2.1	Introduction . . . . .	12
2.1.1	Existence of non-measurable sets . . . . .	12
2.2	$\sigma$ -Algebras . . . . .	14
2.2.1	Metric Spaces . . . . .	15
2.2.2	Product $\sigma$ -Algebras . . . . .	16
2.3	Measures . . . . .	19
2.4	Outer Measures . . . . .	22
2.5	Premeasures . . . . .	25
2.6	Borel Measures on $\mathbb{R}$ . . . . .	27
2.6.1	Lebesgue-Stieltjes Measure . . . . .	29
2.6.2	The Lebesgue Measure . . . . .	31
2.6.3	Null Sets Under Lebesgue Measure . . . . .	32
<b>3</b>	<b>Integration</b>	<b>33</b>
3.1	Measurable Functions . . . . .	33
3.1.1	Lebesgue-Measurable Functions . . . . .	34
3.2	Simple Functions . . . . .	37
3.3	Integration of Nonnegative Functions . . . . .	39
3.4	Fatou's Lemma and Standard Counterexamples . . . . .	42
3.5	Integration of Complex-Valued Functions . . . . .	43
3.5.1	Riemann Integration . . . . .	49
3.5.2	The Gamma Function . . . . .	50
3.6	Modes of Convergence . . . . .	51
3.7	Product Measures . . . . .	53
3.8	Then $n$ -Dimensional Lebesgue Integral . . . . .	58
3.9	Integration in Polar Coordinates . . . . .	60

<b>4</b>	<b>Signed Measures and Differentiation</b>	<b>62</b>
4.1	Signed Measures . . . . .	62
4.2	The Lebesgue-Radon-Nikodym Theorem . . . . .	66
4.3	Complex Measures . . . . .	69
4.4	Differentiation on Euclidean Space . . . . .	71
4.4.1	The Fundamental Differentiation Theorem . . . . .	73
4.4.2	Differentiation of Regular Measures . . . . .	75
4.5	Functions of Bounded Variation . . . . .	76
4.5.1	Decomposition of Complex Borel Measures . . . . .	82

# 1 Preliminaries

Lectures by Prof. Juhi Jang. We use [Folland \(1999\)](#). This book is littered with errors, and I do my best to correct them along the way. Basic knowledge of set theory, real numbers, and extended reals is assumed.

## 1.1 Motivation

**Example 1.1** (Failures of the Riemann integral). *Consider the sequence of Riemann-integrable functions  $(f_n)$  such that  $f_n \rightarrow f$ , and suppose there exists some Riemann-integrable  $g$  such that  $|f_n| \leq g$  everywhere. If  $(f_n)$  is difficult to understand, maybe we can approximate with  $g$ . Is it true that  $f$  is Riemann-integrable? Not in general:  $f$  may be extremely irregular. We have the following example.*

*Order the rationals  $\{r_k\} = \mathbb{Q} \cap [0, 1]$ , and define*

$$h_k(x) := \begin{cases} 1, & x = r_k \\ 0, & \text{else.} \end{cases}$$

*Let  $f_n(x) = \sum_{k=1}^n h_k(x)$ . For each  $n$ , the number of discontinuities is finite, so  $f_n$  is Riemann-integrable, but we claim its pointwise limit is not. Indeed,  $f_n \rightarrow f$  for*

$$f(x) := \begin{cases} 1, & x \in \mathbb{Q} \\ 0, & \text{else.} \end{cases}$$

*The limits of the upper and lower sums will always differ. As we will see,  $f$  is actually Lebesgue-integrable w.r.t. the Lebesgue measure.*

**Example 1.2** (Generalized functions). *Measure theory will also yield some generalization of functions. Consider the sequence  $(f_n)$  for indicator function*

$$f_n(x) = n \cdot \chi_{[0, 1/n]}(x).$$

*This is certainly Riemann-integrable for every  $n$ . What about the limit? Take any  $\phi \in C_c^\infty(\mathbb{R})$ , a smooth function with compact support. One can show that as  $n \rightarrow \infty$ , we have*

$$\int_{\mathbb{R}} f_n(x) \phi(x) \, dx \rightarrow \phi(0).$$

*Interestingly, however, there is no function  $f$  such that for arbitrary  $\phi \in C_c^\infty(\mathbb{R})$ ,*

$$\phi(0) = \int_{\mathbb{R}} f(x) \phi(x) \, dx.$$

*To better understand this situation, we need “measures,” or generalized functions. In particular, we will need to study the Dirac measure  $\delta_0(x)$ .*

**Example 1.3** (Convergence properties and completeness). *Recall that a bounded sequence  $(x_n)$  in  $\mathbb{R}^d$  has a convergent subsequence by completeness. We’d like to have a similar statement for integrable functions. To do so, the notion of convergence must be changed. Often times, pointwise convergence will fail to give this property, but a different notion of convergence using the Lebesgue measure will give this property.*

## 1.2 Sets

We follow the prologue of [Folland \(1999\)](#). Consider a family of sets  $\{E_n\}_{n=1}^{\infty}$  indexed by  $\mathbb{N}$ . We define

$$\limsup E_n = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} E_n, \quad \liminf E_n = \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} E_n.$$

One can verify that

$$\begin{aligned} \limsup E_n &= \{x : x \in E_n \text{ for infinitely many } n\} \\ \liminf E_n &= \{x : x \in E_n \text{ for all but finitely many } n\} \end{aligned}$$

Let  $f : X \rightarrow Y$ . For any indexed collection of subsets  $\{E_\alpha\}$ , it holds that

$$f^{-1}\left(\bigcup_{\alpha} E_{\alpha}\right) = \bigcup_{\alpha} f^{-1}(E_{\alpha}), \quad f^{-1}\left(\bigcap_{\alpha} E_{\alpha}\right) = \bigcap_{\alpha} f^{-1}(E_{\alpha}), \quad f^{-1}(E^c) = (f^{-1}(E))^c.$$

That is, the inverse image commutes with union, intersection, and complement.

## 1.3 Orderings

**Definition 1.4** (Partial order). A *partial ordering* on a nonempty set  $X$  is a relation on  $X$  satisfying reflexivity, transitivity, and antisymmetry. A *total (or linear) ordering* on  $X$  is a partial ordering on  $X$  where the additional property holds: for all  $x, y \in X$ , either  $x \sim y$  or  $y \sim x$ . We will typically denote partial orderings by  $\leq$ .

**Definition 1.5** (Order isomorphic). Two posets  $(X, \leq), (Y, \leq)$  with order are called *order isomorphic* if there exists a bijection  $X \rightarrow Y$  such that  $x_1 \leq x_2$  iff  $f(x_1) \leq f(x_2)$ .

**Definition 1.6**. For poset  $(X, \leq)$ , a *maximal (resp. minimal) element*  $x \in X$  is such that the only  $y \in X$  satisfying  $x \leq y$  (resp.  $x \geq y$ ) is  $x$  itself. Such elements may not exist, and they may not be unique unless the ordering is total.

If  $E \subset X$ , an *upper (resp. lower) bound* for  $E$  is an element  $x \in X$  such that  $y \leq x$  (resp.  $x \leq y$ ) for all  $y \in E$ . Such a bound need not be an element of  $E$ .

**Definition 1.7** (Well ordered). If  $X$  is totally ordered by  $\leq$ , and every nonempty subset  $E \subset X$  has a (necessarily unique) minimal element,  $X$  is said to be *well ordered* by  $\leq$ . In this case, we say  $\leq$  is called a *well ordering* on  $X$ .

**Principle 1.8** (Hausdorff Maximal Principle). Every partially ordered set has a maximal totally ordered subset. This is, if  $X$  is partially ordered by  $\leq$ , there is a set  $E \subset X$  that is totally ordered by  $\leq$ , such that no subset of  $X$  that properly includes  $E$  is totally ordered by  $\leq$ .

An equivalent statement is Zorn's lemma.

**Principle 1.9** (Zorn's Lemma). If  $X$  is a partially ordered set and every totally ordered subset of  $X$  has an upper bound, then  $X$  has a maximal element.

*Proof that Hausdorff's Principle is equivalent to Zorn's Lemma.* Assume Hausdorff's, and let  $X$  be a poset such that every totally ordered subset has an upper bound. Choose the maximal totally ordered subset  $C$ , and take its upper bound  $x \in C$  (in  $C$  by total ordering). We claim that  $x$  is maximal in  $X$ ; indeed, if not, then  $C \cup \{y\}$  also forms a totally ordered subset for some  $y \notin C$ , contradicting maximality of  $C$ .

To see the other direction, take the collection  $Q$  of all totally ordered subset of  $X$ , which is partially ordered by set inclusion. Take  $C$  to be a totally ordered subset of  $Q$ , a collection of subsets, each totally ordered by  $\leq$ , which are together totally ordered by  $\subset$ . Define  $U = \bigcup_{S \in C} S$ , the union of sets in  $C$ . It's clear that  $U$  is a superset of any set in  $C$ . It is also easy to see that  $U$  is itself totally ordered by  $\leq$ . For any  $x, y \in U$ ,  $x \in S_1$  and  $y \in S_2$ ; since  $Q$  is totally ordered, however, w.l.o.g.  $S_1 \subset S_2$ , so  $x, y \in S_2$  and  $x, y$  are comparable. Hence,  $U$  is an upper bound for  $C$  in  $Q$ . Since  $C$  was arbitrary, we have satisfied the condition for Zorn's Lemma, and  $Q$  has a maximal element, giving Hausdorff's Maximal Principle.  $\square$

**Principle 1.10** (Well Ordering Principle). *Every nonempty set  $X$  can be well ordered.*

*Proof.* Let  $\mathcal{W}$  be the collection of well orderings of subsets of  $X$ , and define a partial ordering  $\leq_{\mathcal{W}}$  on  $\mathcal{W}$  as follows. If  $\leq_1$  and  $\leq_2$  are well orderings on the subsets  $E_1$  and  $E_2$ , then  $\leq_1$  precedes  $\leq_2$  in the partial ordering if (i)  $\leq_2$  extends  $\leq_1$ , meaning  $E_1 \subset E_2$  and  $\leq_2$  agrees with  $\leq_1$  on  $E_1$ , and (ii) if  $x \in E_2 \setminus E_1$  then  $y \leq_2 x$  for all  $y \in E_1$ . One can check that this is a partial ordering.

We aim to apply Zorn's lemma to  $\mathcal{W}$ . Let  $C$  be a totally ordered subset of  $\mathcal{W}$ ; we must produce some well-ordering  $u = (U, \leq_U) \in \mathcal{W}$  such that  $x \leq_{\mathcal{W}} u$  for all  $x \in C$ . Start by letting  $C$  be a chain and defining

$$S = \{E \subset X : E \text{ is well-ordered by element of } C\}.$$

Now take the union of all these subsets,

$$U = \bigcup_{E \in S} E.$$

Now define  $\leq_U$  by  $x \leq_U y$  whenever  $x, y \in K$  and  $x \leq_K y$  for some  $(K, \leq_K) \in C$ .

To see that  $u \in \mathcal{W}$ , take an element and nonempty subset  $a \in A \subset U$ . By definition of  $U$ , we have  $a \in (W, \leq_W) \in C$  for some  $W$ , and there is a minimal element  $m \in W \cap S$  since  $W$  is well-ordered. Now for arbitrary  $a' \in A$ , if  $a' \in W$ , it follows that  $m \leq_W a'$ , and if  $a' \notin W$ , then  $a' \in (W', \leq_{W'}) \in C$ . But since  $C$  is a chain we have  $(W, \leq_W) \leq_{\mathcal{W}} (W', \leq_{W'})$ , and so  $m \leq_{W'} a'$ . Since we take  $\leq_U$  as the union of orders,  $m \leq_U a'$ , and  $m$  is minimal in  $A$ . That is,  $u \in \mathcal{W}$ .

To see that  $u$  is an upper bound for  $C$  in  $\mathcal{W}$ , note that for any  $(W, \leq_W) \in C$ ,  $W \subset U$ . Next, note that by definition of  $\leq_U$ ,  $\leq_U$  agrees with  $\leq_W$ . Finally, for any  $w' \in U \setminus W$  and  $w \in W$ , we have  $w' \in W'$  for some  $W'$  where  $W' \setminus W \neq \emptyset$ . Now by total ordering of  $C$ , it must be so that  $W \leq_{\mathcal{W}} W'$ , so  $w \leq_U w'$ . Hence,  $C$  is an upper bound for  $C$ .

The hypothesis of Zorn's Lemma is satisfied, so  $\mathcal{W}$  has a maximal element, and this must be a well ordering on  $X$ . Indeed, if  $\leq$  is a well ordering on a proper subset  $E$  of  $X$  and  $x_0 \in X \setminus E$ , we can extend  $\leq$  to a well ordering on  $E \cup \{x_0\}$  by declaring that  $x \leq x_0$  for all  $x \in E$ .  $\square$

**Principle 1.11** (Axiom of Choice). *If  $\{X_\alpha\}_{\alpha \in A}$  is a nonempty collection of nonempty sets, then  $\prod_{\alpha \in A} X_\alpha$  is nonempty.*

*Proof by well-ordering principle.* Let  $X = \bigcup_{\alpha \in A} X_\alpha$ . Pick a well ordering on  $A$ , and for  $\alpha \in A$ , let  $f(\alpha)$  be the minimal element of  $X_\alpha$ . Then  $f \in \prod_{\alpha \in A} X_\alpha$ .  $\square$

**Corollary 1.12.** *If  $\{X_\alpha\}_{\alpha \in A}$  is a disjoint collection of nonempty sets, there is a set  $Y \subset \bigcup_{\alpha \in A} X_\alpha$  such that  $Y \cap X_\alpha$  contains precisely one element for each  $\alpha \in A$ .*

*Proof.* Take  $Y = f(A)$  where  $f \in \prod_{\alpha \in A} X_\alpha$ . □

While we have deduced the Axiom of Choice from the Hausdorff principle, the two are in fact equivalent. The following principle is also useful in some proofs, though not mentioned in Folland.

**Principle 1.13** (Axiom of Dependent Choice). *Let  $\mathcal{R}$  be a binary relation on a nonempty set  $S$ . Suppose that*

$$\forall a \in S, \exists b \in S : a\mathcal{R}b.$$

*Then there exists a sequence  $(x_n) \in S^{\mathbb{N}}$  such that for all  $n \in \mathbb{N}$ ,  $x_n\mathcal{R}x_{n+1}$ .*

**Proposition 1.14.** *The Axiom of Choice implies the Axiom of Dependent Choice.*

## 1.4 More About Well-Ordered Sets

We skip a review of cardinality in Folland (1999). [\[Dutch: Finish.\]](#)

## 1.5 The Extended Reals

Those familiar with the basic properties of  $\overline{\mathbb{R}}$  may skip to [Section 1.5.7](#). What follows here may serve as a reference.

### 1.5.1 The Standard Topology

Recall that the standard topology on  $\mathbb{R}$  is the topology whose basis is all open intervals. Considering  $\overline{\mathbb{R}}$ , let

$$\mathcal{B}_1 = \{(a, b) : a, b \in \mathbb{R}, a < b\}, \quad \mathcal{B}_2 = \{[-\infty, a) : a \in \mathbb{R}\}, \quad \mathcal{B}_3 = \{(b, +\infty] : b \in \mathbb{R}\}.$$

Now let  $\overline{\mathcal{B}} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3$ . Since this collection covers  $\overline{\mathbb{R}}$  and satisfies the intersection property, it is a basis for a topology  $\overline{\mathcal{T}}$ , which we refer to as the standard topology on  $\overline{\mathbb{R}}$ . This topology is also compatible with the standard topology  $\mathcal{T}$  on  $\mathbb{R}$ , in the following sense.

**Proposition 1.15.** *Let  $\mathcal{T}$  and  $\overline{\mathcal{T}}$  be the standard topologies for  $\mathbb{R}$  and  $\overline{\mathbb{R}}$ , respectively. Let  $E \subset \overline{\mathbb{R}}$ . Then  $E \in \overline{\mathcal{T}}$  if and only if  $E \cap \mathbb{R} \in \mathcal{T}$ .*

It turns out that while  $\mathbb{R}$  is not compact,  $\overline{\mathbb{R}}$  is compact. It also turns out that  $\overline{\mathbb{R}}$  is metrizable. The standard topology on  $\overline{\mathbb{R}}$  is induced by the metric  $\bar{d} : \overline{\mathbb{R}} \times \overline{\mathbb{R}} \rightarrow \mathbb{R}$ .

$$\bar{d}(x, y) = |f(x) - f(y)|, \quad \text{where} \quad f(x) = \begin{cases} 1 & x = \infty \\ \frac{x}{1+|x|} & x \in \mathbb{R} \\ -1 & x = -\infty. \end{cases}$$

**Proposition 1.16.**  *$(\overline{\mathbb{R}}, \bar{d})$  is a compact metric space.*

*Proof.* Construct homeomorphism  $f : [-\infty, \infty] \rightarrow [-1, 1]$ . □

### 1.5.2 Infinite Limits and Limits at Infinity

Most of our work in this section amounts to studying familiar definitions with the special case where  $\pm\infty$  is involved. Many of the claims have mechanical proofs, which are omitted.

**Proposition 1.17.** *Let  $\overline{\mathbb{R}}$  have the standard topology. Then*

- (i) *If  $U$  is a neighborhood of  $+\infty$  in  $\overline{\mathbb{R}}$ , then there exists  $M \in \mathbb{R}$  such that  $(M, +\infty] \subset U$ .*
- (ii) *If  $A \subset \mathbb{R}$  and  $A$  is not bounded above in  $\mathbb{R}$ , then  $+\infty$  is a limit point of  $A$  with respect to  $\overline{\mathbb{R}}$ .*

We now characterize what it means for certain limits to hold in  $\overline{\mathbb{R}}$ , using the general definition of convergence in topological spaces.

**Proposition 1.18** (Infinite limits, limits at infinity). (i) *Let  $A \subset \mathbb{R}$ , let  $p \in \mathbb{R}$  be a limit point of  $A$  with respect to  $\mathbb{R}$ , and let  $f : A \rightarrow \overline{\mathbb{R}}$  be a function. Then  $\lim_{x \rightarrow p} f(x) = +\infty$  if and only if for every  $L \in \mathbb{R}$ , there exists some  $\delta > 0$  such that  $0 < |x - p| < \delta$  and  $x \in A$  together imply that  $f(x) > L$ .*

(ii) *Let  $B \subset \mathbb{R}$  that is not bounded above in  $\mathbb{R}$ , and let  $g : B \rightarrow \overline{\mathbb{R}}$  be a function. Let  $q$  be a real number. Then  $\lim_{x \rightarrow \infty} g(x) = q$  if and only if for every  $\varepsilon > 0$ , there exists some  $M \in \mathbb{R}$  such that  $x > M$  and  $x \in B$  together imply that  $|g(x) - q| < \varepsilon$ .*

(iii) *Let  $C \subset \mathbb{R}$  that is not bounded above in  $\mathbb{R}$ , let  $h : C \rightarrow \overline{\mathbb{R}}$  be a function. Then  $\lim_{x \rightarrow \infty} h(x) = \infty$  if and only if for every  $N \in \mathbb{R}$ , there exists some  $P \in \mathbb{R}$  such that  $x > P$  and  $x \in C$  together imply that  $h(x) > N$ .*

### 1.5.3 One-Sided Limits and Types of Discontinuities

We consider real-valued functions defined on some interval  $(a, b)$  of  $\mathbb{R}$ . The discussion can be easily extended to the case where  $a$  or  $b$  are infinite, but we keep things simple by restricting to the latter case.

**Definition 1.19.** *Let  $f : (a, b) \rightarrow \mathbb{R}$  be a function and assume  $x \in [a, b)$ . We write  $f(x+) = q$  if whenever  $(t_n)$  is a sequence in  $(x, b)$  which converges to  $x$  we have  $f(t_n) \rightarrow q$ . We call this the limit from the right. The limit from the left is analogous.*

**Remark 1.20.** *Note that  $f(x+)$  and  $f(x-)$  may or may not exist. For  $x \in (a, b)$ ,  $\lim_{t \rightarrow x} f(t) = q$  is equivalent to  $f(x+) = f(x-) = q$ . That is,  $f$  is continuous at  $x \in (a, b)$  if and only if  $f(x+) = f(x-) = f(x)$ .*

**Definition 1.21** (Types of discontinuities). *Let  $f : (a, b) \rightarrow \mathbb{R}$  be a function and let  $x \in (a, b)$ . Assume that  $f(x+)$  and  $f(x-)$  both exist.*

- (i) *If  $f(x+) = f(x-)$  but  $f$  is not continuous at  $x$ , we say that  $f$  has a removable discontinuity at  $x$ .*
- (ii) *If  $f(x+) \neq f(x-)$ , we say that  $f$  has a jump discontinuity at  $x$ .*

*In either case, we say that  $f$  has a discontinuity of the first kind at  $x$ . If one of  $f(x+)$  or  $f(x-)$  does not exist, then  $f$  is said to have an essential discontinuity, or discontinuity of the second kind, at  $x$ .*



### 1.5.4 Monotonic Functions and Sequences

**Theorem 1.22.** Let  $f : (a, b) \rightarrow \mathbb{R}$  be monotonically increasing on  $(a, b)$  and let  $x \in (a, b)$ . Then

$$f(x-) = \sup\{f(t) : t \in (a, x)\}, \quad f(x+) = \inf\{f(t) : t \in (x, b)\}.$$

In particular,  $f(x-)$  and  $f(x+)$  both exist in  $\mathbb{R}$ , and  $f(x-) \leq f(x) \leq f(x+)$ .

**Corollary 1.23.** Let  $f : (a, b) \rightarrow \mathbb{R}$  be monotonically increasing and assume that  $a < x < y < b$ . Then  $f(x+) \leq f(y-)$ .

**Corollary 1.24.** Any monotonic function  $f : (a, b) \rightarrow \mathbb{R}$  has no discontinuities of the second kind.

**Theorem 1.25.** Let  $f : (a, b) \rightarrow \mathbb{R}$  be monotonically increasing and denote by  $E$  the subset of  $(a, b)$  consisting of point at which  $f$  is discontinuous. Then  $E$  is at most countable.

**Theorem 1.26.** Let  $(s_n)_{n=1}^\infty$  be a monotonically increasing sequence in  $\overline{\mathbb{R}}$ . Then  $\lim_{n \rightarrow \infty} s_n$  exists in  $\overline{\mathbb{R}}$  and is equal to  $\sup_n s_n$ . In particular, a monotonically increasing sequence of real number converges in  $\mathbb{R}$  if and only if it is bounded.

### 1.5.5 Upper and Lower Limits

Let  $(s_n)_1^\infty$  be a sequence of real numbers. For each  $n \in \mathbb{N}$ , define  $d_n = \sup\{s_k : k \in \mathbb{N}, k \geq n\}$ . Then  $(d_n)_1^\infty$  is a monotonically decreasing sequence in  $\overline{\mathbb{R}}$ . The following thus makes sense.

**Definition 1.27.** Let  $(s_n)_1^\infty$  be a sequence in  $\overline{\mathbb{R}}$ . Then the upper and lower limits of  $(s_n)_1^\infty$  are defined by

$$\begin{aligned} \limsup s_n &= \lim_{n \rightarrow \infty} \left( \sup_{k \geq n} s_k \right) = \inf_{n \in \mathbb{N}} \left( \sup_{k \geq n} s_k \right) \\ \liminf s_n &= \lim_{n \rightarrow \infty} \left( \inf_{k \geq n} s_k \right) = \sup_{n \in \mathbb{N}} \left( \inf_{k \geq n} s_k \right). \end{aligned}$$

We briefly characterize such limits.

**Proposition 1.28.** Let  $(s_n)_1^\infty$  and  $(t_n)_1^\infty$  be sequences in  $\overline{\mathbb{R}}$ .

- (i) If  $s_n \leq t_n$  for all  $n$  and  $\lim_{n \rightarrow \infty} s_n = +\infty$ , then  $\lim_{n \rightarrow \infty} t_n = +\infty$ .
- (ii) If  $(s_n)_1^\infty$  and  $(t_n)_1^\infty$  converge in  $\overline{\mathbb{R}}$  to  $s$  and  $t$  respectively, and if  $s_n \leq t_n$  for each  $n \in \mathbb{N}$ , then  $s \leq t$ .

**Theorem 1.29.** Let  $(s_n)_1^\infty$  be a sequence in  $\overline{\mathbb{R}}$ . Let  $S^*$  denote the set of its subsequential limits in  $\overline{\mathbb{R}}$ , and denote  $\alpha = \limsup_{n \rightarrow \infty} s_n$ . Let  $x \in \overline{\mathbb{R}}$ . Then

- (i)  $S^*$  is nonempty, and  $\alpha = \sup S^* \in S^*$ . In particular, there is a subsequence of  $S$  that converges to  $\alpha$ .
- (ii)  $\alpha < x$  if and only if there exists  $y < x$  such that  $s_n \leq y$  for all but finitely many  $n \in \mathbb{N}$ . In particular, if  $\alpha < x$ , then  $s_n < x$  for all but finitely many  $n \in \mathbb{N}$ . And if  $s_n < x$  for all but finitely many  $n \in \mathbb{N}$ , then  $\alpha \leq x$ .

(iii)  $\alpha > x$  if and only if  $s_n > x$  for infinitely many  $n \in \mathbb{N}$ .

*Proof.* (i) To see that  $S^*$  is nonempty, note that  $(\overline{\mathbb{R}}, \bar{d})$  is compact; therefore it is sequentially compact, which by definition means that  $S^*$  cannot be empty. Furthermore,  $S^*$  is closed in  $\overline{\mathbb{R}}$  by a more general result about the set of subsequential limits in a metric space. To see that  $\sup S^* \in S^*$ , we consider three cases.

- (i)  $\sup S^* = -\infty$ . Then  $S^*$  cannot contain any elements of  $\overline{\mathbb{R}}$  except possibly  $-\infty$ ; on the other hand,  $S^*$  is nonempty, so  $S^* = \{-\infty\}$  and the claim holds.
- (ii)  $\sup S^* \in \mathbb{R}$ . Then  $S^* \cap \mathbb{R}$  is a closed subset of  $\mathbb{R}$  which is bounded above in  $\mathbb{R}$ , and consequently  $\sup S^* = \sup(S^* \cap \mathbb{R}) \in S^* \cap \mathbb{R} \subset S^*$ .
- (iii)  $\sup S^* = +\infty$ . If  $+\infty \notin S^*$ , then there exists a neighborhood of  $+\infty$  of the form  $(M, +\infty]$  which does not intersect  $S^*$ ; this implies that  $S^*$  is bounded above, contradicting the assumption that  $\sup S^* = +\infty$ .

Next, we prove that  $\alpha \geq \sup S^*$ . Let  $s$  be any element of  $S^*$ , and let  $(s_{n_k})_{k=1}^\infty$  denote the subsequence of  $(s_n)_1^\infty$  that converges to  $s$  in  $\overline{\mathbb{R}}$ . Denote  $d_n = \sup_{k \geq n} s_k$ . Then for each  $k \in \mathbb{N}$ , we have

$$s_{n_k} \leq d_{n_k} \leq d_k.$$

Taking limits, we conclude that

$$s = \lim_{k \rightarrow \infty} s_{n_k} \leq \lim_{k \rightarrow \infty} d_k = \alpha.$$

Thus  $\alpha$  is an upper bound for  $S^*$ , so  $\sup S^* \leq \alpha$ .

To prove the opposite inequality, we argue by contradiction. Assume that  $\sup S^* < \alpha$ , and let  $\beta \in \mathbb{R}$  such that  $\sup S^* < \beta < \alpha$ . We claim that there exists  $N \in \mathbb{N}$  such that  $n \geq N$  implies  $s_n \leq \beta$ . Indeed, if no such  $N$  exists, then there exists a subsequence  $(s_{n_k})$  whose terms all exceed  $\beta$ . This subsequence will in turn have a further subsequence  $(s_{n_{k_\ell}})$  which converges in  $\overline{\mathbb{R}}$ ; its subsequential limit  $s$  will be at least  $\beta$ . But then  $s \in S^*$  and  $s \geq \beta > \sup S^*$ , which is a contradiction. Hence,  $\limsup_{n \rightarrow \infty} s_n \leq \beta < \alpha$ . This is a contradiction.

(ii) The statement  $\limsup_{n \rightarrow \infty} s_n < x$  holds if and only if there exists  $N \in \mathbb{N}$  such that  $\sup_{k \geq N} s_k < x$ , i.e.  $s_k \leq y := \sup_{j \geq N} s_j < x$  for all  $k \geq N$ . This is true if and only if only finitely many  $s_k$  possibly exceed  $x$ . (iii) follows from (i) since we can find a subsequence  $(s_{n_k})_1^\infty$  that converges to  $\alpha$ . We finish by choosing  $N \in \mathbb{N}$  such that  $k \geq N$  implies  $s_{n_k} > x$ .  $\square$

**Theorem 1.30.** Let  $(s_n)_1^\infty$  be a sequence in  $\overline{\mathbb{R}}$ . Then  $\lim_{n \rightarrow \infty} s_n = s$  if and only if  $\limsup_{n \rightarrow \infty} s_n = \liminf_{n \rightarrow \infty} s_n = s$ .

*Proof.* Let  $S^*$  denote the set of subsequential limits of  $(s_n)$  in  $\overline{\mathbb{R}}$ . Let  $(t_k)_1^\infty$  and  $(u_k)_1^\infty$  be subsequences  $(s_n)_1^\infty$  that converge to  $\limsup_{n \rightarrow \infty} s_n = \sup S^*$  and  $\liminf_{n \rightarrow \infty} s_n = \inf S^*$ , respectively.

If  $\lim s_n = s$ , then every subsequence of  $s_n$  tends to  $s$ ; in particular this is true of the sequences  $(t_k)_1^\infty$  and  $(u_k)_1^\infty$ . By uniqueness of limits,  $s = \limsup_{n \rightarrow \infty} s_n = \liminf_{n \rightarrow \infty} s_n$ . For the opposite direction, we argue by contradiction. Assume that  $s = \limsup_{n \rightarrow \infty} s_n = \liminf_{n \rightarrow \infty} s_n$  but that  $\lim_{n \rightarrow \infty} s_n = s$  fails. Then  $\sup S^* = \inf S^* = s$ , so  $S^* = \{s\}$ . But there exists a neighborhood  $U$  of  $s$  and a subsequence  $(s_{n_k})_1^\infty$  such that  $\{s_{n_k}\}_1^\infty$  lies entirely outside  $U$ . Since  $\overline{\mathbb{R}}$  is sequentially compact,  $(s_{n_k})_1^\infty$  has a further subsequence  $(s_{n_{k_\ell}})_1^\infty$  which converges to some point  $p$  of  $\overline{\mathbb{R}}$ , which must be outside  $U$  and therefore cannot be equal to  $s$ . But then  $p \in S^* = \{s\}$ , which is impossible.  $\square$

**Proposition 1.31.** Let  $(a_n)_1^\infty$  and  $(b_n)_1^\infty$  be sequences of real numbers. Then it holds that

$$\limsup_{n \rightarrow \infty} (a_n + b_n) \leq \limsup_{n \rightarrow \infty} a_n + \limsup_{n \rightarrow \infty} b_n,$$

provided that the RHS is not of the form  $\infty - \infty$ .<sup>1</sup>

### 1.5.6 Limits of Some Special Sequences

**Theorem 1.32.** The following limits hold:

- (i) If  $p > 0$ , then  $\lim_{n \rightarrow \infty} 1/n^p = 0$ .
- (ii) If  $q > 0$ , then  $\lim_{n \rightarrow \infty} \sqrt[n]{q} = 1$ .
- (iii)  $\lim_{n \rightarrow \infty} \sqrt[n]{n} = 1$ .
- (iv) If  $|x| < 1$ , then  $\lim_{n \rightarrow \infty} x^n = 0$ .

*Proof.* (i) Choose  $\varepsilon > 0$ , and let  $N$  be any integer larger than  $\varepsilon^{-1/p}$ . Then  $n \geq N$  implies  $n^{-p} \leq N^{-p} < (\varepsilon^{-1/p})^p = \varepsilon$ . (ii) If  $q = 1$ , then  $\sqrt[n]{q} = 1$  for all  $n$ . If  $q > 1$ , put  $x_n = \sqrt[n]{q} - 1$ . Then  $x_n > 0$  and  $1 + nx_n \leq (1 + x_n)^n = (\sqrt[n]{q})^n = q$  by the Binomial theorem. It follows that  $x_n \in (0, (q-1)/n)$  for all  $n$ , and  $x_n \rightarrow 0$ , so  $\sqrt[n]{q} \rightarrow 1$ . Alternatively, if  $q \in (0, 1)$ , then  $1/q > 1$ , so

$$\sqrt[n]{q} = \frac{1}{\sqrt[n]{1/q}} \rightarrow \frac{1}{1} = 1.$$

(iii) Put  $x_n = \sqrt[n]{n} - 1$ ; then

$$n = (1 + x_n)^n \geq \binom{n}{2} x_n^2 = \frac{n(n-1)}{2} x_n^2.$$

Thus  $0 \leq x_n \leq \sqrt{\frac{2}{n-1}}$  for  $n \geq 2$ , giving  $x_n \rightarrow 0$  and  $\sqrt[n]{n} \rightarrow 1$ . (iv) Since  $|x^n| = |x|^n$ , we may assume  $x \in (0, 1)$  without loss of generality. In this case,  $x = (1+p)^{-1}$  for some  $p > 0$ , and  $(1+p)^n \geq np$ , so

$$0 < x^n = (1-p)^{-n} \leq \frac{1}{p} \cdot \frac{1}{n}.$$

The sequence on the right tends to zero by (i), so we are done. □

### 1.5.7 Additional points from Folland

We visit some additional notes on  $\overline{\mathbb{R}}$  from the preface to [Folland \(1999\)](#). One can define the notion of a limit supremum and limit infimum for functions  $f : \mathbb{R} \rightarrow \overline{\mathbb{R}}$ , for instance,

$$\limsup_{x \rightarrow a} f(x) = \inf_{\delta > 0} \left( \sup_{0 < |z-a| < \delta} f(z) \right).$$

---

<sup>1</sup>Here, we use the usual conventions involving addition and subtraction of  $\pm\infty$ .

Uncountable sums are occasionally relevant. For an arbitrary set  $X$  and function  $f : X \rightarrow [0, \infty]$ , we define

$$\sum_{x \in X} f(x) = \sup \left\{ \sum_{x \in F} f(x) : F \subset X, F \text{ finite} \right\}.$$

**Proposition 1.33.** *Given  $f : X \rightarrow [0, \infty]$ , let  $A = \{x : f(x) > 0\}$ . If  $A$  is uncountable, then  $\sum_{x \in X} f(x) = \infty$ . If  $A$  is countably infinite, then  $\sum_{x \in X} f(x) = \sum_{n=1}^{\infty} f(g(n))$  where  $g : \mathbb{N} \rightarrow A$  is any bijection and the sum on the right is an ordinary infinite series.*

*Proof.* We have  $A = \bigcup_1^{\infty} A_n$  where  $A_n = \{x : f(x) > 1/n\}$ . If  $A$  is uncountable, there must be some uncountable  $A_n$ , and  $\sum_{x \in F} f(x) > \text{card}(F)/n$  for any finite subset  $F \subset A_n$ . It follows that  $\sum_{x \in X} f(x) = \infty$ .

If  $A$  is countably infinite,  $g : \mathbb{N} \rightarrow A$  is a bijection, and for  $B_N := g(\{1, \dots, N\})$ , every finite subset  $F$  of  $A$  is contained in some  $B_N$ . Hence,

$$\sum_{x \in F} f(x) \leq \sum_{n=1}^N f(g(n)) \leq \sum_{x \in X} f(x).$$

Taking the supremum over  $N$ , we find

$$\sum_{x \in F} f(x) \leq \sum_{n=1}^{\infty} f(g(n)) \leq \sum_{x \in X} f(x),$$

and then taking the supremum over  $F$ , we obtain the desired result.  $\square$

If a function  $f : \mathbb{R} \rightarrow \mathbb{R}$  is increasing, then  $f$  has right- and left-hand limits at each point:

$$f(a+) = \lim_{x \searrow a} f(x) = \inf_{x > a} f(x), \quad f(a-) = \lim_{x \nearrow a} f(x) = \sup_{x < a} f(x).$$

Moreover, the limiting values  $f(\infty) = \sup_{\mathbb{R}} f(x)$  and  $f(-\infty) = \inf_{\mathbb{R}} f(x)$  exist (and are possibly equal to  $\pm\infty$ ). We say  $f$  is right continuous if  $f(a) = f(a+)$  for all  $a \in \mathbb{R}$  and left continuous if  $f(a) = f(a-)$  for all  $a \in \mathbb{R}$ .

## 2 Measures

### 2.1 Introduction

#### 2.1.1 Existence of non-measurable sets

We want to measure arbitrary sets. We naively try to do this by constructing a map

$$\mu : 2^{\mathbb{R}^n} \rightarrow [0, \infty] \subset \overline{\mathbb{R}}.$$

We also want this map to have some nice properties:

(i) For a countable collection of sets  $E_1, E_2, \dots$  which are disjoint,

$$\mu\left(\bigcup_i E_i\right) = \sum_i \mu(E_i).$$

(ii) If sets  $E$  and  $F$  are congruent (can be mapped to each other by translations, rotations, reflections), then  $\mu(E) = \mu(F)$ .

(iii)  $\mu(Q) = 1$  where  $Q$  is the half-open<sup>2</sup> unit cube,  $Q = [0, 1)^n$ .

We may want to construct a  $\mu$  by approximating arbitrary sets  $U$  with several sets like  $Q$ . We will see that there is no such  $\mu$  satisfying these desiderata. This is due to the existence of non-measurable sets.

**Lemma 2.1.** *There exists a set  $N \subset [0, 1]$  such that for any  $\mu$  satisfying (i - iii),  $\mu(N)$  is ill-defined.*

*Proof.* Consider the following equivalence relation on  $\mathbb{R}$ :

$$x \sim y \iff x - y \in \mathbb{Q}.$$

One can see that  $\mathbb{R}$  decomposes into uncountably many  $[x]$  in the partition. Indeed,  $x$  can only be related to countably many  $y$ . By the Axiom of Choice, we can find some  $N \subset [0, 1)$  such that  $N$  contains exactly one representative of every equivalence class.

Now let  $\{r_j\}$  be an enumeration of the rational numbers  $\mathbb{Q} \cap (-1, 1)$  and consider the set

$$M = \bigcup_{j \in \mathbb{N}} \{x + r_j \in \mathbb{R} : x \in N\} = \bigcup_{i \in \mathbb{N}} N_{r_j}.$$

Delaying a proof, we claim that (1)  $[0, 1) \subset M \subset [-1, 2)$ , (2)  $N_{r_i} \cap N_{r_j} = \emptyset$  if  $r_i \neq r_j$ . Assuming this claim, we have

$$\mu(M) = \sum_{j \in \mathbb{N}} \mu(N_{r_j}) = \sum_{j \in \mathbb{N}} \mu(N), \tag{1}$$

and by monotonicity,

$$\mu([0, 1)) = 1 \leq \mu(M) \leq \mu([-1, 2)) \stackrel{(i)}{=} \mu([-1, 0)) + \mu([0, 1)) + \mu([1, 2)) \stackrel{(ii, iii)}{=} 3.$$

But this is of course a contradiction because  $\mu(N) > 0$ , so [Eq. \(1\)](#) implies  $\mu(M) = \infty$ .

We now return to proving (1) and (2). For (1), first note that  $N \subset [0, 1)$ , so  $M = \bigcup_j N + r_j \subset [-1, 2)$  since each  $r_j \in (-1, 1)$ . Next, let  $x \in [0, 1)$ . Then by the definition of  $N$ , there exists some  $\tilde{x} \in N$  such that  $x \sim \tilde{x}$ . We have  $x - \tilde{x} \in \mathbb{Q} \cap (-1, 1)$ , and  $x - \tilde{x} = r_{j_0}$  for some  $j_0$  and this implies  $x \in N + r_{j_0} = N_{r_{j_0}} \subset M$ . For (2), assume there exists  $x, y \in N$  such that

$$x + r_i = y + r_j.$$

Then  $x - y \in \mathbb{Q}$ , so  $x \sim y$ , and because  $N$  contains exactly one representative, we have  $x = y$ , and  $r_i = r_j$  necessarily.  $\square$

---

<sup>2</sup>Choosing half-open has the benefit of easy partitions. Since we can decompose  $\tilde{Q} = [0, 2)^n$  into  $2^n$  sets with the same measure as  $Q$ , so that  $\mu(\tilde{Q}) = 2^n$ .

Our motivating question becomes: how can we rectify this issue? Are we asking too much from  $\mu$ ? Or are we asking for too large of a domain? To better understand this situation, we cite but do not prove an illustrative result.

**Theorem 2.2** (Banach-Tarski, 1924). *Suppose  $U, V$  are two bounded open sets in  $\mathbb{R}^n$ ,  $n \geq 3$ . Then there exists  $k \in \mathbb{N}$  and sets  $E_1, \dots, E_k \subset \mathbb{R}^n$ ,  $F_1, \dots, F_k \subset \mathbb{R}^n$  such that*

- *the  $E_i$  are pairwise disjoint, and the  $F_i$  are pairwise disjoint;*
- *$\bigcup E_i = U$  and  $\bigcup F_i = V$ ;*
- *for each  $j$ ,  $E_j$  is congruent to  $F_j$ .*

In particular, we can cut a small set  $U$  into finitely many pieces and rearrange them to build a very large open set  $V$ . The consequence for us is that we cannot have some  $\mu : 2^{\mathbb{R}^n} \rightarrow [0, \infty]$  which assigns positive values to bounded open sets and satisfies (i) for finite sequences of sets. We will ultimately restrict the domain in our definition of a measure.

## 2.2 $\sigma$ -Algebras

We attempt to resolve the issues found above through the  $\sigma$ -algebra.

**Definition 2.3** (Algebra). *Let  $X$  be a nonempty set. Then a non-empty collection of subsets  $\mathcal{A} \subset 2^X$  is called an algebra if*

- (i) *for a finite collection  $E_1, \dots, E_n \in \mathcal{A}$ , we have  $\bigcup E_i \in \mathcal{A}$ ;*
- (ii) *if  $E \in \mathcal{A}$ , then the complement  $E^c \in \mathcal{A}$ .*

**Definition 2.4** ( $\sigma$ -Algebra). *An algebra closed under countable unions is called a  $\sigma$ -algebra.*

**Remark 2.5.** *Some obvious statements are:*

- *any algebra contains  $\emptyset$  and  $X$  itself;*
- *any algebra (resp.  $\sigma$ -algebra) is closed under finite (resp. countable) intersections;*
- *given the presence of condition (ii) in the definition of algebra, we can actually relax the additional condition of  $\sigma$ -algebra to hold for only disjoint unions. This can be seen by letting*

$$F_k := E_k \setminus \left( \bigcup_{i < k} E_i \right) = E_k \cap \left( \bigcup_{i < k} E_i \right)^c = \left( E_k^c \cup \left( \bigcup_{i < k} E_i \right) \right)^c.$$

*We can construct all  $F_k$  like so to be disjoint, and then  $\bigsqcup F_k = \bigcup E_k$ .*

**Lemma 2.6.** *Let  $\mathcal{A}$  be an algebra over  $X$ . Then the following statements are equivalent.*

- (i) *For any countable sequence of sets  $E_1, E_2, \dots \in \mathcal{A}$ , we have  $\bigcup E_i \in \mathcal{A}$ .*

(ii) For any countable sequence of disjoint sets,  $\bigsqcup E_i \in \mathcal{A}$ .

(iii) For any countable increasing sequence of sets (meaning  $E_i \subset E_{i+1}$ ), we have  $\bigcup E_i \in \mathcal{A}$ .

*Proof.* Homework 1. □

**Example 2.7.** (i) Let  $X$  be any set. Then  $2^X$  and  $\{\emptyset, X\}$  are  $\sigma$ -algebras.

(ii) If  $\{\mathcal{A}_\alpha\}_{\alpha \in A}$  is a family of  $\sigma$ -algebras, then  $\bigcap_\alpha \mathcal{A}_\alpha$  is also a  $\sigma$ -algebra.

**Proposition 2.8.** If  $\mathcal{E} \subset 2^X$ , there exists a unique, minimal  $\sigma$ -algebra  $\mathcal{M}(\mathcal{E})$  containing  $\mathcal{E}$ , which is obtained by taking intersection of all  $\sigma$ -algebras containing  $\mathcal{E}$ .  $\mathcal{M}(\mathcal{E})$  is the  $\sigma$ -algebra generated by  $\mathcal{E}$ .

**Lemma 2.9.** If  $\mathcal{E} \subset \mathcal{M}(\mathcal{F})$ , then  $\mathcal{M}(\mathcal{E}) \subset \mathcal{M}(\mathcal{F})$ .

*Proof.* Note that  $\mathcal{M}(\mathcal{F})$  is a  $\sigma$ -algebra containing  $\mathcal{E}$ , but  $\mathcal{M}(\mathcal{E})$  is the intersection of all  $\sigma$ -algebras containing  $\mathcal{E}$ . □

[Dutch: can be constructed with transfinite induction]

### 2.2.1 Metric Spaces

**Definition 2.10** (Borel  $\sigma$ -algebra). Let  $(X, \varrho)$  be a metric space. Then the  $\sigma$ -algebra generated by all open sets in  $X$  is called the Borel  $\sigma$ -algebra on  $X$ . We denote this algebra  $\mathcal{B}_X$ , and call its members Borel sets. This is defined, more generally, for any topological space  $(X, \zeta)$ .

**Remark 2.11.** By definition of the  $\sigma$ -algebra, the Borel  $\sigma$ -algebra will contain all open sets and closed sets, thus countable intersections of open sets, and countable unions of closed sets.

**Definition 2.12.** For concinnity, we define the following.

- A  $G_\delta$  set is a countable intersection of open sets.
- An  $F_\sigma$  set is a countable union of closed sets.
- A  $G_{\delta,\sigma}$  set is a countable union of  $G_\delta$  sets.
- An  $F_{\sigma,\delta}$  set is a countable intersection of  $F_\sigma$  sets.

**Proposition 2.13.** The Borel  $\sigma$ -algebra on  $\mathbb{R}$  is generated by each of the following:

- (i) the open intervals  $\mathcal{E}_1 = \{(a, b) : a < b\}$ ,
- (ii) the closed intervals  $\mathcal{E}_2 = \{[a, b] : a < b\}$ ,
- (iii) the half-open intervals  $\mathcal{E}_3 = \{(a, b] : a < b\}$  or  $\mathcal{E}_4 = \{[a, b) : a < b\}$ ,
- (iv) the open rays  $\mathcal{E}_5 = \{(a, \infty) : a \in \mathbb{R}\}$  or  $\mathcal{E}_6 = \{(-\infty, a) : a \in \mathbb{R}\}$ ,
- (v) the closed rays  $\mathcal{E}_7 = \{[a, \infty) : a \in \mathbb{R}\}$  or  $\mathcal{E}_8 = \{(-\infty, a] : a \in \mathbb{R}\}$ .

*Proof.* We prove (i). We first show that (1)  $\mathcal{E}_i \subset \mathcal{B}_{\mathbb{R}}$ , so  $\mathcal{M}(\mathcal{E}_i) \subset \mathcal{B}_{\mathbb{R}}$ . Then we show that (2) all open sets are contained in  $\mathcal{M}(\mathcal{E}_i)$ . To see (1) for  $\mathcal{E}_1$ , note that all sets in  $\mathcal{E}_1$  are open. To see (2), note that all open sets in  $\mathbb{R}$  can be written as a countable union of disjoint open intervals.

We prove (ii). All elements of  $\mathcal{E}_2$  are closed and their complements are open. Hence,  $\mathcal{E}_2 \subset \mathcal{B}_{\mathbb{R}}$ , and  $\mathcal{M}(\mathcal{E}_2) \subset \mathcal{B}_{\mathbb{R}}$ . On the other hand, any open interval  $(a, b)$  can be written

$$(a, b) = \bigcup_{n \in \mathbb{N}} \left[ a + \frac{1}{n}, b - \frac{1}{n} \right].$$

Hence,  $\mathcal{E}_1 \subset \mathcal{E}_2$ , so by (i),  $\mathcal{B}_{\mathbb{R}} = \mathcal{M}(\mathcal{E}_1) \subset \mathcal{M}(\mathcal{E}_2)$ .

We prove (iii). Note that  $\mathcal{E}_3, \mathcal{E}_4$  consist of  $G_{\delta}$  sets

$$(a, b] = \bigcap_{n \in \mathbb{N}} \left( a, b + \frac{1}{n} \right).$$

On the other hand,

$$(a, b) = \bigcup_{n \in \mathbb{N}} \left( a, b - \frac{1}{n} \right].$$

The rest is left as an exercise. □

**Remark 2.14.** We used the fact that any open subset of  $\mathbb{R}$  can be written as a countable disjoint union of open intervals.

*Proof.* Let  $U$  be open in  $\mathbb{R}$  and  $x \in U$ . Let  $I_x$  be the largest open interval such that  $x \in I_x \subset U$ . That is, take  $I_x = (a_x, b_x)$  for  $a_x = \inf\{a < x : (a, x) \subset U\}$  and  $b_x = \sup\{b > x : (x, b) \subset U\}$ . It is clear that

$$U = \bigcup_{x \in U} I_x,$$

but we also claim this can be written as a disjoint union. Suppose  $I_x \cap I_y \neq \emptyset$ . Then  $I_x \cup I_y \subset U$  is an open interval in  $U$ . Since  $I_x$  is maximal,  $I_x = I_x \cap I_y = I_y$ . That is, any two distinct intervals  $I_x$  must be disjoint. Choosing  $\mathcal{I}$  to be the collection of distinct  $I_x$ , we have

$$U = \bigsqcup_{I_x \in \mathcal{I}} I_x.$$

Moreover, since the intervals are not singletons, they must contain rationals. But since  $\mathbb{Q}$  is countable,  $\mathcal{I}$  is countable. □

### 2.2.2 Product $\sigma$ -Algebras

We will want to understand  $\mathbb{R}^n$ , not just  $\mathbb{R}$ . This motivates a definition for the product algebra.

**Definition 2.15.** Let  $\{X_{\alpha}\}_{\alpha \in A}$  be an indexed collection of nonempty sets. We define the product of sets  $X = \prod_{\alpha \in A} X_{\alpha}$  as the set of mappings  $\psi : A \rightarrow \bigcup_{\alpha \in A} X_{\alpha}$  such that  $\psi(\alpha) \in X_{\alpha}$  for all  $\alpha$ . We also define  $\pi_{\alpha} : X \rightarrow X_{\alpha}$  to be the projection map, sending  $\psi$  to  $\psi(\alpha)$ .



If  $\mathcal{M}_\alpha$  is a  $\sigma$ -algebra on  $X_\alpha$  for each  $\alpha \in A$ , then the product  $\sigma$ -algebra on  $\prod_{\alpha \in A} X_\alpha$  is the  $\sigma$ -algebra generated by the set

$$\{\pi_\alpha^{-1}(E_\alpha) : E_\alpha \in \mathcal{M}_\alpha, \alpha \in A\}.$$

We denote this  $\sigma$ -algebra by  $\bigotimes_{\alpha \in A} \mathcal{M}_\alpha$ .

**Proposition 2.16.** *If  $A$  is countable, then  $\bigotimes_{\alpha \in A} \mathcal{M}_\alpha$  is the  $\sigma$ -algebra generated by*

$$G = \left\{ \prod_{\alpha \in A} E_\alpha : E_\alpha \in \mathcal{M}_\alpha \right\}.$$

*Proof.* We want to show

$$\bigotimes_{\alpha \in A} \mathcal{M}_\alpha = \mathcal{M}(G).$$

If  $E_\alpha \in \mathcal{M}_\alpha$ , then  $\pi_\alpha^{-1}(E_\alpha) = \prod_{\beta \in A} E_\beta$  such that  $E_\beta = X_\beta$  if  $\beta \neq \alpha$ . This is contained in  $\mathcal{M}(G)$ . On the other hand, note that

$$\prod_{\alpha \in A} E_\alpha = \bigcap_{\alpha \in A} \pi_\alpha^{-1}(E_\alpha).$$

Since  $\bigotimes_{\alpha \in A} \mathcal{M}_\alpha$  is closed under countable intersection, we have  $\prod_{\alpha \in A} E_\alpha \in \bigotimes_{\alpha \in A} \mathcal{M}_\alpha$ . It follows that  $G \subset \bigotimes_{\alpha \in A} \mathcal{M}_\alpha$  and hence  $\mathcal{M}(G) \subset \bigotimes_{\alpha \in A} \mathcal{M}_\alpha$ .  $\square$

**Proposition 2.17.** *Suppose  $\mathcal{M}_\alpha$  is generated by  $\mathcal{E}_\alpha$ ,  $\alpha \in A$ . Then  $\bigotimes_{\alpha \in A} \mathcal{M}_\alpha$  is generated by*

$$\mathcal{F}_1 = \{\pi_\alpha^{-1}(E_\alpha) : E_\alpha \in \mathcal{E}_\alpha, \alpha \in A\}.$$

*And if  $A$  is countable and  $X_\alpha \in \mathcal{E}_\alpha$  for all  $\alpha \in A$ , then  $\bigotimes_{\alpha \in A} \mathcal{M}_\alpha$  is generated by*

$$\mathcal{F}_2 = \left\{ \prod_{\alpha \in A} E_\alpha : E_\alpha \in \mathcal{E}_\alpha \right\}.$$

*Proof.* To see the first point, note that

$$\mathcal{F}_1 \subset \{\pi_\alpha^{-1}(E_\alpha) : E_\alpha \in \mathcal{M}_\alpha, \alpha \in A\},$$

so  $\mathcal{M}(\mathcal{F}_1) \subset \bigotimes_{\alpha \in A} \mathcal{M}_\alpha$ .

Next, for each  $\alpha \in A$ , one can check that

$$Y = \{E \subset X_\alpha : \pi_\alpha^{-1}(E) \in \mathcal{M}(\mathcal{F}_1)\}$$

is a  $\sigma$ -algebra on  $X_\alpha$  (easy check), and it contains  $\mathcal{E}_\alpha$ . Therefore  $Y \supset \mathcal{M}_\alpha$ . Thus,  $\pi_\alpha^{-1}(E) \in \mathcal{M}(\mathcal{F}_1)$  for all  $E \in \mathcal{M}_\alpha, \alpha \in A$ . Hence,  $\bigotimes_{\alpha \in A} \mathcal{M}_\alpha \subset \mathcal{M}(\mathcal{F}_1)$ .

The second claim follows from the first, as it now suffices to show  $\mathcal{M}(\mathcal{F}_1) = \mathcal{M}(\mathcal{F}_2)$ . If we let  $\pi_\alpha^{-1}(E_\alpha) \in \mathcal{F}_1$ , then  $\pi_\alpha^{-1}(E_\alpha) = \prod_{\beta \in A} E_\beta$  where  $E_\beta = X_\beta$  for all  $\beta \neq \alpha$ , and  $X_\beta \in \mathcal{E}_\beta$  by assumption. In particular,  $\mathcal{F}_1 \subset \mathcal{F}_2$ , so  $\mathcal{M}(\mathcal{F}_1) \subset \mathcal{M}(\mathcal{F}_2)$ . Next, let  $\prod E_\alpha \in \mathcal{F}_2$ , and note that  $\prod E_\alpha = \bigcap \pi_\alpha^{-1}(E_\alpha)$ . Since  $\mathcal{M}(\mathcal{F}_1)$  is closed under countable intersections, and  $A$  is countable,  $\prod E_\alpha \in \mathcal{M}(\mathcal{F}_1)$ .  $\square$

**Lemma 2.18.** *Let  $(X, \varrho)$  be a separable metric space. Any open set  $O \subset X$  can be written as a countable union of open balls with rational radii.*

**Proposition 2.19.** *Let  $X_1, \dots, X_n$  be metric spaces and  $X = \prod_{j=1}^n X_j$  equipped with the product metric. Then  $\bigotimes_j \mathcal{B}_{X_j} \subset \mathcal{B}_X$ . Moreover, if each  $X_j$  is separable, then  $\bigotimes_j \mathcal{B}_{X_j} = \mathcal{B}_X$ .*

*Proof.* See [Stack Exchange](#). □

*Proof.* By the most recent proposition,  $\bigotimes \mathcal{B}_{X_j}$  is generated by the sets  $\pi_j^{-1}(U_j)$  where  $U_j$  is an open set in  $X_j$ . Since these sets are open in the product topology (which is induced by the product metric),  $\bigotimes \mathcal{B}_{X_j} \subset \mathcal{B}_X$ .

Next, suppose that for each  $j$ ,  $C_j$  is a countable dense set in  $X_j$ . Moreover, let  $\mathcal{E}_j$  denote the set of balls in  $X_j$  with rational radius and center in  $C_j$ . Every open set in  $X_j$  is a union of members of  $\mathcal{E}_j$ —in fact a countable union since  $\mathcal{E}_j$  itself is countable. It follows that  $\mathcal{B}_{X_j} = \mathcal{M}(\mathcal{E}_j)$ . Since  $[n]$  is countable, we have that

$$\bigotimes \mathcal{B}_{X_j} = \mathcal{M} \left( \left\{ \prod_{j=1}^n E_j : E_j \in \mathcal{E}_j \cup X_j \right\} \right) = \mathcal{M} \left( \left\{ \prod_{j=1}^n E_j : E_j \in \mathcal{E}_j \right\} \right).$$

In the remainder of the proof, we show that  $\mathcal{B}_X$  is generated by the same set.

Note that the set of points in  $X$  whose  $j$ th coordinate is in  $C_j$  for all  $j$  is a countable dense subset of  $X$ , and the balls of radius  $r$  in the  $X$  are products of balls of radius  $r$  in the  $X_j$ 's. The set of open balls can then be generated by countable unions in  $\{\prod_{j=1}^n E_j : E_j \in \mathcal{E}_j\}$ , so  $\mathcal{B}_X$  is generated by this set. □

**Corollary 2.20.**  $\mathcal{B}_{\mathbb{R}^n} = \bigotimes_{i=1}^n \mathcal{B}_{\mathbb{R}}$ .

**Definition 2.21.** A collection  $\mathcal{E} \subset 2^X$  is an elementary family if

- (i)  $\emptyset \in \mathcal{E}$ ,
- (ii) if  $E, F \in \mathcal{E}$ , then  $E \cap F \in \mathcal{E}$ ,
- (iii) if  $E \in \mathcal{E}$ , then  $E^c$  is a finite disjoint union of members of  $\mathcal{E}$ .

**Proposition 2.22.** *If  $\mathcal{E}$  is an elementary family, the collection  $\mathcal{A}$  of finite disjoint unions of members of  $\mathcal{E}$  is an algebra.*

*Proof.* Let  $A, B \in \mathcal{E}$ . Since  $B^c = \sqcup_{j=1}^J C_j$  for  $C_j$  in  $\mathcal{E}$ , then

$$A \cup B = (A \setminus B) \sqcup B = \left( \bigcup A \cap C_j \right) \sqcup B,$$

which is a finite disjoint union of members in  $\mathcal{E}$ . So  $A \cup B \in \mathcal{A}$ . This argument may be extended to show that any finite union of sets in  $\mathcal{E}$  is in  $\mathcal{A}$ . It follows that  $\mathcal{A}$  is closed under finite union.

Now let  $A = \bigcup_{i=1}^n A_i \in \mathcal{A}$  for disjoint  $\{A_i\}_{i=1}^n \subset \mathcal{E}$ . We have that  $A_m^c = \sqcup_{j=1}^{J_m} B_m^j$ , and the collections  $\{B_m^j\}_{j=1}^{J_m}$  have disjoint members of  $\mathcal{E}$ . Then

$$A^c = \left( \bigcup_{m=1}^n A_m \right)^c = \bigcap_{m=1}^n \left( \bigcup_{j=1}^{J_m} B_m^j \right) = \bigcup \left\{ B_1^{j_1} \cap \dots \cap B_n^{j_n} : j_m \in [J_m], \forall m \in [n] \right\}.$$

That is,  $A^c \in \mathcal{A}$ . □

## 2.3 Measures

**Definition 2.23.** Let  $X$  be a set and  $\mathcal{M}$  be a  $\sigma$ -algebra on  $X$ . A measure on  $\mathcal{M}$ , or  $(X, \mathcal{M})$ , is a function  $\mu : \mathcal{M} \rightarrow [0, \infty]$  such that

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

(ii) for a countable disjoint sequence  $\{E_i\}_{i=1}^{\infty} \subset \mathcal{M}$ , it holds that  $\mu(\cup_i E_i) = \sum_i \mu(E_i)$ .

**Definition 2.24.** The set  $(X, \mathcal{M})$  is called a measurable space, and the sets in  $\mathcal{M}$  are called measurable sets. The tuple  $(X, \mathcal{M}, \mu)$  is called a measure space. We say that  $\mu$  is a finite measure if  $\mu(X) < \infty$ ; finiteness implies that for all  $E \in \mathcal{M}$ ,  $\mu(E) < \infty$ . If  $\mu(X) = 1$ , then  $\mu$  is called a probability measure.

**Remark 2.25.** Some important points.

(i) We call property (ii) countable additivity.

(ii) If  $E_1, \dots, E_n$  are disjoint sets in  $\mathcal{M}$ , then  $\mu(\cup_{i=1}^n E_i) = \sum_{i=1}^n \mu(E_i)$ . This is called finite additivity, and is implied by countable additivity, since  $E = E \cup \emptyset \cup \emptyset \cup \dots$ .

(iii) If  $X = \bigcup_{j=1}^{\infty} E_j$  for  $E_j \in \mathcal{M}$  and  $\mu(E_j) < \infty$  for all  $j$ , then  $\mu$  is called  $\sigma$ -finite. For example,  $\mathbb{R}$  is  $\sigma$ -finite.

(iv) Suppose that for each  $E \in \mathcal{M}$  such that  $\mu(E) = \infty$ , there exists an  $F \in \mathcal{M}$  with  $F \subset E$  and  $0 < \mu(F) < \infty$ . If  $\mu$  satisfies this property, we call it semifinite.

**Example 2.26 (Measures).** • Let  $X$  be an infinite set and  $\mathcal{M} = 2^X$ . Define

$$\mu(E) = \begin{cases} 0 & E \text{ finite} \\ \infty & E \text{ infinite.} \end{cases}$$

This is finitely additive, but not countably additive. So  $\mu$  is a finitely additive measure but not a measure by our definition.

• Take  $X$  to be any nonempty set,  $\mathcal{M} = 2^X$ . Consider any function  $f : X \rightarrow [0, \infty]$ , and define

$$\mu(E) = \sum_{x \in E} f(x) = \sup \left\{ \sum_{x \in F} f(x) : F \subset E, F \text{ finite} \right\}.$$

Also set  $\mu(\emptyset) = 0$ . Then  $\mu$  is a measure. As an exercise, one should check that (i)  $\mu$  is semifinite if and only if  $f(x) < \infty$  for all  $x \in X$ , and (ii)  $\mu$  is  $\sigma$ -finite if and only if  $\mu$  is semifinite and  $\{x : f(x) > 0\}$  is countable.

• Thinking more about the second example above, if we let  $f(x) \equiv 1$ , then  $\mu$  is called the counting measure. If we have

$$f(x) = \begin{cases} 1 & x = x_0 \\ 0 & x \neq x_0, \end{cases}$$

then

$$\mu(E) = \begin{cases} 1 & x_0 \in E \\ 0 & \text{else.} \end{cases}$$

This measure is called the point mass, or Dirac measure at  $x_0$ .

**Theorem 2.27.** Let  $(X, \mathcal{M}, \mu)$  be a measure space. Then

(i) (Monotonicity) If  $E, F \in \mathcal{M}$  and  $E \subset F$ , then  $\mu(E) \leq \mu(F)$ .

(ii) (Subadditivity) If  $\{E_j\}_{j=1}^{\infty} \subset \mathcal{M}$ , then

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

(iii) (Continuity from below) If  $\{E_j\}_{j=1}^{\infty} \subset \mathcal{M}$  and  $E_1 \subset E_2 \subset \cdots$ , then

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

(iv) (Continuity from above) If  $\{E_j\}_{j=1}^{\infty} \subset \mathcal{M}$  and  $E_1 \supset E_2 \supset \cdots$  and  $\mu(E_1) < \infty$ , then

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

*Proof.* (i) Write  $F = E \sqcup (F \setminus E)$ . Then  $\mu(F) = \mu(E) + \mu(F \setminus E) \geq \mu(E)$ .

(ii) Let  $F_1 = E_1$ ,  $F_k = E_k \setminus \bigcup_{j=1}^{k-1} E_j \subset E_k$ . By construction, all  $F_k$  are disjoint. Hence,

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

(iii) Suppose we have an increasing sequence  $\{E_j\}_{j=1}^{\infty}$ . Using a similar approach, and taking  $E_0 = \emptyset$ , we have

$$\begin{aligned} \mu\left(\bigcup_{j=1}^{\infty} E_j\right) &= \mu\left(\bigcup_{j=1}^{\infty} (E_j \setminus E_{j-1})\right) = \sum_{j=1}^{\infty} \mu(E_j \setminus E_{j-1}) \\ &= \lim_{n \rightarrow \infty} \left( \sum_{j=1}^n \mu(E_j \setminus E_{j-1}) \right) \\ &= \lim_{n \rightarrow \infty} \mu(E_n). \end{aligned}$$

- (iv) Let  $F_j = E_1 \setminus E_j$ , so that we obtain an increasing sequence  $F_1 \subset F_2 \subset \dots$ , and because  $E_j \subset E_1$ , we have  $\mu(E_1) = \mu(F_j) + \mu(E_j)$ . We have

$$\bigcup_{j=1}^{\infty} F_j = E_1 \setminus \bigcap_{j=1}^{\infty} E_j, \quad \text{and} \quad E_1 = \left( \bigcap_{j=1}^{\infty} E_j \right) \sqcup \left( \bigcup_{j=1}^{\infty} F_j \right).$$

That is,

$$\mu(E_1) = \mu \left( \bigcap_{j=1}^{\infty} E_j \right) + \mu \left( \bigcup_{j=1}^{\infty} F_j \right),$$

and by (iii),

$$\mu \left( \bigcup_{j=1}^{\infty} F_j \right) = \lim_{n \rightarrow \infty} \mu(F_n) = \lim_{n \rightarrow \infty} (\mu(E_1) - \mu(E_n)).$$

Since  $\mu(E_1) < \infty$ ,

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

□

**Remark 2.28.** It turns out that the assumption  $\mu(E_1) < \infty$  is necessary in (iv). Consider the counting measure in  $\mathbb{R}$  and  $E_j = B_{1/j}(0)$ . Then

$$\bigcap_{j=1}^{\infty} E_j = \{0\}, \quad \mu(\{0\}) = 1.$$

but  $\mu(E_j) = \infty$  for all  $j$ .

**Definition 2.29.** (i) If  $(X, \mathcal{M}, \mu)$  is a measure space, a set  $E \in \mathcal{M}$  with  $\mu(E) = 0$  is called a null set. By subadditivity, a countable union of null sets is a null set.

(ii) If a statement about points  $x \in X$  is true except for the set  $S$  such that  $S \subset N$  for a null set  $N$ , we say the statement is true almost everywhere (or a.e.). To be more specific, we call  $N$  a  $\mu$ -null set or say that the statement holds  $\mu$ -almost-everywhere.

**Example 2.30.**  $f(x) = 1/x$  is defined a.e. on  $\mathbb{R}$  with respect to the Lebesgue measure.

**Definition 2.31.** A measure whose domain includes all subset of null sets is called complete.

**Theorem 2.32.** Let  $(X, \mathcal{M}, \mu)$  be a measure space. Let

$$\mathcal{N} = \{N \in \mathcal{M} : \mu(N) = 0\}, \quad \overline{\mathcal{M}} = \{E \cup F : E \in \mathcal{M}, F \subset N \in \mathcal{N}\}.$$

Then  $\overline{\mathcal{M}}$  is a  $\sigma$ -algebra, and there exists a unique extension  $\bar{\mu}$  of  $\mu$  to a complete measure on  $\overline{\mathcal{M}}$ . We call  $\bar{\mu}$  the completion of  $\mu$ .

*Proof.* Closure under countable union is straightforward. To see complements, let  $E \cup F \in \overline{\mathcal{M}}$ , for  $F \subset N$ . We may assume  $E \cap F = \emptyset$ , just by taking the set difference. So we have  $E \cap N^c = E$  and  $N \cap F = F$ . Then

$$E \cup F = (E \cup N) \cap (N^c \cup F).$$

So

$$(E \cup F)^c = (E \cup N)^c \cup (N \setminus F),$$

and this lies in  $\overline{\mathcal{M}}$ .

Choose any  $E \cup F \in \overline{\mathcal{M}}$ . We define the extension  $\bar{\mu}$  to be

$$\bar{\mu}(E \cup F) = \mu(E), \quad \text{for each } E \cup F \in \overline{\mathcal{M}}.$$

Of course, the representation for the set  $E \cup F$  is not unique, so we need to check if this is well-defined. Suppose  $E_1 \cup F_1 = E_2 \cup F_2 \in \overline{\mathcal{M}}$ ,  $F_1 \subset N_1$  and  $F_2 \subset N_2$ . Clearly  $E_1 \subset E_2 \cup F_2 \subset E_2 \cup N_2$ . So

$$\mu(E_1) \leq \mu(E_2) + \mu(N_2) = \mu(E_2).$$

This goes in both directions, so  $\bar{\mu}$  is well-defined. It remains to show that  $\bar{\mu}$  is complete and unique (Homework 2).  $\square$

## 2.4 Outer Measures

We are interested in finding a practical way to construct measures. We also want some approximate way to measure sets that is easy to use. We address both desires with the notion of an outer measure.

**Definition 2.33** (Outer measure). *An outer measure on a nonempty set  $X$  is a function  $\mu^* : 2^X \rightarrow [0, \infty]$  such that*

- (i)  $\mu^*(\emptyset) = 0$ ,
- (ii)  $\mu^*(A) \leq \mu^*(B)$  if  $A \subset B$ ,
- (iii)  $\mu^*\left(\bigcup_{j=1}^{\infty} A_j\right) \leq \sum_{j=1}^{\infty} \mu^*(A_j)$ .

**Proposition 2.34.** *Let  $\mathcal{E} \subset 2^X$  and  $\varrho : \mathcal{E} \rightarrow [0, \infty]$  be such that (i)  $\emptyset \in \mathcal{E}$ , (ii)  $X \in \mathcal{E}$ , (iii)  $\varrho(\emptyset) = 0$ . For any subset  $A \subset X$ , define*

$$\mu^*(A) := \inf \left\{ \sum_{j=1}^{\infty} \varrho(E_j) : E_j \in \mathcal{E}, A \subset \bigcup_{j=1}^{\infty} E_j \right\}.$$

*Then  $\mu^*$  is an outer measure.*

*Proof.* First note that for any  $A \subset X$ , there exists a countable sequence  $\{E_j\}_{j=1}^{\infty} \subset \mathcal{E}$  such that  $A \subset \bigcup E_j$  because  $X \in \mathcal{E}$ . So the infimum exists and  $\mu^*$  is well-defined. (1) Note that  $\mu^*(\emptyset) = 0$  since we can cover  $\emptyset$  by taking  $E_j = \emptyset$  for all  $j$ . (2) Let  $A \subset B$ . Then any collection  $E_j \in \mathcal{E}$  with  $B \subset \bigcup E_j$  also satisfies

$A \subset \bigcup E_j$ , so  $\mu^*(A) \leq \mu^*(B)$ . (3) Let  $\{A_j\}_{j=1}^\infty \subset 2^X$  and  $\varepsilon > 0$ . For each  $j$  there exists  $\{E_k^j\}_{k=1}^\infty \subset \mathcal{E}$  such that  $A_j \subset \bigcup_k E_k^j$  and  $\sum_k \varrho(E_k^j) \leq \mu^*(A_j) + \varepsilon 2^{-j}$ . It follows that  $\bigcup_j A_j \subset \bigcup_{j,k} E_k^j$ , and

$$\mu^*\left(\bigcup_j A_j\right) \leq \sum_{j,k} \varrho(E_k^j) \leq \sum_j \mu^*(A_j) + \varepsilon.$$

Since  $\varepsilon$  is arbitrary, this completes the proof.  $\square$

Given a function  $\varrho : \mathcal{E} \rightarrow [0, \infty]$ , we have seen that we can construct an outer measure  $\mu^* : 2^X \rightarrow [0, \infty]$ . Now, starting with the outer measure, we want to come up with a measure  $\mu$ . We start by finding a  $\sigma$ -algebra  $\mathcal{A}$  such that the restriction  $\mu = \mu^*|_{\mathcal{A}} : \mathcal{A} \rightarrow [0, \infty]$  forms a measure.

**Definition 2.35** ( $\mu^*$ -measurability). *Given an outer measure  $\mu^* : 2^X \rightarrow [0, \infty]$ , we call  $A \subset X$  a  $\mu^*$ -measurable set if it holds that for all  $E \subset X$ ,*

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

**Remark 2.36.** *We spend some time thinking about how to show  $\mu^*$ -measurability and what the implications of this notion are.*

- For any  $A$  and  $E$ , since  $E \subset (E \cap A) \cup (E \cap A^c)$ , properties of the outer measure give that

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

*It therefore suffices to show the reverse inequality to prove that  $A$  is  $\mu^*$ -measurable. If  $\mu^*(E) = \infty$ , then the reverse inequality is trivial. Hence, we see that  $A$  is  $\mu^*$ -measurable if and only if*

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

*for all  $E \subset X$  with  $\mu^*(E) < \infty$ .*

- Suppose  $A$  is  $\mu^*$ -measurable, and say  $A \subset E$ . Then  $E \cap A$  and  $E \cap A^c$  form a disjoint partition of  $E$ . In particular we have that

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

*That is,  $\mu^*(A) = \mu^*(E) - \mu^*(E \cap A^c)$ .*

- Suppose  $A, B$  are disjoint and  $A$  is  $\mu^*$ -measurable. Then by taking  $E = A \sqcup B$ , we have

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

**Theorem 2.37** (Caratheodory). *Let  $\mu^*$  be an outer measure on  $X$ . The collection  $\mathcal{M}$  of  $\mu^*$ -measurable sets is a  $\sigma$ -algebra, and the restriction  $\mu^*|_{\mathcal{M}}$  is a complete measure on  $\mathcal{M}$ .*

*Proof.* As a first step, we show  $\mathcal{M}$  is a  $\sigma$ -algebra. The definition of  $\mu^*$ -measurable is symmetric in  $A, A^c$ , so  $A \in \mathcal{M}$  implies  $A^c \in \mathcal{M}$ . To show closure under finite union, take  $A, B \in \mathcal{M}$  and  $E \subset X$ . Then

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

where the last step follows because  $A \cup B = (A \cap B) \cup (A \cap B^c) \cup (A^c \cap B)$ . Therefore,

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

and  $A \cup B \in \mathcal{M}$ .

The second step is to show that  $\mathcal{M}$  is a  $\sigma$ -algebra. By our previous arguments, it's sufficient to show  $\mathcal{M}$  is closed under disjoint countable unions. Let  $\{A_j\}_{j=1}^\infty$  be a countable sequence of disjoint sets in  $\mathcal{M}$  and define

$$B_n = \bigcup_{j=1}^n A_j,$$

so  $\{B_n\}_{n=1}^\infty$  is increasing. Moreover,  $B := \bigcup_{j=1}^\infty A_j = \bigcup_{n=1}^\infty B_n$ . Consider a test set  $E \subset X$ . Since  $A_j$  is  $\mu^*$ -measurable,

$$\begin{aligned}\mu^*(E \cap B_n) &= \mu^*(E \cap B_n \cap A_n) + \mu^*(E \cap B_n \cap A_n^c) \\ &= \mu^*(E \cap A_n) + \mu^*(E \cap B_{n-1}) \\ &= \mu^*(E \cap A_n) + \mu^*(E \cap A_{n-1}) + \mu^*(E \cap B_{n-2}) \\ &= \sum_{j=1}^n \mu^*(E \cap A_j).\end{aligned}$$

Next, we note that

$$\begin{aligned}\mu^*(E) &= \mu^*(E \cap B_n) + \mu^*(E \cap B_n^c) \\ &= \sum_{j=1}^n \mu^*(E \cap A_j) + \mu^*(E \cap B_n^c) \\ &\geq \sum_{j=1}^n \mu^*(E \cap A_j) + \mu^*(E \cap B^c).\end{aligned}$$

Since  $n$  was arbitrary, we can take the limit on the RHS to obtain

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



The intermediate inequalities are therefore equalities, and we have  $B = \bigcup_{j=1}^{\infty} A_j \in \mathcal{M}$ .

The third step is to show that  $\mu^*|_{\mathcal{M}}$  is a complete measure. In the previous step, take  $E = B$ . Then

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

That is,  $\mu^*$  is countably additive, and a measure. It remains to show completeness. Let  $A \subset X$  with  $\mu^*(A) = 0$ . For any  $E \subset X$ , monotonicity implies

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

All inequalities must be equalities, so  $A \in \mathcal{M}$ . That is, any set whose outer measure is zero is included in  $\mathcal{M}$ . That is,  $\mathcal{M}$  includes all possible null sets and  $\mu^*|_{\mathcal{M}}$  is a complete measure.  $\square$

## 2.5 Premeasures

**Definition 2.38.** Let  $\mathcal{A} \subset 2^X$  be an algebra (not nec. a  $\sigma$ -algebra). Let  $\mu_0 : \mathcal{A} \rightarrow [0, \infty]$  satisfy

(i)  $\mu_0(\emptyset) = 0$ ,

(ii) for a sequence  $\{A_i\}_{i=1}^{\infty}$  of disjoint sets and  $\bigcup_i A_i \in \mathcal{A}$ , then  $\mu_0(\bigcup_i A_i) = \sum_i \mu_0(A_i)$ .

If both of these conditions are satisfied, we call  $\mu_0$  a premeasure.

**Remark 2.39.** Every premeasure is finitely additive. Take  $A_i = \emptyset$  for  $i$  large.

**Proposition 2.40.** If  $\mu_0$  is a premeasure on an algebra  $\mathcal{A}$ . Define the outer measure

$$\mu^*(E) = \inf \left\{ \sum_{j=1}^{\infty} \mu_0(A_j) : A_j \in \mathcal{A}, E \subset \bigcup_{j=1}^{\infty} A_j \right\}.$$

Then the following hold.

(i)  $\mu^*|_{\mathcal{A}} = \mu_0$ ;

(ii) every set in  $\mathcal{A}$  is  $\mu^*$ -measurable.

*Proof.* To see the first point, let  $E \in \mathcal{A}$ . Then, for an arbitrary cover  $\{A_j\}$ , we have  $E \subset \bigcup_j A_j$ . We define

$$B_n = E \cap \left( A_n \setminus \bigcup_{j=1}^{n-1} A_j \right) \in \mathcal{A}.$$

Note that  $B_n$ 's are disjoint,  $\bigcup_n B_n = E$ , and  $B_n \subset A_n$ . Since  $\mu_0$  is a premeasure,

$$\mu_0(E) = \sum_{n=1}^{\infty} \mu_0(B_n) \leq \sum_{n=1}^{\infty} \mu_0(A_n),$$

where the last inequality uses monotonicity, which can be shown for premeasures. That is,  $\mu_0(E) \leq \mu^*(E)$ . In order to see the other inequality, consider the covering  $E \subset A_1 \in \mathcal{A}$ , for  $A_1 := E$ . We immediately see  $\mu^*(E) \leq \mu_0(E)$ .

To see the second point, let  $A \in \mathcal{A}$  and  $E \subset X$ . Take arbitrary  $\varepsilon > 0$ . By the definition of  $\mu^*(E)$ , there exists a sequence  $\{B_j\}_{j=1}^\infty \subset \mathcal{A}$  such that  $E \subset \bigcup_j B_j$  and

$$\sum_{j=1}^\infty \mu_0(B_j) \leq \mu^*(E) + \varepsilon.$$

Since  $\mu_0$  is finitely additive,

$$\mu_0(B_j) = \mu_0(B_j \cap A) + \mu_0(B_j \cap A^c).$$

Now, we have

$$\mu^*(E) + \varepsilon \geq \sum_{j=1}^\infty \mu_0(B_j) = \sum_{j=1}^\infty \mu_0(B_j \cap A) + \sum_{j=1}^\infty \mu_0(B_j \cap A^c).$$

We notice that the left and right terms introduce covers for  $E \cap A$  and  $E \cap A^c$ , respectively. Hence, we have

$$\mu^*(E) + \varepsilon \geq \mu^*(E \cap A) + \mu^*(E \cap A^c).$$

Since  $\varepsilon > 0$  is arbitrary, we have one inequality. The other is clear from the definition of outer measure.  $\square$

**Theorem 2.41.** *Let  $\mathcal{A} \subset 2^X$  be an algebra. Let  $\mu_0$  be a premeasure on  $\mathcal{A}$ . Consider the generated  $\sigma$ -algebra  $\mathcal{M} = \mathcal{M}(\mathcal{A})$ . There exists a measure  $\mu$  on  $\mathcal{M}$ , such that  $\mu|_{\mathcal{A}} = \mu_0$ . In particular,  $\mu = \mu^*|_{\mathcal{M}}$ , where*

$$\mu^*(E) = \inf \left\{ \sum_{j=1}^\infty \mu_0(A_j) : A_j \in \mathcal{A}, E \subset \bigcup_{j=1}^\infty A_j \right\}.$$

*If  $\nu$  is another measure on  $\mathcal{M}$  that extends  $\mu_0$ , then  $\nu(E) \leq \mu(E)$  for all  $E \in \mathcal{M}$  with equality when  $\mu(E) < \infty$ . If  $\mu_0$  is  $\sigma$ -finite, then  $\mu^*|_{\mathcal{M}}$  is the unique extension of  $\mu_0$  to a measure on  $\mathcal{M}$ .*

*Proof.* The first assertion follows from the previous proposition and Caratheodory's Theorem. Take  $\mu^* : 2^X \rightarrow [0, \infty]$  to be the canonical outer measure, which restricts to  $\mu_0$  on  $\mathcal{A}$ . Moreover, by the previous proposition,  $\mathcal{A}$  is contained in the set  $\mathcal{C}$  of  $\mu^*$ -measurable sets. By Caratheodory's theorem,  $\mathcal{C}$  is a  $\sigma$ -algebra and  $\mu^*|_{\mathcal{C}}$  is a complete measure on  $\mathcal{C}$ . Hence,  $\mathcal{M} = \mathcal{M}(\mathcal{A}) \subset \mathcal{C}$ , and  $\mu^*|_{\mathcal{M}}$  is a measure on  $\mathcal{M}$ , which restricts to  $\mu_0$  on  $\mathcal{A}$ .

We now focus on the second claim. Let  $E \in \mathcal{M}$  and consider a cover  $E \subset \bigcup_{j=1}^\infty A_j$  for  $A_j \in \mathcal{A}$ . Then

$$\nu(E) \leq \sum_{j=1}^\infty \nu(A_j) = \sum_{j=1}^\infty \mu_0(A_j).$$

Since  $\{A_j\}_j$  was an arbitrary cover, this gives  $\nu(E) \leq \mu(E)$ .

Now we make an interesting observation. If we set  $A = \bigcup_{j=1}^\infty A_j$ , then by continuity from below and consistency on  $\mathcal{A}$ ,

$$\nu(A) = \lim_{n \rightarrow \infty} \nu \left( \bigcup_{j=1}^n A_j \right) = \lim_{n \rightarrow \infty} \mu \left( \bigcup_{j=1}^n A_j \right) = \mu(A).$$

This observation becomes useful. If  $\mu(E) < \infty$ <sup>3</sup>, we can choose  $A_j$ 's so that  $\mu(A) < \mu(E) + \varepsilon$ , and hence  $\mu(A \setminus E) < \varepsilon$  and

$$\mu(E) \leq \mu(A) = \nu(A) = \nu(E) + \nu(A \setminus E) \leq \nu(E) + \mu(A \setminus E) \leq \nu(E) + \varepsilon.$$

Since  $\varepsilon > 0$  was arbitrary, we have  $\mu(E) = \nu(E)$ .

Finally, suppose  $X = \bigcup_{j=1}^{\infty} A_j$  with  $\mu_0(A_j) < \infty$ , where we can assume the  $A_j$ 's are disjoint. Then for any  $E \in \mathcal{M}$ , we have

$$\mu(E) = \sum_{j=1}^{\infty} \mu(E \cap A_j) = \sum_{j=1}^{\infty} \nu(E \cap A_j) = \nu(E),$$

so  $\nu = \mu$ . □

**Remark 2.42.** On sets where  $\mu(E) = \infty$ , then the extension may not be unique. We (not so rigorously) construct a counterexample here. Let  $\mathcal{A}$  be generated by  $(-\infty, 0]$  and all intervals of the form  $(a, b)$  with  $0 < a, b$ . Let  $\mu_0((a, b)) = b - a$ . One can check that this is a premeasure. Then let  $\nu((c, d)) = \int_c^d \varphi(x) dx$  with  $\varphi \in C(\mathbb{R})$ ,  $\varphi \geq 0$  and  $\varphi(x) = 1$  for  $x \geq 0$  and  $\int_{-\infty}^{\infty} \varphi(x) dx = 0$ . One can check that no matter what  $\varphi$  is, this is an extension.

## 2.6 Borel Measures on $\mathbb{R}$

We will now use some of the tools we have to construct important measures. Specifically, our goal will be to construct a measure  $\mu$  such that  $\mu((a, b)) = b - a$  on  $\mathcal{B}_{\mathbb{R}}$ . Later on, we will call one such measure the Lebesgue measure, but for now we consider a more general family of measures whose domain is  $\mathcal{B}_{\mathbb{R}}$ .

**Definition 2.43** (Borel measure). For a measure  $\mu : \mathcal{B}_{\mathbb{R}} \rightarrow [0, \infty]$ , we call  $\mu$  a Borel measure.

**Definition 2.44** ( $h$ -interval). Consider all sets of the form  $(a, b]$  or  $(a, \infty)$  or  $\emptyset$  where  $-\infty \leq a < b < \infty$ . We refer to such sets as  $h$ -intervals.

**Remark 2.45.** The intersection of two  $h$ -intervals is an  $h$ -interval. Complements of  $h$ -intervals in  $\mathbb{R}$  are either  $h$ -intervals, or a disjoint union of two  $h$ -intervals. The collection of  $h$ -intervals forms an elementary family. The collection  $\mathcal{A}$  of finite disjoint unions of  $h$ -intervals is an algebra, and  $\mathcal{M}(\mathcal{A}) = \mathcal{B}_{\mathbb{R}}$ .

**Proposition 2.46.** Let  $F : \mathbb{R} \rightarrow \mathbb{R}$  be increasing and right-continuous. If  $(a_j, b_j]$  for  $j \in \mathbb{N}$  are disjoint  $h$ -intervals, define the function  $\mu_0$  with the expression

$$\mu_0 \left( \bigcup_1^n (a_j, b_j] \right) = \sum_1^n F(b_j) - F(a_j),$$

and let  $\mu_0(\emptyset) = 0$ . We claim that  $\mu_0$  is well-defined and a premeasure on the algebra  $\mathcal{A}$ .

*Proof.* First suppose that  $\bigcup_1^n (a_j, b_j] = (a, b]$ , then we must have, after relabeling index  $j$ ,

$$a = a_1 < b_1 = a_2 < \cdots < b_n = b.$$

---

<sup>3</sup>Note how we use finiteness of  $\mu(E)$  here.

Hence,

$$\mu_0 \left( \bigcup_1^n (a_j, b_j] \right) = \sum_1^n F(b_j) - F(a_j) = F(b) - F(a) = \mu_0((a, b]).$$

Now, more generally, if  $\bigcup_1^m I_i = \bigcup_1^m J_j$  are two disjoint unions of  $h$ -intervals, we can partition both further into

$$\tilde{I}_{ij} = I_i \cap J_j,$$

so that

$$\sum_i \mu_0(I_i) = \sum_{i,j} \mu_0(\tilde{I}_{ij}) = \sum_j \mu_0(J_j).$$

So  $\mu_0$  is well-defined, and we note that by construction,  $\mu_0$  is also finitely additive.

We now aim to show that  $\mu_0$  is a premeasure. The only remaining property to prove is: if  $\{I_j\}_{j=1}^\infty$  is a countable disjoint sequence in  $\mathcal{A}$  with  $\bigcup_1^\infty I_j \in \mathcal{A}$ , then  $\mu_0(\bigcup_1^\infty I_j) = \sum_1^\infty \mu_0(I_j)$ . Since we can write  $(a, \infty)$  as a countable union of disjoint intervals  $(a, b]$ , we may assume each  $I_j$  is of the form  $(a, b]$ . By definition of  $\mathcal{A}$ , there is a finite disjoint collection of  $h$ -intervals  $\{J_i\}_{i=1}^m$  such that  $\bigcup I_j = \bigcup_{i=1}^m J_i$ . In particular, we can partition  $\{I_j\}$  into finitely many subsequences of the form  $\{J_i \cap I_j\}_{j=1}^\infty$  such that the union of the intervals in each subsequence is a single  $h$ -interval  $J_i$ . Considering each subsequence separately and using the finite additivity of  $\mu_0$ , we may assume that  $\bigcup_1^\infty I_j$  is an  $h$ -interval  $I = (a, b]$ . In this case, we have that for any  $n \in \mathbb{N}$ ,

$$\mu_0(I) = \mu_0 \left( \bigcup_1^n I_j \right) + \mu_0 \left( I \setminus \bigcup_1^n I_j \right) \geq \mu_0 \left( \bigcup_1^n I_j \right) = \sum_1^n \mu_0(I_j).$$

Letting  $n \rightarrow \infty$ , we have  $\mu_0(I) \geq \sum_1^\infty \mu_0(I_j)$ .

To prove the reverse inequality, first suppose that  $a, b$  finite, and let  $\varepsilon > 0$ . Since  $F$  is right-continuous, there exists  $\delta > 0$  such that  $F(a + \delta) - F(a) < \varepsilon$ , and if  $I_j = (a, b_j]$ , then for each  $j$  there exists  $\delta_j > 0$  such that  $F(b_j + \delta_j) - F(b_j) < \varepsilon 2^{-j}$ . The open intervals cover the compact set  $[a + \delta, b]$ , so there is a finite subcover. By discarding any  $(a_j, b_j + \delta_j)$  that is contained in a larger one and relabeling the index  $j$ , we may assume two additional conditions:

- (i) the intervals  $(a_1, b_1 + \delta_1), \dots, (a_N, b_N + \delta_N)$  cover  $[a + \delta, b]$ ,
- (ii)  $b_j + \delta_j \in (a_{j+1}, b_{j+1} + \delta_{j+1})$  for  $j = 1, \dots, N - 1$ .

It follows that

$$\begin{aligned} \mu_0(I) &\leq F(b) - F(a + \delta) + \varepsilon \\ &\leq F(b_N + \delta_N) - F(a_1) + \varepsilon \\ &= F(b_N + \delta_N) - F(a_N) + \sum_1^{N-1} (F(a_{j+1}) - F(a_j)) + \varepsilon \\ &\leq F(b_N + \delta_N) - F(a_N) + \sum_1^{N-1} (F(b_j + \delta_j) - F(a_j)) + \varepsilon \\ &< \sum_1^N (F(b_j) + \varepsilon 2^{-j} - F(a_j)) + \varepsilon < \sum_1^\infty \mu_0(I_j) + 2\varepsilon. \end{aligned}$$

Since  $\varepsilon > 0$  is arbitrary, this concludes the proof for the case where  $a$  and  $b$  are finite. When  $a = -\infty$ , for any  $M < \infty$ , the intervals  $(a_j, b_j + \delta_j)$  cover  $[-M, b]$ , so the same reasoning gives  $F(b) - F(-M) \leq \sum_1^\infty \mu_0(I_j) + 2\varepsilon$ . If  $b = \infty$ , then for any  $M < \infty$  we likewise obtain  $F(M) - F(a) \leq \sum_1^\infty \mu_0(I_j) + 2\varepsilon$ . The desired result then follows by letting  $\varepsilon \rightarrow 0$  and  $M \rightarrow \infty$ .  $\square$

**Theorem 2.47.** *If  $F : \mathbb{R} \rightarrow \mathbb{R}$  is an increasing, right continuous function, there is a unique Borel measure  $\mu_F$  on  $\mathbb{R}$  such that  $\mu_F((a, b]) = F(b) - F(a)$  for all  $a, b$ . If  $G$  is another such function, we have  $\mu_F = \mu_G$  if and only if  $F - G$  is constant. Conversely, if  $\mu$  is a Borel measure on  $\mathbb{R}$  that is finite on all bounded Borel sets and we define*

$$F(x) = \begin{cases} \mu((0, x]) & x > 0 \\ 0 & x = 0 \\ -\mu((-x, 0]) & x < 0, \end{cases}$$

*then  $F$  is increasing and right continuous and  $\mu = \mu_F$ .*

*Proof.* To start, by the previous proposition, each  $F$  induces a premeasure on the algebra  $\mathcal{A}$  of finite disjoint unions of  $h$ -intervals. Also, it is clear that  $F$  and  $G$  induce the same premeasure if and only if  $F - G$  is constant, and that these premeasures are  $\sigma$ -finite. The first two assertions therefore follow from [Theorem 2.41](#).

As for the last claim, assume  $\mu$  is a Borel measure that is finite on all bounded Borel sets. The monotonicity of  $\mu$  implies the monotonicity of  $F$ , and the continuity of  $\mu$  from above and below implies right continuity of  $F$  for  $x \geq 0$  and  $x < 0$ . Now on  $\mathcal{A}$ , it is clear that  $\mu = \mu_F$ , and hence  $\sigma$ -finiteness and the uniqueness result in [Theorem 2.41](#) imply that  $\mu = \mu_F$  on  $\mathcal{B}_{\mathbb{R}}$ .  $\square$

**Remark 2.48.** *A few points regarding the theory developed in this subsection.*

- (i) *The theory could be equally well be developed using intervals of the form  $[a, b)$  and left-continuous functions  $F$ .*
- (ii) *If  $\mu$  is a finite Borel measure on  $\mathbb{R}$ , then  $\mu = \mu_F$  where  $F(x) = \mu((-\infty, x])$ , and we call  $F$  the cumulative distribution function of  $\mu$ . This differs from the function defined in [Theorem 2.47](#) by the constant  $\mu((-\infty, 0])$ .*
- (iii) *By the Caratheodory theorem, each increasing right-continuous function  $F$  gives us, not only a Borel measure  $\mu_F$ , but a complete measure  $\bar{\mu}_F$  whose domain includes  $\mathcal{B}_{\mathbb{R}}$ . In fact,  $\bar{\mu}_F$  is just the completion of  $\mu_F$  ([Folland, 1999, Exercise 22a](#)) or [Theorem 2.51](#). One can show that the domain will always be strictly larger than that of  $\mathcal{B}_{\mathbb{R}}$ . We call  $\bar{\mu}_F$  the Lebesgue-Stieltjes measure associated to  $F$ .*

### 2.6.1 Lebesgue-Stieltjes Measure

We now investigate the properties of the Lebesgue-Stieltjes measures. For the remainder of the subsection, fix a complete Lebesgue-Stieltjes measure  $\mu$  associated to the increasing right-continuous function  $F$  and denote the domain  $\sigma$ -algebra of  $\mu$  by  $\mathcal{M}_{\mu}$ .

**Lemma 2.49.** For any  $E \in \mathcal{M}_\mu$ , then

$$\mu(E) = \inf \left\{ \sum_1^\infty \mu((a_j, b_j)) : E \subset \bigcup_1^\infty (a_j, b_j) \right\}.$$

*Proof.* By construction of the complete measure,

$$\begin{aligned} \mu(E) &= \inf \left\{ \sum_1^\infty (F(b_j) - F(a_j)) : E \subset \bigcup_1^\infty (a_j, b_j] \right\} \\ &= \inf \left\{ \sum_1^\infty \mu((a_j, b_j]) : E \subset \bigcup_1^\infty (a_j, b_j] \right\}. \end{aligned}$$

Now define

$$\nu(E) = \inf \left\{ \sum_1^\infty \mu((a_j, b_j)) : E \subset \bigcup_1^\infty (a_j, b_j) \right\}.$$

We want to show  $\mu(E) = \nu(E)$ . To see that  $\mu(E) \leq \nu(E)$ , start by supposing that  $E \subset \bigcup_1^\infty (a_j, b_j)$ . Each  $(a_j, b_j)$  can be written as a countable disjoint union of  $h$  intervals  $\{I_j^k\}_k$ . Specifically,  $I_j^k = (c_j^k, c_j^{k+1}]$  where  $\{c_j^k\}_k$  is any sequence such that  $c_j^1 = a_j$  and  $c_j^k$  increases to  $b_j$  as  $k \rightarrow \infty$ . Thus,

$$E \subset \bigcup_{j,k} I_j^k.$$

In particular, we have

$$\sum_1^\infty \mu((a_j, b_j)) = \sum_{j,k} \mu(I_j^k) \geq \mu(E).$$

Since this covering was arbitrary,  $\nu(E) \geq \mu(E)$ .

On the other hand, given  $\varepsilon > 0$ , there exists  $\{(a_j, b_j]\}_1^\infty$  where  $E \subset \bigcup_1^\infty (a_j, b_j]$  and  $\sum_1^\infty \mu((a_j, b_j]) \leq \mu(E) + \varepsilon$ . Moreover, for each  $j$ , there exists some  $\delta_j > 0$  such that  $F(b_j + \delta_j) - F(b_j) < \varepsilon 2^{-j}$ . Then  $E \subset \bigcup_1^\infty (a_j, b_j + \delta_j)$  and

$$\sum_1^\infty \mu((a_j, b_j + \delta_j)) \leq \sum_1^\infty \mu((a_j, b_j]) + \varepsilon \leq \mu(E) + 2\varepsilon.$$

Since the cover was arbitrary,  $\nu(E) \leq \mu(E)$ . □

**Theorem 2.50.** Let  $E \in \mathcal{M}_\mu$ . Then

$$\begin{aligned} \mu(E) &= \inf \{ \mu(U) : U \supset E \text{ and } U \text{ is open} \} \\ &= \sup \{ \mu(K) : K \subset E \text{ and } K \text{ is compact} \}. \end{aligned}$$

*Proof.* By the previous lemma, for any  $\varepsilon > 0$ , there exists intervals  $(a_j, b_j)$  such that  $E \subset \bigcup_1^\infty (a_j, b_j)$  and  $\sum_1^\infty \mu((a_j, b_j)) \leq \mu(E) + \varepsilon$ . If we define  $U = \bigcup_1^\infty (a_j, b_j)$ , then  $U$  is open,  $U \supset E$ , and  $\mu(U) \leq \mu(E) + \varepsilon$ . On the other hand,  $\mu(U) \geq \mu(E)$  whenever  $U \supset E$ , so the first equality is valid.

For the second equality, first suppose that  $E$  is bounded. If  $E$  is closed, then  $E$  is compact and the equality is obvious. Otherwise, given  $\varepsilon > 0$ , we can choose an open set  $U \supset \overline{E} \setminus E$  such that  $\mu(U) \leq \mu(\overline{E} \setminus E) + \varepsilon$ , by the earlier paragraph. Let  $K = \overline{E} \setminus U$ , so  $K$  compact and  $K \subset E$ . Also,

$$\begin{aligned}\mu(K) &= \mu(E) - \mu(E \cap U) = \mu(E) - [\mu(U) - \mu(U \setminus E)] \\ &\geq \mu(E) - \mu(U) + \mu(\overline{E} \setminus E) \geq \mu(E) - \varepsilon.\end{aligned}$$

If  $E$  is unbounded, let  $E_i = E \cap (j, j + 1]$ . The preceding argument, for any  $\varepsilon > 0$ , there exists compact  $K_j \subset E_j$  such that  $\mu(K_j) \geq \mu(E_j) - \varepsilon 2^{-j}$ . Let  $H_n = \bigcup_{-n}^n K_j$ , so  $H_n$  is compact,  $H_n \subset E$ , and for any  $n \in \mathbb{N}$ ,  $\mu(H_n) \geq \mu(\bigcup_{-n}^n E_j) - \varepsilon$ . Since  $\mu(E) = \lim_{n \rightarrow \infty} \mu(\bigcup_{-n}^n E_j)$ , the result follows.  $\square$

**Theorem 2.51.** *If  $E \subset \mathbb{R}$ , then the following are equivalent.*

- (i)  $E \in \mathcal{M}_\mu$ ,
- (ii)  $E = V \setminus N_1$  where  $V$  is a  $G_\delta$  set and  $\mu(N_1) = 0$ ,
- (iii)  $E = H \cup N_2$  where  $H$  is an  $F_\sigma$  set and  $\mu(N_2) = 0$ .

*Proof.* It's clear that (b) and (c) imply (a) since  $\mu$  is complete on  $\mathcal{M}_\mu$ . Suppose that  $E \in \mathcal{M}_\mu$  and  $\mu(E) < \infty$ . By the theorem above, for  $j \in \mathbb{N}$ , we can choose an open  $U_j \supset E$  and compact  $K_j \subset E$  such that

$$\mu(U_j) - 2^{-j} \leq \mu(E) \leq \mu(K_j) + 2^{-j}.$$

Let  $V = \bigcap_{j=1}^\infty U_j$  and  $H = \bigcup_{j=1}^\infty K_j$ . Then  $H \subset E \subset V$  and  $\mu(V) = \mu(H) = \mu(E) < \infty$ . Hence,  $\mu(V \setminus E) = \mu(E \setminus H) = 0$ . The result is thus proved when  $\mu(E) < \infty$ ; an extension to the general case is left to [Folland \(1999, Exercise 25\)](#), and leverages  $\sigma$ -finiteness.  $\square$

**Remark 2.52.** *This theorem is really saying that  $\mathcal{M}_\mu$  consists of Borel sets mod null sets.*

**Proposition 2.53.** *If  $E \in \mathcal{M}_\mu$  and  $\mu(E) < \infty$ , then for every  $\varepsilon > 0$ , there is a set  $A$  that is a finite union of open intervals such that  $\mu(E \Delta A) < \varepsilon$ .*

*Proof.* Use [Lemma 2.49](#).  $\square$

### 2.6.2 The Lebesgue Measure

We now turn to a special case of the complete measure  $\mu_F$ , associated with the identity  $F(x) = x$ . We call this the Lebesgue measure  $m$  with domain  $\mathcal{L}$ . A set  $L \in \mathcal{L}$  is called Lebesgue measurable. In a slight abuse of terminology, we sometimes call the restriction  $m|_{\mathcal{B}_\mathbb{R}}$  the Lebesgue measure, though this will always be specified or clear from context. We will see that this measure space satisfies very nice properties.

**Theorem 2.54.** *Let  $E \in \mathcal{L}$ . Then  $E + s \in \mathcal{L}$  and  $rE \in \mathcal{L}$  for all  $r, s \in \mathbb{R}$ , where*

$$E + s = \{x + s : x \in E\}, \quad rE = \{r \cdot x : x \in E\}.$$

*Moreover,  $m(E + s) = m(E)$  and  $m(rE) = |r| \cdot m(E)$ .*

*Proof.* Since the collection of open intervals is invariant under translations and dilations, so is  $\mathcal{B}_{\mathbb{R}}$ . For  $E \in \mathcal{B}_{\mathbb{R}}$ , define

$$m_s(E) = m(E + s), \quad m^r(E) = m(rE).$$

Notice that for finite unions of intervals,  $m_s = m$  and  $m^r = |r| \cdot m$ . Since these measures agree on the algebra of finite unions of intervals, they agree on  $\mathcal{B}_{\mathbb{R}}$  by [Theorem 2.41](#). In particular, all null sets in  $\mathcal{B}_{\mathbb{R}}$  are preserved under scaling and transformation. By the characterization of elements of  $\mathcal{L}$  as a union of Borel sets and a Lebesgue null set, measures on  $\mathcal{L}$  are preserved under translations and dilations.  $\square$

**Remark 2.55.** (i) *The Lebesgue measure of  $\mathbb{Q}$  is zero. More generally, if the set  $E$  is countable,  $m(E) = 0$ .*

(ii) *Let  $\{q_j\}$  be an enumeration of  $\mathbb{Q} \cap [0, 1]$  and  $\bigcup_i B_{r_j}(q_j)$  where  $r_j = \varepsilon/2^j$ . Now take the intersection  $U = (0, 1) \cap (\bigcup_i B_{r_j}(q_j))$ . So  $U$  is open and dense in  $[0, 1]$ . Note, however,  $m(U) \leq \sum_1^\infty \varepsilon 2^{-j+1} = \varepsilon$ . Moreover, the set  $K = [0, 1] \setminus U$  is closed, nowhere dense<sup>4</sup>, but  $m(K) \geq 1 - \varepsilon$ . The point here is that a “topologically large” set can have a very small measure and a “topologically small” set can have a very large measure.*

(iii) *Any nonempty open set has a positive Lebesgue measure. Indeed, nonempty open sets will contain some nontrivial open interval by definition, and open intervals have positive measure.*

(iv) *The Lebesgue null sets include not only all countable sets, but also some sets with cardinality of the continuum. We will address this fact below.*

### 2.6.3 Null Sets Under Lebesgue Measure

We start with the Cantor set.

**Definition 2.56** (Cantor set, Cantor function). *Define  $C$  as the set of all  $x \in [0, 1]$  that have a base-3 expansion*

$$x = \sum_j a_j \cdot 3^{-j} \quad \text{with } a_j \in \{0, 2\} \quad \forall j \in \mathbb{N}.$$

*The set  $C$  can be obtained by removing the open middle third  $(1/3, 2/3)$  from  $[0, 1]$ , then the open middle thirds of the remaining intervals  $(1/9, 2/9)$  and  $(7/9, 8/9)$ , etc.*

*We also define the Cantor function  $f : C \rightarrow [0, 1]$  mapping*

$$\sum a_j 3^{-j} \mapsto \sum \left(\frac{a_j}{2}\right) 2^{-j}.$$

*In the image, we have a base-2 expansion  $f(x) = \sum_1^\infty b_j 2^{-j}$  where  $b_j = a_j/2$ .*

**Proposition 2.57.** *Let  $C$  be the Cantor set. Then*

(i)  *$C$  is compact, nowhere dense, and totally disconnected (i.e. the only connected subsets are single points). Moreover,  $C$  has no isolated points.*

(ii)  $m(C) = 0$ .

---

<sup>4</sup>The closure has empty interior.



(iii)  $\text{card}(C) = \mathfrak{c}$ , the cardinality of the continuum.

*Proof.* The first claim follows from the base-3 representation, and will appear in the homework. We make a comment, though: disconnected and no isolated points means that for any point  $x \in C$  and any  $\varepsilon > 0$ , we want to show that  $B_\varepsilon(x) \not\subset C$  but  $B_\varepsilon(x) \cap C \supsetneq \{x\}$ .

By definition,

$$m(C) = 1 - \frac{1}{3} - 2 \left( \frac{1}{3} \right)^3 - \dots = 1 - \sum_0^\infty \frac{2^j}{3^{j+1}} = 1 - \frac{1}{3} \left( \frac{1}{1 - 2/3} \right) = 0.$$

Next, consider the Cantor function  $f : C \rightarrow [0, 1]$ . Note that this function is onto, so  $\text{card}(C) = \mathfrak{c}$ .  $\square$

**Remark 2.58.** We make some relevant and not-so-relevant remarks.

- (i) Consider  $[0, 1]$  and scale it by a factor of  $1/n$ . We need  $n$  copies to recover the set, and this is really because the set has dimension 1. Now if you think about  $[0, 1]^2$  and scale it by a factor of  $[0, 1]^2$  and scale it by a factor of  $1/n$ , we need  $n^2$  copies to recover the set. Now for the Cantor set, if we scale it by a factor of  $1/3$ , we need 2 copies. Note that  $2 = 3^{\ln 2 / \ln 3}$ , and so we can think of this set having dimension  $\ln 2 / \ln 3$ . We can use this idea to construct a more general definition of dimension.
- (ii) Going back to the Cantor function, we can extend it to be an increasing function  $\bar{f} : [0, 1] \rightarrow [0, 1]$  by defining  $\bar{f}$  as a constant on each interval removed from  $C$ . Since  $\bar{f}$  is increasing and does not have a jump, it is continuous. This function is called the Cantor-Lebesgue function, or the Devil's staircase. Notice that on all intervals in  $[0, 1] \setminus C$ ,  $\bar{f}$  is constant and hence differentiable almost everywhere.
- (iii) There exist generalizations of the Cantor set, as seen in [Folland \(1999, Exercise 32\)](#).
- (iv) Recall that not every Lebesgue measurable set is a Borel set. The Cantor set is a Borel set as an intersection of closed sets, but the Cantor set does have non-Borel subsets. One can argue this by using transfinite induction to show that

$$\text{card}(\mathcal{L}) = \text{card}(2^{[0,1]}) > \mathfrak{c}, \quad \text{and also that} \quad \text{card}(\mathcal{B}_{\mathbb{R}}) = \mathfrak{c}.$$

This appears in the Notes and References of [Folland \(1999, Chapter 2\)](#).

## 3 Integration

With measures in hand, we want to build a robust notion of the integral.

### 3.1 Measurable Functions

Recall that any mapping  $f : X \rightarrow Y$  induces a map  $f^{-1} : 2^Y \rightarrow 2^X$  by taking  $f^{-1}(E) = \{x \in X : f(x) \in E\}$ . This operation preserves unions, intersections, and complements. Hence, if  $\mathcal{N}$  is a  $\sigma$ -algebra on  $Y$ , the collection  $\{f^{-1}(E) : E \in \mathcal{N}\}$  is a  $\sigma$ -algebra on  $X$ . We start by using these ideas to introduce the notion of a measurable function.

**Definition 3.1.** Let  $(X, \mathcal{M})$ ,  $(Y, \mathcal{N})$  be measurable spaces. We say the map  $f : X \rightarrow Y$  is  $(\mathcal{M}, \mathcal{N})$ -measurable (or just measurable when the spaces are understood) if  $f^{-1}(E) \in \mathcal{M}$  for all  $E \in \mathcal{N}$ . Note that this condition implies  $\{f^{-1}(E) : E \in \mathcal{N}\} \subset \mathcal{M}$ .

**Example 3.2.** Consider the functions  $f : (X, 2^X) \rightarrow (Y, \mathcal{N})$  and  $g : (X, \mathcal{M}) \rightarrow (Y, \{\emptyset, Y\})$ . These are trivial examples of measurable functions.

It will be useful to know that the condition for this definition can be relaxed into a sufficient condition for measurability of a function.

**Proposition 3.3.** If  $\mathcal{N}$  is generated by  $\mathcal{E}$ , then  $f : X \rightarrow Y$  is  $(\mathcal{M}, \mathcal{N})$ -measurable if and only if  $f^{-1}(E) \in \mathcal{M}$  for all  $E \in \mathcal{E}$ .

*Proof.* One direction is trivial. To see the other direction, consider the collection  $\{E \subset Y : f^{-1}(E) \in \mathcal{M}\}$ . By the properties of the inverse operator, this is a  $\sigma$ -algebra. This collection also contains  $\mathcal{E}$ , so it must contain  $\mathcal{N}$ .  $\square$

**Corollary 3.4.** If  $X$  and  $Y$  are topological spaces, then every continuous function  $f : X \rightarrow Y$  is  $(\mathcal{B}_X, \mathcal{B}_Y)$ -measurable.

*Proof.* Let  $\mathcal{E} \subset \mathcal{B}_Y$  be all open sets in  $Y$ . By definition of  $\mathcal{B}_Y$ ,  $\mathcal{E}$  generates  $\mathcal{B}_Y$ . By continuity, every set  $f^{-1}(E)$  for  $E \in \mathcal{E}$  is open in  $X$ , so  $f^{-1}(E) \in \mathcal{B}_X$ . Invoking the proposition above, the proof is complete.  $\square$

### 3.1.1 Lebesgue-Measurable Functions

We now restrict to a regime of interest.

**Definition 3.5.** For functions  $f : (X, \mathcal{M}) \rightarrow \mathbb{R}$  (resp.  $f : (X, \mathcal{M}) \rightarrow \mathbb{C}$ ), we say a function is  $\mathcal{M}$ -measurable if it is  $(\mathcal{M}, \mathcal{B}_{\mathbb{R}})$ -measurable (resp.  $(\mathcal{M}, \mathcal{B}_{\mathbb{C}})$ -measurable). In particular, we say a function  $f : \mathbb{R} \rightarrow \mathbb{C}$  is Lebesgue measurable if  $f$  is  $(\mathcal{L}, \mathcal{B}_{\mathbb{C}})$ -measurable. We say  $f : \mathbb{R} \rightarrow \mathbb{C}$  is Borel measurable if  $f$  is  $(\mathcal{B}_{\mathbb{R}}, \mathcal{B}_{\mathbb{C}})$ -measurable.

**Example 3.6.** Why do we not ask for functions to satisfy  $(\mathcal{L}, \mathcal{L})$ -measurability? This is quite a strong property. A simple example to consider is the Cantor set  $C$  and Cantor function  $f$ . We have seen that  $f$  is a bijection into the Cantor set. One can show that  $f$  is Lebesgue measurable. However, for any  $A \subset [0, 1]$ , one can cook up a non-measurable subset  $F \subset A$  such that  $f(F) \subset C$  is a subset of a null set, hence a null set, but  $f^{-1}(f(F)) = F$  is non-measurable. So the inverse image of a Lebesgue measurable set by a measurable function need not remain Lebesgue measurable.

We want to obtain some nice characterizations of function measurability. Since we have shown that the Borel  $\sigma$ -algebra is generated by the various collections  $\mathcal{E} \subset \mathcal{B}_{\mathbb{R}}$ , we obtain the following simple yet powerful proposition.

**Proposition 3.7.** If  $(X, \mathcal{M})$  is a measurable space, and  $f : X \rightarrow \mathbb{R}$ , TFAE:

- (i)  $f$  is  $\mathcal{M}$ -measurable;

- (ii)  $f^{-1}((a, \infty)) \in \mathcal{M}$  for all  $a \in \mathbb{R}$ ;
- (iii)  $f^{-1}([a, \infty)) \in \mathcal{M}$  for all  $a \in \mathbb{R}$ ;
- (iv)  $f^{-1}((-\infty, a)) \in \mathcal{M}$  for all  $a \in \mathbb{R}$ ;
- (v)  $f^{-1}((-\infty, a]) \in \mathcal{M}$  for all  $a \in \mathbb{R}$ .

*Proof.* Rays generate  $\mathcal{B}_{\mathbb{R}}$ . □

**Definition 3.8.** If  $f$  is a function on  $(X, \mathcal{M})$  and  $E \in \mathcal{M}$ , we say  $f$  is measurable on  $E$  if  $f^{-1}(B) \cap E \in \mathcal{M}$  for all Borel sets  $B$ . This is equivalent to the requirement that  $f|_E$  be  $\mathcal{M}_E$ -measurable for  $\mathcal{M}_E = \{F \cap E : F \in \mathcal{M}\}$ .

What if we want to consider functions  $f : X \rightarrow \mathbb{R}^n$ ? Let  $X$  be a set and  $\{(Y_\alpha, \mathcal{G}_\alpha)\}_{\alpha \in A}$  be a family of measurable spaces and  $f_\alpha : X \rightarrow Y_\alpha$  for each  $\alpha \in A$ . Then there exists a unique (smallest)  $\sigma$ -algebra on  $X$  such that all  $f_\alpha$  are measurable, ie. the  $\sigma$ -algebra generated by the collection

$$\{f_\alpha^{-1}(E_\alpha) : E_\alpha \in \mathcal{G}_\alpha, \forall \alpha \in A\}.$$

We call this the  $\sigma$ -algebra generated by the family  $\{f_\alpha\}_{\alpha \in A}$ . For instance, if  $X = \prod_{\alpha \in A} Y_\alpha$  is the product set and  $f_\alpha = \pi_\alpha : X \rightarrow Y_\alpha$  is the coordinate map, then the  $\sigma$ -algebra generated by  $\{f_\alpha\}_{\alpha \in A}$  is the product  $\sigma$ -algebra.

**Proposition 3.9.** Let  $(X, \mathcal{M})$  and  $(Y_\alpha, \mathcal{G}_\alpha)$  for  $\alpha \in A$  be measurable spaces. Let  $Y = \prod_{\alpha \in A} Y_\alpha$ , let  $\mathcal{G} = \bigotimes_{\alpha \in A} \mathcal{G}_\alpha$ , and let  $\pi_\alpha : Y \rightarrow Y_\alpha$  be coordinate maps. Then  $f : X \rightarrow Y$  is  $(\mathcal{M}, \mathcal{G})$ -measurable if and only if  $f_\alpha = \pi_\alpha \circ f : X \rightarrow Y_\alpha$  is  $(\mathcal{M}, \mathcal{G}_\alpha)$ -measurable for all  $\alpha \in A$ .

*Proof.* The coordinate maps  $\pi_\alpha$  are measurable. Hence, if  $f : X \rightarrow Y$  is measurable, then so is the composition  $f_\alpha$ . Conversely, if each  $f_\alpha$  is measurable, then for  $E_\alpha \in \mathcal{G}_\alpha$ , we have

$$f_\alpha^{-1}(E_\alpha) \in \mathcal{M}, \quad \text{and} \quad f_\alpha^{-1}(E_\alpha) = (\pi_\alpha \circ f)^{-1}(E_\alpha) = f^{-1}(\pi_\alpha^{-1}(E_\alpha)).$$

But  $\pi_\alpha^{-1}(E_\alpha)$  are exactly the generators of  $\mathcal{G}$ , so  $f$  is measurable. □

**Corollary 3.10.** A function  $f : X \rightarrow \mathbb{C}$  is  $\mathcal{M}$ -measurable if and only if  $\Re f$  and  $\Im f$  are  $\mathcal{M}$ -measurable. Similarly,  $f : X \rightarrow \mathbb{R}^n$  is  $\mathcal{M}$ -measurable if and only if  $f_i$  are all  $\mathcal{M}$ -measurable.

For functions with singularities, it is useful to allow these functions to take values in  $\overline{\mathbb{R}}$ . We define Borel sets in  $\overline{\mathbb{R}}$  by

$$\mathcal{B}_{\overline{\mathbb{R}}} = \{E \subset \overline{\mathbb{R}} : E \cap \mathbb{R} \in \mathcal{B}_{\mathbb{R}}\}.$$

This coincides with the usual definition of the Borel  $\sigma$ -algebra if we make  $\overline{\mathbb{R}}$  into a metric space with metric  $\rho(x, y) = |A(x) - A(y)|$ , where  $A(x) = \arctan(x)$ . It is easily verified that  $\mathcal{B}_{\overline{\mathbb{R}}}$  is generated by the rays  $(a, \infty]$  or  $[-\infty, a)$  ( $a \in \mathbb{R}$ ), and we define  $f : X \rightarrow \overline{\mathbb{R}}$  to be  $\mathcal{M}$ -measurable if  $f$  is  $(\mathcal{M}, \mathcal{B}_{\overline{\mathbb{R}}})$ -measurable. See Exercise 1.

**Proposition 3.11.** Let  $f, g : X \rightarrow \mathbb{C}$  be  $\mathcal{M}$ -measurable. Then  $f + g$  and  $f \cdot g$  are also  $\mathcal{M}$ -measurable.

**Remark 3.12.** This proposition is true for  $\overline{\mathbb{R}}$ -valued functions as well

*Proof of Proposition 3.11.* We can view  $f + g$  as a composition. Define  $F : X \rightarrow \mathbb{C} \times \mathbb{C}$ ,  $x \mapsto (f(x), g(x))$ . Also define  $\phi : \mathbb{C} \times \mathbb{C} \rightarrow \mathbb{C}$  sending  $(z_1, z_2) \mapsto z_1 + z_2$  and  $\psi : \mathbb{C} \times \mathbb{C} \rightarrow \mathbb{C}$  by  $(z_1, z_2) \mapsto z_1 z_2$ . Note that both  $\phi$  and  $\psi$  are  $(\mathcal{B}_{\mathbb{C} \times \mathbb{C}}, \mathcal{B}_{\mathbb{C}})$ -measurable since they are continuous. Next, by our previous proposition,  $F$  is  $(\mathcal{M}, \mathcal{B}_{\mathbb{C} \times \mathbb{C}})$ -measurable. Therefore,  $f + g = \phi \circ F$  and  $f \cdot g = \psi \circ F$  are both  $\mathcal{M}$ -measurable.  $\square$

Thinking back on the motivations of measure theory, we want to consider measurability of functions under some limiting process.

**Proposition 3.13.** Let  $\{f_j\}$  be a sequence of  $\overline{\mathbb{R}}$ -valued measurable functions on  $(X, \mathcal{M})$ . Then the functions

$$g_1 = \sup_j f_j, \quad g_2 = \inf_j f_j, \quad g_3 = \limsup_j f_j, \quad g_4 = \liminf_j f_j$$

are measurable. Moreover, if the pointwise limit exists for every  $x \in X$ , then  $\lim_j f_j$  is also measurable.

**Remark 3.14.** Riemann-integrability is not in general preserved under such limiting processes. Consider  $\{r_j\} = \mathbb{Q}$ , partial unions  $S_n := \bigcup_1^n \{r_j\}$ , and  $f_n := \chi_{S_n} \rightarrow \chi_{\mathbb{Q}}$ .

*Proof of Proposition 3.13.* Consider  $x \in g_1^{-1}((a, \infty])$ . This is true if and only if  $\sup_j f_j(x) > a$ , which is true if and only if there exists some  $j_0$  such that  $f_{j_0}(x) > a$ . In particular, we have

$$g_1^{-1}((a, \infty]) = \bigcup_{j=1}^{\infty} f_j^{-1}((a, \infty]).$$

This is a countable union of measurable sets, so measurable. That is,  $g_1$  is measurable by Proposition 3.7. Similarly,

$$g_2^{-1}([-\infty, a)) = \bigcup_{j=1}^{\infty} f_j^{-1}([-\infty, a)).$$

For  $g_3$ , note that  $\limsup$  can be written using only the supremum and infimum operators. For  $h_k(x) = \sup_{j \geq k} f_j(x)$ ,  $h_k$  is measurable for each  $k$ . Hence,

$$g_3(x) = \limsup_j f_j(x) = \inf_k \sup_{j \geq k} f_j(x) = \inf_k h_k(x),$$

which is measurable. A similar argument works for  $g_4$ .

Finally,  $\lim_{j \rightarrow \infty} f_j(x)$  exists if and only if the  $\limsup$  and  $\liminf$  at  $x$  exist and  $\limsup_j f_j(x) = \liminf_j f_j(x)$ . Hence, if the limit exists for all  $x$ , the limit is equivalent to the  $\liminf$ , which is measurable.  $\square$

**Corollary 3.15.** If  $f, g : X \rightarrow \overline{\mathbb{R}}$  are measurable. Then so are  $\max(f, g)$  and  $\min(f, g)$ .

**Corollary 3.16.** If  $\{f_j\}$  is a sequence of complex-valued measurable functions and  $f(x) = \lim_{j \rightarrow \infty} f_j(x)$  exists for all  $x$ , then  $f$  is measurable.

## 3.2 Simple Functions

**Definition 3.17.** For  $f : X \rightarrow \overline{\mathbb{R}}$ , define  $f^+(x) = \max(f(x), 0)$  and  $f^-(x) = \max(-f(x), 0)$  as the positive and negative parts of  $f$ , respectively.

**Remark 3.18.** A note on this definition.

- Note that  $f^+, f^- \geq 0$  and that  $f = f^+ - f^-$ .
- If  $f$  is measurable, then so are  $f^+$  and  $f^-$ . Moreover, if  $f^+$  and  $f^-$  are both measurable, then  $f$  is measurable.
- For  $f : X \rightarrow \mathbb{C}$ , we can come up with an analogous decomposition, the polar decomposition:

$$f = (\operatorname{sgn} f)|f|, \quad \text{where} \quad \operatorname{sgn} z = \begin{cases} \frac{z}{|z|} & \text{if } |z| \neq 0 \\ 0 & \text{else.} \end{cases}$$

It holds that  $f$  is measurable if and only if  $\operatorname{sgn} f$  and  $|f|$  are measurable. Indeed, if  $f$  is measurable, since  $|\cdot|$  is continuous,  $|f|$  is measurable. Next,  $\operatorname{sgn}$  is continuous except at the origin. If  $U \subset \mathbb{C}$  is open, then  $\operatorname{sgn}^{-1}(U)$  is either open or of the form  $V \cup \{0\}$  where  $V$  is open. So  $\operatorname{sgn}$  is Borel-measurable and therefore  $\operatorname{sgn} f$  is measurable. The other direction is clear by previous propositions.

- For a set  $E \subset X$ , then the characteristic function  $\chi_E$  is measurable if and only if  $E \in \mathcal{M}$ .

**Definition 3.19.** A simple function on  $X$  is a finite linear combination (with complex coefficients) of characteristic functions of measurable sets  $E \in \mathcal{M}$ . We do not allow the coefficients to take values  $\pm\infty$ .

**Remark 3.20.** One can characterize simple functions in the following way. Suppose  $f : X \rightarrow \mathbb{C}$  is simple; it takes only finitely many values  $\operatorname{range}(f) = \{z_1, \dots, z_n\}$ . Also assume that  $f$  is measurable. Then

$$E_j := f^{-1}(\{z_j\}) \in \mathcal{M}.$$

Moreover, we can write

$$f = \sum_{j=1}^n z_j \chi_{E_j}, \quad \text{all } E_j \text{ disjoint.}$$

We call this the standard representation of  $f$ .

It turns out that we can approximate measurable functions using simple functions, and this will be a very useful result for integration.

**Theorem 3.21.** Let  $(X, \mathcal{M})$  be a measurable space.

- If  $f : X \rightarrow [0, \infty]$  is a nonnegative measurable function, then there exists a sequence  $\{\phi_n\}_n$  of simple functions such that  $0 \leq \phi_1 \leq \phi_2 \leq \dots \leq f$  such that  $\phi_n \rightarrow f$  pointwise,  $\phi_n \rightrightarrows f$  (uniformly) on any set on which  $f$  is bounded.
- Similarly, if  $f : X \rightarrow \mathbb{C}$  is measurable, then there exists a sequence of simple functions  $\{\psi_n\}_n$  such that  $0 \leq |\psi_1| \leq |\psi_2| \leq \dots \leq |f|$  and  $\psi_n \rightarrow f$  pointwise and  $\psi_n \rightrightarrows f$  on any set where  $f$  is bounded.

*Proof.* We use an explicit construction. For  $n \in \mathbb{N}$ , and  $0 \leq k \leq 2^{2n}-1$ . Let  $E_n^k = f^{-1}((k2^{-n}, (k+1)2^{-n}])$  and  $F_n = f^{-1}((2^n, \infty])$ . Now define

$$\phi_n = \left( \sum_{k=0}^{2^{2n}-1} k2^{-n} \cdot \chi_{E_n^k} \right) + 2^n \chi_{F_n}.$$

This is a simple function.

It remains to check the convergence claims. First notice that for all  $x \in X$ ,  $\phi_n(x)$  is monotonically increasing in  $n$ . Moreover, for each  $n$ , we have  $0 \leq f - \phi_n \leq 2^{-n}$  on the set where  $f \leq 2^n$ . Hence, for any fixed  $x$ , we have pointwise convergence, and for any set where  $f$  is bounded, we have uniform convergence.

The second claim follows by writing  $f = g + ih$  and applying the first result of  $g^+$ ,  $g^-$ ,  $h^+$  and  $h^-$  to get approximations  $\psi_{g,n}^+$ ,  $\psi_{g,n}^-$ ,  $\psi_{h,n}^+$ ,  $\psi_{h,n}^-$ . We can then let  $\psi_n = (\psi_{g,n}^+ - \psi_{g,n}^-) + i(\psi_{h,n}^+ - \psi_{h,n}^-)$ .  $\square$

**Remark 3.22.** Consider the series

$$\sum_{k=0}^{2^{2n}-1} k2^{-n} \mu \left( f^{-1} \left( (k2^{-n}, (k+1)2^{-n}) \right) \right).$$

This corresponds to a Riemann sum for the integral

$$\int_0^{2^n} \mu(\{x : f(x) > t\}) dt,$$

where  $2^{-n}$  serves as the interval size, and  $\mu$  the underestimate within this interval. In particular,

$$\int f d\mu = \int_0^\infty \mu(\{x : f(x) > t\}) dt,$$

as we will see.

**Proposition 3.23.** The following implications are valid if and only if  $\mu$  is a complete measure.

- (i) If  $f$  is measurable and  $f = g$ ,  $\mu$ -a.e., then  $g$  is measurable.
- (ii) If  $f_n$  is measurable for  $n \in \mathbb{N}$  and  $f_n \rightarrow f$ ,  $\mu$ -a.e., then  $f$  is measurable.

*Proof.* Exercise.  $\square$

**Proposition 3.24.** Let  $(X, \mathcal{M}, \mu)$  be a measure space and  $(X, \overline{\mathcal{M}}, \overline{\mu})$  be its completion. If  $f$  is  $\overline{\mathcal{M}}$ -measurable function on  $X$  into  $[0, \infty]$  or  $\mathbb{C}$ , then there exists an  $\mathcal{M}$ -measurable function  $g$  such that  $f = g$ ,  $\overline{\mu}$ -a.e.

*Proof.* Recall that  $\overline{\mathcal{M}}$  differs from  $\mathcal{M}$  only by null sets. If  $f = \chi_E$  where  $E \in \overline{\mathcal{M}}$ , then  $E = F \cup N$  where  $F \in \mathcal{M}$  and  $\overline{\mu}(N) = 0$ . So, take  $g = \chi_F$ , and note that (1)  $f = g$ ,  $\overline{\mu}$ -a.e. and that (2)  $g$  is  $\mathcal{M}$ -measurable. We can say something similar about simple functions.

In the general case, we aim to approximate with simple functions. Choose a sequence  $\{\phi_n\}_n$  of  $\overline{\mathcal{M}}$ -measurable simple functions which converge to  $f$ . For each  $n$ , let  $\psi_n$  be an  $\mathcal{M}$ -measurable simple

function such that  $\phi_n = \psi_n$ ,  $\bar{\mu}$ -a.e.; these only differ on a set  $N_n \in \overline{\mathcal{M}}$  where  $\bar{\mu}(N_n) = 0$ . Then note that  $\bigcup_1^\infty N_n$  is still a null set in  $\overline{\mathcal{M}}$ . Since  $\bar{\mu}$  is a completion of  $\mu$ , there is some set  $N \in \mathcal{M}$  such that  $\bigcup_1^\infty N_n \subset N$  and  $\mu(N) = 0$ . Set

$$g = \lim_{n \rightarrow \infty} \chi_{(X \setminus N)} \psi_n,$$

which is  $\mathcal{M}$ -measurable by [Proposition 3.13](#), and  $g = f$  on  $N^c$ . □

### 3.3 Integration of Nonnegative Functions

Fix a measure space  $(X, \mathcal{M}, \mu)$ .

**Definition 3.25.** We define  $L^+$  to be the space of all measurable functions  $X \rightarrow [0, \infty]$ . If  $\phi$  is a simple function with standard representation  $\phi = \sum_{j=1}^n a_j \chi_{E_j}$ , then we define the integral of  $\phi$  with respect to  $\mu$  by the expression

$$\int \phi \, d\mu = \sum_{j=1}^n a_j \mu(E_j).$$

As a convention,  $0 \cdot \infty = 0$  within the sum.

**Remark 3.26.** Some notes on this definition.

- Note that  $\int \phi \, d\mu = \infty$  is allowed because  $\mu(E_j)$  need not be finite.
- Since  $\phi \in L^+$ , there is no ambiguity arising from indefinite sign  $\infty - \infty$ .
- Different notations include

$$\int \phi = \int \phi \, d\mu = \int \phi(x) \, d\mu(x) = \int \phi(x) \mu(dx) = E_\mu[\phi].$$

- There is an immediate definition for integrals over domains. If  $A \in \mathcal{M}$ , then  $\phi \cdot \chi_A$  is also simple. We define the integral over  $A$  as

$$\int_A \phi \, d\mu = \int \phi \cdot \chi_A \, d\mu.$$

- Such integrals also have several notations:

$$\int_A \phi \, d\mu = \int_A \phi = \int_A \phi(x) \, d\mu(x), \quad \text{and} \quad \int \phi = \int_X \phi.$$

We now discuss several properties of these integrals.

**Proposition 3.27.** Let  $\phi, \psi \in L^+$  be simple functions.

- (i) If  $c \geq 0$ , then  $\int c\phi = c \int \phi$ .
- (ii)  $\int(\phi + \psi) = \int \phi + \int \psi$ .
- (iii) If  $\phi \leq \psi$ , then  $\int \phi \leq \int \psi$ .

(iv) For fixed  $\phi \in L^+$ , the map  $A \mapsto \int_A \phi \, d\mu$  is a measure on  $\mathcal{M}$ .

*Proof.* The proof of (a) is straightforward. To see (b), let  $\phi = \sum_1^n a_j \chi_{E_j}$  and  $\psi = \sum_1^m b_k \chi_{F_k}$ . We write  $E_{jk} = E_j \cap F_k$ , and note that we have disjoint unions  $\bigsqcup_k E_{jk} = E_j$  and  $\bigsqcup_j E_{jk} = F_k$ . Then

$$\phi = \sum_{j,k} a_j \chi_{E_{jk}}, \quad \psi = \sum_{j,k} b_k \chi_{E_{jk}}.$$

Hence,

$$\begin{aligned} \int \phi + \int \psi &= \sum_j a_j \mu(E_j) + \sum_k b_k \mu(F_k) \\ &= \sum_{j,k} (a_j + b_k) \mu(E_j \cap F_k) = \int (\phi + \psi). \end{aligned}$$

To see (c), note that  $\phi \leq \psi$  implies  $a_j \leq b_k$  wherever  $E_j \cap F_k \neq \emptyset$ . Hence,

$$\begin{aligned} \int \phi &= \sum_{j,k} a_j \mu(E_{jk}) \\ &\leq \sum_{j,k} b_k \mu(E_{jk}) = \int \psi. \end{aligned}$$

Finally, for (d), we must check:

- $\int_{\emptyset} \phi = 0$ ;
- $\sigma$ -additivity.

The first point follows from the observation that

$$\int_{\emptyset} \phi = \sum_j a_j \mu(E_j \cap \emptyset).$$

To see the second point, let  $\{A_j\}$  be a disjoint sequence in  $\mathcal{M}$  and  $A = \bigcup_j A_j$ . Now,

$$\begin{aligned} \int_A \phi &= \sum_j a_j \mu(A \cap E_j) = \sum_j a_j \mu \left( \bigcup_k (A_k \cap E_j) \right) \\ &= \sum_j a_j \sum_k \mu(A_k \cap E_j) \\ &= \sum_k \sum_j a_j \mu(A_k \cap E_j) = \sum_k \int_{A_k} \phi. \end{aligned}$$

□

**Definition 3.28.** Let  $f \in L^+$  but not necessarily simple. We define the integral of  $f$  via the expression

$$\int f \, d\mu = \sup \left\{ \int \phi \, d\mu : 0 \leq \phi \leq f, \phi \text{ simple} \right\}.$$



**Remark 3.29.** We make some comments on this definition before proceeding.

- If  $f$  is simple, this agrees with the definition of the integral of a simple function.
- If  $f \leq g$ , then  $\int f \leq \int g$ .
- For  $c \geq 0$ ,  $\int cf = c \int f$ .

**Theorem 3.30** (Monotone Convergence Theorem). If  $\{f_n\} \subset L^+$  satisfies  $f_j \leq f_{j+1}$  for all  $j$ , and  $f = \lim_{n \rightarrow \infty} f_n (= \sup_n f_n)$ , then

$$\int f = \lim_{n \rightarrow \infty} \int f_n.$$

*Proof.* For any  $x$ ,  $\{f_n(x)\}$  is an increasing sequence. Hence,  $\lim_n f_n(x) = f(x)$  equals the supremum, so  $f$  is measurable. Since  $f_n \leq f$  pointwise, we have

$$\lim_{n \rightarrow \infty} \int f_n \leq \int f.$$

It remains to show the reverse inequality. Using the definition of  $f$ , let  $\phi$  be any simple function  $0 \leq \phi \leq f$ . Now for any  $\alpha \in (0, 1)$ , consider

$$E_n = \{x : f_n(x) \geq \alpha \phi(x)\}.$$

Since  $f_n$  is increasing,  $\{E_n\}$  is an increasing sequence of measurable sets whose union is  $X$ . We also have that for any  $n$ ,

$$\int f_n \geq \int_{E_n} f_n \geq \alpha \int_{E_n} \phi.$$

On the other hand, since  $\int_{(\cdot)} \phi : \mathcal{M} \rightarrow [0, \infty]$  is a measure, continuity of measures implies that,

$$\lim_{n \rightarrow \infty} \alpha \int_{E_n} \phi = \alpha \int \phi.$$

That is,

$$\lim_{n \rightarrow \infty} \int f_n \geq \alpha \int \phi.$$

Note that  $\alpha$  and  $\phi$  are arbitrary, so by taking  $\alpha \rightarrow 1$  and the supremum over all simple functions such that  $0 \leq \phi \leq f$ , we have that

$$\lim_{n \rightarrow \infty} \int f_n \geq \int f.$$

□

**Remark 3.31.** The definition of  $\int f \, d\mu$  was via all possible simple functions  $\phi$ , a huge family. Now, MCT tells us that it is sufficient to compute  $\lim \int \phi_n$  where  $\phi_n$  are simple and  $\phi_n \nearrow f$ . This often allows us to establish proofs by looking only at simple functions.

**Theorem 3.32.** If  $\{f_n\} \subset L^+$  is a finite or countably infinite sequence, and  $f = \sum_n f_n$ , then

$$\int f = \sum_n \int f_n.$$

*Proof.* Let  $f_1, f_2 \in L^+$ . Then we can find  $\phi_j^1$  and  $\phi_j^2$  approximating  $f_1$  and  $f_2$ , respectively. Next,

$$\int f_1 + f_2 \stackrel{\text{MCT}}{=} \lim_j \int \phi_j^1 + \phi_j^2 = \lim_j \left( \int \phi_j^1 + \int \phi_j^2 \right) = \lim_j \int \phi_j^1 + \lim_j \int \phi_j^2 \stackrel{\text{MCT}}{=} \int f_1 + \int f_2.$$

By induction, for any finite sum,  $\int \sum f_n = \sum \int f_n$ . Now, consider the partial sum  $g_N = \sum_1^N f_n$  and apply MCT to this function. We obtain

$$\int \sum_1^\infty f_n = \sum_1^\infty \int f_n.$$

□

**Proposition 3.33.** *If  $f \in L^+$ , then  $\int f = 0$  if and only if  $f = 0$  a.e.*

*Proof.* First assume that  $f$  is simple. Then  $f = \sum_1^n a_j \chi_{E_j}$ ,  $a_j \geq 0$ . Observe that  $\int f = 0$  holds if and only if for each  $j$ ,  $a_j = 0$  or  $\mu(E_j) = 0$ . Now let  $f \in L^+$  be given. Let  $\phi$  be a simple function such that  $0 \leq \phi \leq f$ . Assuming  $f = 0$  a.e., we have that  $\phi = 0$  a.e. Hence,

$$\int f = \sup_{0 \leq \phi \leq f} \int \phi = 0.$$

Going the other direction, if  $\{x : f(x) > 0\}$  is not a null set, then we can write

$$\{x : f(x) > 0\} = \bigcup_1^\infty E_n \quad \text{where} \quad E_n = \{x : f(x) > 1/n\}.$$

One such  $E_n$  must not be a null set. But then  $f$  is bounded below by a simple function,  $f > \frac{1}{n} \chi_{E_n}$ . In particular,  $\int f > 0$ , which contradicts the assumption that  $\int f = 0$ . □

**Corollary 3.34.** *If  $\{f_n\} \subset L^+$ ,  $f \in L^+$ , and  $f_n(x) \nearrow f(x)$  a.e. in  $X$ , then*

$$\int f = \lim_{n \rightarrow \infty} \int f_n.$$

*Proof.* Suppose  $f_n(x)$  increases to  $f(x)$  for  $x \in E$  and  $\mu(E^c) = 0$ . We define  $g_n := f_n \chi_E \nearrow g := f \chi_E$  for every  $x$ . Since  $f_n = g_n$  a.e. and  $g = f$  a.e., the previous proposition gives

$$\int f = \int g \stackrel{\text{MCT}}{=} \lim_n \int g_n = \lim_n \int f_n.$$

□

### 3.4 Fatou's Lemma and Standard Counterexamples

What happens when we drop the monotonicity assumption?

**Example 3.35** (Escape to infinity). *Define  $f_j = \chi_{[j, j+1)}$ . It is clear  $f_j \in L^+$  and  $\int f_j \, d\mu = 1$ . Note that  $f_j(x) \rightarrow 0$  for every  $x$ . Hence,  $\int \lim_j f_j = 0$ , but  $\lim_j \int f_j = 1$ .*

**Example 3.36** (Concentration/blowup). Define  $f_j(x) = j\chi_{[0,1/j]}$ . So  $f_j \in L^+$  with  $\int f_j \, dm = 1$ . Note that  $f_j(x) \rightarrow 0$  for all  $x > 0$ , and  $f_j(0) \rightarrow \infty$ . Again,  $0 \neq 1$ .

In both of these examples, we have the inequality

$$\underbrace{0}_{\int \lim} < \underbrace{1}_{\lim \int}.$$

One can make a more general statement, which gives a partial characterization of our counterexamples.

**Lemma 3.37** (Fatou). *If  $\{f_n\} \subset L^+$ , then*

$$\int \liminf f_n \leq \liminf \int f_n.$$

*Proof.* Recall that  $\liminf f_n = \sup_k \inf_{n \geq k}$ . For each  $k \geq 1$ , we have

$$\inf_{n \geq k} f_n \leq f_j, \quad \text{for all } j \geq k.$$

Now,

$$\int \inf_{n \geq k} f_n \leq \int f_j \quad \text{for all } j \geq k, \quad \text{and} \quad \int \inf_{n \geq k} f_n \leq \inf_{j \geq k} \int f_j.$$

But note that  $\inf_{n \geq k} f_n \nearrow \liminf f_n$  as  $k \rightarrow \infty$ , so by MCT, we have

$$\int \liminf f_n = \lim_{k \rightarrow \infty} \int \inf_{n \geq k} f_n \leq \liminf \int f_n.$$

□

**Corollary 3.38.** *Let  $\{f_n\} \subset L^+$ ,  $f \in L^+$ , and  $f_n \rightarrow f$  a.e. Then*

$$\int f \leq \liminf \int f_n.$$

*Proof.* We have  $f = \lim_{n \rightarrow \infty} f_n$  a.e., so  $f = \liminf f_n$  a.e. Using the proposition about integrals of functions that are equivalent except for a null set, we have

$$\int f = \int \liminf f_n \leq \liminf \int f_n.$$

□

### 3.5 Integration of Complex-Valued Functions

**Definition 3.39.** *If  $f : X \rightarrow \mathbb{R}$  is measurable and at least one of  $\int f^+$  or  $\int f^-$  is finite, we define*

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

*If both  $\int f^+$  and  $\int f^-$  are finite, we say  $f$  is integrable.<sup>5</sup>*

---

<sup>5</sup>One must not confused the two notions introduced here: the integral  $\int f$  is a quantity in  $\overline{\mathbb{R}}$ , and this quantity may exist even when  $f$  is not integrable.

**Proposition 3.40.** A function  $f : X \rightarrow \mathbb{R}$  is integrable if and only if  $\int |f| < \infty$ .

**Proposition 3.41.** The set of all integrable, real-valued functions on  $X$  is a real vector space, and the integral is a linear functional on this vector space.

*Proof.* To start, note that  $|af + bg| \leq |a||f| + |b||g|$ , so linear combinations will be integrable. To show linearity of the mapping  $\int(\cdot) : V \rightarrow \mathbb{R}$ , we check the definition. We have  $\int af = a \int f$  using the definition and our result for functions in  $L^+$ . Now let  $h = f + g$  for integrable  $f, g$ . We have

$$h = h^+ - h^- = f^+ + g^+ - f^- - g^-.$$

We also have

$$h^+ + f^- + g^- = h^- + f^+ + g^+ \in L^+.$$

By the theorem about summing functions in  $L^+$ , we have

$$\int h^+ + \int f^- + \int g^- = \int h^- + \int f^+ + \int g^+,$$

and

$$\int h = \int h^+ - \int h^- = \int f^+ - \int f^- + \int g^+ - \int g^- = \int f + \int g.$$

□

**Definition 3.42.** If  $f : X \rightarrow \mathbb{C}$  is measurable, we say that  $f$  is integrable if  $\int |f| < \infty$ . For  $E \in \mathcal{M}$ , we say that  $f$  is integrable on  $E$  if  $\int_E |f| < \infty$ . For integrable  $f$ , we define

$$\int f = \int \Re f + i \int \Im f.$$

**Remark 3.43.** For integrable  $f : X \rightarrow \mathbb{C}$ , by definition,

$$\Re \int f = \int \Re f, \quad \Im \int f = \int \Im f.$$

**Proposition 3.44.** The function  $f : X \rightarrow \mathbb{C}$  is integrable if and only if  $\Re f$  and  $\Im f$  are integrable.

*Proof.* This follows from the fact that

$$|f| \leq |\Re f| + |\Im f| \leq 2|f|.$$

□

**Proposition 3.45.** The space of complex-valued integrable functions is a complex vector space, and the integral is a complex linear functional on this vector space.

*Proof.* Using previous proposition. □

We denote the vector space of complex-valued integrable functions under measure  $\mu$  by  $L^1(\mu)$  (or  $L^1(X, \mu)$  or  $L^1(X)$  or  $L^1$ ). We will eventually redefine  $L^1(\mu)$  by quotienting out equivalencies almost everywhere. For now, though, take  $L^1(\mu)$  as described.

**Proposition 3.46.** *If  $f \in L^1$ , then  $|\int f| \leq \int |f|$ .*

*Proof.* For real-valued functions,

$$\left| \int f \right| = \left| \int f^+ - \int f^- \right| \leq \int f^+ + \int f^- = \int f^+ + f^- = \int |f|.$$

If  $f$  is complex-valued and  $\int f = 0$ , the claim is trivial. If  $\int f \neq 0$ , let  $\alpha = \overline{\text{sgn}(\int f)}$ . Then

$$\int f = \text{sgn} \left( \int f \right) \left| \int f \right|.$$

Then

$$\alpha \int f = \overline{\text{sgn} \left( \int f \right)} \text{sgn} \left( \int f \right) \left| \int f \right| \implies \int \alpha f = \left| \int f \right|.$$

Since  $\int \alpha f$  is real then,

$$\left| \int f \right| = \Re \int \alpha f = \int \Re(\alpha f) \leq \int |\alpha f| = \int |f|.$$

□

**Proposition 3.47.** (i) *If  $f \in L^1$ , then  $\{x : f(x) \neq 0\}$  is  $\sigma$ -finite.*

(ii) *If  $f, g \in L^1$ , then  $\int_E f = \int_E g$  for all  $E \in \mathcal{M}$  if and only if  $\int |f - g| = 0$  if and only if  $f = g$  a.e.*

*Proof.* Part (i) is left as an exercise. For (ii), suppose that  $\int |f - g| = 0$ . Then

$$\left| \int_E f - \int_E g \right| = \left| \int_E f - g \right| \leq \int |f - g| \chi_E \leq \int |f - g| = 0.$$

That is,  $\int_E f = \int_E g$  for all  $E \in \mathcal{M}$ . On the other hand, suppose  $\int |f - g| > 0$ . Then let

$$u = \Re(f - g), \quad v = \Im(f - g).$$

Then at least one of  $u^+, u^-, v^+, v^-$  has to be nonzero on a set of positive measure. W.l.o.g. let  $E = \{x : u^+ > 0\}$ , which has nonzero measure. Then

$$\Re \left( \int_E f - \int_E g \right) = \int_E u^+ > 0,$$

where the first equality holds because  $u^- = 0$  on  $E$ . The other equivalence follows from our previous proposition about the equality of integrals of functions in  $L^1$ . □

We may now define  $L^1$  properly.

**Definition 3.48.** *Define  $L^1(\mu)$  to be the set of equivalence classes of a.e.-defined integrable (complex-valued) functions defined on  $X$ , where  $f \sim g$  if and only if  $f = g$   $\mu$ -a.e.*

**Remark 3.49.**  $L^1(\mu)$  is still a complex vector space.

- Although we will henceforth view  $L^1(\mu)$  as a space of equivalence classes, each class consisting of purely integrable functions, we shall write  $f \in L^1(\mu)$  to mean that  $f$  is equal a.e. to an integrable function. This is a standard abuse of notation, the first example of which appears in the Dominated Convergence Theorem.
- [Proposition 3.24](#) yields a natural one-to-one correspondence between  $L^1(\mu)$  and  $L^1(\bar{\mu})$ .
- $L^1$  is a metric space with metric  $\rho(f, g) := \int |f - g|$ . We denote this  $(L^1, \rho)$ .
- More generally,  $(L^1(\mu), \|\cdot\|_{L^1})$  is a normed vector space where  $\|f\|_{L^1} := \int |f|$ . The induced metric is exactly  $\rho$ .
- We refer to convergence with respect to the metric  $\rho$  as convergence in  $L^1$ ; that is,  $f_n \rightarrow f$  in  $L^1$  if and only if  $\int |f_n - f| \rightarrow 0$ .

**Theorem 3.50** (Dominated Convergence Theorem). Let  $\{f_n\}$  be a sequence in  $L^1$  such that

- (i)  $f_n \rightarrow f$  a.e. ( $f_n$  and  $f$  defined except for set of measure zero),
- (ii) there exists some  $\mathbb{R}$ -valued  $g \in L^1$ ,  $g \geq 0$  such that  $|f_n| \leq g$  for all  $n \in \mathbb{N}$ .

Then  $f \in L^1$  and  $\int f = \lim_{n \rightarrow \infty} \int f_n$ .<sup>6</sup>

*Proof.* Note that all  $f_n$  are  $\mathcal{M}$ -measurable, and by [Proposition 3.24](#),  $f$  is  $\mathcal{M}$ -measurable. By [Proposition 3.23](#), there is an  $\mathcal{M}$ -measurable function  $f'$  such that  $f' = f$   $\bar{\mu}$ -a.e. We may take  $f$  to be defined except on the null set where  $f(x) \neq f'(x)$ .

Since  $|f| \leq g$  a.e., we have  $f \in L^1$ . Now, we may assume  $f_n$  and  $f$  are real-valued, otherwise taking imaginary and real parts. So  $g + f_n \geq 0$  a.e. and  $g - f_n \geq 0$  a.e. By the corollary to Fatou's lemma,

$$\int g + \int f \leq \liminf_n \int g + f_n = \int g + \liminf_n \int f_n.$$

We also have

$$\int g - \int f \leq \liminf_n \int g - f_n = \int g - \limsup_n \int f_n.$$

That is,

$$\liminf_n \int f_n \geq \int f \geq \limsup_n \int f_n.$$

We have  $\int f = \lim_{n \rightarrow \infty} \int f_n$ . □

**Theorem 3.51.** Suppose  $\{f_j\}$  is a sequence in  $L^1$  such that  $\sum_1^\infty \int |f_j| < \infty$ . Then  $\sum_1^\infty f_j$  converges a.e. to a function in  $L^1$  and

$$\int \sum_1^\infty f_j = \sum_1^\infty \int f_j.$$

---

<sup>6</sup>As pointed out in [Remark 3.49](#), we really mean that  $f$  is equal a.e. to some Lebesgue integrable function in  $L^1$ . It could very well be that  $f$  is not even measurable if  $\mathcal{M}$  is not complete.

*Proof.* Note that  $|f_j| \in L^+$ , so

$$\int \sum_1^\infty |f_j| = \sum_1^\infty \int |f_j| < \infty.$$

Hence, we can define  $g = \sum_1^\infty |f_j| \in L^1$ , and note that  $\sum_1^\infty |f_j| < \infty$  holds a.e.<sup>7</sup> For such  $x$  not contained in the null set, the  $N$ th partial sum  $F_N(x) = \sum_1^N f_j(x)$  converges to  $\sum_1^\infty f_j(x)$ . We also note that

$$\left| \sum_1^N f_j \right| \leq g.$$

Applying DCT, we have

$$\int \sum_1^\infty f_j = \lim_N \int F_N = \sum_1^\infty \int f_j.$$

□

We can use this theorem to prove a nice approximation result.

**Theorem 3.52.** *Let  $f \in L^1(\mu)$  and  $\varepsilon > 0$ . Then we can find a simple function  $\phi$  such that  $\|f - \phi\|_{L^1} < \varepsilon$ . It follows that simple functions are dense in  $(L^1(\mu), \rho)$ . Also, if  $\mu$  is a Lebesgue-Stieltjes measure on  $\mathbb{R}$ , then the sets  $E_j$  in the representation  $\phi = \sum_1^N a_j \chi_{E_j}$  can be chosen as a finite union of open intervals. Moreover, there exists a continuous function  $g$ , vanishing outside a bounded interval, such that  $\|f - g\|_{L^1} < \varepsilon$ .*

*Proof.* Take  $\{\phi_n\}$  as in [Theorem 3.21](#). If  $f$  maps into  $[0, \infty]$ , then

$$\phi_n := \sum_{k=0}^{2^{2n}-1} k 2^{-n} \chi_{E_n^k} + 2^n \chi_{F_n}, \quad E_n^k := f^{-1}\left((k 2^{-n}, (k+1) 2^{-n}]\right), \quad F_n := f^{-1}\left((2^n, \infty]\right).$$

For general  $f$ , we consider

$$(\Re f)^\pm, \quad (\Im f)^\pm.$$

Since  $|\phi_n - f| \leq 2|f|$  and  $|\phi_n - f| \rightarrow 0$ , we can apply DCT. This gives

$$\int |\phi_n - f| < \varepsilon, \quad \text{for sufficiently large } n.$$

To see the second claim, write the standard representation  $\phi_n = \sum_1^N a_j \chi_{E_j}$  where the sets  $E_j$  are disjoint and  $a_j \neq 0$ . Then

$$\mu(E_j) = \frac{1}{|a_j|} \int_{E_j} |\phi_n| \leq \frac{1}{|a_j|} \int |f| < \infty.$$

Moreover, if  $E$  and  $F$  are measurable sets, we have  $\mu(E \Delta F) = \int |\chi_E - \chi_F|$ . Hence, since  $\mu$  is a Lebesgue-Stieltjes measure on  $\mathbb{R}$ , and by [Proposition 2.53](#), we can approximate  $\chi_{E_j}$  arbitrarily close in the  $L^1$  metric by finite sums of functions  $\chi_{I_k}$  where  $I_k$  are open intervals. Finally, if  $I_k = (a, b)$ , we can approximate

---

<sup>7</sup>In the homework, we show that if  $f \in L^+$  and  $\int f < \infty$ , then  $f^{-1}(\{\infty\})$  is a null set and  $f^{-1}((0, \infty])$  is  $\sigma$ -finite.

$\chi_{I_k}$  in the  $L^1$  metric by continuous functions that vanish outside of  $(a, b)$ . For example, given  $\varepsilon > 0$ , take  $g : \mathbb{R} \rightarrow \mathbb{R}$  to be the function

$$g(x) = \begin{cases} 0 & x \in (-\infty, a] \cup [b, \infty) \\ 1 & x \in [a + \varepsilon, b - \varepsilon] \\ \frac{x-a}{a+\varepsilon} & x \in (a, a + \varepsilon) \\ \frac{x-(b-\varepsilon)}{b} & x \in (b - \varepsilon, b). \end{cases}$$

More generally, we can approximate any simple function  $\mathbb{R} \rightarrow \mathbb{C}$  by a continuous function arbitrarily close in  $L^1$ .  $\square$

**Theorem 3.53.** Suppose  $f : X \times [a, b] \rightarrow \mathbb{C}$  for  $-\infty < a < b < \infty$ , and that  $f(\cdot, t) : X \rightarrow \mathbb{C}$  is integrable for each  $t \in [a, b]$ . Let

$$F(t) = \int_X f(x, t) \, d\mu(x).$$

(i) Suppose  $\exists g \in L^1(\mu)$  such that  $|f(x, t)| \leq g(x)$  for all  $x, t$ . If  $\lim_{t \rightarrow t_0} f(x, t) = f(x, t_0)$  for all  $x$ , then

$$\lim_{t \rightarrow t_0} F(t) = F(t_0).$$

In particular, if  $f$  is dominated by some  $g \in L^1$  and  $f(x, \cdot) : [a, b] \rightarrow \mathbb{C}$  is continuous for all  $x$ , then  $F$  is continuous.

(ii) Suppose  $\partial f / \partial t$  exists and  $\exists g \in L^1$  such that  $|\frac{\partial f}{\partial t}(x, t)| \leq g(x)$  for all  $x, t$ . Then  $F$  is differentiable and

$$F'(t) = \int \frac{\partial f}{\partial t}(x, t) \, d\mu(x).$$

*Proof.* The first statment is almost trivial; we can apply DCT to any sequence of functions  $\{f(\cdot, t_n)\}_n$  for  $t_n \rightarrow t_0$ . To see the second claim, write

$$\frac{\partial f}{\partial t}(x, t_0) = \lim_{n \rightarrow \infty} h_n(x), \quad \text{where} \quad h_n(x) = \frac{f(x, t_n) - f(x, t_0)}{t_n - t_0}.$$

Notice that as the limit of measurable functions,  $\frac{\partial f}{\partial t}(x, t_0)$  is measurable. By the Mean Value Theorem,

$$|h_n(x)| < \sup_{t \in [a, b]} \left| \frac{\partial f}{\partial t}(x, t) \right| \leq g(x).$$

By DCT,

$$F'(t_0) = \lim_{n \rightarrow \infty} \frac{F(t_n) - F(t_0)}{t_n - t_0} = \lim_{n \rightarrow \infty} \int h_n(x) \, d\mu(x) = \int \frac{\partial f}{\partial t}(x, t_0) \, d\mu(x).$$

$\square$



### 3.5.1 Riemann Integration

In the special case where the measure  $\mu$  is the Lebesgue measure on  $\mathbb{R}$ , the integral is called the Lebesgue integral. If a function is integrable with respect to the Lebesgue measure, we call it Lebesgue-integrable. We now return to the theory of the Riemann-Stieltjes integrals, and compare with measure-theoretic integration.

**Definition 3.54** (Riemann integral). A *partition*  $P$  of  $[a, b]$  is a finite sequence

$$P = \{t_j\}_{j=0}^n, \text{ such that } a = t_0 < \cdots < t_n = b.$$

Associated with each partition, we define the upper and lower sums for a bounded real-valued function on  $[a, b]$  as

$$S_P f = \sum_j M_j(t_j - t_{j-1}), \quad \text{and} \quad s_P f = \sum_j m_j(t_j - t_{j-1})$$

where  $M_j = \sup_{t \in [t_{j-1}, t_j]} f(t), \quad \text{and} \quad m_j = \inf_{t \in [t_{j-1}, t_j]} f(t).$

We define the upper integral  $\bar{I}_a^b(f)$  and lower integral  $\underline{I}_a^b(f)$  as

$$\bar{I}_a^b(f) = \inf_P S_P(f), \quad \underline{I}_a^b(f) = \sup_P s_P(f).$$

If  $\bar{I}_a^b(f) = \underline{I}_a^b(f)$ , we denote the value by  $\int_a^b f(x) dx$ , and call  $f$  Riemann integrable on  $[a, b]$ .

**Theorem 3.55.** Let  $f$  be a bounded real-valued function on  $[a, b]$ .

- (i) If  $f$  is Riemann-integrable, then  $f$  is Lebesgue-integrable (and hence integrable on  $[a, b]$  since it is bounded), and

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

- (ii) The function  $f$  is Riemann-integrable if and only if the set

$$\{x \in [a, b] : f \text{ is discontinuous at } x\}$$

has measure zero.

*Proof.* We prove (a) and leave (b) to homework. Consider the simple functions

$$G_P = \sum_1^n M_j \chi_{(t_{j-1}, t_j]}, \quad g_P = \sum_1^n m_j \chi_{(t_{j-1}, t_j]}.$$

Then

$$\int G_P dm = S_P f, \quad \int g_P dm = s_P f.$$

Let  $\{P_k\}$  be a sequence of partitions with mesh size tending to zero, obtained by refining the previous partition ( $P_k \subset P_{k+1}$ ), such that  $g_{P_k}$  is increasing,  $G_{P_k}$  is decreasing, and  $S_{P_k}f, s_{P_k}f \rightarrow \int_a^b f \, dx$ . Now let

$$g = \lim_k g_{P_k}, \quad G = \lim_k G_{P_k}.$$

Then  $g \leq f \leq G$  and by DCT, we have that

$$\int_{[a,b]} g \, dm = \int_a^b f \, dx = \int_{[a,b]} G \, dm.$$

Hence

$$\int (G - g) \, dm = 0$$

and since  $G - g \geq 0$ , [Proposition 3.33](#) implies that  $G = g = f$  a.e. Since  $G$  is measurable and  $m$  is complete, [Proposition 3.23](#) implies that  $f$  is measurable and

$$\int_{[a,b]} f \, dm = \int_a^b f \, dx.$$

□

Henceforth, we will generally use the notation  $\int_a^b f \, dx$  for Lebesgue integrals.

### 3.5.2 The Gamma Function

**Definition 3.56.** Let  $z \in \mathbb{C}$  with  $\Re z > 0$ . We define

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} \, dt.$$

Consider the function  $f_z : (0, \infty) \rightarrow \mathbb{C}$ , defined by  $f_z(t) = t^{z-1} e^{-t}$ . We can write

$$t^{z-1} = \exp((z-1) \ln(t)),$$

and  $|t^{z-1}| = t^{\Re z - 1}$ , so  $|f_z(t)| \leq t^{\Re z - 1}$ . In particular,  $|f_z(t)| \leq C_z e^{-t/2}$  for  $t \geq 1$ . Since  $\int_0^1 t^a \, dt < \infty$  for  $a > -1$ , and  $\int_1^\infty e^{-t/2} \, dt < \infty$ , we see that  $f_z \in L^1((0, \infty))$  for  $\Re z > 0$ . The Gamma function is therefore well-defined.

Next, note that

$$\int_\varepsilon^N t^z e^{-t} \, dt = -t^z e^{-t} \Big|_\varepsilon^N + \int_\varepsilon^N z t^{z-1} e^{-t} \, dt.$$

Letting  $\varepsilon \rightarrow 0$  and  $N \rightarrow \infty$ , we have  $\Gamma(z+1) = z\Gamma(z)$ . We can use this identity to come up with an extension of  $\Gamma$  to all of  $\mathbb{C}$ . For  $-1 < \Re z < 0$ , define  $\Gamma(z) = \Gamma(z+1)/z$ . The result is a function defined on all of  $\mathbb{C}$  except for singularities at the nonpositive integers, where the above expression involves division by zero. We conclude by noting that  $\Gamma(1) = \int_0^\infty e^{-t} \, dt = -e^{-t} \Big|_0^\infty = 1$ , so  $n$ -fold application of the recurrence relation shows that  $\Gamma(n+1) = n!$ . The Gamma function will often appear in constants, for example the volume of the ball in  $\mathbb{R}^d$ ,

### 3.6 Modes of Convergence

**Definition 3.57** (Convergence in measure). Let  $f_n$  be measurable complex-valued functions on  $(X, \mathcal{M}, \mu)$ . We say that  $\{f_n\}$  is Cauchy in measure if  $\forall \varepsilon > 0$ ,

$$\mu(\{x : |f_n(x) - f_m(x)| \geq \varepsilon\}) \rightarrow 0, \quad \text{as } n, m \rightarrow \infty.$$

We say that  $f_n$  converges in measure to  $f$  if  $\forall \varepsilon > 0$ ,

$$\mu(\{x : |f_n(x) - f(x)| \geq \varepsilon\}) \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

**Proposition 3.58.** Convergence in measure implies Cauchy in measure.

*Proof.* We note that for any  $n, m$ , the triangle inequality gives

$$|f_n(x) - f(x)| + |f(x) - f_m(x)| \geq |f_n(x) - f_m(x)|.$$

If  $|f_n(x) - f_m(x)| \geq \varepsilon$ , then it must also hold that  $|f_n(x) - f(x)| \geq \varepsilon/2$  or  $|f(x) - f_m(x)| \geq \varepsilon/2$ . Hence,

$$\{x : |f_n(x) - f_m(x)| \geq \varepsilon\} \subset \{x : |f_n(x) - f(x)| \geq \varepsilon/2\} \cup \{x : |f(x) - f_m(x)| \geq \varepsilon/2\}.$$

Since the measure of the RHS terms goes to zero, so must the measure of the LHS.  $\square$

**Example 3.59.** Consider the Lebesgue measure.

- (i) Let  $f_n = \frac{1}{n}\chi_{(0,n)}$ . It's clear that  $f_n \rightarrow 0$  pointwise, and uniformly, but not in  $L^1$ . Also note that  $f_n$  converges to 0 in measure.
- (ii) Let  $f_n = \chi_{(n,n+1)}$ . We have pointwise convergence to 0, but not uniform convergence, and not convergence in  $L^1$ . This is not Cauchy in measure, and not convergent in measure.
- (iii) Let  $f_n = n\chi_{[0,1/n]}$ . This converges pointwise a.e., not truly pointwise, not uniformly, and not in  $L^1$ . This also converges in measure.
- (iv) Define  $f_1 = \chi_{[0,1]}$ ,  $f_2 = \chi_{[0,1/2]}$ ,  $f_3 = \chi_{[1/2,1]}$ ,  $f_4 = \chi_{[0,1/4]}$ ,  $f_5 = \chi_{[1/4,1/2]}$ . In general, take

$$f_n := \chi_{[j/2^k, (j+1)/2^k]}, \quad \text{where } n = 2^k + j, \quad 0 \leq j < 2^k.$$

It's clear that  $f_n \rightarrow 0$  in  $L^1$  and in measure. In particular,

$$\int |f_n| = 2^{-k}, \quad \text{for } 2^k \leq n < 2^{k+1}.$$

We observe that for each  $x \in [0, 1]$ , it holds that  $f_n(x) = 0$  for infinitely many  $n$  and  $f_n(x) = 1$  for infinitely many  $n$ . That is,  $f_n$  does not converge pointwise for any  $x \in [0, 1]$ .

**Proposition 3.60.** If  $f_n \rightarrow f$  in  $L^1$ . Then  $f_n \rightarrow f$  in measure.

---

<sup>8</sup>We do not assume that  $f \in L^1$ , but as we will see this is true without loss of generality.

*Proof.* We have

$$\mu(\{x : |f_n(x) - f(x)| \geq \varepsilon\}) \leq \frac{1}{\varepsilon} \int |f_n - f| \rightarrow 0.$$

□

**Remark 3.61.** The converse is not true in general, as can be seen in examples (1) and (3).

**Theorem 3.62.** Suppose  $\{f_n\}_n$  is Cauchy in measure. Then there exists a measurable function  $f$  such that  $f_n \rightarrow f$  in measure, and there exists a subsequence  $\{f_{n_j}\}_j$  that converges pointwise to  $f$  a.e. Moreover, if  $f_n \rightarrow g$  in measure, then  $f = g$  a.e.

*Proof.* We start by constructing the subsequence  $\{f_{n_j}\}_j$ , and use it to show convergence in measure. By definition of Cauchy in measure, for any  $\varepsilon > 0$ ,

$$\mu(\{x : |f_n(x) - f_m(x)| \geq \varepsilon\}) \rightarrow 0, \quad \text{as } n, m \rightarrow \infty.$$

We can choose a subsequence  $\{f_{n_j}\}$  such that for all  $j$ ,

$$\mu(\{x : |f_{n_j}(x) - f_{n_{j+1}}(x)| \geq 2^{-j}\}) \leq 2^{-j}.$$

Define  $E_j := \{x : |f_{n_j}(x) - f_{n_{j+1}}(x)| \geq 2^{-j}\}$ <sup>9</sup>, and

$$F_k = \bigcup_{j=k}^{\infty} E_j.$$

We now have

$$\mu(F_k) \leq \sum_{j=k}^{\infty} \mu(E_j) \leq \sum_{j=k}^{\infty} 2^{-j} = 2^{1-k}.$$

If  $x \notin F_k$ , then for any  $i, j \geq k$ ,

$$|f_{n_j}(x) - f_{n_i}(x)| \leq \sum_{\ell=j}^{i-1} |f_{n_{\ell+1}}(x) - f_{n_{\ell}}(x)| \leq \sum_{\ell=j}^{i-1} 2^{-\ell} = 2^{1-j}.$$

Therefore, letting  $g_j(x) = f_{n_j}(x)$ , we have  $g_j$  is pointwise Cauchy on  $F_k^c$ , and hence  $g_j$  converges pointwise in  $\mathbb{C}$ . Let  $F = \bigcap_{k=1}^{\infty} F_k = \limsup_j E_j$ . By Exercise 8 on page 27,  $\mu(F) \leq \limsup_j \mu(E_j) = 0$ . Recalling that we define  $g_j = f_{n_j}$ , set

$$f(x) = \begin{cases} \lim_{j \rightarrow \infty} f_{n_j}(x) & x \notin F \\ 0 & \text{else.} \end{cases}$$

So  $f_{n_j} \rightarrow f$  a.e. We also see that  $f$  is measurable:  $f|_{F^c}$  is measurable, so by Exercise 5 on page 48 of Folland,  $f$  is measurable.

---

<sup>9</sup>Choosing the subsequence does not require AoC. One can consider the choice function defined by selecting  $n_j$  to be the smallest natural number such that for all  $i, j \geq n_j$ ,  $\mu(\{s : |f_i(x) - f_j(x)| \geq 2^{-j}\}) \leq 2^{-j}$ . The selection of  $n_{j+1}$  must be greater than or equal to  $n_j$  by definition, so this gives a subsequence.

We now show that  $f_n \rightarrow f$  in measure. First note that for  $x \notin F_k$ , we have  $|f_{n_j}(x) - f(x)| < 2^{1-j}$  for any  $j \geq k$ . Since  $\mu(F_k) \rightarrow 0$  as  $k \rightarrow \infty$ , we see that  $f_{n_j} \rightarrow f$  in measure. On the other hand, we have

$$\{x : |f_n(x) - f(x)| \geq \varepsilon\} \subset \{x : |f_n(x) - f_{n_j}(x)| \geq \varepsilon/2\} \cup \{x : |f_{n_j}(x) - f(x)| \geq \varepsilon/2\},$$

since at least one of  $|f_n(x) - f_{n_j}(x)|$  or  $|f_{n_j}(x) - f(x)|$  must be greater than or equal to  $\varepsilon/2$ . The sets on the RHS have small measure when  $n, j$  are large, so we have convergence  $f_n \rightarrow f$  in measure.

Finally, suppose that  $f_n \rightarrow g$  in measure. Consider the set  $\{x : |f(x) - g(x)| \geq \varepsilon\}$ . By a similar observation, for any  $n$ ,

$$\{x : |f(x) - g(x)| \geq \varepsilon\} \subset \{x : |f(x) - f_n(x)| \geq \varepsilon/2\} \cup \{x : |f_n(x) - g(x)| \geq \varepsilon/2\}.$$

So  $\mu(\{x : |f(x) - g(x)| \geq \varepsilon\}) = 0$  for any  $\varepsilon > 0$ . It follows that  $f = g$  a.e.  $\square$

**Corollary 3.63.** *If  $f_n \rightarrow f$  in  $L^1$ , then there exists a subsequence  $\{f_{n_j}\}$  such that  $f_{n_j} \rightarrow f$  a.e.*

*Proof.*  $f_n \xrightarrow{L^1} f$  implies  $f_n \rightarrow f$  in measure, which implies  $f_n$  is Cauchy in measure, which implies the existence of a subsequence  $\{f_{n_j}\}$  such that  $f_{n_j} \rightarrow f$  a.e.  $\square$

If  $f_n \rightarrow f$  a.e., it does not follow that  $f_n \rightarrow f$  in measure, as example (ii) demonstrates. This conclusion does hold on a finite measure space, however.

**Theorem 3.64 (Egoroff).** *Suppose that  $\mu(X) < \infty$ , and  $f_1, f_2, \dots$  and  $f$  are measurable functions  $X \rightarrow \mathbb{C}$  such that  $f_n \rightarrow f$  a.e. Then for every  $\varepsilon > 0$ , there exists  $E \subset X$  such that  $\mu(E) < \varepsilon$  and  $f_n \rightrightarrows f$  on  $E^c$ .*

The type of convergence involved in the conclusion of this theorem is sometimes called almost uniform convergence. It is not hard to see that almost uniform convergence implies a.e. convergence and convergence in measure (Exercise 39).

*Proof.* First assume  $f_n \rightarrow f$  everywhere on  $X$ . One can see that this is without loss of generality. For  $k, n \in \mathbb{N}$ , let

$$E_n(k) := \bigcup_{m=n}^{\infty} \{x : |f_m(x) - f(x)| \geq k^{-1}\}.$$

Then for fixed  $k$ ,  $E_n(k)$  decreases as  $n$  increases, and since  $f_n \rightarrow f$ , we have  $\bigcap_{n=1}^{\infty} E_n(k) = \emptyset$ . Since  $\mu(X) < \infty$ ,  $\mu(E_1(k)) < \infty$ , and we conclude that  $\mu(E_n(k)) \rightarrow 0$  as  $n \rightarrow \infty$ . Given  $\varepsilon > 0$  and  $k \in \mathbb{N}$ , choose  $n_k$  large enough that  $\mu(E_{n_k}(k)) < \varepsilon 2^{-k}$  and let  $E = \bigcup_{k=1}^{\infty} E_{n_k}(k)$ . Then  $\mu(E) < \varepsilon$ . Moreover, for  $x \notin E$  and  $n > n_k$ , we have  $|f_n(x) - f(x)| < k^{-1}$ . So  $f_n \rightarrow f$  uniformly on  $E^c$ .  $\square$

### 3.7 Product Measures

Let  $(X, \mathcal{M}, \mu)$  and  $(Y, \mathcal{N}, \nu)$  be measure spaces. We have already discussed the product  $\sigma$ -algebra  $\mathcal{M} \otimes \mathcal{N}$  on  $X \times Y$ ; we now construct a measure on  $\mathcal{M} \otimes \mathcal{N}$ . To start, define a (measurable) rectangle to be a set of the form  $A \times B$  where  $A \in \mathcal{M}$  and  $B \in \mathcal{N}$ . We have

$$(A \times B) \cap (E \times F) = (A \cap E) \times (B \cap F), \quad (A \times B)^c = (X \times B^c) \cup (A^c \times B).$$

Using our proposition about elementary families, the collection  $\mathcal{A}$  of finite disjoint unions of rectangles is an algebra, and by [Proposition 2.17](#), the  $\sigma$ -algebra it generates is  $\mathcal{M} \otimes \mathcal{N}$ .

To start, suppose  $A \times B$  is a rectangle that is a (finite or countable) disjoint union of rectangles  $A_j \times B_j$ . Then for  $x \in X$  and  $y \in Y$ ,

$$\chi_A(x)\chi_B(y) = \chi_{A \times B}(x, y) = \sum \chi_{A_j \times B_j}(x, y) = \sum \chi_{A_j}(x)\chi_{B_j}(y).$$

Integrating with respect to  $x$  and recalling [Theorem 3.32](#), we have

$$\mu(A)\chi_B(y) = \int \chi_A(x)\chi_B(y) d\mu(x) = \sum_j \int \chi_{A_j}(x)\chi_{B_j}(y) d\mu(x) = \sum_j \mu(A_j)\chi_{B_j}(y),$$

where we adopt the usual convention that  $0 \cdot \infty = 0$ . In the same way, integration in  $y$  then yields

$$\mu(A)\mu(B) = \sum_j \mu(A_j)\nu(B_j).$$

It follows that if  $E \in \mathcal{A}$  is the disjoint union of rectangles  $A_1 \times B_1, \dots, A_n \times B_n$  and we set

$$\pi(E) = \sum_1^n \mu(A_j)\nu(B_j),$$

then  $\pi$  is well-defined on  $\mathcal{A}$  (since any two representations of  $E$  as a finite disjoint union of rectangles have a common refinement), and  $\pi$  is a premeasure on  $\mathcal{A}$ . By [Theorem 2.41](#),  $\pi$  generates an outer measure on  $X \times Y$  whose restriction to  $\mathcal{M} \otimes \mathcal{N}$  is a measure that extends  $\pi$ . We call this measure the product of  $\mu$  and  $\nu$  and denote it by  $\mu \times \nu$ .

If  $\mu, \nu$  are  $\sigma$ -finite, then  $X = \bigcup_1^\infty A_j$ ,  $Y = \bigcup_1^\infty B_j$  where  $\mu(A_j) < \infty$  and  $\nu(B_j) < \infty$ . Hence,  $X \times Y = \bigcup_{j,k} A_j \times B_k$  and  $(\mu \times \nu)(A_j \times B_k) < \infty$ . So the product measure is  $\sigma$ -finite. In this case, by [Theorem 2.41](#),  $\mu \times \nu$  is the unique measure on  $\mathcal{M} \otimes \mathcal{N}$  such that  $(\mu \times \nu)(A \times B) = \mu(A)\nu(B)$  for all rectangles  $A \times B$ .

The same construction works for any finite number of factors. Suppose that  $(X_j, \mathcal{M}_j, \mu_j)$  are measure spaces for  $j \in [n]$ . If we define a rectangle to be a set of the form  $A_1 \times \dots \times A_n$  with  $A_j \in \mathcal{M}_j$ , then the collection  $\mathcal{A}$  of finite disjoint unions of rectangles is an algebra, and the same procedure as above produces a measure  $\mu_1 \times \dots \times \mu_n$  on  $\bigotimes_1^n \mathcal{M}_j$  such that for rectangles  $A_1 \times \dots \times A_n$ ,

$$\mu_1 \times \dots \times \mu_n(A_1 \times \dots \times A_n) = \prod_1^n \mu_j(A_j).$$

Moreover, if the  $\mu_j$ 's are  $\sigma$ -finite so that the extension from  $\mathcal{A}$  to  $\bigotimes_1^n \mathcal{M}_j$  is uniquely determined, then the obvious associativity property holds.<sup>10</sup>

All of the following results have obvious extensions to products with  $n$  factors, but for simplicity, we return to the case of two measure spaces  $(X, \mathcal{M}, \mu)$  and  $(Y, \mathcal{N}, \nu)$ . If  $E \subset X \times Y$  for  $x \in X$  and  $y \in Y$ , we define the  $x$ -section  $E_x$  and  $y$ -section  $E^y$  of  $E$  by

$$E_x = \{y \in Y : (x, y) \in E\}, \quad E^y = \{x \in X : (x, y) \in E\}.$$

<sup>10</sup>For example, if we identify  $X_1 \times X_2 \times X_3$  with  $(X_1 \times X_2) \times X_3$ , we have  $\mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3 = (\mathcal{M}_1 \otimes \mathcal{M}_2) \otimes \mathcal{M}_3$  (these two being formed by different kinds rectangles), and  $\mu_1 \times \mu_2 \times \mu_3 = (\mu_1 \times \mu_2) \times \mu_3$  (since they agree on sets of the form  $A_1 \times A_2 \times A_3$ , and hence in general by uniqueness). Details are left to the reader (Exercise 45).

Also, if  $f$  is a function on  $X \times Y$ , we define the  $x$ -section  $f_x$  and  $y$ -section  $f^y$  of  $f$  by

$$f_x(y) = f^y(x) = f(x, y).$$

Thus, for example  $(\chi_E)_x = \chi_{E_x}$  and  $(\chi_E)^y = \chi_{E^y}$ .

**Proposition 3.65.** (i) If  $E \in \mathcal{M} \otimes \mathcal{N}$ , then  $E_x \in \mathcal{N} \forall x \in X$  and  $E^y \in \mathcal{M} \forall y \in Y$ .

(ii) If  $f$  is  $(\mathcal{M} \otimes \mathcal{N})$ -measurable, then  $f_x$  is  $\mathcal{N}$ -measurable  $\forall x \in X$  and  $f^y$  is  $\mathcal{M}$ -measurable  $\forall y \in Y$ .

*Proof.* Let  $\mathcal{R}$  be the set of all subsets  $E$  of  $X \times Y$  such that  $E_x \in \mathcal{N}$  for all  $x$  and  $E^y \in \mathcal{M}$  for all  $y$ . Then  $\mathcal{R}$  contains all rectangles. Since  $(\bigcup_1^\infty E_j)_x = \bigcup_1^\infty (E_j)_x$  and  $(E^c)_x = (E_x)^c$ , and likewise for  $y$ -sections,  $\mathcal{R}$  is a  $\sigma$ -algebra. Therefore,  $\mathcal{R} \supset \mathcal{M} \otimes \mathcal{N}$ , giving (i). Next, note that (ii) follows from (i) because  $(f_x)^{-1}(B) = (f^{-1}(B))_x$  and  $(f^y)^{-1}(B) = (f^{-1}(B))^y$ .  $\square$

**Example 3.66.** Let  $X = Y = [0, 1]$  and  $\mathcal{M} = \mathcal{N} = \mathcal{B}_{[0,1]}$ . Let  $m$  be the Lebesgue measure and  $\nu$  be the counting measure. Let

$$D = \{(x, x) : x \in [0, 1]\}$$

be the diagonal of  $X \times Y$ . We see that

$$\iint \chi_D(x, y) d\nu(x) dm(y) = \int 1 dm = 1, \quad \text{but} \quad \iint \chi_D(x, y) dm(y) d\nu(x) = 0.$$

What about the integral over  $X \times Y$  with respect to  $(m \times \nu)$ ? What is  $(m \times \nu)(D)$ ? Looking at the definition of the product measure, we cover  $D$  by finitely many or countably many rectangles. One can argue that at least one of the rectangles in the cover must satisfy  $m(I)\nu(J) = \infty$ .

As a technical tool for the upcoming proof, we define a monotone class on a space  $X$  to be a subset  $\mathcal{C}$  of  $2^X$  that is closed under countable increasing unions and countable decreasing intersections. Clearly, every  $\sigma$ -algebra is a monotone class. Also the intersection of any family of monotone classes is a monotone class. So for any  $\mathcal{E} \subset 2^X$ , there is a unique smallest monotone class containing  $\mathcal{E}$ , called the monotone class generated by  $\mathcal{E}$ .

**Lemma 3.67** (Monotone Class Lemma). If  $\mathcal{A} \subset 2^X$  is an algebra, then the monotone class  $\mathcal{C}$  generated by  $\mathcal{A}$  is equal to  $\mathcal{M}(\mathcal{A})$ , the  $\sigma$ -algebra generated by  $\mathcal{A}$ .

*Proof.*  $\mathcal{M}(\mathcal{A})$  is a monotone class, so  $\mathcal{C} \subset \mathcal{M}(\mathcal{A})$ . It suffices to show that  $\mathcal{C}$  is a  $\sigma$ -algebra. For  $E \in \mathcal{C}$ , define

$$\mathcal{C}(E) = \{F \in \mathcal{C} : E \setminus F, F \setminus E, E \cap F \in \mathcal{C}\}.$$

Then  $\emptyset, E \in \mathcal{C}(E)$ , and  $E \in \mathcal{C}(F)$  if and only if  $F \in \mathcal{C}(E)$ . Moreover,  $\mathcal{C}(E)$  is a monotone class. To see this, let  $\{F_j\} \subset \mathcal{C}(E)$  satisfying  $F_1 \subset F_2 \subset \dots$ , and let  $F = \bigcup F_j$ . Then  $\{E \setminus F_j = E \cap F_j^c\} \subset \mathcal{C}$  is decreasing,  $\{F_j \setminus E\}$  is increasing, and  $\{E \cap F_j\}$  is increasing. Next,

$$E \setminus F = E \cap \left(\bigcup F_j\right)^c = E \cap \left(\bigcap_1^\infty F_j^c\right) = \bigcap_1^\infty (E \cap F_j^c) \in \mathcal{C}.$$

Similarly,  $F \setminus E = \bigcup_1^\infty (F_j \setminus E) \in \mathcal{C}$ , and  $E \cap F = \bigcup_1^\infty E \cap F_j \in \mathcal{C}$ . Since  $\mathcal{C}$  is closed under countable increasing unions and countable decreasing intersections,  $\mathcal{C}(E)$  is closed under countable increasing unions. We can similarly show that  $\mathcal{C}(E)$  is closed under countable decreasing intersections. That is,  $\mathcal{C}(E)$  is a monotone class.

If  $E \in \mathcal{A}$ , note that  $F \in \mathcal{C}(E)$  for all  $F \in \mathcal{A}$  because  $\mathcal{A}$  is an algebra. That is,  $\mathcal{A} \subset \mathcal{C}(E)$  and hence  $\mathcal{C} \subset \mathcal{C}(E)$ . Therefore, if  $F \in \mathcal{C}$ , then  $F \in \mathcal{C}(E)$  for all  $E \in \mathcal{A}$ ; this means that  $E \in \mathcal{C}(F)$  for all  $E \in \mathcal{A}$ . That is,  $\mathcal{A} \subset \mathcal{C}(F)$ , and hence  $\mathcal{C} \subset \mathcal{C}(F)$ . We conclude that if  $E, F \in \mathcal{C}$ , then  $E \setminus F$  and  $E \cap F$  are in  $\mathcal{C}$ . Since  $X \in \mathcal{A} \subset \mathcal{C}$ ,  $\mathcal{C}$  is an algebra.

But then if  $\{E_j\}_1^\infty \subset \mathcal{C}$ , we have  $\bigcup_1^\infty E_j \in \mathcal{C}$  for all  $n$ , and since  $\mathcal{C}$  is closed under countable increasing unions, it follows that  $\bigcup_1^\infty E_j \in \mathcal{C}$ . So  $\mathcal{C}$  is a  $\sigma$ -algebra.  $\square$

**Theorem 3.68.** Let  $(X, \mathcal{M}, \mu)$  and  $(Y, \mathcal{N}, \nu)$  be  $\sigma$ -finite measure spaces. If  $E \in \mathcal{M} \otimes \mathcal{N}$ , then the functions  $x \mapsto \nu(E_x)$  and  $y \mapsto \mu(E^y)$  are measurable on  $X$  and  $Y$ , respectively. Moreover,

$$(\mu \times \nu)(E) = \int \nu(E_x) d\mu(x) = \int \mu(E^y) d\nu(y).$$

*Proof.* First assume that  $\mu$  and  $\nu$  are finite. Let  $\mathcal{C}$  be the set of all  $E \in \mathcal{M} \otimes \mathcal{N}$  for which the conclusions of the theorem are true. If  $E = A \times B$  is a rectangle, then  $\nu(E_x) = \chi_A(x)\nu(B)$  and  $\mu(E^y) = \mu(A)\chi_B(y)$ , so clearly  $E \in \mathcal{C}$ . By additivity, it follows that finite disjoint unions of rectangles are in  $\mathcal{C}$ . Recall that these sets generate  $\mathcal{M} \otimes \mathcal{N}$ , however. Therefore, by the Monotone Class Lemma, it suffices to show that  $\mathcal{C}$  is a monotone class.

If  $\{E_n\}$  is an increasing sequence in  $\mathcal{C}$ , and  $E = \bigcup_1^\infty E_n$ , then the functions  $f_n(y) := \mu((E_n)^y)$  are measurable and increase pointwise to  $f(y) := \mu(E^y)$ . Hence,  $f$  is measurable and by MCT,

$$\int \mu(E^y) d\nu(y) = \lim_{n \rightarrow \infty} \int \mu((E_n)^y) d\nu(y) = \lim_{n \rightarrow \infty} (\mu \times \nu)(E_n) = (\mu \times \nu)(E). \quad (2)$$

Likewise,  $(\mu \times \nu)(E) = \int \nu(E_x) d\mu(x)$ , so  $E \in \mathcal{C}$ . Similarly, if  $\{E_n\}$  is a decreasing sequence in  $\mathcal{C}$  and  $E = \bigcap_1^\infty E_n$ , the function  $y \mapsto \mu((E_1)^y)$  is in  $L^1(\nu)$  since  $\mu((E_1)^y) \leq \mu(X) < \infty$  and  $\nu(Y) < \infty$ . By continuity from above and DCT, Eq. (2) again holds. We are done for finite measure spaces.

Finally, if  $\mu$  and  $\nu$  are  $\sigma$ -finite, we can write  $X \times Y$  as the union of an increasing sequence  $\{X_j \times Y_j\}$  of rectangles of finite measure. If  $E \in \mathcal{M} \otimes \mathcal{N}$ , the preceding argument applied to  $E \cap (X_j \times Y_j)$  for each  $j$  gives

$$(\mu \times \nu)(E \cap (X_j \times Y_j)) = \int \chi_{X_j}(x) \nu(E_x \cap Y_j) d\mu(x) = \int \chi_{Y_j}(y) \mu(E^y \cap X_j) d\nu(y).$$

An application of MCT then yields the result.  $\square$

**Theorem 3.69** (Fubini-Tonelli). Let  $(X, \mathcal{M}, \mu)$  and  $(Y, \mathcal{N}, \nu)$  be  $\sigma$ -finite.

- (i) (Tonelli) If  $f \in L^+(X \times Y, \mu \times \nu)$ , then the functions  $g(x) = \int f_x d\nu$  and  $h(y) = \int f^y d\mu$  are in  $L^+(X, \mu)$  and  $L^+(Y, \nu)$ , respectively, and

$$\int f d(\mu \times \nu) = \int \left( \int f(x, y) d\nu(y) \right) d\mu(x) = \int \left( \int f(x, y) d\mu(x) \right) d\nu(y). \quad (3)$$



(ii) (Fubini) If  $f \in L^1(\mu \times \nu)$ , then  $f_x \in L^1(\nu)$  for a.e.  $x \in X$ ,  $f^y \in L^1(\mu)$  for a.e.  $y \in Y$ , the a.e.-defined functions  $g(x) = \int f_x d\nu$  and  $h(y) = \int f^y d\mu$  are in  $L^1(\mu)$  and  $L^1(\nu)$ , respectively, and Eq. (3) holds.

*Proof.* Tonelli's theorem reduces to Theorem 3.68 in the case that  $f$  is a characteristic function, and it therefore holds for nonnegative simple functions by linearity. If  $f \in L^+(X \times Y, \mu \times \nu)$ , let  $\{f_n\}$  be a sequence of simple functions that increase pointwise to  $f$  as in Theorem 3.21. MCT implies that the corresponding  $g_n$  and  $h_n$  increase to  $g$  and  $h$ , respectively, so that  $g$  and  $h$  are measurable. MCT also implies that

$$\begin{aligned} \int g d\mu &= \lim_{n \rightarrow \infty} \int g_n d\mu = \lim_{n \rightarrow \infty} \int f_n d(\mu \times \nu) = \int f d(\mu \times \nu) \\ \int h d\nu &= \lim_{n \rightarrow \infty} \int h_n d\nu = \lim_{n \rightarrow \infty} \int f_n d(\mu \times \nu) = \int f d(\mu \times \nu), \end{aligned}$$

which is Eq. (3). This establishes Tonelli's theorem, and also shows that if  $f \in L^+(X \times Y, \mu \times \nu)$  and  $\int f d(\mu \times \nu) < \infty$ , then  $g < \infty$  a.e. and  $h < \infty$  a.e. In particular,  $f_x \in L^1(Y, \nu)$  for a.e.  $x$  and  $f^y \in L^1(X, \mu)$  for a.e.  $y$ . If  $f \in L^1(X \times Y, \mu \times \nu)$ , then the conclusion of Fubini's theorem follows by applying these results to the positive and negative parts of the real and imaginary parts of  $f$ .  $\square$

**Remark 3.70.** A few notes are in order.

- We shall usually omit parentheses in the iterated integrals in Eq. (3). Thus,

$$\int \left( \int f(x, y) d\mu(x) \right) d\nu(y) = \int \int f(x, y) d\mu(x) d\nu(y) = \int \int f d\mu d\nu.$$

- The hypothesis of  $\sigma$ -finiteness is necessary; see Exercise 46 in Folland.
- The hypothesis  $f \in L^+(X \times Y, \mu \times \nu)$  or  $f \in L^1(X \times Y, \mu \times \nu)$  is necessary, in two respects. First, it is possible for  $f_x$  and  $f^y$  to be measurable for all  $x, y$  and for the iterated integrals  $\int \int f d\mu d\nu$  and  $\int \int f d\nu d\mu$  to exist even if  $f$  is not  $(\mathcal{M} \otimes \mathcal{N})$ -measurable. However, the iterated integrals need not then be equal; see Exercise 47. Second, if  $f$  is not nonnegative, it is possible for  $f_x$  and  $f^y$  to be integrable for all  $x$  and  $y$  and for the iterated integrals  $\int \int f d\mu d\nu$  and  $\int \int f d\nu d\mu$  to exist even if  $\int |f| d(\mu \times \nu) = \infty$ . Again, the iterated integrals need not be equal; see Exercise 48.
- Typical usage of this theorem consists of: since  $|f| \geq 0$ , we can apply Tonelli's theorem to check that  $|f| \in L^1$ , then we can apply Fubini to exchange the order of integration, show that  $\int |f| d(\mu \times \nu) < \infty$ , then invoke Fubini's Theorem.

Even if  $\mu$  and  $\nu$  are complete,  $\mu \times \nu$  is almost never complete.<sup>11</sup> If one wishes to work with complete measures, one can consider the completion of  $\mu \times \nu$ . In this setting, the relationship between measurability of a function on  $X \times Y$  and measurability of its  $x$ -sections and  $y$ -sections is not so simple. However, Fubini-Tonelli is still valid when suitably reformulated.

<sup>11</sup>Indeed, let  $A$  be a  $\mu$ -null set and  $E$  be a non-measurable set in  $Y$ . Then  $A \times E \subset A \times Y$ , which is a null set, but  $A \times E \notin \mathcal{M} \otimes \mathcal{N}$ . This will motivate our definition of the Lebesgue measure in higher dimensions.

**Theorem 3.71** (Fubini-Tonelli, complete measures). Let  $(X, \mathcal{M}, \mu)$  and  $(Y, \mathcal{N}, \nu)$  be complete,  $\sigma$ -finite measure spaces, and let  $(X \times Y, \mathcal{L}, \lambda)$  be the completion of  $(X \times Y, \mathcal{M} \otimes \mathcal{N}, \mu \times \nu)$ . If  $f$  is  $\mathcal{L}$ -measurable and either (a)  $f \geq 0$  or (b)  $f \in L^1(\lambda)$ , then  $f_x$  is  $\mathcal{N}$ -measurable for a.e.  $x$  and  $f^y$  is  $\mathcal{M}$ -measurable for a.e.  $y$ , and in case (b)  $f_x$  and  $f^y$  are also integrable for a.e.  $x$  and  $y$ . Moreover,  $x \mapsto \int f_x d\nu$  and  $y \mapsto \int f^y d\mu$  are measurable, and in case (b) also integrable, and

$$\int f d\lambda = \int \int f(x, y) d\mu(x) d\nu(y) = \int \int f(x, y) d\nu(y) d\mu(x).$$

*Proof.* (Folland, 1999, Exercise 49). □

### 3.8 Then $n$ -Dimensional Lebesgue Integral

**Definition 3.72.** Let  $(\mathbb{R}^n, \mathcal{L}^n, m^n)$  denote the completion of the measure space  $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n}, m \times \cdots \times m)^{12}$ . We call this the  $n$ -dimensional Lebesgue measure. When there is no danger of confusion, we dispense with the superscript  $n$ , and usually write  $\int f(x) dx$  for  $\int f dm$ . If  $E = \prod_1^n E_j$  is a rectangle in  $\mathbb{R}^n$ , then we refer to a set  $E_j$  as a side of  $E$ .

We now examine some regularity properties.

**Theorem 3.73.** Let  $E \in \mathcal{L}^n$ . Then

- (i)  $m(E) = \inf\{m(U) : U \supset E, U \text{ open}\} = \sup\{m(K) : E \supset K, K \text{ compact}\}.$
- (ii)  $E = A_1 \cup N_1 = A_2 \setminus N_2$ , where  $A_1$  is an  $F_\sigma$ ,  $A_2$  is a  $G_\delta$  set, and  $m(N_1) = m(N_2) = 0$ .
- (iii) If  $m(E) < \infty$ , then for any  $\varepsilon > 0$ , there is a finite collection  $\{R_j\}_1^N$  of disjoint rectangles whose sides are intervals such that  $m(E \Delta \bigcup_1^N R_j) < \varepsilon$ .

*Proof.* See Folland (1999, Theorem 2.40). □

**Theorem 3.74.** If  $f \in L^1(m)$  and  $\varepsilon > 0$ , there is a simple function  $\phi = \sum_1^N a_j \chi_{R_j}$ , where each  $R_j$  is a product of intervals, such that  $\int |f - \phi| < \varepsilon$ , and there is a continuous function  $g$  that vanishes outside a bounded set such that  $\int |f - g| < \varepsilon$ .

*Proof.* As in the proof of Theorem 3.52, approximate  $f$  by simple functions, then use Theorem 3.73 to approximate the latter by functions  $\phi$  of the desired form. Finally, approximate such  $\phi$ 's by continuous functions by applying an obvious generalization of the argument in the proof of Theorem 3.52. □

**Theorem 3.75.** The Lebesgue measure is translation-invariant. More precisely, for  $a \in \mathbb{R}^n$ , define  $\tau_a : \mathbb{R}^n \rightarrow \mathbb{R}^n$  by  $\tau_a(x) = x + a$ . Then

- (i) If  $E \in \mathcal{L}^n$ , then  $\tau_a(E) \in \mathcal{L}^n$  and  $m(\tau_a(E)) = m(E)$ .
- (ii) If  $f : \mathbb{R}^n \rightarrow \mathbb{C}$  is Lebesgue measurable, then so is  $f \circ \tau_a$ . Moreover, if either  $f \geq 0$  or  $f \in L^1(m)$ , then  $\int (f \circ \tau_a) dm = \int f dm$ .

---

<sup>12</sup>Equivalently, one can take the completion of  $(\mathbb{R}^n, \mathcal{L}^{\otimes n}, m \times \cdots \times m)$ . It is sometimes easier to use the former definition.

*Proof.* Since  $\tau_a$  and the inverse  $\tau_{-a}$  are continuous, they preserve the class of Borel sets. We may therefore compare the measures  $m$  and  $(m \circ \tau_a)$  over  $\mathcal{B}_{\mathbb{R}^n}$ . For a rectangle  $E$ , the formula  $m(\tau_a(E)) = m(E)$  follows from the one-dimensional result by definition of the product measure. It then follows for general Borel sets since  $m$  is induced by a premeasure, and hence determined by its action on rectangles. In particular, the collection of Borel sets  $E$  such that  $m(E) = 0$  is invariant under  $\tau_a$ , and assertion (i) follows.

If  $f$  is Lebesgue measurable and  $B$  is a Borel set in  $\mathbb{C}$ , we have  $f^{-1}(B) = E \cup N$ , where  $E$  is a Borel set and  $m(N) = 0$ . But  $\tau_a^{-1}(E)$  is Borel and  $m(\tau_a^{-1}(N)) = 0$ , so  $(f \circ \tau_a)^{-1}(B) \in \mathcal{L}^n$ , and  $f \circ \tau_a$  is Lebesgue measurable. The equality  $\int (f \circ \tau_a) d\mu = \int f d\mu$  reduces to the equality  $m(\tau_{-a}(E)) = m(E)$  when  $f = \chi_E$ . It is then true for simple functions by linearity, and hence for nonnegative measurable functions by the definition of the integral. Taking positive and negative parts of real and imaginary parts then yields the result for  $f \in L^1(m)$ .  $\square$

Now recall that a linear transformation is invertible if and only if it has a decomposition into a composition of finitely many elementary linear transformations. We use this fact to establish the following theorem.

**Theorem 3.76.** *Suppose  $T \in \text{GL}_n(\mathbb{R})$ .*

(i) *If  $f : \mathbb{R} \rightarrow \mathbb{C}$  is Lebesgue measurable, then so is  $f \circ T$ . If  $f \geq 0$  or  $f \in L^1(m)$ , then*

$$\int f dx = |\det T| \int f \circ T dx.$$

(ii) *If  $E \in \mathcal{L}^n$ , then  $T(E) \in \mathcal{L}^n$  and  $m(T(E)) = |\det T| \cdot m(E)$ .*

*Proof.* See Folland.  $\square$

**Corollary 3.77.** *The Lebesgue measure is invariant under rotations. That is, for a linear map  $T$  satisfying  $TT^* = I$ , it holds that  $m(T(E)) = m(E)$ .<sup>13</sup>*

We can actually generalize this “change of variables” result beyond linear maps. Let  $G = (g_1, \dots, g_n)$  be a map from some open set  $\Omega \subset \mathbb{R}^n$  into  $\mathbb{R}^n$  whose components are all  $C^1$ ; i.e.  $g_j$  all have continuous first order partial derivatives. We denote by  $D_x G$  the linear map defined by the matrix  $\left( (\partial g_i / \partial x_j)(x) \right)$  of partial derivatives at  $x$ . Observe that if  $G$  is linear, then  $D_x G = G$  for all  $x$ .  $G$  is called a  $C^1$  diffeomorphism if  $G$  is injective and  $D_x G$  is invertible for all  $x \in \Omega$ . In this case, the inverse function theorem guarantees that  $G^{-1} : G(\Omega) \rightarrow \Omega$  is also a  $C^1$  diffeomorphism and that  $D_x(G^{-1}) = [D_{G^{-1}(x)} G]^{-1}$  for all  $x \in G(\Omega)$ .

**Theorem 3.78.** *Suppose that  $\Omega$  is an open set in  $\mathbb{R}^n$  and  $G : \Omega \rightarrow \mathbb{R}^n$  is a  $C^1$  diffeomorphism.*

(i) *If  $f$  is a Lebesgue measurable function on  $G(\Omega)$ , then  $f \circ G$  is Lebesgue measurable on  $\Omega$ . If  $f \geq 0$ , or  $f \in L^1(G(\Omega), m)$ , then*

$$\int_{G(\Omega)} f(x) dx = \int_{\Omega} f \circ G(x) |\det D_x G| dx.$$

---

<sup>13</sup>Here,  $T^*$  denotes the transpose of  $T$ .

(ii) If  $E \subset \Omega$  and  $E \in \mathcal{L}^n$ , then  $G(E) \in \mathcal{L}^n$  and  $m(G(E)) = \int_E |\det D_x G| dx$ .

*Proof.* For those interested, be sure to have [Folland errata](#) handy. □

### 3.9 Integration in Polar Coordinates

The most important nonlinear coordinate systems in  $\mathbb{R}^2$  and  $\mathbb{R}^3$  are polar coordinates ( $x = r \cos \theta$ ,  $y = r \sin \theta$ ) and spherical coordinates ( $x = r \sin \phi \cos \theta$ ,  $y = r \sin \phi \sin \theta$ ,  $z = r \cos \phi$ ). [Theorem 3.78](#), applied to these coordinates, yields the familiar formulas (loosely stated)  $dx dy = r dr d\theta$  and  $dx dy dz = r^2 \sin \phi dr d\theta d\phi$ .

Similar coordinate systems exist in higher dimensions, but they become increasingly complicated as the dimensions increase (see Exercise 65). For most purposes, however, it suffices to know that the Lebesgue measure is effectively the product of the measure  $r^{n-1} dr$  on  $(0, \infty)$  and a certain “surface measure” on the unit sphere ( $d\theta$  for  $n = 2$  and  $\sin \phi d\theta d\phi$  for  $n = 3$ ).

**Definition 3.79.** We denote the unit sphere  $S^{n-1} = \{x \in \mathbb{R}^n : |x| = 1\}$ . If  $x \in \mathbb{R}^n \setminus \{0\}$ , the polar coordinates of  $x$  are

$$r = |x| \in (0, \infty), \quad x' = \frac{x}{|x|} \in S^{n-1}.$$

The map  $\Phi(x) = (r, x')$  is a continuous bijection  $\mathbb{R}^n \setminus \{0\} \rightarrow (0, \infty) \times S^{n-1}$  whose (continuous) inverse is  $\Phi^{-1}(r, x') = rx'$ . Indeed,  $\Phi$  is a diffeomorphism. We denote by  $m_*$  the Borel measure on  $(0, \infty) \times S^{n-1}$  induced by  $\Phi$  from the Lebesgue measure on  $\mathbb{R}^n$ , that is,  $m_*(E) := m(\Phi^{-1}(E))$ . Moreover, we define the measure  $\rho = \rho_n$  on  $(0, \infty)$  by  $\rho(E) = \int_E r^{n-1} dr$ .

**Theorem 3.80.** There is a unique Borel measure  $\sigma = \sigma_{n-1}$  on  $S^{n-1}$  such that  $m_* = \rho \times \sigma$ . If  $f$  is Borel measurable on  $\mathbb{R}^n$  and  $f \geq 0$  or  $f \in L^1(m)$ , then

$$\int_{\mathbb{R}^n} f dx = \int_0^\infty \int_{S^{n-1}} f(rx') r^{n-1} d\sigma(x') dr. \quad (4)$$

*Proof.* [Equation \(4\)](#), when  $f$  is a characteristic function of a set, is merely a restatement of the equation  $m_* = \rho \times \sigma$ , and it follows for general  $f$  by the usual linearity and approximation arguments. Hence, we need only construct  $\sigma$ .

If  $E$  is a Borel set in  $S^{n-1}$ , for  $a > 0$ , let

$$E_a = \Phi^{-1}((0, a] \times E) = \{rx' : 0 < r \leq a, x' \in E\}.$$

If  $m_* = \rho \times \sigma$  is to hold for  $E_1$ , then we must have

$$m(E_1) = m_*((0, 1] \times E) = \int_0^1 \int_E r^{n-1} d\sigma(x') dr = \sigma(E) \int_0^1 r^{n-1} dr = \frac{\sigma(E)}{n}.$$

We therefore define  $\sigma(E) := n \cdot m(E_1)$ . Since the map  $E \mapsto E_1$  takes Borel sets to Borel sets and commutes with unions, intersections, and complements, it is clear that  $\sigma$  is a Borel measure on  $S^{n-1}$ . Also, since  $E_a$  is the image of  $E_1$  under the map  $x \mapsto ax$ , it follows from [Theorem 3.76](#) that  $m(E_a) =$

$a^n m(E_1)$ , and hence, if  $0 < a < b$ ,

$$\begin{aligned} m_*((a, b] \times E) &= m(E_b \setminus E_a) = \frac{b^n - a^n}{n} \sigma(E) \\ &= \sigma(E) \int_a^b r^{n-1} dr = \rho \times \sigma((a, b] \times E). \end{aligned}$$

Now fix  $E \in \mathcal{B}_{S^{n-1}}$  and let  $\mathcal{A}_E$  be the collection of finite disjoint unions of sets of the form  $(a, b] \times E$  or  $(a, \infty) \times E$ . Since such sets form an elementary family,  $\mathcal{A}_E$  is an algebra on  $(0, \infty) \times E$  that generates the  $\sigma$ -algebra  $\mathcal{M}_E = \{A \times E : A \in \mathcal{B}_{(0, \infty)}\}$ . By the preceding calculation we have  $m_* = \rho \times \sigma$  on  $\mathcal{A}_E$ , and hence by the uniqueness assertion of [Theorem 2.41](#),  $m_* = \rho \times \sigma$  on  $\mathcal{M}_E$ . But  $\bigcup \{\mathcal{M}_E : E \in \mathcal{B}_{S^{n-1}}\}$  is precisely the set of Borel rectangles in  $(0, \infty) \times S^{n-1}$ , so another application of the uniqueness theorem shows that  $m_* = \rho \times \sigma$  on all Borel sets.  $\square$

Of course, [Eq. \(4\)](#) can be extended to Lebesgue measurable functions by considering the completion of the measure  $\sigma$ . Details are left to the reader.

**Corollary 3.81.** *If  $f$  is a measurable function on  $\mathbb{R}^n$ , nonnegative or integrable, such that  $f(x) = g(|x|)$  for some function  $g$  on  $(0, \infty)$ , then*

$$\int f(x) dx = \sigma(S^{n-1}) \int_0^\infty g(r) r^{n-1} dr.$$

**Corollary 3.82.** *Let  $c$  and  $C$  denote positive constants, and let  $B = \{x \in \mathbb{R}^n : |x| < c\}$ . Suppose that  $f$  is a measurable function on  $\mathbb{R}^n$ .*

- (i) *If  $|f(x)| \leq C|x|^{-a}$  on  $B$  for some  $a < n$ , then  $f \in L^1(B)$ . However, if  $|f(x)| \geq C|x|^{-n}$  on  $B$ , then  $f \notin L^1(B)$ .*
- (ii) *If  $|f(x)| \leq C|x|^{-a}$  on  $B^c$  for some  $a > n$ , then  $f \in L^1(B^c)$ . However, if  $|f(x)| \geq C|x|^{-n}$  on  $B^c$ , then  $f \notin L^1(B^c)$ .*

*Proof.* Apply [Corollary 3.81](#) to  $|x|^{-a} \chi_B$  and  $|x|^{-a} \chi_{B^c}$ .  $\square$

We shall compute  $\sigma(S^{n-1})$  shortly. Of course, we know that  $\sigma(S^1) = 2\pi$ ; this is just the definition of  $2\pi$  as the ratio of the circumference of a circle to its radius. Armed with this fact, we can compute a very important integral.

**Proposition 3.83.** *If  $a > 0$ ,*

$$\int_{\mathbb{R}^n} \exp(-a|x|^2) dx = \left(\frac{\pi}{a}\right)^{n/2}.$$

*Proof.* Denote the integral on the left by  $I_n$ . For  $n = 2$ , by [Corollary 3.81](#), we have

$$I_2 = 2\pi \int_0^\infty r e^{-ar^2} dr = \left(\frac{\pi}{2}\right) e^{-ar^2} \Big|_0^\infty = \frac{\pi}{a}.$$

Since  $\exp(-a|x|^2) = \prod_{j=1}^n \exp(-x_j^2)$ , Tonelli's theorem implies that  $I_n = (I_1)^n$ . In particular,  $I_1 = (I_2)^{1/2}$ , so  $I_n = (I_2)^{n/2} = (\pi/2)^{n/2}$ .  $\square$

Once we know this result, the device used in its proof can be turned around to compute  $\sigma(S^{n-1})$  for all  $n$  in terms of the gamma function.

**Proposition 3.84.**  $\sigma(S^{n-1}) = \frac{2\pi^{n/2}}{\Gamma(n/2)}.$

*Proof.* By [Corollary 3.81](#) and [Proposition 3.83](#) and the substitution  $s = r^2$ , we have

$$\begin{aligned}\pi^{n/2} &= \int_{\mathbb{R}^n} e^{-|x|^2} dx = \sigma(S^{n-1}) \int_0^\infty r^{n-1} e^{-r^2} dr \\ &= \frac{\sigma(S^{n-1})}{2} \int_0^\infty s^{(n/2)-1} e^{-s} ds = \frac{\sigma(S^{n-1})}{2} \Gamma\left(\frac{n}{2}\right).\end{aligned}$$

□

**Corollary 3.85.** If  $B^n = \{x \in \mathbb{R}^n : |x| < 1\}$ , then  $m(B^n) = \frac{\pi^{n/2}}{\Gamma(\frac{1}{2}n+1)}.$

*Proof.* By definition of  $\sigma$ , we have  $m(B^n) = n^{-1}\sigma(S^{n-1})$ . Moreover,  $\frac{1}{2}n\Gamma(\frac{1}{2}n) = \Gamma(\frac{1}{2}n+1)$  by the functional equation for the Gamma function. The equality follows immediately. □

Recall that  $\Gamma(n) = (n-1)!$ . Now, we can also evaluate at the half integers.

**Proposition 3.86.**  $\Gamma(n + \frac{1}{2}) = (n - \frac{1}{2})(n - \frac{3}{2}) \cdots (\frac{1}{2})\sqrt{\pi}.$

*Proof.* We have  $\Gamma(n + \frac{1}{2}) = (n - \frac{1}{2})(n - \frac{3}{2}) \cdots (\frac{1}{2})\Gamma(\frac{1}{2})$  by the functional equation, and by [Proposition 3.83](#) and the substitution  $s = r^2$ ,

$$\Gamma\left(\frac{1}{2}\right) = \int_0^\infty s^{-1/2} e^{-s} ds = 2 \int_0^\infty e^{-r^2} dr = \int_{-\infty}^\infty e^{-r^2} dr = \sqrt{\pi}.$$

□

An amusing consequence of [Proposition 3.86](#) and the formula  $\Gamma(n) = (n-1)!$  is that the surface measure of the unit sphere and the Lebesgue measure of the unit ball in  $\mathbb{R}^n$  are always rational multiples of integer powers of  $\pi$ , and the power of  $\pi$  increases by 1 when  $n$  increases by 2.

## 4 Signed Measures and Differentiation

Recall that for  $f \in L^+$ , we can define a new measure  $\nu : A \mapsto \int_A f d\mu$ . Moreover, if  $f \in L^1$ , the measure is finite. In some sense,  $\nu$  has density  $f$  with respect to  $\mu$ . We would like to formalize this notion.

### 4.1 Signed Measures

**Definition 4.1.** Let  $(X, \mathcal{M})$  be a measurable space. A signed measure is a function  $\nu : \mathcal{M} \rightarrow [-\infty, \infty]$  such that

- (i)  $\nu(\emptyset) = 0$ ,

- (ii)  $\nu$  takes on at most one of the values of  $\{\pm\infty\}$ ,
- (iii) if  $\{E_j\}$  is a countable disjoint sequence in  $\mathcal{M}$ , then  $\nu(\bigcup_j E_j) = \sum_j \nu(E_j)$ , where the sum converges absolutely if  $\nu(\bigcup_j E_j) < \infty$ .

**Remark 4.2.** Measures are signed measures. For emphasis, we often refer to measures as positive measures.

**Example 4.3.** We have two examples.

- (i) Let  $\mu$  and  $\nu$  be positive measures such that one of  $\mu$  or  $\nu$  is finite. Then  $\mu - \nu$  is a signed measure.
- (ii) Let  $f : X \rightarrow [-\infty, \infty]$  with at least one of  $\int f^+ d\mu < \infty$  or  $\int f^- d\mu < \infty$ , in which case we call  $f$  an extended  $\mu$ -integrable function. Then

$$\nu(E) := \int_E f d\mu$$

is a signed measure.<sup>14</sup> To see this, decompose  $f$  into  $f^+ - f^-$ , so that

$$\nu(E) = \nu_1(E) - \nu_2(E), \quad \nu_1(E) = \int_E f^+ d\mu, \quad \nu_2(E) = \int_E f^- d\mu.$$

We will see that every signed measure takes on this general form.

**Proposition 4.4.** Let  $\nu$  be a signed measure on  $(X, \mathcal{M})$ . If  $\{E_j\} \subset \mathcal{M}$  is an increasing sequence, then we have

$$\nu\left(\bigcup_1^\infty E_j\right) = \lim_{j \rightarrow \infty} \nu(E_j),$$

and if  $\{F_j\} \subset \mathcal{M}$  is a decreasing sequence and  $\nu(E_1) < \infty$ , then

$$\nu\left(\bigcap_1^\infty E_j\right) = \lim_{j \rightarrow \infty} \nu(E_j).$$

*Proof.* Exercise. □

**Definition 4.5.** Let  $\nu$  be a signed measure on  $(X, \mathcal{M})$ . Then a set  $E \in \mathcal{M}$  is called positive (resp. negative, resp. null) for  $\nu$  if

$$\begin{aligned} \nu(F) &\geq 0 \quad \text{for all } F \in \mathcal{M} : F \subset E \\ (\text{resp. } &\leq 0 \quad \text{for all } F \in \mathcal{M} : F \subset E) \\ (\text{resp. } &= 0 \quad \text{for all } F \in \mathcal{M} : F \subset E). \end{aligned}$$

**Lemma 4.6.** Any measurable subset of a positive set is positive, and the union of any countable family of positive sets is positive. This statement continues to hold if we replace the word ‘positive’ with ‘negative’ or ‘null.’

---

<sup>14</sup>Recall from our definitions that, even though  $f$  may not be integrable, we defined the quantity  $\int f d\mu$  in the more general case where at least one of  $\int f^+ d\mu < \infty$  or  $\int f^- d\mu < \infty$ .



*Proof.* The first statement follows directly from the definition of a positive set. Suppose a countable family  $\{P_n\}$  of positive sets is given. Set

$$Q_n := P_n \setminus \bigcup_{j < n} P_j \subset P_n, \text{ for } n \geq 2, \quad Q_1 = P_1.$$

Since  $Q_n \subset P_n$ ,  $Q_n$  is positive. If  $E \subset \bigcup_1^\infty P_j$ , then

$$\nu(E) = \nu\left(E \cap \left(\bigcup_1^\infty P_j\right)\right) = \sum_1^\infty \nu(E \cap Q_j) \geq 0.$$

□

**Lemma 4.7** (Hahn decomposition). *If  $\nu$  is a signed measure on  $(X, \mathcal{M})$ , then there exists a positive set  $P$  and a negative set  $N$  for  $\nu$  such that  $P \cup N = X$  and  $P \cap N = \emptyset$ . If  $P', N'$  is another pair satisfying this property, then  $P \Delta P', N \Delta N'$  are null for  $\nu$ .*

*Proof.* W.l.o.g., we may assume that  $\nu$  does not assume  $+\infty$ . Otherwise consider  $-\nu$ . Let

$$\mathcal{P} = \{E \in \mathcal{M} : E \text{ positive for } \nu\}, \quad m = \sup_{E \in \mathcal{P}} \nu(E).$$

Now let  $\{P_j\}$  be a maximizing sequence such that  $\nu(P_j) \rightarrow m$  as  $j \rightarrow \infty$ . Then by our lemma,  $P = \bigcup_j P_j$  is also positive, and by the lemma before that,  $\nu(P) = m < \infty$ .

We now claim that  $N := X \setminus P$  is negative. Suppose not. So there is a subset  $Q \in \mathcal{M}$ ,  $Q \subset N$  such that  $\nu(Q) > 0$ . **Case 1.** Suppose there is a subset  $Q$  of strictly positive measure. Then  $P \sqcup Q$  is positive, with measure greater than  $P$ . This is a contradiction. **Case 2.** Suppose no such  $Q$  exists. So we have a subset  $Q$  of postive measure, and there is a subset  $C \in \mathcal{M}$ ,  $C \subset Q$  such that  $\nu(C) < 0$ . Therefore, if we set  $B = Q \setminus C$ , we have

$$\nu(B) = \nu(Q) - \nu(C) > \nu(Q).$$

So we have found a subset of  $Q$  with measure greater than  $Q$ . Let  $n_1$  be the smallest integer such that there exists a  $B \subset N$  with  $\nu(B) > 1/n_1$ , and let  $A_1$  be such a set. Inductively, let  $n_j$  be the smallest integer such that there exists some  $B \subset A_{j-1}$  with  $\nu(B) > \nu(A_{j-1}) + 1/n_j$ , and let  $A_j$  be such a set. By the Axiom of Dependent Choice, there is a sequence  $\{A_j\}_1^\infty$  satisfying the relation for all  $j$ .

Take  $A = \bigcap_j A_j$ . Now,

$$\infty > \nu(A) = \lim_{j \rightarrow \infty} \nu(A_j) > \sum_j 1/n_j.$$

So  $n_j \rightarrow \infty$  as  $j \rightarrow \infty$ . There is some  $B \subset A$  such that  $\nu(B) > \nu(A) + 1/n$  for some  $n \in \mathbb{N}$ . For sufficiently large  $j$ , we have  $n < n_j$  and  $B \subset A \subset A_{j-1}$  by definition. This contradicts the construction of  $n_j$  and  $A_j$ , however. So  $N$  must be negative.

Now, if we have another pair  $P', N'$ , note that  $P \setminus P' \subset P$  and  $P \setminus P' \subset N'$ , so  $P \setminus P'$  is both positive and negative. Similarly, we can show that  $P' \setminus P$  is null. Hence,  $P \Delta P'$  is null. A similar argument works for  $N \Delta N'$ . □

**Definition 4.8.** *Let  $(X, \mathcal{M})$  be a measurable space with signed measure  $\nu$ . The decomposition  $X = P \cup N$  for  $P, N$  taken from [Lemma 4.7](#) is called a Hahn decomposition for  $\nu$ . Such decompositions are not unique, but as we will see, they furnish a unique representation of  $\nu$  by two positive measures.*



**Definition 4.9.** We say that two signed measures  $\mu$  and  $\nu$  on  $(X, \mathcal{M})$  are mutually singular and write  $\mu \perp \nu$  if there exists  $E, F \in \mathcal{M}$  such that  $E \cap F = \emptyset$ ,  $E \cup F = X$ ,  $E$  is null for  $\mu$ , and  $F$  is null for  $\nu$ .

**Theorem 4.10** (Jordan decomposition). If  $\nu$  is a signed measure, then there exist unique positive measures  $\nu^+, \nu^-$  such that  $\nu = \nu^+ - \nu^-$  and  $\nu^+ \perp \nu^-$ .

*Proof.* We can write a Hahn decomposition  $X = P \cup N$ . Now for any measurable set  $A \in \mathcal{M}$ , define

$$\nu^+(A) = \nu(A \cap P), \quad \nu^-(A) = -\nu(A \cap N).$$

Now, notice that  $\nu = \nu^+ - \nu^-$ , and  $\nu^+ \perp \nu^-$ .

To see uniqueness, let  $\nu = \mu^+ - \mu^-$  be another decomposition into positive measures such that  $\mu^+ \perp \mu^-$ . Let  $E, F \in \mathcal{M}$  such that  $E \cap F = \emptyset$ ,  $E \cup F = X$ , and  $\mu^+(F) = \mu^-(E) = 0$ . Since  $\mu^+$  and  $\mu^-$  are positive measures, it follows that  $E, F$  is a Hahn decomposition for  $\nu$ . Hence,  $P \Delta E$  is null for  $\nu$ , and for any  $A \in \mathcal{M}$ , we have

$$\mu^+(A) = \mu^+(A \cap E) = \nu(A \cap E) = \nu(A \cap P) = \nu^+(A).$$

Likewise,  $\nu^- = \mu^-$ . □

**Definition 4.11.** We call  $\nu = \nu^+ - \nu^-$  from [Theorem 4.10](#) the Jordan decomposition of  $\nu$ , and  $\nu^+, \nu^-$  the positive and negative variations of  $\nu$ . We define  $|\nu| = \nu^+ + \nu^-$  to be the total variation of  $\nu$ .

**Example 4.12.** If  $\nu(E) = \int_E f \, d\mu$ , then

$$\nu^+(E) = \int_E f^+ \, d\mu, \quad \nu^-(E) = \int_E f^- \, d\mu, \quad |\nu|(E) = \int_E |f| \, d\mu.$$

**Remark 4.13.** (i)  $E \in \mathcal{M}$  is  $\nu$ -null if and only if  $|\nu|(E) = 0$ .

(ii)  $\nu \perp \mu \iff |\nu| \perp \mu \iff \nu^+ \perp \mu \text{ and } \nu^- \perp \mu$ .

(iii) If  $\nu$  omits  $\infty$ , then  $\nu^+(X) = \nu(P) < \infty$ , so  $\nu^+$  is a finite measure. An analogous statement about  $\nu^-$  is true when  $\nu$  omits  $-\infty$ . Similarly, if the range of  $\nu$  is contained in  $\mathbb{R}$ , then  $\nu$  is bounded.

(iv) Let  $\nu$  be a signed measure. We have  $\nu(E) = \int_E f \, d\mu$  where  $\mu = |\nu|$ ,  $f = \chi_P - \chi_N$ , for Hahn decomposition  $X = P \cup N$  for  $\nu$ . Indeed,

$$\begin{aligned} \int_E f \, d\mu &= \int_E \chi_P - \chi_N \, d\mu = |\nu|(P \cap E) - |\nu|(N \cap E) \\ &= \nu^+(P \cap E) - \nu^-(N \cap E) = \nu^+(E) - \nu^-(E) = \nu(E). \end{aligned}$$

In other words, we can always write  $\nu$  in the form of an integral, as proposed above.

(v) We call a signed measure  $\nu$  finite (resp.  $\sigma$ -finite) if  $|\nu|$  is finite (resp.  $\sigma$ -finite).

**Definition 4.14.** For a measurable function  $f$ , if the quantity  $\int f \, d\nu^+ - \int f \, d\nu^-$  is defined, we define

$$\int f \, d\nu := \int f \, d\nu^+ - \int f \, d\nu^-.$$

We also define

$$L^1(\nu) := L^1(\nu^+) \cap L^1(\nu^-).$$

If  $f \in L^1(\nu)$ , we say that  $f$  is  $\nu$ -integrable.<sup>15</sup>

## 4.2 The Lebesgue-Radon-Nikodym Theorem

**Definition 4.15.** Let  $\nu$  be a signed measure on  $(X, \mathcal{M})$  and  $\mu$  be a positive measure on  $(X, \mathcal{M})$ . We say that  $\nu$  is absolutely continuous with respect to  $\mu$  and write  $\nu \ll \mu$  if  $\nu(E) = 0$  for all  $E \in \mathcal{M}$  where  $\mu(E) = 0$ .

**Remark 4.16.** (i) One can verify that  $\nu \ll \mu \iff |\nu| \ll \mu \iff \nu^+ \ll \mu$  and  $\nu^- \ll \mu$ .

(ii) Absolute continuity and mutual singularity are essentially the direct ‘opposite cases.’ Indeed, if  $\nu \perp \mu$  and  $\nu \ll \mu$ , then  $\nu = 0$ . To see this, note that  $\nu \perp \mu$  implies  $|\nu| \perp \mu$ . Let  $E, F \in \mathcal{M}$  such that  $E \sqcup F = X$ ,  $E$  is null for  $\mu$  and  $F$  is null for  $|\nu|$ . We have  $\mu(E) = |\nu|(F) = 0$ , and since  $|\nu| \ll \mu$ ,  $|\nu|(E) = 0$ . Hence,  $|\nu| = 0$  and  $\nu = 0$ .

**Theorem 4.17.** Let  $\nu$  be a finite signed measure and  $\mu$  be a positive measure on  $(X, \mathcal{M})$ . Then  $\nu \ll \mu$  if and only if  $\forall \varepsilon > 0, \exists \delta > 0$  such that  $|\nu(E)| < \varepsilon$  whenever  $\mu(E) < \delta$ .

*Proof.* ( $\implies$ ) The first direction is clear. If  $\mu(E) = 0$ , then for any  $\delta > 0$ ,  $\mu(E) < \delta$ . Hence,  $|\nu(E)| < \varepsilon$  for all  $\varepsilon > 0$ , and  $\nu(E) = 0$ .

( $\impliedby$ ) Now suppose  $\nu \ll \mu$ , so  $|\nu| \ll \mu$ . Since  $|\nu(E)| \leq |\nu|(E)$  for any set  $E$ , it suffices to show the  $\varepsilon$ - $\delta$  property for  $|\nu|$ . Suppose this does not hold; there exists  $\varepsilon > 0$  such that for all  $n \in \mathbb{N}$ , there exists  $E_n \in \mathcal{M}$  such that

$$\mu(E_n) < 2^{-n}, \quad \text{but} \quad |\nu|(E_n) \geq \varepsilon.$$

Let  $F_k = \bigcup_{n=k}^{\infty} E_n$  and  $F = \bigcap_{k=1}^{\infty} F_k$ . We have

$$\mu(F_k) < \sum_{n=k}^{\infty} 2^{-n} = 2^{1-k}, \quad \text{so} \quad \mu(F) = 0.$$

By finiteness of  $\nu$ , however,

$$|\nu|(F) = \lim_{n \rightarrow \infty} |\nu|(F_k) \geq \varepsilon.$$

In particular,  $|\nu| \not\ll \mu$ , and we have a contradiction. □

If  $\mu$  is a measure and  $f$  is an extended  $\mu$ -integrable function, the signed measure  $\nu$  defined by  $\nu(E) = \int_E f \, d\mu$  is clearly absolutely continuous with respect to  $\mu$ ; it is finite if and only if  $f \in L^1(\mu)$ . For any complex-valued  $f \in L^1(\mu)$ , the preceding theorem can be applied to  $\Re f$  and  $\Im f$ , and we obtain the following useful result.

**Corollary 4.18.** If  $f \in L^1(\mu)$ , then for every  $\varepsilon > 0$  there exists  $\delta > 0$  such that  $|\int_E f \, d\mu| < \varepsilon$  whenever  $\mu(E) < \delta$ .

<sup>15</sup>As before, the object  $\int f \, d\nu$  exists even when  $f \notin L^1(\nu)$ .

We introduce some notation:

We write  $d\nu = f d\mu$  to mean that  $\nu(E) = \int_E f d\mu \quad \forall E \in \mathcal{M}$ .

Sometimes, in a slight abuse of language, we shall refer to “the signed measure  $f d\mu$ .” We now come to the main theorem of this section, which gives a complete picture of the structure of signed measures relative to a given positive measure. First, a technical lemma.

**Lemma 4.19.** *Suppose that  $\nu$  and  $\mu$  are finite positive measures on  $(X, \mathcal{M})$ . Either  $\nu \perp \mu$ , or there exists some  $\varepsilon > 0$  and  $E \in \mathcal{M}$  such that  $\mu(E) > 0$  and  $\nu \geq \varepsilon\mu$  on  $E$  (that is,  $E$  is a positive set for  $\nu - \varepsilon\mu$ ).*

*Proof.* Let  $X = P_n \cup N_n$  be a Hahn decomposition for  $\nu - n^{-1}\mu$ , and let  $P = \bigcup_1^\infty P_n$  and  $N = \bigcap_1^\infty N_n = P^c$ . Then  $N$  is a negative set for  $\nu - n^{-1}\mu$  for all  $n$ ; i.e.,  $0 \leq \nu(N) \leq n^{-1}\mu(N)$  for all  $n$ , so  $\nu(N) = 0$ . If  $\mu(P) = 0$ , then  $\nu \perp \mu$ . If  $\mu(P) > 0$ , then  $\mu(P_n) > 0$  for some  $n$ , and  $P_n$  is a positive set for  $\nu - n^{-1}\mu$ .  $\square$

**Theorem 4.20** (Lebesgue-Radon-Nikodym). *Let  $\nu$  be a  $\sigma$ -finite signed measure and  $\mu$  a  $\sigma$ -finite positive measure on  $(X, \mathcal{M})$ . There exist unique  $\sigma$ -finite signed measures  $\lambda, \rho$  such that*

$$\lambda \perp \mu, \quad \rho \ll \mu, \quad \nu = \lambda + \rho.$$

*Moreover, there is an extended  $\mu$ -integrable function  $f : X \rightarrow \mathbb{R}$  such that  $d\rho = f d\mu$ , and any two such functions are equal  $\mu$ -a.e.*

*Proof. Case 1.* Suppose that  $\nu$  and  $\mu$  are finite positive measures. We start by proving that there are unique finite positive measures  $\lambda, \rho$  satisfying the desired properties. Let

$$\mathcal{F} = \left\{ f : X \rightarrow [0, \infty] : \int_E f d\mu \leq \nu(E) \quad \forall E \in \mathcal{M} \right\},$$

which is nonempty since it contains 0. Also, if  $f, g \in \mathcal{F}$ , then  $h = \max(f, g) \in \mathcal{F}$ , for if we let  $A = \{x : f(x) > g(x)\}$ , then for any  $E \in \mathcal{M}$  we have

$$\int_E h d\mu = \int_{E \cap A} f d\mu + \int_{E \setminus A} g d\mu \leq \nu(E \cap A) + \nu(E \setminus A) = \nu(E).$$

Now let  $a = \sup\{\int f d\mu : f \in \mathcal{F}\}$ , noting that  $a \leq \nu(X) < \infty$ , and choose a sequence  $\{f_n\} \in \mathcal{F}$  such that  $\int f_n d\mu \rightarrow a$ . Let  $g_n = \max(f_1, \dots, f_n)$  and  $f = \sup_n f_n$ . Then  $g_n \in \mathcal{F}$ ,  $g_n$  increases pointwise to  $f$ , and  $\int g_n d\mu \geq \int f_n d\mu$ . It follows  $\lim \int f_n d\mu = a$  and by MCT,  $f \in \mathcal{F}$  and  $\int f d\mu = a$ . In particular,  $f < \infty$  a.e., so we may take  $f$  to be  $\mathbb{R}$ -valued everywhere.

We now claim that the measure  $d\lambda = d\nu - f d\mu$  (which is positive since  $f \in \mathcal{F}$ ) is singular with respect to  $\mu$ . If not, then by Lemma 4.19, then there exists some  $\varepsilon > 0$  and  $E \in \mathcal{M}$  such that  $\mu(E) > 0$  and  $\lambda \geq \varepsilon\mu$  on  $E$ . But then  $\varepsilon d\mu \leq d\lambda = d\nu - f d\mu$ ; that is,  $(f + \varepsilon\chi_E) d\mu \leq d\nu$ , so  $f + \varepsilon\chi_E \in \mathcal{F}$  and

$$\int (f + \varepsilon\chi_E) d\mu = a + \varepsilon\mu(E) > a,$$

contradicting the definition of  $a$ .

Thus, the existence of measures  $\lambda$ ,  $f$ , and  $d\rho := f d\mu$  is proved. As for uniqueness, if it also holds that  $d\nu = d\lambda' + f' d\mu$  for finite positive measures  $\lambda'$  and  $f' d\mu$ , we have  $d\lambda - d\lambda' = (f' - f) d\mu$ . But  $\lambda - \lambda' \perp \mu$ , and  $(f' - f) d\mu \ll \mu$  (generalization of Exercise 9). Hence,  $d\lambda - d\lambda' = (f - f') d\mu = 0$ , so  $\lambda = \lambda'$  and  $f = f'$   $\mu$ -a.e. by Proposition 3.47. That is,  $\lambda$  and  $\rho$  are the unique finite positive measures satisfying our desired properties.

**Case 2.** Suppose that  $\nu$  and  $\mu$  are  $\sigma$ -finite positive measures. Then  $X$  is a countable disjoint union of  $\mu$ -finite sets and a countable disjoint union of  $\nu$ -finite sets; by taking intersections of these, we obtain a disjoint sequence  $\{A_j\} \subset \mathcal{M}$  such that  $\mu(A_j), \nu(A_j) < \infty$  for all  $j$ , and  $X = \bigcup_1^\infty A_j$ . Define  $\mu_j(E) = \mu(E \cap A_j)$  and  $\nu_j(E) = \nu(E \cap A_j)$ . By the reasoning above, for each  $j$ , we have  $d\nu_j = d\lambda_j + f_j d\mu_j$  where  $\lambda_j \perp \mu_j$ . Since  $\mu_j(A_j^c) = \nu_j(A_j^c) = 0$ , we have  $\lambda_j(A_j^c) = \nu_j(A_j^c) - \int_{A_j^c} f_j d\mu_j = 0$ , and we may assume that  $f_j = 0$  on  $A_j^c$ . Let  $\lambda = \sum_1^\infty \lambda_j$  and  $f = \sum_1^\infty f_j$ . Then one can verify that  $d\nu = d\lambda + f d\mu$ , and  $\lambda \perp \mu$  (again, Exercise 9). Moreover, since each  $d\lambda_j$  is finite, and each  $f_j d\mu$  is finite, so  $d\lambda$  and  $f d\mu$  are  $\sigma$ -finite. Uniqueness follows as before, now taking care of  $\sigma$ -finiteness. **General case.** If  $\nu$  is a signed measure (that is not necessarily positive), we apply the preceding argument to  $\nu^+$  and  $\nu^-$ , then subtract the results.  $\square$

**Definition 4.21.** Let  $\nu$  be a  $\sigma$ -finite signed measure and  $\mu$  be a  $\sigma$ -finite positive measure on  $(X, \mathcal{M})$ . From Theorem 4.20, we have a decomposition  $\nu = \lambda + \rho$  where  $\lambda \perp \mu$  and  $\rho \ll \mu$ . This is called the Lebesgue decomposition of  $\nu$  with respect to  $\mu$ .

**Remark 4.22.** • In the case where  $\nu \ll \mu$ , Theorem 4.20 says that  $d\nu = f d\mu$  for some extended  $\mu$ -integrable  $f : X \rightarrow \mathbb{R}$ . This result is usually known as the Radon-Nikodym theorem, and  $f$  is called the Radon-Nikodym derivative of  $\nu$  with respect to  $\mu$ . It is standard to denote it by  $f = d\nu/d\mu$ . We therefore often write

$$d\nu = \frac{d\nu}{d\mu} d\mu.$$

Strictly speaking,  $d\nu/d\mu$  should be constructed as the class of functions equal to  $f$   $\mu$ -a.e.

- The formulas suggested by  $d\nu/d\mu$  are generally correct. For example, in the case where  $\nu_1, \nu_2$  are signed measures such that  $\nu_1, \nu_2 \ll \mu$ , and  $\nu_1 + \nu_2$  is a signed measure,<sup>16</sup> one can see that

$$d(\nu_1 + \nu_2)/d\mu = (d\nu_1/d\mu) + (d\nu_2/d\mu).$$

**Proposition 4.23** (Chain rule). Suppose that  $\nu$  is a  $\sigma$ -finite signed measure and  $\mu, \lambda$  are  $\sigma$ -finite positive measures on  $(X, \mathcal{M})$  such that  $\nu \ll \mu$  and  $\mu \ll \lambda$ .

- (i) If  $g \in L^1(\nu)$ , then  $g \cdot (d\nu/d\mu) \in L^1(\mu)$  and

$$\int g d\nu = \int g \frac{d\nu}{d\mu} d\mu.$$

- (ii) We have  $\nu \ll \lambda$ , and

$$\frac{d\nu}{d\lambda} = \frac{d\nu}{d\mu} \frac{d\mu}{d\lambda} \quad \lambda\text{-a.e.}$$

<sup>16</sup>This allows us to avoid cases where  $\nu_1(E) = \infty, \nu_2(E) = -\infty$ .

*Proof.* By considering  $\nu^+$  and  $\nu^-$  separately, we may assume that  $\nu \geq 0$ . The equation  $\int g \, d\mu = \int g (\, d\nu / d\mu) \, d\mu$  is true when  $g = \chi_E$  by definition of  $d\nu / d\mu$ . It is therefore true for simple functions by linearity, then for nonnegative measurable functions by MCT, and finally for functions in  $L^1(\nu)$  by linearity again. This gives (a).

Now, since  $\nu$  is  $\sigma$ -finite, we can take a countable disjoint union  $E = \bigsqcup_i E_i$  such that  $\nu(E_i) < \infty$  for all  $i$ . Hence,

$$\int_E \frac{d\nu}{d\lambda} \, d\lambda = \nu(E) = \sum_1^\infty \nu(E_i) = \sum_1^\infty \int_{E_i} \frac{d\nu}{d\mu} \, d\mu = \sum_1^\infty \int_{E_i} \frac{d\nu}{d\mu} \frac{d\mu}{d\lambda} \, d\lambda.$$

Since we assumed  $\nu \geq 0$  and  $\mu, \lambda$  are positive, we can assume  $(d\nu / d\mu)(d\mu / d\lambda) \geq 0$ . We may therefore conclude

$$\int_E \frac{d\nu}{d\lambda} \, d\lambda = \int_E \frac{d\nu}{d\mu} \frac{d\mu}{d\lambda} \, d\lambda.$$

Hence, by  $\sigma$ -finiteness of  $\lambda$  and [Proposition 3.47](#), we may conclude (b).  $\square$

The following corollary can be seen as a variant of the inverse function theorem.

**Corollary 4.24.** *Let  $\mu, \lambda$  be positive measures,  $\mu \ll \lambda$  and  $\lambda \ll \mu$ . Then  $(d\lambda / d\mu)(d\mu / d\lambda) = 1$  a.e. (with respect to either  $\lambda$  or  $\mu$ ).*

**Example 4.25.** *An anti-example. Let  $\mu$  be the Lebesgue measure and  $\nu$  the point mass at 0 in  $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ . Clearly,  $\nu \perp \mu$ . The Radon-Nikodym derivative does not exist, but the theory of distributions allows us to think about this (currently vague) idea. Using this theory, one finds that the “derivative” of  $\nu$ , if it existed, would be the Dirac  $\delta$ -function.*

We conclude this section with a simple but important observation.

**Proposition 4.26.** *If  $\mu_i$  for  $i \in [n]$  are positive measures on  $(X, \mathcal{M})$ , then there is a positive measure  $\mu$  such that  $\mu_j \ll \mu$  for all  $j$ . In particular, take  $\mu = \sum_1^n \mu_j$ .*

**Remark 4.27.** *More generally, we can compare  $\mu_1, \dots, \mu_n$  by comparing their densities  $f_1, \dots, f_n \in L^+(\mu)$ , and going through the Lebesgue-Radon-Nikodym theorem.*

### 4.3 Complex Measures

**Definition 4.28.** *A complex measure on a measurable space  $(X, \mathcal{M})$  is a map  $\nu : \mathcal{M} \rightarrow \mathbb{C}$  such that*

- (i)  $\nu(\emptyset) = 0$ ;
- (ii) *if  $\{E_j\}$  is a countable sequence of disjoint sets in  $\mathcal{M}$ , then  $\nu(\bigcup_1^\infty E_j) = \sum_1^\infty \nu(E_j)$ , where the series converges absolutely.*

**Remark 4.29.** (i) *In particular, infinite values are not allowed; a positive measure is a complex measure only if it is finite.*

- (ii) *For example, if  $\mu$  is a positive measure and  $f \in L^1(\mu)$ , then  $f \, d\mu$  is a complex measure.*

(iii) If  $\nu$  is a complex measure, we shall write  $\nu_r$  and  $\nu_i$  for the real and imaginary parts of  $\nu$ . Thus  $\nu_r$  and  $\nu_i$  are signed measures that do not assume values  $\pm\infty$ ; hence they are finite, and the range of  $\nu$  is a bounded subset of  $\mathbb{C}$ .

The notions we have developed for signed measures generalize easily to complex measures.

**Definition 4.30.** We define  $L^1(\nu)$  to be  $L^1(\nu_r) \cap L^1(\nu_i)$ , and for  $f \in L^1(\nu)$ , we define  $\int f d\nu := \int f d\nu_r + i \int f d\nu_i$ . If  $\nu$  and  $\mu$  are complex measures, we say that  $\nu \perp \mu$  if  $\nu_a \perp \mu_b$  for  $a, b = r, i$ . If  $\lambda$  is a positive measure, we say that  $\nu \ll \lambda$  if  $\nu_r \ll \lambda$  and  $\nu_i \ll \lambda$ .

The theorems of the previous section generalize; one merely has to apply them separately to real and imaginary parts, whose positive and negative parts are necessarily finite.

**Theorem 4.31** (Lebesgue-Radon-Nikodym). If  $\nu$  is a complex measure and  $\mu$  is a  $\sigma$ -finite positive measure on  $(X, \mathcal{M})$ , there exists a complex measure  $\lambda$  and a function  $f \in L^1(\mu)$  such that  $\lambda \perp \mu$  and  $d\nu = d\lambda + f d\mu$ . If also  $\lambda' \perp \mu$  and  $d\nu = d\lambda' + f' d\mu$ , then  $\lambda = \lambda'$  and  $f = f'$   $\mu$ -a.e.

As before, if  $\nu \ll \mu$ , we denote the  $f$  in [Theorem 4.31](#) by  $d\nu/d\mu$ .

**Definition 4.32.** The total variation of a complex measure  $\nu$  is the positive measure  $|\nu|$  determined by the property that if  $d\nu = f d\mu$  where  $\mu$  is a  $\sigma$ -finite positive measure, then  $d|\nu| := |f| d\mu$ .

*Total variation is well-defined.* To see this, we first observe that every  $\nu$  is of the form  $f d\mu$  for some finite measure  $\mu$  and some  $f \in L^1(\mu)$ ; indeed, we can take  $\mu = |\nu_r| + |\nu_i|$  and use [Theorem 4.31](#) to obtain  $f$ . As a first case, suppose  $d\nu = f_1 d\mu_1 = f_2 d\mu_2$ , where  $\mu_1$  and  $\mu_2$  are  $\sigma$ -finite.<sup>17</sup> Let  $\rho = \mu_1 + \mu_2$ . Then by [Proposition 4.23](#), we have

$$f_1 \frac{d\mu_1}{d\rho} d\rho = d\nu = f_2 \frac{d\mu_2}{d\rho} d\rho,$$

so that  $f_1(d\mu_1/d\rho) = f_2(d\mu_2/d\rho)$   $\rho$ -a.e. Since  $\mu_1, \mu_2$  are positive measures,  $d\mu_j/d\rho$  is nonnegative, and we have

$$|f_1| \frac{d\mu_1}{d\rho} = \left| f_1 \frac{d\mu_1}{d\rho} \right| = \left| f_2 \frac{d\mu_2}{d\rho} \right| = |f_2| \frac{d\mu_2}{d\rho} \quad \rho\text{-a.e.},$$

and thus

$$|f_1| d\mu_1 = |f_1| \frac{d\mu_1}{d\rho} d\rho = |f_2| \frac{d\mu_2}{d\rho} d\rho = |f_2| d\mu_2.$$

Hence, the definition of  $|\nu|$  is independent of the choice of  $\mu$  and  $f$ .

This definition agrees with our previous definition when  $\nu$  is a signed measure, for in that case  $d\nu = (\chi_P - \chi_N) d(\nu^+ + \nu^-)$  where  $X = P \cup N$  is a Hahn decomposition, and  $|\chi_P - \chi_N| = 1$ .<sup>18</sup>  $\square$

**Proposition 4.33.** Let  $\nu$  be a complex measure on  $(X, \mathcal{M})$ .

(i)  $|\nu(E)| \leq |\nu|(E)$  for all  $E \in \mathcal{M}$ ;

<sup>17</sup>Folland ignores the case where these measures are not  $\sigma$ -finite, but it turns out to be without loss of generality. It follows from the fact that if  $\mu$  is positive and  $f \in L^1(\mu)$ , then  $\{x : f(x) \neq 0\}$  is  $\sigma$ -finite; an explanation on [Stack Exchange](#).

<sup>18</sup>Note that when  $\nu$  is signed and finite  $\chi_P - \chi_N \in L^1(|\nu|)$ , so the definitions coincide. When  $\nu$  is signed but not finite,  $\chi_P - \chi_N$  is nevertheless extended  $|\nu|$ -integrable, but the new definition of total variation does not apply.

- (ii)  $\nu \ll |\nu|$ , and  $d\nu/d|\nu|$  has absolute value 1  $|\nu|$ -a.e.;
- (iii)  $L^1(\nu) = L^1(|\nu|)$ , and if  $f \in L^1(\nu)$ , then  $|\int f d\nu| \leq \int |f| d|\nu|$ .

*Proof.* Suppose  $d\nu = f d\mu$  as in the definition of  $|\nu|$ . Then,

$$|\nu(E)| = \left| \int_E f d\mu \right| \leq \int_E |f| d\mu = |\nu|(E).$$

This proves part (a) and shows that  $\nu \ll |\nu|$ . If  $g = d\nu/d|\nu|$ , then we have  $f d\mu = d\nu = g d|\nu| = g|f| d\mu$ , so  $g|f| = f$   $\mu$ -a.e. and hence  $|\nu|$ -a.e. But clearly  $|f| > 0$   $|\nu|$ -a.e. by definition of  $|\nu|$ , so  $|g| = 1$   $|\nu|$ -a.e. Part (c) is left to Exercise 18.  $\square$

**Proposition 4.34.** *If  $\nu_1, \nu_2$  are complex measures on  $(X, \mathcal{M})$ , then  $|\nu_1 + \nu_2| \leq |\nu_1| + |\nu_2|$ .*

*Proof.* By Proposition 4.26, we can write  $\nu_j = f_j d\mu$ , with the same  $\mu$ , for  $j \in \{1, 2\}$ . Hence,

$$d(\nu_1 + \nu_2) = d\nu_1 + d\nu_2 = f_1 d\mu + f_2 d\mu = (f_1 + f_2) d\mu.$$

Hence,

$$d|\nu_1 + \nu_2| = |f_1 + f_2| d\mu \leq |f_1| d\mu + |f_2| d\mu = d|\nu_1| + d|\nu_2|.$$

$\square$

## 4.4 Differentiation on Euclidean Space

In this section, we examine the Radon-Nikodym derivative of a signed or complex measure in the special case where  $(X, \mathcal{M}) = (\mathbb{R}^n, \mathcal{B}_{\mathbb{R}})$  and  $\mu = m$  is the Lebesgue measure. Here, one can define the pointwise derivative of  $\nu$  with respect to  $m$  in the following way. Let  $B(r, x)$  be the open ball of radius  $r$  about  $x$  in  $\mathbb{R}^n$ ; consider the limit

$$F(x) = \lim_{r \rightarrow 0} \frac{\nu(B(r, x))}{m(B(r, x))}$$

when it exists. If  $\nu \ll m$ , so that  $d\nu = f dm$ , then  $\nu(B(r, x))/m(B(r, x))$  is simply the average value of  $f$  on  $B(r, x)$ , so one would hope that  $F = f$   $m$ -a.e. This will turn out to be the case when  $\nu$  satisfies general regularity properties.

For the remainder of this section, terms such as “integrable” and “almost everywhere” refer to the Lebesgue measure unless otherwise stated. We begin with a technical lemma.

**Lemma 4.35** (Vitali Covering Lemma). *Let  $\mathcal{C}$  be a collection of open balls in  $\mathbb{R}^n$ , and let  $U = \bigcup_{B \in \mathcal{C}} B$ . If  $c < m(U)$ , there exist disjoint  $B_1, \dots, B_k \in \mathcal{C}$  such that  $\sum_1^k m(B_j) > 3^{-n}c$ .*

*Proof.* If  $c < m(U)$ , there is a compact  $K \subset U$  with  $m(K) > c$ , and finitely many balls  $A_1, \dots, A_m$  that cover  $K$ . Let  $B_1$  be the largest of the  $A_j$ 's (that is, choose  $B_1$  to have maximal radius, breaking ties arbitrarily), let  $B_2$  be the largest of the  $A_j$ 's that are disjoint from  $B_1$ ,  $B_3$  be the largest of the  $A_j$ 's that are disjoint from  $B_1$  and  $B_2$ , and so on until the list of  $A_j$ 's is exhausted. According to this construction, if  $A_i$  is not one of the  $B_j$ 's, there is a  $j$  such that  $A_j \cap B_j \neq \emptyset$ , and if  $j$  is the smallest integer with this

property, the radius of  $A_i$  is at most that of  $B_j$ . Hence  $A_i \subset B_j^*$ , where  $B_j^*$  is the ball concentric with  $B_j$  whose radius is three times that of  $B_j$ . But then  $K \subset \bigcup_1^k B_j^*$ , so

$$c < m(K) \leq \sum_1^k m(B_j^*) = 3^n \sum_1^k m(B_j).$$

□

**Definition 4.36.** A measurable function  $f : \mathbb{R}^n \rightarrow \mathbb{C}$  is called *locally integrable* (with respect to the Lebesgue measure) if  $\int_K |f(x)| dx < \infty$  for every bounded measurable set  $K \subset \mathbb{R}^n$ . We denote the space of locally integrable functions by  $L_{\text{loc}}^1$ . If  $f \in L_{\text{loc}}^1$ ,  $x \in \mathbb{R}^n$ , and  $r > 0$ , we define  $A_r f(x)$  to be the average value of  $f$  on  $B(r, x)$ :

$$A_r f(x) = \frac{1}{m(B(r, x))} \int_{B(r, x)} f(y) dy.$$

**Lemma 4.37.** If  $f \in L_{\text{loc}}^1$ , then  $A_r f(x)$  is jointly continuous in  $r$  and  $x$  ( $r > 0$ ,  $x \in \mathbb{R}^n$ ).

*Proof.* We know that  $m(B(r, x)) = cr^n$  where  $c = m(B(1, 0))$ , and  $m(S(r, x)) = 0$  where  $S(r, x) = \{y : |y - x| = r\}$ . Moreover, as  $r \rightarrow r_0$  and  $x \rightarrow x_0$ , it holds that  $\chi_{B(r, x)} \rightarrow \chi_{B(r_0, x_0)}$  pointwise on  $\mathbb{R}^n \setminus S(r_0, x_0)$ . Hence,  $\chi_{B(r, x)} \rightarrow \chi_{B(r_0, x_0)}$  a.e., and  $|\chi_{B(r, x)}| \leq \chi_{B(r_0+1, x_0)}$  if  $r < r_0 + 1/2$  and  $|x - x_0| < 1/2$ . By DCT, it follows that  $\int_{B(r, x)} f(y) dy$  is continuous in  $r$  and  $x$ , and hence so is  $A_r f(x) = c^{-1}r^{-n} \int_{B(r, x)} f(y) dy$ . □

**Definition 4.38.** For  $f \in L_{\text{loc}}^1$ , we define the Hardy-Littlewood maximal function  $Hf$  by

$$Hf(x) = \sup_{r>0} A_r |f|(x) = \sup_{r>0} \left( \frac{1}{m(B(r, x))} \int_{B(r, x)} |f(y)| dy \right).$$

Note that  $Hf$  is measurable, as  $(Hf)^{-1}((a, \infty]) = \bigcup_{r>0} (A_r |f|)^{-1}((a, \infty])$  is measurable for any  $a \in \mathbb{R}$ , by the lemma above.

**Theorem 4.39** (Maximal Theorem). *There is a constant  $C > 0$  such that for all  $f \in L^1$  and  $\alpha > 0$ ,*

$$m(\{x : Hf(x) > \alpha\}) \leq \frac{C}{\alpha} \int |f(x)| dx.$$

*Proof.* Let  $E_\alpha = \{x : Hf(x) > \alpha\}$ . For each  $x \in E_\alpha$  we can choose  $r_x > 0$  such that  $A_{r_x} |f|(x) > \alpha$ . The balls  $B(r_x, x)$  cover  $E_\alpha$ , so by Vitali Covering Lemma, if  $c < m(E_\alpha)$ , there exist  $x_1, \dots, x_k \in E_\alpha$  such that the balls  $B_j = B(r_{x_j}, x_j)$  are disjoint and  $\sum_1^k m(B_j) > 3^{-n}c$ . But then

$$c < 3^{-n} \sum_1^k m(B_j) \leq \frac{3^n}{\alpha} \sum_1^k \int_{B_j} |f(y)| dy \leq \frac{3^n}{\alpha} \int_{\mathbb{R}^n} |f(y)| dy,$$

where the second inequality follows from manipulating the definition of  $A_{r_x} |f|(x)$ . Letting  $c \rightarrow m(E_\alpha)$ , we obtain the desired result. Note that the constant is the same for any  $f \in L^1$ . □



#### 4.4.1 The Fundamental Differentiation Theorem

With the tools above in hand, we now present three successively sharper versions of the fundamental differentiation theorem. In the proofs, we shall use the notion of limit superior for real-valued functions of a real variable,

$$\limsup_{r \rightarrow R} \phi(r) = \lim_{\varepsilon \rightarrow 0} \sup_{0 < |r-R| < \varepsilon} \phi(r) = \inf_{\varepsilon > 0} \sup_{0 < |r-R| < \varepsilon} \phi(r),$$

and the easily verified fact that for  $c \in \mathbb{R}$ ,

$$\lim_{r \rightarrow R} \phi(r) = c \iff \limsup_{r \rightarrow R} |\phi(r) - c| = 0.$$

**Theorem 4.40.** *If  $f \in L^1_{\text{loc}}$ , then  $\lim_{r \rightarrow 0} A_r f(x) = f(x)$  for a.e.  $x \in \mathbb{R}^n$ .*

*Proof.* It suffices to show that for  $N \in \mathbb{N}$ ,  $A_r f(x) \rightarrow f(x)$  for a.e.  $x$  with  $|x| < N$ . Indeed, one can then just take the union of null sets corresponding to each  $B(N, x)$ . But for  $|x| \leq N$  and  $r \leq 1$ , the values  $A_r f(x)$  depend only on the values  $f(y)$  for  $|y| \leq N+1$ , so replacing  $f$  with  $f \cdot \chi_{B(N+1,0)}$ , we may assume that  $f \in L^1$ .

Given  $\varepsilon > 0$ , [Theorem 3.74](#) gives a continuous integrable function  $g$  such that  $\int |g(y) - f(y)| dy < \varepsilon$ . Continuity of  $g$  implies that for every  $x \in \mathbb{R}^n$  and  $\delta > 0$  there exists  $r > 0$  such that  $|g(y) - g(x)| < \delta$  whenever  $|y - x| < r$ , and hence

$$|A_r g(x) - g(x)| = \frac{1}{m(B(r, x))} \left| \int_{B(r, x)} (g(y) - g(x)) dy \right| < \delta.$$

Therefore  $A_r g(x) \rightarrow g(x)$  as  $r \rightarrow 0$  for every  $x$ , so

$$\begin{aligned} \limsup_{r \rightarrow 0} |A_r f(x) - f(x)| &= \limsup_{r \rightarrow 0} |A_r(f - g)(x) + (A_r g - g)(x) + (g - f)(x)| \\ &\leq H(f - g)(x) + 0 + |f - g|(x). \end{aligned}$$

Hence, if we define

$$E_\alpha = \{x : \limsup_{r \rightarrow 0} |A_r f(x) - f(x)| > \alpha\}, \quad F_\alpha = \{x : |f - g|(x) > \alpha\},$$

we have

$$E_\alpha \subset F_{\alpha/2} \cup \{x : H(f - g)(x) > \alpha/2\}.$$

But  $(\alpha/2)m(F_{\alpha/2}) \leq \int_{F_{\alpha/2}} |f(x) - g(x)| dx < \varepsilon$ , so by the maximal theorem,

$$m(E_\alpha) \leq \frac{2\varepsilon}{\alpha} + \frac{2C\varepsilon}{\alpha}.$$

Since  $\varepsilon$  is arbitrary,  $m(E_\alpha) = 0$  for all  $\alpha > 0$ . To conclude, note that  $\lim_{r \rightarrow 0} A_r f(x) = f(x)$  holds whenever  $x \notin \bigcup_1^\infty E_{1/n}$ .  $\square$

This result can be rephrased as follows: if  $f \in L^1_{\text{loc}}$ , then

$$\lim_{r \rightarrow 0} \frac{1}{m(B(r, x))} \int_{B(r, x)} (f(y) - f(x)) dy = 0 \text{ for a.e. } x. \quad (5)$$

Actually, something stronger is true. Eq. (5) remains true if one replaces the integrand by its absolute value. That is, let us define the Lebesgue set  $L_f$  of  $f$  to be

$$L_f = \left\{ x : \lim_{r \rightarrow 0} \frac{1}{m(B(r, x))} \int_{B(r, x)} |f(y) - f(x)| dy = 0 \right\}.$$

**Theorem 4.41.** *If  $f \in L^1_{\text{loc}}$ , then  $m((L_f)^c) = 0$ .*

*Proof.* For each  $c \in \mathbb{C}$ , we can apply Theorem 4.40 to  $g_c(x) = |f(x) - c|$  to conclude that, except on a Lebesgue null set  $E_c$ , we have

$$\lim_{r \rightarrow 0} \frac{1}{m(B(r, x))} \int_{B(r, x)} |f(y) - c| dy = |f(x) - c|.$$

Let  $D$  be a countable dense subset of  $\mathbb{C}$ , and let  $E = \bigcup_{c \in D} E_c$ . Then  $m(E) = 0$ , and if  $x \notin E$ , then for any  $\varepsilon > 0$ , we can choose some  $c \in D$  with  $|f(x) - c| < \varepsilon$ , so that  $|f(y) - f(x)| < |f(y) - c| + \varepsilon$ . It follows that

$$\limsup_{r \rightarrow 0} \frac{1}{m(B(r, x))} \int_{B(r, x)} |f(y) - f(x)| dy \leq |f(x) - c| + \varepsilon < 2\varepsilon.$$

Since  $\varepsilon$  is arbitrary, the desired result follows.  $\square$

Finally, we consider families of sets more general than balls.

**Definition 4.42.** *A family  $\{E_r\}_{r>0}$  of Borel subsets of  $\mathbb{R}^n$  is said to shrink nicely to  $x \in \mathbb{R}^n$  if*

- (i)  $E_r \subset B(r, x)$  for each  $r$ ;
- (ii) *there is a constant  $\alpha > 0$ , independent of  $r$ , such that  $m(E_r) > \alpha m(B(r, x))$ .*

The sets  $E_r$  need not contain  $x$  itself. For example, if  $U$  is any Borel subset of  $B(1, 0)$  such that  $m(U) > 0$ , and  $E_r = \{x + ry : y \in U\}$ , the  $\{E_r\}$  shrinks nicely to  $x$ . Here, then, is the final version of the differentiation theorem.

**Theorem 4.43** (Lebesgue Differentiation Theorem). *Suppose  $f \in L^1_{\text{loc}}$ . For every  $x$  in the Lebesgue set of  $f$ —in particular, for almost every  $x$ —we have*

$$\lim_{r \rightarrow 0} \frac{1}{m(E_r)} \int_{E_r} |f(y) - f(x)| dy = 0 \quad \text{and} \quad \lim_{r \rightarrow 0} \frac{1}{m(E_r)} \int_{E_r} f(y) dy = f(x)$$

*for every family  $\{E_r\}_{r>0}$  that shrinks nicely to  $x$ .*

*Proof.* For some  $\alpha > 0$ , we have

$$\begin{aligned} \frac{1}{m(E_r)} \int_{E_r} |f(y) - f(x)| dy &\leq \frac{1}{m(E_r)} \int_{B(r, x)} |f(y) - f(x)| dy \\ &\leq \frac{1}{\alpha m(B(r, x))} \int_{B(r, x)} |f(y) - f(x)| dy. \end{aligned}$$

The first equality therefore follows from Theorem 4.41, and one sees immediately that it implies the second one by writing the latter in the form of Eq. (5).  $\square$

#### 4.4.2 Differentiation of Regular Measures

We now return to the study of measures.

**Definition 4.44.** A positive Borel measure  $\nu$  on  $\mathbb{R}^n$  will be called regular if

- (i)  $\nu(K) < \infty$  for every compact  $K$ ;
- (ii)  $\nu(E) = \inf\{\nu(U) : U \text{ open}, E \subset U\}$  for every  $E \in \mathcal{B}_{\mathbb{R}^n}$ .<sup>19</sup>

A signed or complex measure  $\nu$  is called regular if  $|\nu|$  is regular.

**Remark 4.45.** We make some observations.

- (i) By (i), every regular measure is  $\sigma$ -finite.
- (ii) If  $f \in L^+(\mathbb{R}^n)$ , the measure  $f \, dm$  is regular if and only if  $f \in L^1_{\text{loc}}$ .

We explain the second point above. The condition that  $f \in L^1_{\text{loc}}$  is clearly equivalent to (i) in the definition of regularity. Moreover, if it does hold that  $f \in L^1_{\text{loc}}$ , then (ii) in the definition of regularity may be verified as follows. First suppose that  $E$  is a Borel set bounded by the ball  $B$ . Since  $f$  is locally integrable,  $f \cdot \chi_B \in L^1(m)$ . Given  $\delta > 0$ , [Theorem 3.73](#) furnishes an open set  $U \supset E$  such that  $m(U) < m(E) + \delta$  and hence  $m(U \setminus E) < \delta$ . Moreover, we may assume  $U$  is bounded by taking  $U \cap B$ . But then, given  $\varepsilon > 0$ , by [Corollary 4.18](#) there is an open set  $U \supset E$  such that  $\int_{U \setminus E} f \, dm < \varepsilon$  and hence  $\int_U f \, dm < \int_E f \, dm + \varepsilon$ . The case where  $E$  is unbounded follows easily by writing  $E = \bigcup_1^\infty E_j$  where  $E_j$  is bounded and finding an open  $U_j \supset E_j$  such that  $\int_{U_j \setminus E_j} f \, dm < \varepsilon 2^{-j}$ .

**Theorem 4.46.** Let  $\nu$  be a regular signed or complex Borel measure on  $\mathbb{R}^n$ , and let  $d\nu = d\lambda + f \, dm$  be its Lebesgue decomposition. Then for  $m$ -a.e.  $x \in \mathbb{R}^n$ , it holds that

$$\lim_{r \rightarrow 0} \frac{\nu(E_r)}{m(E_r)} = f(x)$$

for every family  $\{E_r\}_{r>0}$  that shrinks nicely to  $x$ .

**Exercise 4.47.** Let  $\nu$  be a signed or complex measure and let  $d\nu = d\lambda + f \, dm$  be its Lebesgue

*Proof.* It is easily verified that  $d|\nu| = d|\lambda| + |f| \, dm$ <sup>20</sup>, so the regularity of  $\nu$  implies the regularity of both  $\lambda$  and  $f \, dm$  (Exercise 26). In particular,  $f \in L^1_{\text{loc}}$ , so in view of [Theorem 4.43](#), it suffices to show that if  $\lambda$  is regular and  $\lambda \perp m$ , then for  $m$ -a.e.  $x$ , it holds that  $\lambda(E_r)/m(E_r) \rightarrow 0$  as  $r \rightarrow 0$  when  $E_r$  shrinks nicely to  $x$ . It also suffices to take  $E_r = B(r, x)$  and to assume that  $\lambda$  is positive, since for some  $\alpha > 0$  we have

$$\left| \frac{\lambda(E_r)}{m(E_r)} \right| \leq \frac{|\lambda|(E_r)}{m(E_r)} \leq \frac{|\lambda|(B(r, x))}{m(E_r)} \leq \frac{|\lambda|(B(r, x))}{\alpha m(B(r, x))}.$$

<sup>19</sup>Condition (ii) is actually implied by condition (i). For  $n = 1$ , this follows from results in our study of Borel measures over  $\mathbb{R}$ . The proof for arbitrary  $n$  appears in [Folland \(1999, Chapter 7\)](#). For now, however, we assume (ii) explicitly.

<sup>20</sup>Exercise. Use the partition given by the fact that  $\lambda \perp m$ .

Assuming  $\lambda \geq 0$ , then, let  $A$  be a Borel set such that  $\lambda(A) = m(A^c) = 0$ , and let

$$F_k = \left\{ x \in A : \limsup_{r \rightarrow 0} \frac{\lambda(B(r, x))}{m(B(r, x))} > \frac{1}{k} \right\}.$$

We shall show that  $m(F_k) = 0$  for all  $k$ , and this will complete the proof.

By regularity of  $\lambda$ , given  $\varepsilon > 0$ , there is an open  $U_\varepsilon \supset A$  such that  $\lambda(U_\varepsilon) < \varepsilon$ . Each  $x \in F_k$  is the center of a ball  $B_x \subset U_\varepsilon$  such that  $\lambda(B_x) > k^{-1}m(B_x)$ . By [Lemma 4.35](#), if  $V_\varepsilon = \bigcup_{x \in F_k} B_x$  and  $c < m(V_\varepsilon)$ , then there exist  $x_1, \dots, x_J$  such that  $B_{x_1}, \dots, B_{x_J}$  are disjoint and

$$c < 3^n \sum_{j=1}^J m(B_{x_j}) \leq 2^n k \sum_{j=1}^J \lambda(B_{x_j}) \leq 3^n k \lambda(V_\varepsilon) \leq 3^n k \lambda(U_\varepsilon) \leq 3^n k \varepsilon.$$

We conclude that  $m(V_\varepsilon) \leq 3^n k \varepsilon$ , and since  $\varepsilon > 0$  is arbitrary,  $m(F_k) = 0$ . □

## 4.5 Functions of Bounded Variation

The theorems of the previous section apply in particular on the real line, where because of the correspondence between regular Borel measures and increasing functions that was established in the first chapter, they yield results about differentiation and integration of functions. As in the section on Borel measures, we adopt the notation that if  $F$  is an increasing right continuous function on  $\mathbb{R}$ ,  $\mu_F$  is the Borel measure determined by the relation  $\mu_F((a, b]) = F(b) - F(a)$ . Also, throughout this section the term “almost everywhere” will always refer to the Lebesgue measure.

Our first result uses the Lebesgue differentiation theorem to prove the a.e. differentiability of increasing functions. This result will become a useful tool as we develop a theory for functions of bounded variation. We will see that functions of bounded variation possess very similar properties.

**Theorem 4.48.** *Let  $F : \mathbb{R} \rightarrow \mathbb{R}$  be increasing, and let  $G(x) = F(x+)$ .*

(i) *The set of points at which  $F$  is discontinuous is countable.*

(ii)  *$F$  and  $G$  are differentiable a.e., and  $F' = G'$  a.e.*

*Proof.* Since  $F$  is increasing, the intervals  $(F(x-), F(x+))$  for  $x \in \mathbb{R}$  are disjoint, and for  $|x| < N$ , they lie in the interval  $(F(-N), F(N))$ . Hence,

$$\sum_{|x| < N} [F(x+) - F(x-)] \leq F(N) - F(-N) < \infty.$$

This implies that  $\{x \in (-N, N) : F(x+) - F(x-)\}$  is countable. As this is true for all  $N$ , (i) is proved.

Next, we observe that  $G$  is increasing and right continuous, and  $G = F$  except perhaps where  $F$  is discontinuous. Moreover,

$$G(x+h) - G(x) = \begin{cases} \mu_G((x, x+h]) & h > 0 \\ -\mu_G((x+h, x]) & h < 0, \end{cases}$$

and the families  $\{(x - r, x]\}$  and  $\{(x, x + r]\}$  shrink nicely to  $x$  as  $r = |h| \rightarrow 0$ . Also,  $\mu_G$  is regular by [Theorem 2.50](#). Thus, an application of [Theorem 4.46](#) to the measure  $\mu_G$  will show that  $G'(x)$  exists for a.e.  $x$ . To complete the proof, it remains to show that if  $H = G - F$ , then  $H'$  exists and equals zero a.e.

Let  $\{x_j\}$  be an enumeration of the points at which  $H \neq 0$ . Then  $H(x_j) > 0$  and as above we have  $\sum_{\{j: |x_j| < N\}} H(x_j) < \infty$  for any  $N$ . Let  $\delta_j$  be the point mass at  $x_j$  and  $\mu = \sum_j H(x_j) \delta_j$ . Then  $\mu$  is finite on compact sets by the preceding sentence, and hence  $\mu$  is regular by [Theorems 2.47](#) and [2.50](#); also  $\mu \perp m$  since  $m(E) = \mu(E^c) = 0$  where  $E = \{x_j\}_1^\infty$ . But then

$$\left| \frac{H(x+h) - H(x)}{h} \right| \leq \frac{H(x+h) - H(x)}{|h|} \leq 4 \frac{\mu((x-2|h|, x+2|h|))}{4|h|},$$

and by [Theorem 4.46](#), this tends to zero as  $h \rightarrow 0$  for a.e.  $x$ . Thus  $H' = 0$  a.e., and we are done.  $\square$

As positive Borel measures on  $\mathbb{R}$  are related to increasing functions, complex Borel measures on  $\mathbb{R}$  are related to so-called functions of bounded variation.

**Definition 4.49.** Let  $F : \mathbb{R} \rightarrow \mathbb{C}$  and  $x \in \mathbb{R}$ . We define the total variation function of  $F$  as

$$T_F(x) = \sup \left\{ \sum_1^n |F(x_j) - F(x_{j-1})| : n \in \mathbb{N}, -\infty < x_0 < \cdots < x_n = x \right\}.$$

We observe that the sums in the definition of  $T_F$  are made bigger if the additional subdivision points  $x_j$  are added. Hence, if  $a < b$ , the definition of  $T_F(b)$  is unaffected if we assume that  $a$  is always one of the subdivision points. It follows that

$$T_F(b) - T_F(a) = \sup \left\{ \sum_1^n |F(x_j) - F(x_{j-1})| : n \in \mathbb{N}, a = x_0 < \cdots < x_n = b \right\}. \quad (6)$$

Thus  $T_F$  is an increasing function with values in  $[0, \infty]$ .

**Definition 4.50.** If  $T_F(\infty) = \lim_{x \rightarrow \infty} T_F(x)$  is finite, we say that  $F$  is of bounded variation on  $\mathbb{R}$ , and denote the space of all such  $F$  as  $BV$ . More generally, the supremum on the RHS of [Eq. \(6\)](#) is called the total variation of  $F$  on  $[a, b]$ . It depends only on the values of  $F$  on  $[a, b]$ , so we may define  $BV([a, b])$  to be the set of all functions on  $[a, b]$  whose total variation on  $[a, b]$  is finite.

**Remark 4.51.** If  $F \in BV$ , then the restriction of  $F$  to  $[a, b]$  is in  $BV([a, b])$  for all  $a, b$ ; conversely, if  $F \in BV([a, b])$  and we set  $F(x) = F(a)$  for  $x < a$  and  $F(x) = F(b)$  for  $x > b$ , then  $F \in BV$ . By this device, the results that we prove for  $BV$  can also be applied to  $BV([a, b])$ .

**Example 4.52.** We leave the verification of the following to [Folland \(1999, Exercise 27\)](#).

- (i) If  $F : \mathbb{R} \rightarrow \mathbb{R}$  is bounded and increasing, then  $F \in BV$  (in fact,  $T_F(x) = F(x) - F(-\infty)$ ).
- (ii) If  $F, G \in BV$  and  $a, b \in \mathbb{C}$ , then  $aF + bG \in BV$ .
- (iii) If  $F$  is differentiable on  $\mathbb{R}$  and  $F'$  is bounded, then  $F \in BV([a, b])$  for  $-\infty < a < b < \infty$ , by the mean value theorem.

(iv) If  $F(x) = \sin(x)$ , then  $F \in \text{BV}([a, b])$  for  $-\infty < a < b < \infty$ , but  $F \notin \text{BV}$ .

(v) If  $F(x) = x \sin(x^{-1})$  for  $x \neq 0$  and  $F(0) = 0$ , then  $F \notin \text{BV}([a, b])$  for  $a \leq 0 < b$  or  $a < 0 \leq b$ .

**Lemma 4.53.** If  $F \in \text{BV}$  is real-valued, then  $T_F + F$  and  $T_F - F$  are increasing.

*Proof.* If  $x < y$  and  $\varepsilon > 0$ , choose  $x_0 < \cdots < x_n = x$  such that

$$\sum_1^n |F(x_j) - F(x_{j-1})| \geq T_F(x) - \varepsilon.$$

Then  $\sum_1^n |F(x_j) - F(x_{j-1})| + |F(y) - F(x)|$  is an approximating sum for  $T_F(y)$ , and  $F(y) = [F(y) - F(x)] + F(x)$ , so we have

$$\begin{aligned} T_F(y) \pm F(y) &\geq \sum_1^n |F(x_j) - F(x_{j-1})| + |F(y) - F(x)| \pm [F(y) - F(x)] \pm F(x) \\ &\geq T_F(x) - \varepsilon \pm F(x). \end{aligned}$$

Since  $\varepsilon$  is arbitrary,  $T_F(y) \pm F(y) \geq T_F(x) \pm F(x)$ . □

We now establish some valuable tools for the further examination of BV.

**Theorem 4.54.** (i)  $F \in \text{BV}$  if and only if  $\Re F \in \text{BV}$  and  $\Im F \in \text{BV}$ .

(ii) If  $F : \mathbb{R} \rightarrow \mathbb{R}$ , then  $F \in \text{BV}$  if and only if  $F$  is the difference of two bounded increasing functions; for  $F \in \text{BV}$  these functions may be taken to be  $\frac{1}{2}(T_F + F)$  and  $\frac{1}{2}(T_F - F)$ .

(iii) If  $F \in \text{BV}$ , then  $F(x+) = \lim_{y \searrow x} F(y)$  and  $F(x-) = \lim_{y \nearrow x} F(y)$  exist for all  $x \in \mathbb{R}$ , as do  $F(\pm\infty) = \lim_{y \rightarrow \pm\infty} F(y)$ .

(iv) If  $F \in \text{BV}$ , the set of points at which  $F$  is discontinuous is countable.

(v) If  $F \in \text{BV}$ , and  $G(x) = F(x+)$ , then  $F'$  and  $G'$  exist a.e. and are equal a.e.

*Proof.* (i) is clear. For (ii), the “if” implication is clear from the fact that BV is a complex vector space. To prove the “only if,” observe that by [Lemma 4.53](#), the equation  $F = \frac{1}{2}(T_F + F) - (T_F - F)$  expresses  $F$  as the difference of two increasing functions. Also, the inequalities

$$T_F(y) \pm F(y) \geq T_F(x) \pm F(x) \quad (y > x)$$

imply that

$$|F(y) - F(x)| \leq T_F(y) - T_F(x) \leq T_F(\infty) - T_F(-\infty) < \infty,$$

so that  $F$ , and hence  $T_F \pm F$ , is bounded. Finally, (c), (d), and (e) follows from (a), (b), and [Theorem 4.48](#). □

**Definition 4.55.** The representation  $F = \frac{1}{2}(T_F + F) - \frac{1}{2}(T_F - F)$  of a real-valued  $F \in \text{BV}$  is called the Jordan decomposition of  $F$ , and  $\frac{1}{2}(T_F + F)$  and  $\frac{1}{2}(T_F - F)$  are called the positive and negative variations of  $F$ .

Since

$$x^+ := \max\{x, 0\} = \frac{1}{2}(|x| + x) \quad \text{and} \quad x^- = \max\{-x, 0\} = \frac{1}{2}(|x| - x) \quad \text{for } x \in \mathbb{R},$$

we have

$$\frac{1}{2}(T_F \pm F)(x) = \sup \left\{ \sum_{j=1}^n [F(x_j) - F(x_{j-1})]^\pm : x_0 < \cdots < x_n = x \right\} \pm \frac{1}{2}F(-\infty).$$

[Theorem 4.54](#) (a,b) leads to the connection between BV and the space of complex Borel measures on  $\mathbb{R}$ . To make this precise, we introduce the “normalized” space NBV defined by

$$\text{NBV} = \{F \in \text{BV} : F \text{ is right continuous and } F(-\infty) = 0\}.$$

We observe that if  $F \in \text{BV}$ , then the function  $G$  defined by  $G(x) = F(x+) - F(-\infty)$  is in NBV and  $G' = F'$  a.e. To see that  $G \in \text{BV}$ , use [Theorem 4.54](#) (a, b): if  $F$  is real and  $F = F_1 - F_2$  where  $F_2$  are bounded and increasing, then  $G(x) = F_1(x+) - [F_2(x+) + F(-\infty)]$ , which is again the difference of two bounded increasing functions.

**Lemma 4.56.** *If  $F \in \text{BV}$ , then  $T_F(-\infty) = 0$ . If  $F$  is also right continuous, then so is  $T_F$ .*

*Proof.* If  $\varepsilon > 0$  and  $x \in \mathbb{R}$ , choose  $x_0 < \cdots < x_n = x$  so that

$$\sum_{j=1}^n |F(x_j) - F(x_{j-1})| \geq T_F(x) - \varepsilon.$$

From [Eq. \(6\)](#) we see that  $T_F(x) - T_F(x_0) \geq T_F(x) - \varepsilon$ , and hence  $T_F(y) \leq \varepsilon$  for  $y \leq x_0$ . Thus  $T_F(-\infty) = 0$ .

Now suppose that  $F$  is right continuous. Given  $x \in \mathbb{R}$  and  $\varepsilon > 0$ , let  $\alpha = T_F(x+) - T_F(x)$ , and choose  $\delta > 0$  so that  $|F(x+h) - F(x)| < \varepsilon$  and  $T_F(x+h) - T_F(x) < \varepsilon$  whenever  $h \in (0, \delta)$ . For any such  $h$ , by [Eq. \(6\)](#), there exist  $x = x_0 < \cdots < x_n = x+h$  such that

$$\sum_{j=1}^n |F(x_j) - F(x_{j-1})| \geq \frac{3}{4}[T_F(x+h) - T_F(x)] \geq \frac{3}{4}\alpha,$$

and hence

$$\sum_{j=2}^n |F(x_j) - F(x_{j-1})| \geq \frac{3}{4}\alpha - |F(x_1) - F(x_0)| \geq \frac{3}{4}\alpha - \varepsilon.$$

Likewise, there exist  $x = t_0 < \cdots < t_m = x_1$  such that  $\sum_{j=1}^m |F(t_j) - F(t_{j-1})| \geq \frac{3}{4}\alpha$ , and hence

$$\begin{aligned} \alpha + \varepsilon &> T_F(x+h) - T_F(x) + T_F(x) - T_F(x) = T_F(x+h) - T_F(x) \\ &\geq \sum_{j=1}^m |F(t_j) - F(t_{j-1})| + \sum_{j=1}^n |F(x_j) - F(x_{j-1})| \\ &\geq \frac{3}{2}\alpha - \varepsilon. \end{aligned}$$

Thus,  $\alpha < 4\varepsilon$ , and since  $\varepsilon$  is arbitrary,  $\alpha = 0$ . □



**Theorem 4.57.** If  $\mu$  is a complex Borel measure on  $\mathbb{R}$  and  $F(x) = \mu((-\infty, x])$ , then  $F \in \text{NBV}$ . Conversely, if  $F \in \text{NBV}$ , there is a unique complex Borel measure  $\mu_F$  such that  $F(x) = \mu_F((-\infty, x])$ , and  $|\mu_F| = \mu_{T_F}$ .

*Proof.* If  $\mu$  is a complex measure, we have  $\mu = \mu_1^+ - \mu_1^- + i(\mu_2^+ - \mu_2^-)$  where the  $\mu_j^\pm$  are finite measures. If  $F_j^\pm(x) = \mu_j^\pm((-\infty, x])$ , then  $F_j^\pm$  is increasing and right continuous,  $F_j^\pm(-\infty) = 0$ , and  $F_j^\pm(\infty) = \mu_j^\pm(\mathbb{R}) < \infty$ . By Theorem 4.54(a,b), the function  $F = F_1^+ - F_1^- + i(F_2^+ - F_2^-)$  is in NBV. Conversely, by Theorem 4.54 and Lemma 4.56, any  $F \in \text{NBV}$  can be written in the this form with the  $F_j^\pm$  increasing and in NBV. Each  $F_j^\pm$  gives rise to a measure  $\mu_j^\pm$  according to Theorem 2.47, so  $F(x) = \mu_F((-\infty, x])$  where  $\mu_F = \mu_1^+ - \mu_1^- + i(\mu_2^+ - \mu_2^-)$ . Uniqueness follows by splitting into real and complex, positive and negative parts, then using the same argument as in Theorem 2.47. The proof that  $|\mu_F| = \mu_{T_F}$  is outlined in Exercise 28.  $\square$

A natural question: which functions in NBV correspond to measures  $\mu$  such that  $\mu \perp m$  or  $\mu \ll m$ ?

**Proposition 4.58.** If  $F \in \text{NBV}$ , then  $F' \in L^1(m)$ . Moreover,  $\mu_F \perp m$  if and only if  $F' = 0$  a.e., and  $\mu_F \ll m$  if and only if  $F(x) = \int_{-\infty}^x F'(t) dt$ .

*Proof.* We first establish that  $\mu_F$  is a regular complex measure. By Theorem 4.57, we have  $|\mu_F| = \mu_{T_F}$ . Since  $F \in \text{NBV}$ ,  $F$  is right continuous, and  $T_F$  is right continuous and increasing, so  $\mu_{T_F}$  can be taken as a Lebesgue-Stieltjes Borel measure. By Theorem 2.50,  $\mu_{T_F}$  is regular, so  $\mu_F$  is regular.

Next, note that the sets  $E_r = (x, x + r]$  shrink nicely to  $x$ , and observe that

$$F'(x) = \lim_{r \rightarrow 0} \frac{F(x+r) - F(x)}{r} = \lim_{r \rightarrow 0} \frac{\mu_F(E_r)}{m(E_r)} = f \text{ a.e.},$$

where  $d\mu_F = d\lambda + f dm$  is the Lebesgue-Radon-Nikodym representation of  $\mu_F$ . Since  $T_F \in \text{NBV}$ ,  $|\mu_F| = \mu_{T_F}$  is finite, so  $f \in L^1(m)$ . It is now clear that  $\mu_F \perp m$  if and only if  $F' = 0$  a.e. Next, if  $\mu_F \ll m$ , we have  $d\mu_F = F' dm$  a.e., which gives  $F(x) = \int_{-\infty}^x F'(t) dt$ . Conversely, if this holds, by splitting into real and complex, positive and negative parts, we have equality on a generating algebra, which gives  $d\mu_F = F' dm$ , and  $\mu_F \ll m$ .  $\square$

The condition  $\mu_F \ll m$  can also be expressed directly in terms of  $F$ , as follows.

**Definition 4.59.** A function  $F : \mathbb{R} \rightarrow \mathbb{C}$  is called *absolutely continuous* if for every  $\varepsilon > 0$ , there exists  $\delta > 0$  such that for any finite set of disjoint intervals  $\{(a_i, b_i)\}_1^n$ ,

$$\sum_1^n (b_j - a_j) < \delta \implies \sum_1^n |F(b_j) - F(a_j)| < \varepsilon. \quad (7)$$

More generally,  $F$  is said to be *absolutely continuous on  $[a, b]$*  if this condition is satisfied whenever the intervals  $(a_j, b_j)$  all lie in  $[a, b]$ .

Clearly, if  $F$  is absolutely continuous, then  $F$  is uniformly continuous (take  $n = 1$  is Eq. (7)). On the other hand, if  $F$  is everywhere differentiable and  $F'$  is bounded, then  $F$  is absolutely continuous, for  $|F(b_j) - F(a_j)| \leq (\max |F'|)(b_j - a_j)$  by the mean value theorem.

**Proposition 4.60.** If  $F \in \text{NBV}$ , then  $F$  is absolutely continuous if and only if  $\mu_F \ll m$ .



*Proof.* If  $\mu_F \ll m$ , we can split  $\mu_F$  into real and complex, positive and negative parts. The functions  $F_j^\pm$  corresponding to these parts will be increasing and right continuous. By [Theorem 4.17](#), we can choose  $\delta$  so that  $m(E) < \delta$  implies  $\mu_{F_j^\pm}(E) < \varepsilon$ . That is, if  $\{(a_j, b_j)\}_1^n$  is a finite set of disjoint intervals such that  $\sum_1^n b_j - a_j < \delta$ , it holds that for  $E = \bigcup_1^n (a_k, b_k]$ , we have  $\mu_{F_j^\pm}(E) = \sum_1^n F_j^\pm(b_k) - F_j^\pm(a_k) < \varepsilon$ . We can then bound the  $\sum_1^n |F(b_k) - F(a_k)|$ .

To prove the converse, suppose that  $E$  is a Borel set such that  $m(E) = 0$ . If  $\varepsilon$  and  $\delta$  are as in the definition of absolute continuity of  $F$ , by [Theorem 2.50](#), we can find a decreasing sequence of open sets  $U_1 \supset U_2 \supset \dots \supset E$  such that  $m(U_1) < \delta$  (and thus  $m(U_j) < \delta$  for all  $j$ ) and  $\mu_F(U_j) \rightarrow \mu_F(E)$ .<sup>21</sup> Now each  $U_j$  is a disjoint union of open intervals  $(a_k^j, b_k^j)$ , and for any  $N \in \mathbb{N}$ ,

$$\sum_{k=1}^N |\mu_F((a_k^j, b_k^j))| \leq \sum_{k=1}^N |F(b_k^j) - F(a_k^j)| < \varepsilon.$$

Letting  $N \rightarrow \infty$ ,  $|\mu_F(U_j)| \leq \varepsilon$  and hence  $|\mu_F(E)| \leq \varepsilon$ . Since  $\varepsilon$  is arbitrary,  $\mu_F(E) = 0$ , which shows that  $\mu_F \ll m$ .  $\square$

**Corollary 4.61.** *If  $f \in L^1(m)$ , then the function  $F(x) = \int_{-\infty}^x f(t) dt$  is in NBV and is absolutely continuous, and  $f = F'$  a.e. Conversely, if  $F \in \text{NBV}$  is absolutely continuous, then  $F' \in L^1(m)$  and  $F(x) = \int_{-\infty}^x F'(t) dt$ .*

*Proof.* This follows from [Propositions 4.58](#) and [4.60](#). We start with the second direction. If  $F \in \text{NBV}$  is absolutely continuous, then  $F' \in L^1(m)$  by [Proposition 4.58](#). Moreover,  $\mu_F \ll m$  by [Proposition 4.60](#), so  $F(x) = \int_{-\infty}^x F'(t) dt$  by [Proposition 4.58](#).

Alternatively, if  $f \in L^1(m)$ , we can split into real and complex, positive and negative parts. All of these are bounded and increasing, and must lie in BV. Using DCT, we also have

$$\lim_{x \searrow y} \int_{-\infty}^x f dt = \int_{-\infty}^y f dt, \quad \lim_{x \rightarrow -\infty} \int_{-\infty}^x f dt = 0.$$

That is,  $F(x) \in \text{NBV}$ . It's clear that  $\mu_F \ll m$ , so  $F$  is absolutely continuous by [Proposition 4.60](#) and  $F' = f$  a.e. by [Proposition 4.58](#).  $\square$

If we consider functions on bounded intervals, we can slightly refine the result above.

**Lemma 4.62.** *If  $F$  is absolutely continuous on  $[a, b]$ , then  $F \in \text{BV}([a, b])$ .*

*Proof.* Let  $\delta$  be as in the definition of absolute continuity, corresponding to  $\varepsilon = 1$ . Let  $N$  be the smallest integer greater than  $\delta^{-1}(b - a)$ . If  $a = x_0 < \dots < x_n = b$ , by inserting more subdivision points if necessary, we can collect the intervals  $(x_{j-1}, x_j)$  into at most  $N$  groups of consecutive intervals such that the sum of the lengths in each group is less than  $\delta$ . The sum  $\sum |F(x_j) - F(x_{j-1})|$  over each group is at most 1, and hence the total variation of  $F$  on  $[a, b]$  is at most  $N$ .  $\square$

**Theorem 4.63** (Fundamental Theorem of Calculus for Lebesgue Integrals). *If  $-\infty < a < b < \infty$  and  $F : [a, b] \rightarrow \mathbb{C}$ , the following are equivalent:*

<sup>21</sup>To see this, split  $\mu_F$  into  $\mu_i^\pm$ . By [Theorem 2.50](#), we can find a sequence of decreasing open sets  $\{U_{i,j}^\pm\}_i$  containing  $E$  such that  $\mu_i^\pm(U_{i,j}^\pm) \rightarrow \mu_i^\pm(E)$  (by finiteness of each variation). Taking  $U_j$  to be the intersection of each  $U_{i,j}^\pm$ , we have the desired limit condition. To ensure that  $m(U_1) < \delta$ , just take intersections with an open set  $O$  containing  $E$  such that  $m(O) < \delta$ .

(i)  $F$  is absolutely continuous on  $[a, b]$ .

(ii)  $F(x) - F(a) = \int_a^x f(t) dt$  for some  $f \in L^1([a, b], m)$ .

(iii)  $F$  is differentiable a.e. on  $[a, b]$ ,  $F' \in L^1([a, b], m)$ , and  $F(x) - F(a) = \int_a^x F'(t) dt$ .

*Proof.* (a  $\implies$  c) We may assume by subtracting a constant from  $F$  that  $F(a) = 0$ . If we set  $F(x) = 0$  for  $x < a$  and  $F(x) = F(b)$  for  $x > b$ , then using [Lemma 4.62](#),  $F \in \text{NBV}$ . We may conclude (c) by [Corollary 4.61](#). (c  $\implies$  b) Trivial. (b  $\implies$  a) Set  $f(t) = 0$  for  $t \notin [a, b]$  and apply [Corollary 4.61](#).  $\square$

### 4.5.1 Decomposition of Complex Borel Measures

The following decomposition of Borel measures on  $\mathbb{R}^n$  is sometimes useful.

**Definition 4.64.** A complex Borel measure  $\mu$  on  $\mathbb{R}^n$  is called *discrete* if there is a countable set  $\{x_j\} \subset \mathbb{R}^n$  and complex numbers  $c_j$  such that  $\sum |c_j| < \infty$  and  $\mu = \sum c_j \delta_{x_j}$ , where  $\delta_x$  is the point mass at  $x$ . On the other hand,  $\mu$  is called *continuous* if  $\mu(\{x\}) = 0$  for all  $x \in \mathbb{R}^n$ .

Any complex measure  $\mu$  can be written uniquely as  $\mu = \mu_d + \mu_c$  where  $\mu_d$  is discrete and  $\mu_c$  is continuous. Indeed, let  $E = \{x : \mu(\{x\}) \neq 0\}$ . For any countable subset  $F$  of  $E$  the series  $\sum_{x \in F} \mu(\{x\})$  converges absolutely (to  $\mu(F)$ ), so  $\{x \in E : |\mu(\{x\})| > k^{-1}\}$  is finite for all  $k$ , and it follows that  $E$  itself is countable. Hence, we may write  $\mu_d(A) = \mu(A \cap E)$  and  $\mu_c(A) = \mu(A \setminus E)$ . Uniqueness is easy to see.

We can decompose this sum even further. Clearly, if  $\mu$  is discrete, then  $\mu \perp m$ ; and if  $\mu \ll m$ , then  $\mu$  is continuous. Next, (assuming regularity and hence  $\sigma$ -finiteness) by [Theorem 4.31](#), a complex Borel measure on  $\mathbb{R}^n$  can be written uniquely as  $\lambda + \rho$  for  $\lambda \perp m$  and  $\rho \ll m$ . We can then uniquely decompose  $\lambda$  into discrete and continuous parts. This gives the unique decomposition

$$\mu = \mu_d + \mu_{sc} + \mu_{ac},$$

where  $\mu_d$  is discrete,  $\mu_{ac}$  is absolutely continuous with respect to  $m$ , and  $\mu_{sc}$  is a “singular continuous” measure, meaning  $\mu_{sc}$  is continuous but  $\mu_{sc} \perp m$ . To see uniqueness, observe that  $\mu_d$  is uniquely determined, and so is  $\mu_{ac}$ .

The existence of nonzero singular continuous measures in  $\mathbb{R}^n$  is evident enough when  $n > 1$ ; the surface measure on the unit sphere discussed in [\[Dutch: add section, and review this comment\]](#) is one example. Their existence when  $n = 1$  is not quite so obvious; they correspond via [Theorem 4.57](#) to nonconstant functions  $F \in \text{NBV}$  such that  $F$  is continuous but  $F' = 0$  a.e. One such function is the Cantor-Lebesgue function. More surprisingly, there exists strictly increasing continuous functions  $F$  such that  $F' = 0$  a.e.; see [Folland \(1999, Exercise 40\)](#).

If  $F \in \text{NBV}$ , it is customary to denote the integral of a function  $g$  with respect to the measure  $\mu_F$  by  $\int g dF$  or  $\int g(x) dF(x)$ ; such integrals are called Lebesgue-Stieltjes integrals. We conclude by presenting an integration-by-parts formula for Lebesgue-Stieltjes integrals; other variants of this result can be found in [Folland \(1999, Exercises 34, 35\)](#).

**Theorem 4.65.** If  $F$  and  $G$  are in NBV and at least one of them is continuous, then for  $-\infty < a < b < \infty$ ,

$$\int_{(a,b]} F dG + \int_{(a,b]} G dF = F(b)G(b) - F(a)G(a).$$

*Proof.*  $F$  and  $G$  are linear combinations of increasing functions in NBV by [Theorem 4.54](#). By a simple calculation, it suffices to assume  $F$  and  $G$  are increasing. Suppose w.l.o.g. that  $G$  is continuous, and let  $\Omega = \{(x, y) : a < x \leq y \leq b\}$ . We use Fubini's to compute  $\mu_F \times \mu_G(\Omega)$  in two ways.

$$\begin{aligned}\mu_F \times \mu_G(\Omega) &= \int_{(a,b]} \int_{(a,y]} dF(x) dG(y) = \int_{(a,b]} [F(y) - F(a)] dG(y) \\ &= \int_{(a,b]} F dG - F(a)[G(b) - F(a)].\end{aligned}$$

Next, since  $G(x) = G(x-)$ ,

$$\begin{aligned}\mu_F \times \mu_G(\Omega) &= \int_{(a,b]} \int_{[x,b]} dG(y) dF(x) = \int_{(a,b]} [G(b) - G(x)] dF(x) \\ &= G(b)[F(b) - F(a)] - \int_{(a,b]} G dF.\end{aligned}$$

Subtracting the two equations gives the result. □

## References

Folland, G. (1999). *Real Analysis: Modern Techniques and Their Applications*. Pure and Applied Mathematics: A Wiley Series of Texts, Monographs and Tracts. Wiley.