Real Analysis

Richard F. Bass

April 8, 2009

1 σ-algebras

Let X be a set. We will use the notation: $A^c = \{x \in X : x \notin A\}$ and $A - B = A \cap B^c$. (The notation $A \setminus B$ is also commonly used.)

Definition 1.1 *An algebra is a collection $\mathcal{A}$ of subsets of X such that*

(a) $\emptyset, X \in \mathcal{A}$;

(b) if $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$;

(c) if $A_1, \ldots, A_n \in \mathcal{A}$, then $\cup_{i=1}^n A_i$ and $\cap_{i=1}^n A_i$ are in $\mathcal{A}$.

$\mathcal{A}$ is a σ-algebra *(or* σ-field*) if in addition*

(d) if $A_1, A_2, \ldots$ are in $\mathcal{A}$, then $\cup_{i=1}^\infty A_i$ and $\cap_{i=1}^\infty A_i$ are in $\mathcal{A}$.

In (d) we allow countable unions and intersections only; we do not allow uncountable unions and intersections.

Example 1.2 Let $X = \mathbb{R}$ and $\mathcal{A}$ be the collection of all subsets of $\mathbb{R}$.

Example 1.3 Let $X = \mathbb{R}$ and let

$$\mathcal{A} = \{A \subset \mathbb{R} : A \text{ is countable or } A^c \text{ is countable}\}.$$

Parts (a) and (b) of the definition are easy. Suppose $A_1, A_2, \ldots$ are all in $\mathcal{A}$. If each of the A_i are countable, then $\cup_i A_i$ is countable, and so in $\mathcal{A}$. If $A_{i_0}^c$ is countable for some i_0, then

$$(\cup A_i)^c = \cap_i A_i^c \subset A_{i_0}^c$$

is countable, and again $\cup_i A_i$ is in $\mathcal{A}$. Since $\cap A_i = (\cup_i A_i^c)^c$, then the countable intersection of sets in $\mathcal{A}$ is again in $\mathcal{A}$.

Example 1.4 Let $X = [0, 1]$ and $\mathcal{A} = \{\emptyset, X, [0, \frac{1}{2}], (\frac{1}{2}, 1]\}$.

Example 1.5 $X = \{1, 2, 3\}$ *and* $\mathcal{A} = \{X, \emptyset, \{1\}, \{2, 3\}\}$.

Example 1.6 Let $X = [0, 1]$, and $B_1, \ldots, B_8$ subsets of X which are pairwise disjoint and whose union is all of X. Let $\mathcal{A}$ be the collection of all finite unions of the B_i's as well as the empty set. (So $\mathcal{A}$ consists of 2^8 elements.)

Note that if we take an intersection of σ-algebras, we get a σ-algebra; this is just a matter of checking the definition.If we have a collection $\mathcal{C}$ of subsets of X, there is at least one σ-algebra containing $\mathcal{C}$, namely, the one consisting of all subsets of X. We can take the intersection of all σ-algebras that contain $\mathcal{C}$; we denote this intersection by $\sigma(\mathcal{C})$. If $\mathcal{A}$ is any σ-algebra containing $\mathcal{C}$, then $\mathcal{A} \supset \sigma(\mathcal{C})$.

If X has some additional structure, say, it is a metric space, then we can talk about open sets. If $\mathcal{G}$ is the collection of open subsets of X, then we call $\sigma(\mathcal{G})$ the Borel σ-algebra on X, and this is often denoted $\mathcal{B}$. We will see later that when X is the real line, that $\mathcal{B}$ is **not** equal to the collection of all subsets of X.

We end this section with the following proposition.

Proposition 1.7 *If $X = \mathbb{R}$, then the Borel σ-algebra is generated by each of the following collection of sets:*

(1) $\mathcal{C}_1 = \{(a, b) : a, b \in \mathbb{R}\}$.

(2) $\mathcal{C}_2 = \{[a, b] : a, b \in \mathbb{R}\}$*;*

(3) $\mathcal{C}_3 = \{(a, b] : a, b \in \mathbb{R}\}$*;*

(4) $\mathcal{C}_4 = \{(a, \infty) : a \in \mathbb{R}\}$*;*

Proof. (1) Let $\mathcal{G}$ be the collection of open sets. Then $\mathcal{C}_1 \subset \mathcal{G} \subset \sigma(\mathcal{G})$. $\sigma(\mathcal{G})$ is the Borel σ-algebra and contains $\mathcal{C}_1$. Since $\sigma(\mathcal{C}_1)$ is the intersection of all σ-algebras containing $\mathcal{C}_1$, then $\sigma(\mathcal{C}_1) \subset \sigma(\mathcal{G})$.

To get the reverse inclusion, if G is open, it is the countable union of open intervals. So $G \in \sigma(\mathcal{C}_1)$, and hence $\mathcal{G} \subset \sigma(\mathcal{C}_1)$. $\sigma(\mathcal{G})$ is the intersection of all σ-algebras containing $\mathcal{G}$; $\sigma(\mathcal{C}_1)$ is one such, so $\sigma(\mathcal{G}) \subset \sigma(\mathcal{C}_1)$.

(2) If $[a, b] \in \mathcal{C}_2$, then $[a, b] = \cap_{n=1}^{\infty}(a - \frac{1}{n}, b + \frac{1}{n}) \in \sigma(\mathcal{G})$. So $\mathcal{C}_2 \subset \sigma(\mathcal{G})$, and by an argument similar to that in (1), we conclude $\sigma(\mathcal{C}_2) \subset \sigma(\mathcal{G})$.

If $(a, b) \in \mathcal{C}_1$, choose $n_0 \geq 2/(b - a)$ and note $(a, b) = \cup_{n=n_0}^{\infty}[a + \frac{1}{n}, b - \frac{1}{n}] \in \sigma(\mathcal{C}_2)$. So the Borel σ-algebra, which is equal to $\sigma(\mathcal{C}_1)$ by part (1), is contained in $\sigma(\mathcal{C}_2)$.

(3) The proof here is similar to (2), using $(a, b] = \cap_{n=1}^{\infty}(a, b + \frac{1}{n})$ and $(a, b) = \cup_{n=n_0}^{\infty}(a, b - \frac{1}{n}]$, provided n_0 is taken large enough.

(4) The proof of this comes from (3), using that $(a, b] = (a, \infty) - (b, \infty)$ and $(a, \infty) = \cup_{n=1}^{\infty}(a, a + n]$. □

2 Measures

Definition 2.1 *A measure on $(X, \mathcal{A})$ is a function $\mu : \mathcal{A} \to [0, \infty]$ such that*

(a) $\mu(A) \geq 0$ *for all* $A \in \mathcal{A}$*;*

(b) $\mu(\emptyset) = 0$*;*

(c) if $A_i \in \mathcal{A}$ are disjoint, then

$$\mu(\cup_{i=1}^\infty A_i) = \sum_{i=1}^\infty \mu(A_i).$$

Example 2.2 X is any set, $\mathcal{A}$ is the collection of all subsets, and $\mu(A)$ is the number of elements in A. This is called *counting measure*.

Example 2.3 $X = \mathbb{R}$, $\mathcal{A}$ the collection of all subsets, $x_1, x_2, \ldots \in \mathbb{R}$, and $a_1, a_2, \ldots > 0$. Set $\mu(A) = \sum_{\{i: x_i \in A\}} a_i$. A particular case of this is if $x_i = i$ and all the $a_i = 1$. We will see later that this allows us to view infinite series as functions on this space.

Example 2.4 $\delta_x(A) = 1$ if $x \in A$ and 0 otherwise. This measure is called *point mass* at x.

We will construct Lebesgue measure on $\mathbb{R}$; this is an extension of the notion of length. However, the construction is a bit lengthy. We will also construct Lebesgue measure on $\mathbb{R}^n$; when $n = 2$, this is an extension of the notion of area, when $n = 3$, of volume.

Proposition 2.5 *The following hold:*

(a) If $A, B \in \mathcal{A}$ with $A \subset B$, then $\mu(A) \le \mu(B)$.

(b) If $A_i \in \mathcal{A}$ and $A = \cup_{i=1}^\infty A_i$, then $\mu(A) \le \sum_{i=1}^\infty \mu(A_i)$.

(c) If $A_i \in \mathcal{A}$, $A_1 \subset A_2 \subset \cdots$, and $A = \cup_{i=1}^\infty A_i$, then $\mu(A) = \lim_{n\to\infty} \mu(A_n)$.

(d) If $A_i \in \mathcal{A}$, $A_1 \supset A_2 \supset \cdots$, $\mu(A_1) < \infty$, and $A = \cap_{i=1}^\infty A_i$, then we have $\mu(A) = \lim_{n\to\infty} \mu(A_n)$.

Proof. (a) Let $A_1 = A$, $A_2 = B - A$, and $A_3 = A_4 = \cdots = \emptyset$. Now use part (c) of the definition of measure.

(b) Let $B_1 = A_1$, $B_2 = A_2 - B_1$, $B_3 = A_3 - (B_1 \cup B_2)$, and so on. The B_i are disjoint and $\cup_{i=1}^\infty B_i = \cup_{i=1}^\infty A_i$. So $\mu(A) = \sum \mu(B_i) \le \sum \mu(A_i)$.

(c) Define the B_i as in (b). Since $\cup_{i=1}^n B_i = \cup_{i=1}^n A_i$, then

$$\begin{aligned}\mu(A) &= \mu(\cup_{i=1}^\infty A_i) = \mu(\cup_{i=1}^\infty B_i) = \sum_{i=1}^\infty \mu(B_i)\\ &= \lim_{n\to\infty} \sum_{i=1}^n \mu(B_i) = \lim_{n\to\infty} \mu(\cup_{i=1}^n B_i) = \lim_{n\to\infty} \mu(\cup_{i=1}^n A_i).\end{aligned}$$

(d) Apply (c) to the sets $A_1 - A_i$, $i = 1, 2, \ldots$. □

Example 2.6 To see that $\mu(A_1) < \infty$ is necessary, let X be the positive integers, μ counting measure, and $A_i = \{i, i+1, \ldots\}$. Then the A_i decrease, $\mu(A_i) = \infty$ for all i, but $\mu(\cap_i A_i) = \mu(\emptyset) = 0$.

Definition 2.7 *A probability or probability measure is a measure such that* $\mu(X) = 1$. *In this case we usually write* $(\Omega, \mathcal{F}, \mathbb{P})$ *instead of* $(X, \mathcal{A}, \mu)$.

3 Construction of Lebesgue measure

Define $m((a,b)) = b - a$. If G is an open set and $G \subset \mathbb{R}$, then $G = \cup_{i=1}^\infty (a_i, b_i)$ with the intervals disjoint. Define $m(G) = \sum_{i=1}^\infty (b_i - a_i)$. If $A \subset \mathbb{R}$, define

$$m^*(A) = \inf\{m(G) : G \text{ open}, A \subset G\}. \tag{3.1}$$

We will show the following.

(A) m^* is not a measure on the collection of all subsets of $\mathbb{R}$.

(B) m^* is a measure on a strictly smaller σ-algebra that strictly contains the Borel σ-algebra.

We will prove these two facts (and a bit more) in a moment, but let's first make some remarks.

A set N is a *null set* with respect to m^* if $m^*(N) = 0$. Let $\mathcal{L}$ be the smallest σ-algebra containing $\mathcal{B}$ and all the null sets. More precisely, let $\mathcal{N}$ be the collection of all sets that are null sets with respect to m^* and let $\mathcal{L} = \sigma(\mathcal{B} \cup \mathcal{N})$. $\mathcal{L}$ is called the *Lebesgue σ-algebra*, and sets in $\mathcal{L}$ are called *Lebesgue measurable*.

As part of our proof of (B) we will show that m^* is a measure on $\mathcal{L}$. *Lebesgue measure* is the measure m^* on $\mathcal{L}$. (A) shows that $\mathcal{L}$ is strictly smaller than the collection of all subsets of $\mathbb{R}$.

It is easy to get lost in the construction of Lebesgue measure, so let us summarize our steps.

First we prove (A), which Proposition 3.1.

We then turn to the construction of Lebesgue measure. It is more convenient for technical reasons to define

$$m^*(A) = \inf\{\sum_{i=1}^{\infty}(b_i - a_i) : A \subset \cup_{i=1}^{\infty}(a_i, b_i]\}. \tag{3.2}$$

There is no real difference between this and (3.1) since it is clear that the difference between an open set and a set of the form $\cup_{i=1}^{\infty}(a_i, b_i]$ is countable, and with either definition of m^*, the measure of a point is 0, so the measure of a set consisting of countably many points is 0. However, when we talk about Lebesgue-Stieltjes measure, then there is a real difference.

We define what it means to be an outer measure (Definition 3.2) and prove that m^* is an outer measure (Proposition 3.3). We then define what it means for a set to be m^*-measurable (Definition 3.4) and prove that the collection of m^*-measurable sets is a σ-algebra and that m^* restricted to this σ-algebra is a measure.

This looks promising, but we do not yet know that enough sets are m^*-measurable. That takes one more step. We show in Proposition 3.6 that the collection of m^*-measurable sets contains the Borel σ-algebra.

Proposition 3.1 *m^* is not a measure on the collection of all subsets of $\mathbb{R}$.*

Proof. Suppose m^* is a measure. Define $x \sim y$ if $x - y$ is rational. This is an equivalence relationship on $[0,1]$. For each equivalence class, pick an element out of that class (by the axiom of choice) Call the collection of such points A. Given a set B, define $B + x = \{y + x : y \in B\}$. Note $m^*(A+q) = m^*(A)$ since this translation invariance holds for intervals, hence for open sets, hence for all sets. Moreover, the sets $A + q$ are disjoint for different rationals q.

Now

$$[0,1] \subset \cup_{q \in [-2,2]}(A+q),$$

where the sum is only over rational q, so $1 \leq \sum_{q \in [-2,2]} m^*(A+q)$, and therefore $m^*(A) > 0$. But

$$\cup_{q \in [-2,2]}(A+q) \subset [-6,6],$$

where again the sum is only over rational q, so if m^* is a measure, then $12 \geq \sum_{q \in [-2,2]} m^*(A+q)$, which implies $m^*(A) = 0$, a contradiction. □

Definition 3.2 *A function n on the collection of all subsets satisfying*

(a) $n(\emptyset) = 0$;

(b) if $A \subset B$, then $n(A) \leq n(B)$;

(c) $n(\cup_{i=1}^\infty A_i) \leq \sum_{i=1}^\infty n(A_i)$.

is called an outer measure.

Proposition 3.3 *m^* defined by (3.2) is an outer measure.*

Proof. (a) and (b) are obvious. To prove (c), let $\varepsilon > 0$. For each i there exist intervals $I_{i1}, I_{i2}, \ldots$, each of the form $(a_{ij}, b_{ij}]$, such that $A_i \subset \cup_{j=1}^\infty I_{ij}$ and $\sum_j m(I_{ij}) \leq m^*(A_i) + \varepsilon/2^i$. Then $\cup_{i=1}^\infty A_i \subset \cup_{i,j} I_{ij}$ and

$$\sum_{i,j} m(I_{ij}) \leq \sum_i m^*(A_i) + \sum_i \varepsilon/2^i = \sum_i m^*(A_i) + \varepsilon.$$

Since ε is arbitrary, $m^*(\cup_{i=1}^\infty A_i) \leq \sum_{i=1}^\infty m^*(A_i)$. □

Definition 3.4 *Let m^* be an outer measure. A set $A \subset X$ is m^*-measurable if*

$$m^*(E) = m^*(E \cap A) + m^*(E \cap A^c) \tag{3.3}$$

for all $E \subset X$.

Theorem 3.5 *If m^* is an outer measure on X, then the collection $\mathcal{A}$ of m^* measurable sets is a σ-algebra and the restriction of m^* to $\mathcal{A}$ is a measure. Moreover, $\mathcal{A}$ contains all the null sets.*

Proof. By Proposition 3.3,

$$m^*(E) \le m^*(E \cap A) + m^*(E \cap A^c)$$

for all $E \subset X$. So to check (3.3) it is enough to show $m^*(E) \ge m^*(E \cap A) + m^*(E \cap A^c)$. This will be trivial in the case $m^*(E) = \infty$.

If $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$ by symmetry and the definition of $\mathcal{A}$. Suppose $A, B \in \mathcal{A}$ and $E \subset X$. Then

$$\begin{aligned} m^*(E) &= m^*(E \cap A) + m^*(E \cap A^c) \\ &= (m^*(E \cap A \cap B) + m^*(E \cap A \cap B^c)) + (m^*(E \cap A^c \cap B) \\ &\quad + m^*(E \cap A^c \cap B^c). \end{aligned}$$

The first three terms on the right have a sum greater than or equal to $m^*(E \cap (A \cup B))$ because $A \cup B \subset (A \cap B) \cup (A \cap B^c) \cup (A^c \cap B)$. Therefore

$$m^*(E) \ge m^*(E \cap (A \cup B)) + m^*(E \cap (A \cup B)^c),$$

which shows $A \cup B \in \mathcal{A}$. Therefore $\mathcal{A}$ is an algebra.

Let A_i be disjoint sets in $\mathcal{A}$, let $B_n = \cup_{i=1}^n A_i$, and $B = \cup_{i=1}^\infty A_i$. If $E \subset X$,

$$\begin{aligned} m^*(E \cap B_n) &= m^*(E \cap B_n \cap A_n) + m^*(E \cap B_n \cap A_n^c) \\ &= m^*(E \cap A_n) + m^*(E \cap B_{n-1}). \end{aligned}$$

Repeating for $m^*(E \cap B_{n-1})$, we obtain

$$m^*(E \cap B_n) \ge \sum_{i=1}^n m^*(E \cap A_i).$$

Since $B_n \in \mathcal{A}$, then

$$m^*(E) = m^*(E \cap B_n) + m^*(E \cap B_n^c) \geq \sum_{i=1}^{n} m^*(E \cap A_i) + m^*(E \cap B^c).$$

Let $n \to \infty$. Recalling that m^* is an outer measure,

$$\begin{aligned} m^*(E) &\geq \sum_{i=1}^{\infty} m^*(E \cap A_i) + m^*(E \cap B^c) \\ &\geq m^*(\cup_{i=1}^{\infty}(E \cap A_i)) + m^*(E \cap B^c) \\ &= m^*(E \cap B) + m(E \cap B^c) \\ &\geq m^*(E). \end{aligned}$$

This shows $B \in \mathcal{A}$.

If we set $E = B$ in this last equation, we obtain

$$m^*(B) = \sum_{i=1}^{\infty} m^*(A_i),$$

or m^* is countably additive on $\mathcal{A}$.

If $m^*(A) = 0$ and $E \subset X$, then

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

which shows $\mathcal{A}$ contains all null sets. □

Define $m(\cup_{i=1}^n (a_i, b_i]) = \sum_{i=1}^n (b_i - a_i)$ if the $(a_i, b_i]$ are disjoint. Note m is well-defined (a set might be expressible as a union of such intervals in more than one way).

The last step in the construction is the following.

Proposition 3.6 *Every set in the Borel σ-algebra is m^*- measurable.*

Proof. Since the collection of m^*-measurable sets is a σ-algebra, it suffices to show that every interval J of the form $(a, b]$ is m^*-measurable. Let E be any set with $m^*(E) < \infty$; we need to show

$$m^*(E) \geq m^*(E \cap J) + m^*(E \cap J^c). \tag{3.4}$$

Choose $I_1, I_2, \ldots$ of the form $(a_i, b_i]$ such that $E \subset \cup_i I_i$ and

$$m^*(E) \geq \sum_i (b_i - a_i) - \varepsilon.$$

Since $E \subset \cup I_i$, we have $m^*(E \cap J) \leq \sum m^*(I_i \cap J)$ and $m^*(E \cap J^c) \leq \sum m^*(I_i \cap J^c)$. Hence we have

$$m^*(E \cap J) + m^*(E \cap J^c) \leq \sum_i [m^*(I_i \cap J) + m^*(I_i \cap J^c)].$$

Now $m^*(I_i \cap J)$ is the length of an interval and $m^*(I_i \cap J^c)$ is the length of two intervals, so

$$m^*(I_i \cap J) + m^*(I_i \cap J) = m^*(I_i).$$

Thus

$$m^*(E \cap J) + m^*(E \cap J^c) \leq \sum m^*(I_i) \leq m^*(E) + \varepsilon.$$

Since ε is arbitrary, this proves (3.4). □

We now drop the asterisks from m^* and call m Lebesgue measure.

4 Examples and related results

Example 4.1 Recall the Cantor set is constructed by taking the interval $[0, 1]$, removing the middle third, removing the middle thirds of each of the two remaining subintervals, and continuing. The Cantor set is what remains; it is closed, uncountable, and every point is a limit point. Moreover, it contains no intervals.

After one stage, the measure of the two intervals is $2(\frac{1}{3})$, after two stages $4(1/9)$, and after n stages, $(2/3)^n$. Since the Cantor set C is the intersection of all these sets, the Lebesgue measure of C is 0.

Suppose we define f_0 to be $1/2$ on the interval $(1/3, 2/3)$, $1/4$ on the interval $(1/9, 2/9)$, $3/4$ on the interval $(7/9, 8/9)$, and so on. Define $f(x) = \inf\{f_0(y) : y \geq x\}$ for $x < 1$. Define $f(1) = 1$. Notice $f = f_0$ on the complement of the Cantor set. f is monotone, so it has only jump discontinuities. But if it has a jump continuity, there is a rational of the form $k/2^n$ with

$k \leq 2^n$ that is not in the range of f. On the other hand, by the construction, each of these values is taken by f_0 for some point in the complement of C, and so is taken by f. The only way this can happen is if f is continuous. This function f is called the Cantor-Lebesgue function. We will use it in examples later on. For now, we can see that it a function that increases only on the Cantor set, which is of Lebesgue measure 0, yet is continuous.

Example 4.2 Let $q_1, q_2, \ldots$ be an enumeration of the rationals, let $\varepsilon > 0$, and let I_i be the interval $(q_i - \varepsilon/2^i, q_i + \varepsilon/2^i)$. Then the measure of I_i is $\varepsilon/2^{i-1}$, so the measure of $\cup_i I_i$ is at most 2ε. (It is not equal to that because there is a lot of overlap.) So the measure of $A = [0,1] - \cup_i I_i$ is larger than $1 - 2\varepsilon$. But A contains no rational numbers.

Example 4.3 Let us follow the construction of the Cantor set, with this difference. Instead of removing the middle third at the first stage, remove the middle fourth, i.e., remove $(3/8, 5/8)$. On each of the two intervals that remain, remove the middle sixteenths. On each of the four intervals that remain, remove the middle interval of length $1/64$, and so on. The total that we removed is

$$\tfrac{1}{4} + 2(\tfrac{1}{16}) + 4(\tfrac{1}{64}) + \cdots = \tfrac{1}{2}.$$

The set that remains contains no intervals, is closed, every point is a limit point, is uncountable, and has measure $1/2$. Such a set is called a fat Cantor set or generalized Cantor set. Of course, other choices that $1/4$, $1/16$, etc. are possible.

Let $A \subset [0,1]$ be a Borel measurable set. We will show that A is almost equal to the countable intersection of open sets and almost equal to the countable union of closed sets.. (A similar argument to what follows is possible for sets that have infinite measure.)

Proposition 4.4 *Suppose $A \subset [0,1]$ is a Borel measurable set.*

(a) There exists a set H that is the countable intersection of open sets which contains A and $m(H - A) = 0$.

(b) There exists a set F that is the countable union of closed sets which is contained in A and $m(A - F) = 0$.

Proof. (a) For each i, there is an open set G_i that contains A and such that $m(G_i - A) < 2^{-i}$. This follows from the fact that $m(A) = m^*(A)$ and the definition of m^*. Then $H_i = \cap_{j \le i} G_j$ will contain A, is open, and since it is contained in G_i, then $m(H_i - A) < 2^{-i}$. Let $H = \cap_{i=1}^{\infty} H_i$. H need not be open, but it is the intersection of countably many open sets. The set H is a Borel set, contains A, and $m(H - A) \le m(H_i - A) < 2^{-i}$ for each i, hence $m(H - A) = 0$.

(b) If $A \subset [0, 1]$, let $F_i = [0, 1] - H_i$, where H_i is a decreasing sequence of open sets containing A^c such that $m(H_i - A^c) < 2^{-i}$. (The H_i are constructed as in the proof of (a), but in terms of A^c instead of A.) Then F_i is an increasing sequence of closed sets, $F_i \subset A$ for each i, and $m(A - F_i) < 2^{-i}$ for each i. Our result follows from letting $F = \cup_i F_i$ since $m(A - F) \le m(A - F_i) < 2^{-i}$ for each i, hence $m(A - F) = 0$.. □

The countable intersections of open sets are sometimes called G_δ sets; the G is for *geoffnet*, the German word for "open" and the δ for *Durchschnitt*, the German word for "intersection." The countable unions of closed sets are called F_σ sets, the F coming from *fermé*, the French word for "closed," and the σ coming from *Summe*, the German word for "union."

Therefore, when trying to understand Lebesgue measure, we can look at G_δ or F_σ sets, which are not so bad, and at null sets, which can be quite bad but don't have positive measure.

Next we prove the Carathéodory extension theorem. We say that a measure μ is *σ-finite* if there exist $E_1, E_2, \ldots$, such that $\mu(E_i) < \infty$ for all i and $X \subset \cup_{i=1}^{\infty} E_i$.

Theorem 4.5 *Suppose $\mathcal{A}_0$ is an algebra and m restricted to $\mathcal{A}_0$ is a measure. Define*

$$m^*(E) = \inf\Big\{\sum_{i=1}^{\infty} m(A_i) : A_i \in \mathcal{A}_0, E \subset \cup_{i=1}^{\infty} A_i\Big\}.$$

Then

(a) $m^*(A) = m(A)$ if $A \in \mathcal{A}_0$;

(b) every set in $\mathcal{A}_0$ is m^*-measurable;

(c) if m is σ-finite, then there is a unique extension to the smallest σ-algebra containing $\mathcal{A}_0$.

Proof. We start with (a). Suppose $E \in \mathcal{A}_0$. We know $m^*(E) \le m(E)$ since we can take $A_1 = E$ and $A_2, A_3, \ldots$ empty in the definition of m^*. If $E \subset \cup_{i=1}^\infty A_i$ with $A_i \in \mathcal{A}_0$, let $B_n = E \cap (A_n - \cup_{i=1}^{n-1} A_i)$. Then the B_n are disjoint, they are each in $\mathcal{A}_0$, and their union is E. Therefore

$$m(E) = \sum_{i=1}^\infty m(B_i) \le \sum_{i=1}^\infty m(A_i).$$

Thus $m(E) \le m^*(E)$.

Next we look at (b). Suppose $A \in \mathcal{A}_0$. Let $\varepsilon > 0$ and let $E \subset X$. Pick $B_i \in \mathcal{A}_0$ such that $E \subset \cup_{i=1}^\infty B_i$ and $\sum_i m(B_i) \le m^*(E) + \varepsilon$. Then

$$\begin{aligned} m^*(E) + \varepsilon &\ge \sum_{i=1}^\infty m(B_i) = \sum_{i=1}^\infty m(B_i \cap A) + \sum_{i=1}^\infty m(B_i \cap A^c) \\ &\ge m^*(E \cap A) + m^*(E \cap A^c). \end{aligned}$$

Since ε is arbitrary, $m^*(E) \ge m^*(E \cap A) + m^*(E \cap A^c)$. So A is m^*-measurable.

Finally, suppose we have two extensions to the smallest σ-algebra containing $\mathcal{A}_0$; let the other extension be called n. We will show that if E is in this smallest σ-algebra, then $m^*(E) = n(E)$.

Since m is σ-finite, we can reduce to the case where m is a finite measure: if $X = \cup_i K_i$ with $m(K_i) < \infty$ and we prove uniqueness for the measure m_i defined by $m_i(A) = m(A \cap K_i)$, then uniqueness for m follows. So we suppose $m(X) < \infty$.

Since E must be m^*-measurable,

$$m^*(E) = \inf\{\sum_{i=1}^\infty m(A_i) : E \subset \cup_{i=1}^\infty A_i, A_i \in \mathcal{A}_0\}.$$

But $m = n$ on $\mathcal{A}_0$, so $\sum_i m(A_i) = \sum_i n(A_i)$. Therefore $n(E) \le \sum_i n(A_i)$, which implies $n(E) \le m^*(E)$.

Since we do not know that n is constructed via an outer measure, we must use a different argument to get the reverse inequality. Let $\varepsilon > 0$ and choose

$A_i \in \mathcal{A}_0$ such that $m^*(E) + \varepsilon \geq \sum_i m(A_i)$ and $E \subset \cup_i A_i$. Let $A = \cup_i A_i$ and $B_k = \cup_{i=1}^k A_i$. Observe $m^*(E) + \varepsilon \geq m^*(A)$, hence $m^*(A - E) < \varepsilon$. We have

$$m^*(A) = \lim_{k\to\infty} m^*(B_k) = \lim_{k\to\infty} n(B_k) = n(A).$$

Then

$$m^*(E) \leq m^*(A) = n(A) = n(E) + n(A-E) \leq n(E) + m(A-E) \leq n(E) + \varepsilon.$$

Since ε is arbitrary, this completes the proof. □

Remarks: (1) Uniqueness implies there is only one possible Lebesgue measure.

(2) We will use the Carathéodory extension theorem in the study of product measures. It is also used in the Riesz representation theorem and in the Daniell-Kolmogorov extension theorem.

We now define Lebesgue-Stieltjes measures. Let $\alpha : \mathbb{R} \to \mathbb{R}$ be nondecreasing and right continuous (i.e., $\alpha(x+) = \alpha(x)$ for all x, where $\alpha(x+) = \lim_{y\to x, y>x} \alpha(y)$). Suppose we define $m_\alpha((a, b]) = \alpha(b) - \alpha(a)$, define

$$m_\alpha(\cup_{i=1}^\infty (a_i, b_i]) = \sum_i (\alpha(b_i) - \alpha(a_i))$$

when the intervals $(a_i, b_i]$ are disjoint, and define

$$m_\alpha^*(A) = \inf\{\sum_i (\alpha(b_i) - \alpha(a_i)) : A \subset \cup_i (a_i, b_i]\}.$$

Very much as in the previous section we can show that m_α^* is a measure on the Borel σ-algebra.

The m_α measure of a point x is $\alpha(x) - \alpha(x-)$, where

$$\alpha(x-) = \lim_{y\to x, y<x} \alpha(x).$$

So $m_\alpha(\{x\})$ is equal to the size of the jump (if any) of α at x.

Lebesgue measure is the special case of m_α when $\alpha(x) = x$.

Given a measure μ on $\mathbb{R}$ such that $\mu(K) < \infty$ whenever K is compact, define $\alpha(x) = \mu((0, x])$ if $x \geq 0$ and $\alpha(x) = -\mu((x, 0])$ if $x < 0$. Then α is nondecreasing, right continuous, and it is not hard to see that $\mu = m_\alpha$.

5 Measurable functions

Suppose we have a set X together with a σ-algebra $\mathcal{A}$.

Definition 5.1 *$f : X \to \mathbb{R}$ is measurable or $\mathcal{A}$-measurable if $\{x : f(x) > a\} \in \mathcal{A}$ for all $a \in \mathbb{R}$.*

Example 5.2 Suppose f is identically constant. Then $\{x : f(x) > a\}$ is either empty or the whole space, so f is measurable.

Example 5.3 Suppose $f(x) = 1$ if $x \in A$ and 0 otherwise. Then $\{x : f(x) > a\}$ is either $\emptyset$, A, or X. So f is measurable if and only if A is in the σ-algebra.

Example 5.4 Suppose X is the real line with the Borel σ-algebra and $f(x) = x$. Then $\{x : f(x) > a\} = (a, \infty)$, and so f is measurable.

Proposition 5.5 *The following are equivalent.*

(a) $\{x : f(x) > a\} \in \mathcal{A}$ *for all* a;

(b) $\{x : f(x) \le a\} \in \mathcal{A}$ *for all* a;

(c) $\{x : f(x) < a\} \in \mathcal{A}$ *for all* a;

(d) $\{x : f(x) \ge a\} \in \mathcal{A}$ *for all* a.

Proof. The equivalence of (a) and (b) and of (c) and (d) follow from taking complements. The remaining equivalences follow from the equations

$$\begin{aligned} \{x : f(x) \ge a\} &= \cap_{n=1}^\infty \{x : f(x) > a - 1/n\}, \\ \{x : f(x) > a\} &= \cup_{n=1}^\infty \{x : f(x) \ge a + 1/n\}. \end{aligned}$$

□

Proposition 5.6 *If X is a metric space, $\mathcal{A}$ contains all the open sets, and f is continuous, then f is measurable.*

Proof. $\{x : f(x) > a\} = f^{-1}(a, \infty)$ is open. □

Proposition 5.7 *If f and g are measurable, so are $f+g$, cf, fg, $\max(f,g)$, and $\min(f,g)$.*

Proof. If $f(x)+g(x) < \alpha$, then $f(x) < \alpha - g(x)$, and there exists a rational r such that $f(x) < r < \alpha - g(x)$. So

$$\{x : f(x) + g(x) < \alpha\} = \bigcup_{r \text{ rational}} (\{x : f(x) < r\} \cap \{x : g(x) < \alpha - r\}).$$

f^2 is measurable since $\{x : f(x)^2 > a) = \{x : f(x) > \sqrt{a}\} \cup \{x : f(x) < -\sqrt{a}\}$. The measurability of fg follows since $fg = \frac{1}{2}[(f+g)^2 - f^2 - g^2]$.

$\{x : \max(f(x), g(x)) > a\} = \{x : f(x) > a\} \cup \{x : g(x) > a\}$, and the argument for $\min(f,g)$ is similar. □

Proposition 5.8 *If f_i is measurable for each i, then so are $\sup_i f_i$, $\inf_i f_i$, $\limsup_{i\to\infty} f_i$, and $\liminf_{i\to\infty} f_i$.*

Proof. The result will follow for lim sup and lim inf once we have the result for the sup and inf by using the definitions. We have $\{x : \sup_i f_i > a\} = \cap_{i=1}^{\infty}\{x : f_i(x) > a\}$, and the proof for $\inf f_i$ is similar. □

Definition 5.9 *We say $f = g$ almost everywhere, written $f = g$ a.e., if $\{x : f(x) \neq g(x)\}$ has measure zero. Similarly, we say $f_i \to f$ a.e., if the set of x where this fails has measure zero.*

We saw in Proposition 5.6 that all continuous functions are Borel measurable. The same is true for monotone functions on the real line.

Proposition 5.10 *If $f : \mathbb{R} \to \mathbb{R}$ is nondecreasing or nonincreasing, then f is Borel measurable.*

Proof. Let us suppose f is nondecreasing. The set $A_a = \{x : f(x) > a\}$ is then either a semi-infinite open interval or semi-infinite closed interval. This can be seen by a picture. To be more careful, given $a \in \mathbb{R}$, let $x_0 = \sup\{y : f(y) \le a\}$. If $f(x_0) = a$, then $A_a = (x_0, \infty)$, while if $f(x_0) \ne a$, then $A_a = [x_0, \infty)$. In each case A_a is a Borel set. □

Proposition 5.11 *Let X be a space, $\mathcal{A}$ a σ-algebra on X, and $f : X \to \mathbb{R}$ a $\mathcal{A}$-measurable function. If A is in the Borel σ-algebra on $\mathbb{R}$, then $f^{-1}(A) \in \mathcal{A}$.*

Proof. Let $\mathcal{B}$ be the Borel σ-algebra on $\mathbb{R}$ and $\mathcal{C} = \{A \in \mathcal{B} : f^{-1}(A) \in \mathcal{A}\}$. If $A_1, A_2, \ldots \in \mathcal{C}$, then since $f^{-1}(\cup_i A_i) = \cup_i f^{-1}(A_i) \in \mathcal{A}$, we have that $\mathcal{C}$ is closed under countable unions. Similarly $\mathcal{C}$ is closed under countable intersections and complements, so $\mathcal{C}$ is a σ-algebra. Since f is measurable, $\mathcal{C}$ contains (a, ∞) for every real a, hence $\mathcal{C}$ contains the σ-algebra generated by these intervals, that is, $\mathcal{C}$ contains $\mathcal{B}$. □

Example 5.12 We want to construct a set that is Lebesgue measurable, but not Borel measurable. Let F be the Cantor-Lebesgue function of Example 4.1 and define

$$f(x) = \inf\{y : F(y) \ge x\}.$$

Although f is not continuous, observe that f is strictly increasing (hence one-to-one) and maps $[0, 1]$ into C, the Cantor set. Since f is nondecreasing, f^{-1} maps Borel measurable sets to Borel measurable sets.

Let A be the non-measurable set we constructed in Proposition 3.1. Let $B = f(A)$. Since $f(A) \subset C$ and $m(C) = 0$, then $f(A)$ is a null set, hence is Lebesgue measurable. On the other hand, $f(A)$ is not Borel measurable, because if it were, then $A = f^{-1}(f(A))$ would be Borel measurable, a contradiction.

6 Integration

In this section we introduce the Lebesgue integral.

Definition 6.1 *If $E \subset X$, define the characteristic function of E by*

$$\chi_E(x) = \begin{cases} 1 & x \in E; \\ 0 & x \notin E. \end{cases}$$

A simple function s is one of the form

$$s(x) = \sum_{i=1}^n a_i \chi_{E_i}(x)$$

for reals a_i and measurable sets E_i.

Proposition 6.2 *Suppose $f \geq 0$ is measurable. Then there exists a sequence of nonnegative measurable simple functions increasing to f.*

Proof. Let $E_{ni} = \{x : (i-1)/2^n \leq f(x) < i/2^n\}$ and $F_n = \{x : f(x) \geq n\}$ for $n = 1, 2, \dots,$ and $i = 1, 2, \dots, n2^n$. Then define

$$s_n = \sum_{i=1}^{n2^n} \frac{i-1}{2^n} \chi_{E_{ni}} + n\chi_{F_n}.$$

It is easy to see that s_n has the desired properties. □

Definition 6.3 *If $s = \sum_{i=1}^n a_i \chi_{E_i}$ is a nonnegative measurable simple function, define the* Lebesgue integral *of s to be*

$$\int s \, d\mu = \sum_{i=1}^n a_i \mu(E_i). \tag{6.1}$$

If $f \geq 0$ is a measurable function, define

$$\int f \, d\mu = \sup\Big\{ \int s \, d\mu : 0 \leq s \leq f, s \text{ simple} \Big\}. \tag{6.2}$$

If f is measurable and at least one of the integrals $\int f^+ \, d\mu$, $\int f^- \, d\mu$ is finite, where $f^+ = \max(f, 0)$ and $f^- = -\min(f, 0)$, define

$$\int f \, d\mu = \int f^+ \, d\mu - \int f^- \, d\mu. \tag{6.3}$$

Finally, if $f = u + iv$ and $\int (|u| + |v|) \, d\mu$ is finite, define

$$\int f \, d\mu = \int u \, d\mu + i \int v \, d\mu. \tag{6.4}$$

A few remarks are in order. A function s might be written as a simple function in more than one way. For example $\chi_{A\cup B} = \chi_A + \chi_B$ is A and B are disjoint. It is clear that the definition of $\int s \, d\mu$ is unaffected by how s is written. Secondly, if s is a simple function, one has to think a moment to verify that the definition of $\int s \, d\mu$ by means of (6.1) agrees with its definition by means of (6.2).

Definition 6.4 *If $\int |f| \, d\mu < \infty$, we say f is integrable.*

The proof of the next proposition follows from the definitions.

Proposition 6.5 *(a) If f is measurable, $a \le f(x) \le b$ for all x, and $\mu(X) < \infty$, then $a\mu(X) \le \int f \, d\mu \le b\mu(X)$;*

(b) If $f(x) \le g(x)$ for all x and f and g are measurable and integrable, then $\int f \, d\mu \le \int g \, d\mu$.

(c) If f is integrable, then $\int cf \, d\mu = c \int f \, d\mu$ for all real c.

(d) If $\mu(A) = 0$ and f is measurable, then $\int f\chi_A \, d\mu = 0$.

The integral $\int f\chi_A \, d\mu$ is often written $\int_A f \, d\mu$. Other notation for the integral is to omit the μ if it is clear which measure is being used, to write $\int f(x) \, \mu(dx)$, or to write $\int f(x) \, d\mu(x)$.

Proposition 6.6 *If f is integrable,*

$$\Big| \int f \Big| \le \int |f|.$$

Proof. For the real case, this is easy. $f \le |f|$, so $\int f \le \int |f|$. Also $-f \le |f|$, so $-\int f \le \int |f|$. Now combine these two facts.

For the complex case, $\int f$ is a complex number. If it is 0, the inequality is trivial. If it is not, then $\int f = re^{i\theta}$ for some r and θ. Then

$$\Big|\int f\Big| = r = e^{-i\theta}\int f = \int e^{-i\theta} f.$$

From the definition of $\int f$ when f is complex, we have $\mathrm{Re}\,(\int f) = \int \mathrm{Re}\,(f)$. Since $|\int f|$ is real, we have

$$\Big|\int f\Big| = \mathrm{Re}\,\Big(\int e^{-i\theta} f\Big) = \int \mathrm{Re}\,(e^{-i\theta} f) \le \int |f|.$$

□

We do not yet have that $\int (f+g) = \int f + \int g$.

7 Limit theorems

One of the most important results concerning Lebesgue integration is the monotone convergence theorem.

Theorem 7.1 *Suppose f_n is a sequence of nonnegative measurable functions with $f_1(x) \le f_2(x) \le \cdots$ for all x and with $\lim_{n\to\infty} f_n(x) = f(x)$ for all x. Then $\int f_n\, d\mu \to \int f\, d\mu$.*

Proof. By Proposition 6.5(b), $\int f_n$ is an increasing sequence of real numbers. Let L be the limit. Since $f_n \le f$ for all n, then $L \le \int f$. We must show $L \ge \int f$.

Let $s = \sum_{i=1}^m a_i \chi_{E_i}$ be any nonnegative simple function less than f and let $c \in (0,1)$. Let $A_n = \{x : f_n(x) \ge cs(x)\}$. Since the $f_n(x)$ increases to $f(x)$ for each x and $c < 1$, then $A_1 \subset A_2 \subset \cdots$, and the union of the A_n is all of X. For each n,

$$\int f_n \geq \int_{A_n} f_n \geq c \int_{A_n} s_n$$
$$= c \int_{A_n} \sum_{i=1}^{m} a_i \chi_{E_i}$$
$$= c \sum_{i=1}^{m} a_i \mu(E_i \cap A_n).$$

If we let $n \to \infty$, by Proposition 2.5(c), the right hand side converges to

$$c \sum_{i=1}^{m} a_i \mu(E_i) = c \int s.$$

Therefore $L \geq c \int s$. Since c is arbitrary in the interval $(0,1)$, then $L \geq \int s$. Taking the supremum over all simple $s \leq f$, we obtain $L \geq \int f$. $\square$

Example 7.2 Let $X = [0, \infty)$ and $f_n(x) = -1/n$ for all x. Then $\int f_n = -\infty$, but $f_n \uparrow f$ where $f = 0$ and $\int f = 0$. The problem here is that the f_n are not nonnegative.

Example 7.3 Suppose $f_n = n\chi_{(0,1/n)}$. Then $f_n \geq 0$, $f_n \to 0$ for each x, but $\int f_n = 1$ does not converge to $\int 0 = 0$. The trouble here is that the f_n do not increase for each x.

Once we have the monotone convergence theorem, we can prove that the Lebesgue integral is linear.

Theorem 7.4 *If f_1 and f_2 are integrable, then*

$$\int (f_1 + f_2) = \int f_1 + \int f_2.$$

Proof. First suppose f_1 and f_2 are nonnegative and simple. Then it is clear from the definition that the theorem holds in this case. Next suppose f_1

and f_2 are nonnegative. Take s_n simple and increasing to f_1 and t_n simple and increasing to f_2. Then $s_n + t_n$ increases to $f_1 + f_2$, so the result follows from the monotone convergence theorem and the result for simple functions. Finally in the general case, write $f_1 = f_1^+ - f_1^-$ and similarly for f_2, and use the definitions and the result for nonnegative functions. □

Proposition 7.5 *Suppose f_n are nonnegative measurable functions. Then*

$$\int \sum_{n=1}^{\infty} f_n = \sum_{n=1}^{\infty} \int f_n.$$

Proof. Let $F_N = \sum_{n=1}^{N} f_n$ and write

$$\begin{aligned}\int \sum_{n=1}^{\infty} f_n &= \int \lim_{N\to\infty} \sum_{n=1}^{N} f_n \\ &= \int \lim_{N\to\infty} F_N = \lim_{N\to\infty} \int F_N \\ &= \lim_{N\to\infty} \sum_{n=1}^{N} \int f_n = \sum_{n=1}^{\infty} \int f_n,\end{aligned} \tag{7.1}$$

using the monotone convergence theorem and the linearity of the integral. □

The next theorem is known as Fatou's lemma.

Theorem 7.6 *Suppose the f_n are nonnegative and measurable. Then*

$$\int \liminf_{n\to\infty} f_n \le \liminf_{n\to\infty} \int f_n.$$

Proof. Let $g_n = \inf_{i\ge n} f_i$. Then g_n are nonnegative and g_n increases to $\liminf f_n$. Clearly $g_n \le f_i$ for each $i \ge n$, so $\int g_n \le \int f_i$. Therefore

$$\int g_n \le \inf_{i\ge n} \int f_i.$$

If we take the limit as $n \to \infty$, on the left hand side we obtain $\int \liminf f_n$ by the monotone convergence theorem, while on the right hand side we obtain $\liminf_n \int f_n$. □

A typical use of Fatou's lemma is the following. Suppose we have $f_n \to f$ and $\sup_n \int |f_n| \le K < \infty$. Then $|f_n| \to |f|$, and by Fatou's lemma, $\int |f| \le K$.

Another very important theorem is the dominated convergence theorem.

Theorem 7.7 *Suppose f_n are measurable functions and $f_n(x) \to f(x)$. Suppose there exists an integrable function g such that $|f_n(x)| \le g(x)$ for all x. Then $\int f_n \, d\mu \to \int f \, d\mu$.*

Proof. Since $f_n + g \ge 0$, by Fatou's lemma,

$$\int (f+g) \le \liminf \int (f_n + g).$$

Since g is integrable,

$$\int f \le \liminf \int f_n.$$

Similarly, $g - f_n \ge 0$, so

$$\int (g - f) \le \liminf \int (g - f_n),$$

and hence

$$-\int f \le \liminf \int (-f_n) = -\limsup \int f_n.$$

Therefore

$$\int f \ge \limsup \int f_n,$$

which with the above proves the theorem. □

Example 7.3 is an example where the limit of the integrals is not the integral of the limit because there is no dominating function g.

If in the monotone convergence theorem or dominated convergence theorem we have only $f_n(x) \to f(x)$ almost everywhere, the conclusion still holds. For if the f_n and f are measurable and $A = \{x : f_n(x) \to f(x)\}$, then $f\chi_A \to f\chi_A$ for each x. And since A^c has measure 0, we see from Proposition 6.5(d) that $\int f\chi_A = \int f$, and similarly with f replaced by f_n.

8 Properties of Lebesgue integrals

Later on we will need the following two propositions.

Proposition 8.1 *Suppose f is measurable and for every measurable set A we have $\int_A f\,d\mu = 0$. Then $f = 0$ almost everywhere.*

Proof. Let $A = \{x : f(x) > \varepsilon\}$. Then

$$0 = \int_A f \geq \int_A \varepsilon = \varepsilon\mu(A)$$

since $f\chi_A \geq \varepsilon\chi_A$. Hence $\mu(A) = 0$. We use this argument for $\varepsilon = 1/n$ and $n = 1, 2, \dots$, so $\mu\{x : f(x) > 0\} = 0$. Similarly $\mu\{x : f(x) < 0\} = 0$. □

Proposition 8.2 *Suppose f is measurable and nonnegative and $\int f\,d\mu = 0$. Then $f = 0$ almost everywhere.*

Proof. If f is not almost everywhere equal to 0, there exists an n such that $\mu(A_n) > 0$ where $A_n = \{x : f(x) > 1/n\}$. But then since f is nonnegative,

$$0 = \int f \geq \int_{A_n} f \geq \frac{1}{n}\mu(A_n),$$

a contradiction. □

We give a result on approximating a function on $\mathbb{R}$ by continuous functions.

Proposition 8.3 *Suppose f is a measurable function from $\mathbb{R}$ to $\mathbb{R}$ that is integrable. Let $\varepsilon > 0$. Then there exists a continuous function that is 0 outside some bounded interval such that*

$$\int |f - g| < \varepsilon.$$

Proof. If we have continuous functions g_1, g_2 such that $\int |f^+ - g_1| < \varepsilon/2$ and $\int |f^- - g_2| < \varepsilon/2$, where $f^+ = \max(f, 0)$ and $f^- = \max(-f, 0)$, then taking $g = g_1 - g_2$ will prove our result. So without loss of generality, we may assume $f \geq 0$.

By monotone convergence $\int f\chi_{[-n,n]}$ increases to $\int f$, so by taking n large enough, the difference of the integrals will be less than $\varepsilon/2$. If we find g such that $\int |f\chi_{[-n,n]} - g| < \varepsilon/2$, then $\int |f - g| < \varepsilon$. Therefore we may assume that f is 0 outside some bounded interval.

We can find simple functions increasing to f whose integrals increase to $\int f$. Let s_m be a simple function such that $s_m \leq f$ and $\int s_m \geq \int f - \varepsilon/2$. If we find g such that $\int |s_m - g| < \varepsilon/2$, then $\int |f - g| < \varepsilon$. So it suffices to consider the case where f is a simple function.

If $f = \sum_{i=1}^p a_i\chi_{A_i}$ and we find g_i continuous such that $\int |a_i\chi_{A_i} - g_i| < \varepsilon/p$, then $\sum_{i=1}^p g_i$ will be the desired function. So we may assume f is a constant times a characteristic function, and by linearity, we may assume f is equal to χ_A for some A contained in a bounded interval $[-n, n]$.

We can choose G open and F closed such that $F \subset A \subset G$ and $m(G-F) < \varepsilon$. We can replace G by $G \cap (-n-1, n+1)$. $G^c \cap [-n-1, n+1]$ and F are compact sets, so there is a minimum distance between them, say, δ. Let $g(x) = \max(0, 1 - \operatorname{dist}(x, F)/\delta)$. Then g is continuous, $0 \leq g \leq 1$, g is 1 on F, g is 0 on G^c, and g is 0 outside of $[-n-1, n+1]$. Therefore

$$|g - \chi_A| \leq \chi_G - \chi_F,$$

so

$$\int |g - \chi_A| \leq \int (\chi_G - \chi_F) = m(G - F) < \varepsilon.$$

□

The method of proof, where one proves a result for characteristic functions, then simple functions, then non-negative functions, and then finally integrable functions is very common.

We finish this section with a comparison of the Lebesgue integral and the Riemann integral. Here we are only looking at bounded functions from $[a, b]$ into $\mathbb{R}$. If we are looking at the Lebesgue integral, we write $\int f$, while, temporarily, if we are looking at the Riemann integral, we write $R(f)$. Recall that the Riemann integral on $[a, b]$ is defined as follows: if P is a partition of $[a, b]$, then

$$U(P, f) = \sum_{i=1}^{n} (\sup_{x_{i=1} \leq x \leq x} f(x)) \, (x_i - x_{i-1})$$

and

$$L(P, f) = \sum_{i=1}^{n} (\inf_{x_{i=1} \leq x \leq x} f(x)) \, (x_i - x_{i-1}).$$

Set $\overline{R}(f) = \inf\{U(P, f) : P \text{ is a partition}\}$ similarly $\underline{R}(f)$. Then the Riemann integral exists if $\overline{R}(f) = \underline{R}(f)$, and the common value is the Riemann integral, which we denote $R(f)$.

Theorem 8.4 *A bounded measurable function f on $[a, b]$ is Riemann integrable if and only if the set of points at which f is discontinuous has Lebesgue measure 0, and in that case, the Riemann integral is equal in value to the Lebesgue integral.*

Proof. If P is a partition, define

$$T_P(x) = \sum_{i=1}^{n} (\sup_{x_{i-1} \leq y \leq x_i} f(y)) \chi_{[x_{i-1}, x_i)}(x),$$

and

$$S_P(x) = \sum_{i=1}^{n} (\inf_{x_{i-1} \leq y \leq x_i} f(y)) \chi_{[x_{i-1}, x_i)}(x).$$

We see that $\int T_P = U(P, f)$ and $\int S_P = L(P, f)$.

If f is Riemann integrable, there exists a sequence of partitions Q_i such that $U(Q_i, f) \downarrow R(f)$ and a sequence Q'_i such that $L(Q'_i, f) \uparrow R(f)$. It is not hard to check that adding points to a partition increases L and decreases U, so if we let $P_i = \cup_{j \leq i}(Q_j \cup Q'_j)$, then P_i is an increasing sequence of partitions, $U(P_i, f) \downarrow R(f)$, $L(P_i, f) \uparrow R(f)$. We see also that $T_{P_i}(x)$ decreases at each point, say, to $T(x)$, and $S_{P_i}(x)$ increases at each point, say, to $S(x)$. Also

$T(x) \geq f(x) \geq S(x)$. Then by dominated convergence (recall that f is bounded)

$$\int (T - S) = \lim_{i\to\infty} \int (T_{P_i} - S_{P_i}) = \lim_{i\to\infty} (U(P_i, f) - L(P_i, f)) = 0.$$

We conclude $T = S = f$ a.e. If x is not in the null set where $T(x) \neq S(x)$ nor in $\cup_i P_i$, which is countable and hence of Lebesgue measure 0, then $T_{P_i}(x) \downarrow f(x)$ and $S_{P_i}(x) \uparrow f(x)$. This implies that f is continuous at such f. Since

$$R(f) = \lim_{i\to\infty} U(P_i, f) = \lim_{i\to\infty} \int T_{P_i} = \int f,$$

we see the Riemann integral and Lebesgue integral agree.

Now suppose that f is continuous a.e. Let $\varepsilon > 0$. Let P_i be the partition where we divide $[a, b]$ into 2^i equal parts. If x is not in the null set where f is discontinuous, nor in $\cup_{i=1}^\infty P_i$, then $T_{P_i}(x) \downarrow f(x)$ and $S_{P_i}(x) \uparrow f(x)$. By dominated convergence,

$$U(P_i, f) = \int T_{P_i} \to \int f$$

and

$$L(P_i, f) = \int S_{P_i} \to \int f.$$

This does it. □

9 Modes of convergence

Definition 9.1 *If μ is a measure, we say a sequence of measurable functions f_n converges to f almost everywhere (written $f_n \to f$ a.e.) if there is a set of measure 0 and for x not in this set we have $f_n(x) \to f(x)$.*

We say f_n converges to f in measure if for each $\varepsilon > 0$

$$\mu(\{x : |f_n(x) - f(x)| > \varepsilon\}) \to 0$$

as $n \to \infty$.

Proposition 9.2 *Suppose μ is a finite measure.*

(a) If $f_n \to f$, a.e., then f_n converges to f in measure.

(b) If $f_n \to f$ in measure, there is a subsequence n_j such that $f_{n_j} \to f$, a.e.

Proof. Let $\varepsilon > 0$. If $A_n = \{x : |f_n(x) - f(x)| > \varepsilon\}$, then $\chi_{A_n} \to 0$ a.e., and by dominated convergence,

$$\mu(A_n) = \int \chi_{A_n}(x)\, \mu(dx) \to 0.$$

This proves (a).

To prove (b), let $n_1 = 1$ and choose $n_j > n_{j-1}$ inductively so that

$$\mu(\{x : |f_{n_j}(x) - f(x)| > 1/j\}) \le 2^{-j}.$$

Let $A_j = \{x : |f_{n_j}(x) - f(x)| > 1/j\}$. Then $\mu(A_j) \le 2^{-j}$, and

$$A = \cap_{k=1}^{\infty} \cup_{j=k}^{\infty} A_j$$

has measure less than $\cup_{j=k} A_j$ for every k, hence less than 2^{-k+1} for every k. Therefore A has measure 0. If $x \notin A$, then $x \notin \cup_{j=k}^{\infty} A_j$ for some k, so $|f_{n_j}(x) - f(x)| \le 1/j$ for $j \ge k$, which means $f_{n_j} \to f$ a.e. on A^c. □

Example 9.3 Part (a) of the above proposition is not true if $\mu(X) = \infty$. Let $X = \mathbb{R}$ and let $f_n = \chi_{(n,n+1)}$.

Example 9.4 For an example where $f_n \to f$ in measure but not almost everywhere, let $X = [0,1]$, let μ be Lebesgue measure, and let $f_n(x) = \chi_{F_n}(x)$, where $F_n = \{y : (\sum_{j=1}^{n} 1/j)(\text{ mod } 1) \le y \le (\sum_{j=1}^{n+1} 1/j)(\text{ mod } 1)\}$. $z(\text{ mod } 1)$ is defined as the fractional part of z (where the largest integer less than z is subtracted from z). Let $f(x) = 0$ for all x.

Then $\mu(F_n) \le 1/n \to 0$, so $f_n \to f$ in measure. But any x will be in infinitely many F_n's, so f_n does not converge to $f(x)$ at any point.

The following is known as Egoroff's theorem.

Theorem 9.5 *If μ is a finite measure, $\varepsilon > 0$, and $f_n \to f$ a.e., then there exists a measurable set A such that $\mu(A) < \varepsilon$ and $f_n \to f$ uniformly on A^c.*

This type of convergence is sometimes known as almost uniform convergence.

Proof. Let

$$E_{nk} = \cup_{m=n}^{\infty}\{x : |f_m(x) - f(x)| > 1/k\}.$$

for fixed k, E_{nk} decreases as n increases, and the intersection $\cap_n E_{nk}$ has measure 0. So $\mu(E_{nk}) \to 0$. Then there exists an integer n_k such that $\mu(E_{n_k k}) < \varepsilon 2^{-k}$. Let $E = \cup_{k=1}^{\infty} E_{n_k k}$. Then $\mu(E) < \varepsilon$, and if $x \notin E$ and $n > n_k$, then $|f_n(x) - f(x)| \le 1/k$. Thus $f_n \to f$ uniformly on E^c. □

10 Product measures

If $A_1 \subset A_2 \subset \cdots$ and $A = \cup_{i=1}^{\infty} A_i$, we write $A_i \uparrow A$. If $A_1 \supset A_2 \supset \cdots$ and $A = \cap_{i=1}^{\infty} A_i$, we write $A_i \downarrow A$.

Definition 10.1 *$\mathcal{M}$ is a monotone class is $\mathcal{M}$ is a collection of subsets of X such that*

(a) if $A_i \uparrow A$ and each $A_i \in \mathcal{M}$, then $A \in \mathcal{M}$;

(b) if $A_i \downarrow A$ and each $A_i \in \mathcal{M}$, then $A \in \mathcal{M}$.

The intersection of monotone classes is a monotone class, and the intersection of all monotone classes containing a given collection of sets is the smallest monotone class containing that collection.

The next theorem, the monotone class lemma, is rather technical, but very useful.

Theorem 10.2 *Suppose $\mathcal{A}_0$ is a algebra, $\mathcal{A}$ is the smallest σ-algebra containing $\mathcal{A}_0$, and $\mathcal{M}$ is the smallest monotone class containing $\mathcal{A}_0$. Then $\mathcal{M} = \mathcal{A}$.*

Proof. A σ-algebra is clearly a monotone class, so $\mathcal{M} \subset \mathcal{A}$. We must show $\mathcal{A} \subset \mathcal{M}$.

Let $\mathcal{N}_1 = \{A \in \mathcal{M} : A^c \in \mathcal{M}\}$. Note $\mathcal{N}_1$ is contained in $\mathcal{M}$, contains $\mathcal{A}_0$, and is a monotone class. So $\mathcal{N}_1 = \mathcal{M}$, and therefore $\mathcal{M}$ is closed under the operation of taking complements.

Let $\mathcal{N}_2 = \{A \in \mathcal{M} : A \cap B \in \mathcal{M} \text{ for all } B \in \mathcal{A}_0\}$. $\mathcal{N}_2$ is contained in $\mathcal{M}$; $\mathcal{N}_2$ contains $\mathcal{A}_0$ because $\mathcal{A}_0$ is an algebra; $\mathcal{N}_2$ is a monotone class because $(\cup_{i=1}^\infty A_i) \cap B = \cup_{i=1}^\infty (A_i \cap B)$, and similarly for intersections. Therefore $\mathcal{N}_2 = \mathcal{M}$; in other words, if $B \in \mathcal{A}_0$ and $A \in \mathcal{M}$, then $A \cap B \in \mathcal{M}$.

Let $\mathcal{N}_3 = \{A \in \mathcal{M} : A \cap B \in \mathcal{M} \text{ for all } B \in \mathcal{M}\}$. As in the preceding paragraph, $\mathcal{N}_3$ is a monotone class contained in $\mathcal{M}$. By the last sentence of the preceding paragraph, $\mathcal{N}_3$ contains $\mathcal{A}_0$. Hence $\mathcal{N}_3 = \mathcal{M}$.

We thus have that $\mathcal{M}$ is a monotone class closed under the operations of taking complements and taking intersections. This shows $\mathcal{M}$ is a σ-algebra, and so $\mathcal{A} \subset \mathcal{M}$. □

Suppose $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ are two measure spaces, i.e., $\mathcal{A}$ and $\mathcal{B}$ are σ-algebras on X and Y, resp., and μ and ν are measures on $\mathcal{A}$ and $\mathcal{B}$, resp. A *rectangle* is a set of the form $A \times B$, where $A \in \mathcal{A}$ and $B \in \mathcal{B}$. Define a set function $\mu \times \nu$ on rectangles by

$$\mu \times \nu(A \times B) = \mu(A)\nu(B).$$

Lemma 10.3 *Suppose $A \times B = \cup_{i=1}^\infty A_i \times B_i$, where $A, A_i \in \mathcal{A}$ and $B, B_i \in \mathcal{B}$ and the $A_i \times B_i$ are disjoint. Then*

$$\mu \times \nu(A \times B) = \sum_{i=1}^\infty \mu \times \nu(A_i \times B_i).$$

Proof. We have

$$\chi_{A \times B}(x, y) = \sum_{i=1}^\infty \chi_{A_i \times B_i}(x, y),$$

and so

$$\chi_A(x)\chi_B(y) = \sum_{i=1}^\infty \chi_{A_i}(x)\chi_{B_i}(y).$$

Holding x fixed and integrating over y with respect to ν, we have, using (7.1),

$$\chi_A(x)\nu(B) = \sum_{i=1}^{\infty} \chi_{A_i}(x)\nu(B_i).$$

Now use (7.1) again and integrate over x with respect to μ to obtain the result. □

Let $\mathcal{C}_0 = \{\text{finite unions of rectangles}\}$. It is clear that $\mathcal{C}_0$ is an algebra. By Lemma 10.3 and linearity, we see that $\mu \times \nu$ is a measure on $\mathcal{C}_0$. Let $\mathcal{A} \times \mathcal{B}$ be the smallest σ-algebra containing $\mathcal{C}_0$; this is called the *product σ-algebra*. By the Carathéodory extension theorem, $\mu \times \nu$ can be extended to a measure on $\mathcal{A} \times \mathcal{B}$.

We will need the following observation. Suppose a measure μ is σ-finite. So there exist E_i which have finite μ measure and whose union is X. If we let $F_n = \cup_{i=1}^n E_i$, then $F_i \uparrow X$ and $\mu(F_n)$ is finite for each n.

If μ and ν are both σ-finite, say with $F_i \uparrow X$ and $G_i \uparrow Y$, then $\mu \times \nu$ will be σ-finite, using the sets $F_i \times G_i$.

The main result of this section is Fubini's theorem, which allows one to interchange the order of integration.

Theorem 10.4 *Suppose $f : X \times Y \to \mathbb{R}$ is measurable with respect to $\mathcal{A} \times \mathcal{B}$. If f is nonnegative or $\int |f(x,y)|\, d(\mu \times \nu)(x,y) < \infty$, then*

(a) the function $g(x) = \int f(x,y)\nu(dy)$ is measurable with respect to $\mathcal{A}$;

(b) the function $h(y) = \int f(x,y)\mu(dx)$ is measurable with respect to $\mathcal{B}$;

(c) we have

$$\begin{aligned} \int f(x,y)\, d(\mu \times \nu)(x,y) &= \int \Big(\int f(x,y)\, d\mu(x) \Big)\, d\nu(y) \\ &= \int \Big(\int f(x,y)\, d\nu(y) \Big)\, \mu(dx). \end{aligned}$$

Proof. First suppose μ and ν are finite measures. If f is the characteristic function of a rectangle, then (a)–(c) are obvious. By linearity, (a)–(c) hold if f is the characteristic function of a set in $\mathcal{C}_0$, the set of finite unions of rectangles.

Let $\mathcal{M}$ be the collection of sets C such that (a)–(c) hold for χ_C. If $C_i \uparrow C$ and $C_i \in \mathcal{M}$, then (c) holds for χ_C by monotone convergence. If $C_i \downarrow C$, then (c) holds for χ_C by dominated convergence. (a) and (b) are easy. So $\mathcal{M}$ is a monotone class containing $\mathcal{A}_0$, so $\mathcal{M} = \mathcal{A} \times \mathcal{B}$.

If μ and ν are σ-finite, applying monotone convergence to $C \cap (F_n \times G_n)$ for suitable F_n and G_n and monotone convergence, we see that (a)–(c) holds for the characteristic functions of sets in $\mathcal{A} \times \mathcal{B}$ in this case as well.

By linearity, (a)–(c) hold for nonnegative simple functions. By monotone convergence, (a)–(c) hold for nonnegative functions. In the case $\int |f| < \infty$, writing $f = f^+ - f^-$ and using linearity proves (a)–(c) for this case, too. □

11 The Radon-Nikodym theorem

Suppose f is nonnegative, measurable, and integrable with respect to μ. If we define ν by

$$\nu(A) = \int_A f \, d\mu, \tag{11.1}$$

then ν is a measure. The only part that needs thought is the countable additivity, and this follows from (7.1) applied to the functions $f\chi_{A_i}$. Moreover, $\nu(A)$ is zero whenever $\mu(A)$ is. We sometimes write $f = d\nu/d\mu$ for (11.1).

Definition 11.1 *A measure ν is called* absolutely continuous *with respect to a measure μ if $\nu(A) = 0$ whenever $\mu(A) = 0$. This is frequently written $\nu \ll \mu$.*

Proposition 11.2 *A finite measure ν is absolutely continuous with respect to μ if and only if for all ε there exists δ such that $\mu(A)$ implies $\nu(A) < \varepsilon$.*

Proof. If the condition given in the statement of the proposition holds, it is clear that $\nu \ll \mu$. Suppose now that $\nu \ll \mu$. If the condition does not hold, there exists E_k such that $\mu(E_k) < 2^{-k}$ but $\nu(E_k) \geq \varepsilon$. Let $F = \cap_{n=1}^{\infty} \cup_{k=n}^{\infty} E_k$. Then

$$\mu(F) = \lim_{n\to\infty} \mu(\cup_{k=n}^{\infty} E_k) \leq \lim_{n\to\infty} \sum_{k=n}^{\infty} 2^{-k} = 0,$$

but

$$\nu(F) = \lim_{n\to\infty} \nu(\cup_{k=n}^{\infty} E_k) \geq \varepsilon;$$

This contradicts the absolute continuity. □

Definition 11.3 *A function $\mu : \mathcal{A} \to (-\infty, \infty]$ is called a signed measure if $\mu(\emptyset) = 0$ and $\mu(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$ whenever the A_i are disjoint and all the A_i are in $\mathcal{A}$.*

Definition 11.4 *Let μ be a signed measure. A set $A \in \mathcal{A}$ is called a positive set for μ if $\mu(B) \geq 0$ whenever $B \subset A$ and $B \in \mathcal{A}$. We define a negative set similarly. A null set A is one where $\mu(B) = 0$ whenever $B \subset A$ is measurable.*

Example 11.5 Suppose m is Lebesgue measure and $\mu(A) = \int_A f\, dm$ for some integrable f. If we let $P = \{x : f(x) \geq 0\}$, then P is easily seen to be a positive set, and if $N = \{x : f(x) < 0\}$, then N is a negative one. The Hahn decomposition which we give below is a decomposition of our space (in this case $\mathbb{R}$) into positive and negative sets. This decomposition is unique, except that $C = \{x : f(x) = 0\}$ could be included in N instead of P, or apportioned partially to P and partially to N. Note, however, that C is a null set. The Jordan decomposition below is a decomposition of μ into μ^+ and μ^-, where $\mu^+(A) = \int_A f^+\, dm$, and similarly $\mu^-(A) = \int_A f^-\, d\mu$.

Note that if μ is a signed measure, then $\mu(\cup_{i=1}^{\infty} A_i) = \lim_{n\to\infty} \mu(\cup_{i=1}^{n} A_i)$. The proof is the same as in the case of positive measures.

Proposition 11.6 *Let μ be a signed measure taking values in $(-\infty, \infty]$. Let E be measurable with $\mu(E) < 0$. Then there exists a subset F of E that is a negative set with $\mu(F) < 0$.*

Proof. If E is a negative set, we are done. If not, there exists a subset with positive measure. Let n_1 be the smallest positive integer such that there exists $E_1 \subset E$ with $\mu(E_1) \geq 1/n$. Let $k \geq 2$. If $F_k = E - (E_1 \cup \cdots \cup E_{k-1})$ is negative, we are done. If not, let n_k be the smallest positive integer such that there exists $E_k \subset F_k$ with $\mu(E_k) \geq 1/n_k$. We continue.

If the construction stops after a finite number of sets, we are done. If not, let $F = \cap_k F_k = E - (\cup_k E_k)$. Since $0 > \mu(E) > -\infty$ and $\mu(E_k) \geq 0$, then

$$\mu(E) = \mu(F) + \sum_{k=1}^{\infty} \mu(E_k).$$

Then $\mu(F) \leq \mu(E) < 0$, so the sum converges. If $G \subset F$ is measurable with $\mu(G) > 0$, then $\mu(G) \geq 1/N$ for some N, which contradicts the construction. Therefore F must be a negative set. □

We write $A \Delta B$ for $(A - B) \cup (B - A)$. The following is known as the Hahn decomposition theorem.

Theorem 11.7 *Let μ be a signed measure taking values in $(-\infty, \infty]$. There exist sets E and F in $\mathcal{A}$ that are disjoint whose union is X and such that E is a negative set and F is a positive set. If E' and F' are another such pair, then $E \Delta E' = F \Delta F'$ is a null set with respect to μ..*

Proof. Let $L = \inf\{\mu(A) : A \text{ is a negative set}\}$. Choose negative sets A_n such that $\mu(A_n) \to L$. Let $E = \cup_{n=1}^{\infty} A_n$. Let $B_n = A_n - (B_1 \cup \cdots \cup B_{n-1})$ for each n. Since A_n is a negative set, so is each B_n. Also, the B_n are disjoint. If $C \subset E$, then

$$\mu(C) = \lim_{n\to\infty} \mu(C \cap (\cup_{i=1}^{n} B_i)) = \lim_{n\to\infty} \sum_{i=1}^{n} \mu(C \cap B_i) \leq 0.$$

So E is a negative set.

Since E is negative,

$$\mu(E) = \mu(A_n) + \mu(E - A_n) \leq \mu(A_n).$$

Letting $n \to \infty$, we obtain $\mu(E) = L$.

Let $F = E^c$. If F were not a positive set, there would exist $B \subset F$ with $\mu(B) < 0$. By Proposition 11.6 there exists a negative set C contained in B with $\mu(C) < 0$. But then $E \cup C$ would be a negative set with $\mu(E \cup C) < \mu(E) = L$, a contradiction.

To prove uniqueness, if E', F' are another such pair of sets and $A \subset E - E' \subset E$, then $\mu(A) \leq 0$. But $A \subset E - E' = F' - F \subset F'$, so $\mu(A) \leq 0$. Therefore $\mu(A) = 0$. The same argument works if $A \subset E' - E$, and any subset of $E \Delta E'$ can be written as the union of A_1 and A_2, where $A_{\subset} E - E'$ and $A_2 \subset E' - E$. □

Let us say two measures μ and ν are mutually singular if there exist two disjoint sets E and F in $\mathcal{A}$ whose union is X with $\mu(E) = \nu(F) = 0$. This is often written $\mu \perp \nu$.

Example 11.8 If μ is Lebesgue measure restricted to $[0, 1/2]$, that is, $\mu(A) = m(A \cap [0, 1/2])$, and ν is Lebesgue measure restricted to $[1/2, 1]$, then μ and ν are mutually singular. We let $E = [0, 1/2]$ and $F = (1/2, 1]$. This example works because the Lebesgue measure of $\{1/2\}$ is 0.

Example 11.9 A more interesting example is the following. Let f be the Cantor-Lebesgue function and let ν be the Lebesgue-Stieltjes measure associated with f. Let μ be Lebesgue measure restricted to $[0, 1]$. Then $\mu \perp \nu$. To see this, we let $E = C$, where C is the Cantor set, and $F = [0, 1] - C$. We already know that $m(E) = 0$ and we need to show $\nu(F) = 0$. To do that, we need to show $\nu(I) = 0$ for every open interval contained in F. This will follow if we show $\nu(J) = 0$ for every interval of the form $J = (a, b]$ contained in F. But f is constant on every such interval, so $f(b) = f(a)$, and therefore $\nu(J) = f(b) - f(a) = 0$.

The following is known as the Jordan decomposition theorem.

Theorem 11.10 *If μ is a signed measure, there exist measures μ^+ and μ^- such that $\mu = \mu^+ - \mu^-$ and μ^+ and μ^- are mutually singular. This decomposition is unique.*

Proof. Let E and F be positive and negative sets for μ and let $\mu^+(A) = \mu(E \cap A)$, $\mu^-(A) = -\mu(A \cap F)$. This gives the desired decomposition.

If $\mu = \nu^+ - \nu^-$ is another such decomposition with ν^+, ν^- mutually singular, let E' and F' be the sets in the definition of mutually singular. Then $X = E' \cup F'$ gives another Hahn decomposition, hence $E \Delta E'$ is a null set with respect to μ. Then for any $A \in \mathcal{A}$,

$$\nu^+(A) = \mu(A \cap E') = \mu(A \cap E) = \mu^+(A),$$

and similarly for ν^-, μ^-. □

The measure $\mu^+ + \mu^-$ is called the total variation measure and is written $|\mu|$.

We now are ready for the Radon-Nikodym theorem.

Theorem 11.11 *Suppose μ is a σ-finite measure and ν is a finite measure such that ν is absolutely continuous with respect to μ. There exists a μ-integrable nonnegative function f such that $\nu(A) = \int_A f\, d\mu$ for all $A \in \mathcal{A}$. Moreover, if g is another such function, then $f = g$ almost everywhere with respect to μ.*

Proof. Let us first prove the uniqueness assertion. For every set A we have

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

By Proposition 8.1 we have $f - g = 0$ a.e. with respect to μ.

Since μ is σ-finite, there exist $F_i \uparrow X$ such that $\mu(F_i) < \infty$ for each i. Let μ_i be the restriction of μ to F_i, that is, $\mu_i(A) = \mu(A \cap F_i)$. Define ν_i, the restriction of ν to F_i, similarly. If f_i is a function such that $\nu_i(A) = \int_A f_i\, d\mu_i$ for all A, the argument of the first paragraph shows that $f_i = f_j$ on F_i if $i \le j$. If we define f by $f(x) = f_i(x)$ if $x \in F_i$, we see that f will be the desired function. So it suffices to restrict attention to the case where μ is finite.

Let

$$\mathcal{F} = \Big\{ g : 0 \le g, \int_A g\, d\mu \le \nu(A) \text{ for all } A \in \mathcal{A} \Big\}.$$

$\mathcal{F}$ is not empty because $0 \in \mathcal{F}$. Let $L = \sup\{\int g\, d\mu : g \in \mathcal{F}\}$, and let g_n be a sequence in $\mathcal{F}$ such that $\int g_n\, d\mu \to L$. Let $h_n = \max(g_1, \ldots, g_n)$.

If g_1 and g_2 are in $\mathcal{F}$, then $h_2 = \max(g_1, g_2)$ is also in $\mathcal{F}$. To see this, let $B = \{x : g_1(x) \geq g_2(x)\}$, and write

$$\begin{aligned} \int_A h_2 \, d\mu &= \int_{A\cap B} h_2 \, d\mu + \int_{A\cap B^c} h_2 \, d\mu \\ &= \int_{A\cap B} g_1 \, d\mu + \int_{A\cap B^c} g_2 \, d\mu \\ &\leq \nu(A \cap B) + \nu(A \cap B^c) \\ &= \nu(A). \end{aligned}$$

By an induction argument, h_n is in $\mathcal{F}$.

The h_n increase, say to f. By monotone convergence $\int f \, d\mu = L$ and

$$\int_A f \, d\mu \leq \nu(A) \tag{11.2}$$

for all A.

Let A be a set where there is strict inequality in (11.2); let ε be chosen sufficiently small so that if π is defined by

$$\pi(B) = \nu(B) - \int_B f \, d\mu - \varepsilon\mu(B),$$

then $\pi(A) > 0$. π is a signed measure; let F be the positive set as constructed in Theorem 11.7. In particular, $\pi(F) > 0$. So for every B

$$\int_{B\cap F} f \, d\mu + \varepsilon\mu(B \cap F) \leq \nu(B \cap F).$$

We then have, using (11.2), that

$$\begin{aligned} \int_B (f + \varepsilon\chi_F) \, d\mu &= \int_B f \, d\mu + \varepsilon\mu(B \cap F) \\ &= \int_{B\cap F^c} f \, d\mu + \int_{B\cap F} f \, d\mu + \varepsilon\mu(B \cap F) \\ &\leq \nu(B \cap F^c) + \nu(B \cap F) = \nu(B). \end{aligned}$$

This says that $f + \varepsilon\chi_F \in \mathcal{F}$. However,

$$L \geq \int (f + \varepsilon\chi_F) \, d\mu = \int f \, d\mu + \varepsilon\mu(F) = L + \varepsilon\mu(F),$$

which implies $\mu(F) = 0$. But then $\nu(F) = 0$, and hence $\pi(F) = 0$, contradicting the fact that F is a positive set for F with $\pi(F) > 0$. □

The proof of the Lebesgue decomposition theorem is almost the same.

Theorem 11.12 *Suppose μ and ν are two finite measures. There exist measures λ, ρ such that $\nu = \lambda + \rho$, ρ is absolutely continuous with respect to μ, and λ and μ are mutually singular.*

Proof. Define $\mathcal{F}$ and L and construct f as in the proof of the Radon-Nikodym theorem. Let $\rho(A) = \int_A f\, d\mu$ and let $\lambda = \nu - \rho$. We have $\int_A f\, d\mu \le \nu(A)$, so $\lambda(A) \ge 0$ for all A. To keep things straight, we record that we have $f = d\rho/d\mu$ and $\rho + \lambda = \nu$. We need to show μ and λ are mutually singular.

Suppose not. Then there exists $F \in \mathcal{A}$ with $\mu(F) > 0$ and $\lambda(F) > 0$, and so $\nu(F) \ge \lambda(F) > 0$. Note that for ε small enough, $(\lambda - \varepsilon\mu)(F) > 0$. We claim that there exist $\varepsilon > 0$ and $E \subset F$ such that E is a positive set with respect to $\lambda - \varepsilon\mu$ and $\mu(E) > 0$. Given the claim, if $A \in \mathcal{A}$,

$$\begin{aligned}\varepsilon \int_A \chi_E\, d\mu &= \varepsilon\mu(A \cap E) \le \nu(A \cap E)\\ &\le \lambda(A \cap E) \le \lambda(A)\\ &= \nu(A) - \int_A f\, d\mu.\end{aligned}$$

This says that

$$\int_A (f + \varepsilon\chi_E)\, d\mu \le \nu(A)$$

for all $A \in \mathcal{A}$, or $f + \varepsilon\chi_E \in \mathcal{F}$. But

$$\int (f + \varepsilon\chi_E)\, d\mu = \int f\, d\mu + \varepsilon\mu(E) > L,$$

a contradiction to the definition of L.

It remains to prove the claim. Let $F = P_n \cup N_n$ be a Hahn decomposition for the measure $\lambda - \frac{1}{n}\mu$ restricted to F, let $P = \cup P_n$, and $N = \cap N_n = F - P$. Then N is a negative set for $\lambda - \frac{1}{n}\mu$ for each n, or $0 \le \lambda(N) \le \frac{1}{n}\mu(N)$; this implies $\lambda(N) = 0$. If $\mu(P) = 0$, then $\mu \perp \lambda$, and we are supposing that is not the case. Therefore $\mu(P) > 0$, hence $\mu(P_n) > 0$ for some n, and P_n is a positive set for $\lambda - \frac{1}{n}\mu$. Now take $\varepsilon = 1/n$ and $E = P_n$. $\square$

12 Differentiation of real-valued functions

In this section we want to look at when $f : \mathbb{R} \to \mathbb{R}$ is differentiable and when the fundamental theorem of calculus holds. Briefly,

(1) Functions of bounded variation are differentiable;

(2) The derivative of $\int_a^x f(y)\,dy$ is equal to f a.e. if f is integrable;

(3) $\int_a^b f'(y)\,dy = f(b) - f(a)$ if f' is absolutely continuous.

Let $E \subset \mathbb{R}$ be a measurable set and let $\mathcal{O}$ be a collection of intervals. We say $\mathcal{O}$ is a *Vitali cover* of E if for each $x \in E$ and each $\varepsilon > 0$ there exists an interval $G \in \mathcal{O}$ containing x whose length is less than ε. m will denote Lebesgue measure.

Lemma 12.1 *Let E have finite measure and let $\mathcal{O}$ be a Vitali cover of E. Given $\varepsilon > 0$ there exists a finite subcollection of disjoint intervals $I_1, \ldots, I_n$ such that $m(E - \cup_{i=1}^n I_n) < \varepsilon$.*

Proof. We may replace each interval in $\mathcal{O}$ by a closed one, since the set of endpoints of a finite subcollection will have measure 0.

Let O be an open set of finite measure containing E. Since $\mathcal{O}$ is a Vitali cover, we may suppose without loss of generality that each set of $\mathcal{O}$ is contained in O. Let $a_0 = \sup\{m(I) : I \in \mathcal{O}\}$. Let I_1 be any element of $\mathcal{O}$ with $m(I_1) \geq a_0/2$. Let $a_1 = \sup\{m(I) : I \in \mathcal{O}, I \text{ disjoint from } I_1\}$,and choose $I_2 \in \mathcal{O}$ disjoint from I_1 such that $m(I_2) \geq a_1/2$. Continue in this way, choosing I_{n+1} disjoint from $I_1, \ldots, I_n$ and in $\mathcal{O}$ with length at least one half as large as any other such interval in $\mathcal{O}$ that is disjoint from $I_1, \ldots, I_n$.

If the process stops at some finite stage, we are done. If not, we generate a sequence of disjoint intervals $I_1, I_2, \ldots$ Since they are disjoint and all contained in O, then $\sum_{i=1}^{\infty} m(I_i) \le m(O) < \infty$. So there exists N such that $\sum_{i=N+1}^{\infty} m(I_i) < \varepsilon/5$.

Let $R = E - \cup_{i=1}^{N} I_i$; we will show $m(R) < \varepsilon$. Let J_n be the interval with the same center as I_n but five times the length. Let $x \in R$. There exists an interval $I \in \mathcal{O}$ containing x with I disjoint from $I_1, \ldots, I_N$. Since $\sum m(I_n) < \infty$, then $\sum a_n \le 2 \sum m(I_n) < \infty$, and $a_n \to 0$. So I must either be one of the I_n for some $n > N$ or at least intersect it, for otherwise we would have chosen I at some stage. Let n be the smallest integer such that I intersects I_n; note $n > N$. We have $m(I) \le a_{n-1} \le 2m(I_n)$. Since x is in I and I intersects I_n, the distance from x to the midpoint of I_n is at most $m(I) + m(I_n)/2 \le (5/2)m(I_n)$. Therefore $x \in J_n$.

Then $R \subset \cup_{i=N+1}^{\infty} J_n$, so $m(R) \le \sum_{i=N+1}^{\infty} m(J_n) = 5 \sum_{i=N+1}^{\infty} m(I_n) < \varepsilon$. □

Given a function f, we define the *derivates* of f at x by

$$D^+ f(x) = \limsup_{h \to 0+} \frac{f(x+h) - f(x)}{h}, \qquad D^- f(x) = \limsup_{h \to 0-} \frac{f(x+h) - f(x)}{h}$$

$$D_+ f(x) = \liminf_{h \to 0+} \frac{f(x+h) - f(x)}{h}, \qquad D_- f(x) = \liminf_{h \to 0-} \frac{f(x+h) - f(x)}{h}.$$

If all the derivates are equal, we say that f is *differentiable* at x and define $f'(x)$ to be the common value.

Theorem 12.2 *Suppose f is nondecreasing on $[a, b]$. Then f is differentiable almost everywhere, f' is integrable, and $\int_a^b f'(x)\, dx \le f(b) - f(a)$.*

Proof. We will show that the set where any two derivates are unequal has measure zero. We consider the set E where $D^+ f(x) > D_- f(X)$, the other sets being similar. Let $E_{u,v} = \{x : D^+ f(x) > u > v > D_- f(x)\}$. If we show $m(E_{u,v}) = 0$, then taking the union of all pairs of rationals with $u > v$ shows $m(E) = 0$.

Let $s = m(E_{u,v})$, let $\varepsilon > 0$, and choose an open set O such that $E_{u,v} \subset O$ and $m(O) < s + \varepsilon$. For each $x \in E_{u,v}$ there exists an arbitrarily small interval $[x - h, x]$ contained in O such that $f(x) - f(x - h) < vh$. Use Lemma 12.1 to choose $I_1, \ldots, I_n$ which are disjoint and whose interiors cover a subset A

of $E_{u,v}$ of measure greater than $s - \varepsilon$. Suppose $I_n = [x_n - h_n, x_n]$. Summing over these intervals,

$$\sum_{n=1}^{N} [f(x_n) - f(x_n - h_n)] < v \sum_{n=1}^{n} h_n < vm(O) < v(s + \varepsilon).$$

Each point $y \in A$ is the left endpoint of an arbitrarily small interval $(y, y + k)$ that is contained in some I_n and for which $f(y + k) - f(y) > uk$. Using Lemma 12.1 again, we pick out a finite collection $J_1, \ldots, J_M$ whose union contains a subset of A of measure larger than $s - 2\varepsilon$. Summing over these intervals yields

$$\sum_{i=1}^{M} [f(y_i + k_i) - f(y_i)] > u \sum k_i > u(s - 2\varepsilon).$$

Each interval J_i is contained in some interval I_n, and if we sum over those i for which $J_i \subset I_n$ we find

$$\sum [f(y_i + k_i) - f(y_i)] \leq f(x_n) - f(x_n - h_n),$$

since f is increasing. Thus

$$\sum_{n=1}^{N} [f(x_n) - f(x_n - h_n)] \geq \sum_{i=1}^{M} [f(y_i + k_i) - f(y_i)],$$

and so $v(s + \varepsilon) > u(s - 2\varepsilon)$. This is true for each ε, so $vs \geq us$. Since $u > v$, this implies $s = 0$.

This shows that

$$g(x) = \lim_{h \to 0} \frac{f(x + h) - f(x)}{h}$$

is defined almost everywhere and that f is differentiable wherever g is finite. Define $f(x) = f(b)$ if $x \geq b$. Let $g_n(x) = n[f(x + 1/n) - f(x)]$. Then $g_n(x) \to g(x)$ for almost all x, and so g is measurable. Since f is increasing, $g_n \geq 0$. By Fatou's lemma

$$\begin{aligned}\int_a^b g &\leq \liminf \int_a^b g_n = \liminf n \int_a^b [f(x + 1/n) - f(x)]dx \\ &= \liminf \left[n \int_b^{b+1/n} f - n \int_a^{a+1/n} f \right] = \liminf \left[f(b) - n \int_a^{a+1/n} f \right] \\ &\leq f(b) - f(a).\end{aligned}$$

For the last inequality, we use the fact that f is increasing. This shows that g is integrable and hence finite almost everywhere. □

A function is of *bounded variation* if $\sup\{\sum_{i=1}^k |f(x_i) - f(x_{i-1})|\}$ is finite, where the supremum is over all partitions $a = x_0 < x_1 < \cdots < x_k = b$ of $[a, b]$.

Lemma 12.3 *If f is of bounded variation on $[a, b]$, then f can be written as the difference of two nondecreasing functions on $[a, b]$.*

Proof. Define

$$P(y) = \sup\Big\{\sum_{i=1}^k [f(x_i) - f(x_{i-1})]^+\Big\}, \qquad N(y) = \sup\Big\{\sum_{i=1}^k [f(x_i) - f(x_{i-1})]^-\Big\},$$

where the supremum is over all partitions $a = x_0 < x_1 < \cdots < x_k = y$ for $y \in [a, b]$. P and N are measurable since they are both increasing. Since

$$\sum_{i=1}^k [f(x_i) - f(x_{i-1})]^+ = \sum_{i=1}^k [f(x_i) - f(x_{i-1})]^- + f(y) - f(a),$$

taking the supremum over all partitions of $[a, y]$ yields

$$P(y) = N(y) + f(y) - f(a).$$

Clearly P and N are nondecreasing in y, and the result follows by solving for $f(y)$. □

From this lemma, we see that functions of bounded variation are differentiable a.e. But the function $\sin(1/x)$ defined on $(0, 1]$ is differentiable everywhere, but is not of bounded variation.

Next we look at when the derivative of $\int_a^x f(t)\,dt$ is equal to $f(x)$ a.e. Define the *indefinite integral* of an integrable function f by

$$F(x) = \int_a^x f(t)\,dt.$$

Lemma 12.4 *If f is integrable, then F is continuous and of bounded variation.*

Proof. The continuity follows from the dominated convergence theorem. The bounded variation follows from

$$\sum_{i=1}^{k} |F(x_i) - F(x_{i-1})| = \sum_{i=1}^{k} \Big| \int_{x_{i-1}}^{x_i} f(t)\,dt \Big| \le \sum_{i=1}^{k} \int_{x_{i-1}}^{x_i} |f(t)|\,dt \le \int_a^b |f(t)|\,dt$$

for all partitions. □

Lemma 12.5 *If f is integrable and $F(x) = 0$ for all x, then $f = 0$ a.e.*

Proof. For any interval, $\int_c^d f = \int_a^d f - \int_a^c f = 0$. By dominated convergence and the fact that any open set is the countable union of disjoint open intervals, $\int_O f = 0$ for any open set O.

If E is any measurable set, take O_n open that such that χ_{O_n} decreases to χ_E a.e. By dominated convergence,

$$\int_E f = \int f\chi_E = \lim \int f\chi_{O_n} = \lim \int_{O_n} f = 0.$$

This with Proposition 8.1 implies f is zero a.e. □

Proposition 12.6 *If f is bounded and measurable, then $F'(x) = f(x)$ for almost every x.*

Proof. By Lemma 12.4, F is continuous and of bounded variation, and so F' exists a.e. Let K be a bound for $|f|$. If

$$f_n(x) = \frac{F(x + 1/n) - F(x)}{1/n},$$

then

$$f_n(x) = n \int_x^{x+1/n} f(t)\,dt,$$

so $|f_n|$ is also bounded by K. Since $f_n \to F'$ a.e., then by dominated convergence,

$$\begin{aligned}\int_a^c F'(x)\,dx &= \lim \int_a^c f_n(x)\,dx = \lim n \int_a^c [F(x+1/n) - F(x)]\,dx \\ &= \lim \Big[n \int_c^{c+1/n} F(x)\,dx - n \int_a^{a+c} F(x)\,dx \Big] \\ &= F(c) - F(a) = \int_a^c f(x)\,dx,\end{aligned}$$

using the fact that F is continuous. So $\int_a^c [F'(x) - f(x)]\,dx = 0$ for all c, which implies $F' = f$ a.e. by Lemma 12.5. □

Theorem 12.7 *If f is integrable, then $F' = f$ almost everywhere.*

Proof. Without loss of generality we may assume $f \geq 0$. Let $f_n(x) = f(x)$ if $f(x) \leq n$ and let $f_n(x) = n$ if $f(x) > n$. Then $f - f_n \geq 0$. If $G_n(x) = \int_a^x [f - f_n]$, then G_n is nondecreasing, and hence has a derivative almost everywhere. By Proposition 12.6, we know the derivative of $\int_a^x f_n$ is equal to f_n almost everywhere. Therefore

$$F'(x) = G_n'(x) + \Big[\int_a^x f_n \Big]' \geq f_n(x)$$

a.e. Since n is arbitrary, $F' \geq f$ a.e. So $\int_a^b F' \geq \int_a^b f = F(b) - F(a)$. On the other hand, by Theorem 12.2, $\int_a^b F'(x)\,dx \leq F(b) - F(a) = \int_a^b f$. We conclude that $\int_a^b [F' - f] = 0$; since $F' - f \geq 0$, this tells us that $F' = f$ a.e. □

Finally, we look at when $\int_a^b F'(y)\,dy = F(b) - F(a)$.

A function is *absolutely continuous* on $[a, b]$ if given ε there exists δ such that $\sum_{i=1}^k |f(x_i') - f(x_i)| < \varepsilon$ whenever $\{x_i, x_i')\}$ is a finite collection of nonoverlapping intervals with $\sum_{i=1}^k |x_i' - x_i| < \delta$.

It is easy to see that absolutely continuous functions are continuous and that the Cantor-Lebesgue function is not absolutely continuous.

Lemma 12.8 *If $F(x) = \int_a^x f(t)\,dt$ for f integrable on $[a,b]$, then F is absolutely continuous.*

Proof. Let $\varepsilon > 0$. Choose a simple function s such that $\int_a^b |f - s| < \varepsilon/2$. Let K be a bound for $|s|$ and let $\delta = \varepsilon/2K$. If $\{(x_i, x_i')\}$ is a collection of nonoverlapping intervals, the sum of whose lengths is less than δ, then set $A = \cup_{i=1}^k (x_i, x_i')$ and note $\int_A |f - s| < \varepsilon/2$ and $\int_A s < K\delta = \varepsilon/2$. □

Lemma 12.9 *If f is absolutely continuous, then it is of bounded variation.*

Proof. Let δ correspond to $\varepsilon = 1$ in the definition of absolute continuity. Given a partition, add points if necessary so that each subinterval has length at most δ. We can then group the subintervals into at most K collections, each of total length less than δ, where K is an integer larger than $(1+b-a)/\delta$. So the total variation is then less than K. □

Lemma 12.10 *If f is absolutely continuous on $[a,b]$ and $f'(x) = 0$ a.e., then f is constant.*

The Cantor-Lebesgue function is an example to show that we need the absolute continuity.

Proof. Let $c \in [a,b]$, let $E = \{x \in [a,c] : f'(x) = 0\}$, and let $\varepsilon > 0$. For each point $x \in E$ there exists arbitrarily small intervals $[x, x+h] \subset [a,c]$ such that $|f(x+h) - f(x)| < \varepsilon h$. By Lemma 12.1 we can find a finite collection of such intervals that cover all of E except for a set of measure less than δ, where δ is the δ in the definition of absolute continuity. If the intervals are $[x_i, y_i]$ with $x_i < y_i \le x_{i+1}$, then $\sum |f(x_{i+1}) - f(y_i)| < \varepsilon$ by the definition of absolute continuity, while $\sum |f(y_i) - f(x_i)| < \varepsilon \sum (y_i - x_i) \le \varepsilon(c - a)$. So adding these two inequalities together,

$$|f(c) - f(a)| = \left| \sum [f(x_{i+1}) - f(y_i)] + \sum [f(y_i) - f(x_i)] \right| \le \varepsilon + \varepsilon(c - a).$$

Since ε is arbitrary, then $f(c) = f(a)$, which implies that f is constant. □

Theorem 12.11 *If F is absolutely continuous, then*

$$F(b) - F(a) = \int_a^b F'(y)\, dy.$$

Proof. Suppose F is absolutely continuous on $[a, b]$. Then F is of bounded variation, so $F = F_1 - F_2$ where F_1 and F_2 are nondecreasing, and F' exists a.e. Since $|F'(x)| \le F_1'(x) + F_2'(x)$, then $\int |F'(x)|\, dx \le F_1(b) + F_2(b) - F_1(a) - F_2(a)$, and hence F' is integrable. If $G(x) = \int_a^x F'(t)\, dt$, then G is absolutely continuous by Lemma 12.8, so $F - G$ is absolutely continuous. Then $(F - G)' = 0$ a.e., and therefore $F - G$ is constant. Thus $F(x) = \int_a^x F'(t)\, dt + F(a)$. If we set $x = b$, we get our result. □

13 L^p spaces

We assume throughout this section that the measure is σ-finite. For $1 \le p < \infty$, define the L^p norm of f by

$$\|f\|_p = \Big(\int |f(x)|^p d\mu \Big)^{1/p}.$$

For $p = \infty$, define the L^∞ norm of f by

$$\|f\|_\infty = \inf\{M : \mu(\{x : |f(x)| \ge M\}) = 0\}.$$

For $1 \le p \le \infty$ the space L^p is the set $\{f : \|f\|_p < \infty\}$.

The L^∞ norm of a function f is the supremum of f provided we disregard sets of measure 0.

It is clear that $\|f\|_p = 0$ if and only if $f = 0$ a.e.

Proposition 13.1 (Hölder's inequality) *If $1 < p, q < \infty$ and $p^{-1} + q^{-1} = 1$, then*

$$\int |f(x)g(x)|\, d\mu \le \|f\|_p \|g\|_q.$$

This also holds if $p = \infty$ and $g = 1$.

Proof. If $M = \|f\|_\infty$, then $\int fg \le M \int |g|$ and the case $p = \infty$ and $q = 1$ follows. So let us assume $1 < p, q < \infty$. If $\|f\|_p = 0$, then $f = 0$ a.e and $\int fg = 0$, so the result is clear if $\|f\|_p = 0$ and similarly if $\|g\|_q = 0$. Let $F(x) = |f(x)|/\|f\|_p$ and $G(x) = |g(x)|/\|g\|_q$. Note $\|F\|_p = 1$ and $\|G\|_q = 1$, and it suffices to show that $\int FG \le 1$.

The second derivative of the function e^x is again e^x, which is positive, and so e^x is convex. Therefore if $0 \le \lambda \le 1$, we have

$$e^{\lambda a + (1-\lambda)b} \le \lambda e^a + (1-\lambda)e^b.$$

If $F(x), G(x) \neq 0$, let $a = p \log F(x)$, $b = q \log G(x)$, $\lambda = 1/p$, and $1 - \lambda = 1/q$. We then obtain

$$F(x)G(x) \le \frac{F(x)^p}{p} + \frac{G(x)^q}{q}.$$

Clearly this inequality also holds if $F(x) = 0$ or $G(x) = 0$. Integrating,

$$\int FG \le \frac{\|F\|_p^p}{p} + \frac{\|G\|_q^q}{q} = \frac{1}{p} + \frac{1}{q} = 1.$$

□

One application of Hölder's inequality is to prove Minkowski's inequality, which is simply the triangle inequality for L^p.

We first need the following lemma:

Lemma 13.2 *If $a, b > 0$ and $1 \le p < \infty$, then*

$$(a+b)^p \le 2^{p-1}a^p + 2^{p-1}b^p.$$

Proof. To prove this, we may without loss of generality assume $a \le b$. The case $a = 0$ is obvious, so we assume $a > 0$. Dividing both sides by a and letting $x = b/a$, the inequality we want is equivalent to

$$(1+x)^p \le 2^{p-1} + 2^{p-1}x^p, \qquad x \ge 1. \tag{13.1}$$

Clearly this inequality is valid for $x = 1$. So to prove (13.1) it suffices to show that the derivative of

$$(1+x)^p - 2^{p-1} - 2^{p-1}x^p$$

is less than equal to 0, or

$$p(1+x)^{p-1} \le 2^{p-1} p x^{p-1}, \qquad x \ge 1.$$

This last inequality hods because $2x \ge 1+x$ when $x \ge 1$. □

Proposition 13.3 (Minkowski's inequality) *If* $1 \le p \le \infty$*, then*

$$\|f+g\|_p \le \|f\|_p + \|g\|_p.$$

Proof. Since $|(f+g)(x)| \le |f(x)| + |g(x)|$, integrating gives the case when $p = 1$. The case $p = \infty$ is also easy. So let us suppose $1 < p < \infty$. If $\|f\|_p$ or $\|g\|_p$ is infinite, the result is obvious, so we may assume both are finite. The inequality Lemma 13.2 with $a = |f(x)|$ and $b = |g(x)|$ yields, after an integration,

$$\int |(f+g)(x)|^p d\mu \le 2^p \int |f(x)|^p d\mu + 2^p \int |g(x)|^p d\mu.$$

So we have $\|f+g\|_p < \infty$. Clearly we may assume $\|f+g\|_p > 0$.

Now write

$$|f+g|^p \le |f|\,|f+g|^{p-1} + |g|\,|f+g|^{p-1}$$

and apply Hölder's inequality with $q = (1 - \frac{1}{p})^{-1}$. We obtain

$$\int |f+g|^p \le \|f\|_p \Big(\int |f+g|^{(p-1)q} \Big)^{1/q} + \|g\|_p \Big(\int |f+g|^{(p-1)q} \Big)^{1/q}.$$

Since $p^{-1} + q^{-1} = 1$, then $(p-1)q = p$, so we have

$$\|f+g\|_p^p \le \Big(\|f\|_p + \|g\|_p \Big) \|f+g\|_p^{p/q}.$$

Dividing both sides by $\|f+g\|_p^{p/q}$ and using the fact that $p - (p/q) = 1$ gives us our result. □

Minkowski's inequality says that L^p is a normed linear space, provided we identify functions that are equal a.e.

We say f_n converges to f in L^p if $\|f_n - f\|_p \to 0$ as $n \to \infty$. The next proposition compares convergence in L^p to convergence in measure. Before we prove this, we prove an easy preliminary result known as Chebyshev's inequality.

Lemma 13.4 *If $1 \leq p < \infty$,*

$$\mu(\{x : |f(x)| \geq a\}) \leq \frac{\|f\|_p^p}{a^p}.$$

Proof. If $A = \{x : |f(x)| \geq a\}$, then

$$\mu(A) \leq \int_A \frac{|f(x)|^p}{a^p} d\mu \leq \frac{1}{a^p} \int |f|^p d\mu.$$

□

Proposition 13.5 *If f_n converges to f in L^p, then it converges in measure.*

Proof. If $\varepsilon > 0$, by Chebyshev's inequality

$$\mu(\{x : |f_n(x) - f(x)| > \varepsilon\}) = \mu(\{x : |f_n(x) - f(x)|^p > \varepsilon^p\}) \leq \frac{\|f_n - f\|_p^p}{\varepsilon^p} \to 0.$$

□

Letting $f_n = n^2 \chi_{(0,1/n)}$ on $[0, 1]$ with the measure being Lebesgue measure gives an example where f_n converges to 0 a.e. and in measure, but does not converge in L^p.

Example 9.4 is an example where f_n converges to 0 in L^p but not a.e.

We next show that L^p is complete. This is often phrased as saying that L^p is a Banach space, i.e., a complete normed linear space.

Proposition 13.6 *If $1 \leq p \leq \infty$, then L^p is complete.*

Proof. We do only the case $p < \infty$; the case $p = \infty$ is easy. Suppose f_n is a Cauchy sequence in L^p. Given $\varepsilon = 2^{-(j+1)}$, there exists n_j such that if $n, m \geq n_j$, then $\|f_n - f_m\|_p \leq 2^{-(j+1)}$. Without loss of generality we may assume $n_j \geq n_{j-1}$ for each j.

Set $n_0 = 0$ and define $f_0 \equiv 0$. If $A_j = \{x : |f_{n_j}(x) - f_{n_{j-1}}(x)| > 2^{-j/2}\}$, then from Lemma 13.4, $\mu(A_j) \le 2^{-jp/2}$. We have

$$\mu(\cap_{j=1}^\infty \cup_{m=j}^\infty A_m) = \lim_{j\to\infty} \mu(\cup_{m=j}^\infty A_m) \le \lim_{j\to\infty} \sum_{m=j}^\infty \mu(A_m) = 0.$$

So except for a set of measure 0, for each x there is a last j for which $x \in \cup_{m=j}^\infty A_m$, hence a last j for which $x \in A_j$. So for each x (except for the null set) there is a j_0 (depending on x) such that if $j \ge j_0$, then $|f_{n_j}(x) - f_{n_{j-1}}(x)| \le 2^{-j}$.

Set

$$g_j(x) = \sum_{m=1}^j |f_{n_m}(x) - f_{n_{m-1}}(x)|.$$

$g_j(x)$ increases for each x, and the limit is finite for almost every x by the preceding paragraph. Let us call the limit $g(x)$. We have

$$\|g_j\|_p \le \sum_{m=1}^j 2^{-j} + \|f_{n_1}\|_p \le 2 + \|f_{n_1}\|_p$$

by Minkowski's inequality, and so by Fatou's lemma, $\|g\|_p \le 2 + \|f_{n_1}\|_p < \infty$. We have

$$f_{n_j}(x) = \sum_{m=1}^j (f_{n_m}(x) - f_{n_{m-1}}(x)).$$

Suppose x is not in the null set where $g(x)$ is infinite. Since $|f_{n_j}(x) - f_{n_k}(x)| \le |g_{n_j}(x) - g_{n_k}(x)| \to 0$ as $j, k \to \infty$, then $f_{n_j}(x)$ is a Cauchy series (in $\mathbb{R}$), and hence converges, say to $f(x)$. We have $\|f - f_{n_j}\|_p = \lim_{m\to\infty} \|f_{n_m} - f_{n_j}\|_p$; this follows by dominated convergence with the function g defined above as the dominating function.

We have thus shown that $\|f - f_{n_j}\|_p \to 0$. Given $\varepsilon = 2^{-(j+1)}$, if $m \ge n_j$, then $\|f - f_m\|_p \le \|f - f_{n_j}\|_p + \|f_m - f_{n_j}\|_p$. This shows that f_m converges to f in L^p norm. □

Next we show:

Proposition 13.7 *The set of continuous functions with compact support is dense in $L^p(\mathbb{R}^d)$.*

Proof. Suppose $f \in L^p$. By dominated convergence $\int |f - f\chi_{-[n,n]}|^p \to 0$ as $n \to \infty$, the dominating function being $|f|^p$. So we may suppose f has compact support. By writing $f = f^+ - f^-$ we may suppose $f \geq 0$. By taking simple functions s_m increasing to f, we have $\int |f - s_m|^p \to 0$ by dominated convergence, so it suffices to consider simple functions. By linearity, it suffices to consider characteristic functions with compact support. Given such a χ_E and $\varepsilon > 0$ we showed in Proposition 8.3 that there exists g continuous with compact support and with values in $[0, 1]$ such that $\int |g - \chi_E| < \varepsilon$. Since $|g - \chi_E| \leq 1$, then $\int |g - \chi_E|^p \leq \int |g - \chi_E| < \varepsilon$. □

The following is very useful.

Proposition 13.8 *For* $1 < p < \infty$ *and* $p^{-1} + q^{-1} = 1$,

$$\|f\|_p = \sup\Big\{\int fg : \|g\|_q \leq 1\Big\}. \tag{13.2}$$

When $p = 1$ *(13.2) holds if we take* $q = \infty$*, and if* $p = \infty$ *(13.2) holds if we take* $q = 1$*.*

Proof. The right hand side of (13.2) is less than the left hand side by Hölder's inequality. So we need only show that the right hand side is greater than the left hand side.

First suppose $p = 1$. Take $g(x) = \operatorname{sgn} f(x)$, where $\operatorname{sgn} a$ is 1 if $a > 0$, is 0 if $a = 0$, and is -1 if $a < 0$. Then g is bounded by 1 and $fg = |f|$. This takes care of the case $p = 1$.

Next suppose $p = \infty$. Since μ is σ-finite, there exist sets F_n increasing up to X such that $\mu(F_n) < \infty$ for each n. If $M = \|f\|_\infty$, let a be any finite real less than M. By the definition of L^∞ norm, the measure of $A = \{x \in F_n : |f(x)| > a\}$ must be positive if n is sufficiently large. Let $g(x) = (\operatorname{sgn} f(x))\chi_A(x)/\mu(A)$. Then the L^1 norm of g is 1 and $\int fg = \int_A |f|/\mu(A) \geq a$. Since a is arbitrary, the supremum on the right hand side must be M.

Now suppose $1 < p < \infty$. We may suppose $\|f\|_p > 0$. Let q_n be a sequence of nonnegative simple functions increasing to f^+, r_n a sequence of nonnegative simple functions increasing to f^-, and $s_n(x) = (q_n(x) - r_n(x))\chi_{F_n}(x)$. Then $s_n(x) \to f(x)$ for each x, $|s_n(x)| \leq |f(x)|$ for each x, s_n is a simple function, and $\|s_n\|_p < \infty$ for each n. If $f \in L^p$, then $\|s_n\|_p \to \|f\|_p$ by dominated convergence. If $\int |f|^p = \infty$, then $\int |s_n|^p \to \infty$ by monotone convergence. For n sufficiently large, $\|s_n\|_p > 0$.

Let

$$g_n(x) = (\operatorname{sgn} f(x)) \frac{|s_n(x)|^{p-1}}{\|s_n\|_p^{p/q}}.$$

Since $(p-1)q = p$, then

$$\|g_n\|_q = \frac{(\int |s_n|^{(p-1)q})^{1/q}}{\|s_n\|_p^{p/q}} = \frac{\|s_n\|_p^{p/q}}{\|s_n\|_p^{p/q}} = 1.$$

On the other hand, since $|f| \geq |s_n|$,

$$\int f g_n = \frac{\int |f|\,|s_n|^{p-1}}{\|s_n\|_p^{p/q}} \geq \frac{\int |s_n|^p}{\|s_n\|_p^{p/q}} = \|s_n\|_p^{p-(p/q)}.$$

Since $p - (p/q) = 1$, then $\int f g_n \geq \|s_n\|_p$, which tends to $\|f\|_p$. □

The above proof also establishes

Corollary 13.9 *For $1 < p < \infty$ and $p^{-1} + q^{-1} = 1$,*

$$\|f\|_p = \sup\Big\{ \int fg : \|g\|_q \leq 1, g \textit{ simple}\Big\}.$$

The space L^p is a normed linear space. We can thus talk about its dual, namely, the set of bounded linear functionals on L^p. The dual of a space Y is denoted Y^*. If H is a bounded linear functional on L^p, we define the norm of H to be $\|H\| = \sup\{H(f) : \|f\|_p \leq 1\}$.

Theorem 13.10 *If $1 < p < \infty$ and $p^{-1} + q^{-1} = 1$, then $(L^p)^* = L^q$.*

What this means is that if H is a bounded linear functional on L^p, then there exists $g \in L^q$ such that $H(f) = \int fg$ and that if $g \in L^q$, then $H(f) = \int fg$ is a bounded linear functional on L^p.

Proof. If $g \in L^q$, then setting $H(f) = \int fg$ for $f \in L^p$ yields a bounded linear functional; the boundedness follows from Hölder's inequality. Moreover, from Hölder's inequality and Proposition 13.8 we see that $\|H\| = \|g\|_q$.

Now suppose we are given a bounded linear functional H on L^p and we must show there exists $g \in L^q$ such that $H(f) = \int fg$. First suppose $\mu(X) < \infty$. Define $\nu(A) = H(\chi_A)$. If A and B are disjoint, then

$$\nu(A \cup B) = H(\chi_{A\cup B}) = H(\chi_A + \chi_B) = H(\chi_A) + H(\chi_B) = \nu(A) + \nu(B).$$

To show ν is countably additive, it suffices to show that if $A_n \uparrow A$, then $\nu(A_n) \to \nu(A)$. But if $A_n \uparrow A$, then $\chi_{A_n} \to \chi_A$ in L^p, and so $\nu(A_n) = H(\chi_{A_n}) \to H(\chi_A) = \nu(A)$; we use here the fact that $\mu(X) < \infty$. Therefore ν is a countably additive signed measure. Moreover, if $\mu(A) = 0$, then $\chi_A = 0$ a.e., hence $\nu(A) = H(\chi_A) = 0$. By writing $\nu = \nu^+ - \nu^-$ and using the Radon-Nikodym theorem for both the positive and negative parts, we see there exists an integrable g such that $\nu(A) = \int_A g$ for all sets A. If $s = \sum a_i \chi_{A_i}$ is a simple function, by linearity we have

$$H(s) = \sum a_i H(\chi_{A_i}) = \sum a_i \nu(A_i) = \sum a_i \int g\chi_{A_i} = \int gs.$$

By Corollary 13.9,

$$\|g\|_q = \sup\Big\{\int gs : \|s\|_p \le 1, s \text{ simple}\Big\} \le \sup\{H(s) : \|s\|_p \le 1\} \le \|H\|.$$

If s_n are simple functions tending to f in L^p, then $H(s_n) \to H(f)$, while by Hölder's inequality $\int s_n g \to \int fg$. We thus have $H(f) = \int fg$ for all $f \in L^p$, and $\|g\|_p \le \|H\|$. By Hölder's inequality, $\|H\| \le \|g\|_p$.

In the case where μ is σ-finite, but not finite, let $F_n \uparrow X$ be such that $\mu(F_n) < \infty$ for each n. Define functionals H_n by $H_n(f) = H(f\chi_{F_n})$. Clearly each H_n is a bounded linear functional on L^p. Applying the above argument, we see there exist g_n such that $H_n(f) = \int fg_n$ and $\|g_n\|_q = \|H_n\| \le \|H\|$. It is easy to see that g_n is 0 if $x \notin F_n$. Moreover, by the uniqueness part of the Radon-Nikodym theorem, if $n > m$, then $g_n = g_m$ on F_m. Define g by setting $g(x) = g_n(x)$ if $x \in F_n$. Then g is well defined. By Fatou's lemma, g is in L^q with a norm bounded by $\|H\|$. Since $f\chi_{F_n} \to f$ in L^p by dominated convergence, then $H_n(f) = H(f\chi_{F_n}) \to H(f)$, since H is a bounded linear functional on L^p. On the other hand $H_n(f) = \int_{F_n} fg_n = \int_{F_n} fg \to \int fg$ by dominated convergence. So $H(f) = \int fg$. Again by Hölder's inequality $\|H\| \le \|g\|_p$. □

14 Fourier transforms

Fourier transforms give a representation of a function in terms of frequencies. We give the basic properties here.

If $f \in L^1(\mathbb{R}^n)$, define the Fourier transform $\widehat{f}$ by

$$\widehat{f}(u) = \int_{\mathbb{R}^n} e^{iu\cdot x} f(x)dx, \qquad u \in \mathbb{R}^n. \tag{14.1}$$

We are using $u \cdot x$ for the standard inner product in $\mathbb{R}^n$. Various books have slightly different definitions. Some put a negative sign before the $iu \cdot x$, some have a 2π either in front of the integral or in the exponent. The basic theory is the same in any case.

Some basic properties of the Fourier transform are given by

Proposition 14.1 *Suppose f and g are in L^1. Then*

(a) $\widehat{f}$ is bounded and continuous;

(b) $\widehat{(f+g)}(u) = \widehat{f}(u) + \widehat{g}(u)$; $\widehat{(af)}(u) = a\widehat{f}(u)$;

(c) if $f_a(x) = f(x+a)$, then $\widehat{f_a}(u) = e^{-iu\cdot a}\widehat{f}(u)$;

(d) if $g_a(x) = e^{ia\cdot x}g(x)$, then $\widehat{g_a}(u) = \widehat{f}(u+a)$;

(e) if $h_a(x) = f(ax)$, then $\widehat{h_a}(u) = a^{-n}\widehat{f}(u/a)$.

Proof. (a) $\widehat{f}$ is bounded because $f \in L^1$ and $|e^{iu\cdot x}| = 1$. We have

$$\widehat{f}(u+h) - \widehat{f}(u) = \int \Big(e^{i(u+h)\cdot x} - e^{iu\cdot x}\Big) f(x)dx.$$

So

$$|\widehat{f}(u+h) - \widehat{f}(u)| \le \int \Big|e^{iu\cdot x}\Big| \Big|e^{ih\cdot x} - 1\Big| \, |f(x)|dx.$$

The integrand is bounded by $2|f(x)|$, which is integrable, and $e^{ih\cdot x} - 1 \to 0$ as $h \to 0$, and thus the continuity follows by dominated convergence.

(b) is obvious. (c) follows because

$$\widehat{f_a}(u) = \int e^{iu\cdot x} f(x+a)dx = \int e^{iu\cdot(x-a)} f(x)dx = e^{-iu\cdot a}\widehat{f}(u)$$

by a change of variables. For (d),

$$\widehat{g_a}(u) = \int e^{iu\cdot x} e^{ia\cdot x} f(x)dx = \int e^{i(u+a)\cdot x} f(x)dx = \widehat{f}(u+a).$$

Finally for (e), by a change of variables,

$$\begin{aligned}\widehat{h_a}(u) &= \int e^{iu\cdot x} f(ax)dx = a^{-n}\int e^{iu\cdot(y/a)} f(y)dy \\ &= a^{-n}\int e^{i(u/a)\cdot y} f(y)dy = a^{-n}\widehat{f}(u/a).\end{aligned}$$

□

One reason for the usefulness of Fourier transforms is that they relate derivatives and multiplication.

Proposition 14.2 *Suppose* $f \in L^1$ *and* $x_j f(x) \in L^1$*, where* x_j *is the* j^{th} *coordinate of* x*. Then*

$$\frac{\partial \widehat{f}}{\partial u_j}(u) = i\int e^{iu\cdot x} x_j f(x)dx.$$

Proof. Let e_j be the unit vector in the j^{th} direction. Then

$$\begin{aligned}\frac{\widehat{f}(u+he_j) - \widehat{f}(u)}{h} &= \frac{1}{h}\int \Big(e^{i(u+he_j)\cdot x} - e^{iu\cdot x}\Big) f(x)dx \\ &= \int e^{iu\cdot x}\Big(\frac{e^{ihx_j}-1}{h}\Big) f(x)dx.\end{aligned}$$

Since

$$\Big|\frac{1}{h}\Big(e^{ihx_j}-1\Big)\Big| \le |x_j|$$

and $x_j f(x) \in L^1$, the right hand side converges to $\int e^{iu\cdot x} i x_j f(x)dx$ by dominated convergence. Therefore the left hand side converges. Of course, the limit is $\partial\widehat{f}/\partial u_j$. □

The convolution of f and g is defined by

$$f * g(x) = \int f(x-y)g(y)dy.$$

By a change of variables, this is the same as $\int f(y)g(x-y)dy$, so $f*g = g*f$.

Proposition 14.3 *(a) If $f, g \in L^1$, then $f * g$ is in L^1 and $\|f * g\|_1 \le \|f\|_1\|g\|_1$.*

(b) The Fourier transform of $f * g$ is $\widehat{f}(u)\widehat{g}(u)$.

Proof. (a) We will show $f * g$ is finite a.e. by showing $\int |f * g(x)|\, dx < \infty$. We have

$$\int |f * g(x)| dx \le \int \int |f(x-y)|\, |g(y)| dy\, dx.$$

Since the integrand is nonnegative, we can apply Fubini and the right hand side is equal to

$$\int \int |f(x-y)| dx\, |g(y)| dy = \int \int |f(x)| dx\, |g(y)| dy = \|f\|_1 \|g\|_1.$$

The first equality here follows by a change of variables. To verify that we can do a change of variables, we reduce to simple functions and then characteristic functions, and then use the translation invariance of Lebesgue measure.

(b) We have

$$\begin{aligned} \widehat{f * g}(u) &= \int e^{iu\cdot x} \int f(x-y)g(y)dy\, dx \\ &= \int \int e^{iu\cdot(x-y)} f(x-y) dx\, e^{iu\cdot y} g(y) dy \\ &= \int \widehat{f}(u) e^{iu\cdot y} g(y) dy = \widehat{f}(u)\widehat{g}(u). \end{aligned}$$

We applied Fubini in the first equality; this is valid because as we saw in (a), the absolute value of the integrand is integrable. □

We want to give a formula for recovering f from $\widehat{f}$. First we need to calculate the Fourier transform of a particular function.

Proposition 14.4 *(a) Suppose $f_1 : \mathbb{R} \to \mathbb{R}$ is defined by*

$$f_1(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Then $\widehat{f_1}(u) = e^{-u^2/2}$.

(b) Suppose $f_n : \mathbb{R}^n \to \mathbb{R}$ is given by

$$f_n(x) = \frac{1}{(2\pi)^{n/2}} e^{-|x|^2/2}.$$

Then $\widehat{f_n}(u) = e^{-|u|^2/2}$.

Proof. (a) may be proved using contour integration, but let's give a real variable proof. Let $g(u) = \int e^{iux} e^{-x^2/2} dx$. Differentiate with respect to u. We may differentiate under the integral sign because $(e^{i(u+h)x} - e^{iux})/h$ is bounded in absolute value by $|x|$ and $|x|e^{-x^2/2}$ is integrable; therefore dominated convergence applies. We then obtain

$$g'(u) = i \int e^{iux} x e^{-x^2/2} dx.$$

By integration by parts this is equal to

$$-u \int e^{iux} e^{-x^2/2} dx = -ug(u).$$

Solving the differential equation $g'(u) = -ug(u)$, we have

$$[\log g(u)]' = \frac{g'(u)}{g(u)} = -u,$$

so $\log g(u) = -u^2/2 + c_1$, and so then

$$g(u) = c_2 e^{-u^2}. \tag{14.2}$$

Since $g(0) = \int e^{-x^2/2} dx = \sqrt{2\pi}$, $c_2 = \sqrt{2\pi}$. Substituting this value of c_2 in (14.2) and dividing both sides by $\sqrt{2\pi}$ proves (a).

For (b), since $f_n(x) = f_1(x_1)\cdots f_1(x_n)$ if $x = (x_1, \ldots, x_n)$,

$$\begin{aligned}\widehat{f_n}(u) &= \int\cdots\int e^{i\sum_j u_j x_j} f_1(x_1)\cdots f_1(x_n)dx_1\cdots dx_n\\ &= \widehat{f_1}(u_1)\cdots\widehat{f_1}(u_n) = e^{-|u|^2/2}.\end{aligned}$$

□

One more preliminary before proving the inversion theorem.

Proposition 14.5 *Suppose* φ *is in* L^1 *and* $\int \varphi(x)dx = 1$. *Let* $\varphi_A(x) = A^{-n}\varphi(x/A)$.

(a) Then $\|f * \varphi_A - f\|_1 \to 0$ *as* $A \to 0$.

(b) If f *is continuous with compact support, then* $f * \varphi_A$ *converges to* f *pointwise.*

Proof. (a) Let $\varepsilon > 0$. Choose g continuous with compact support so that $\|f - g\|_1 < \varepsilon$. Let $h = f - g$. A change of variables shows that $\|\varphi_A\|_1 = \|\varphi\|_1$. Observe

$$\|f * \varphi_A - f\|_1 \le \|g * \varphi_A - g\|_1 + \|h * \varphi_A - h\|_1$$

and

$$\|h * \varphi_A - h\|_1 \le \|h\|_1 + \|h * \varphi_A\|_1 \le \|h\|_1 + \|h\|_1\|\varphi_A\|_1 < \varepsilon(1 + \|\varphi\|_1).$$

So since ε is arbitrary, it suffices to show that $g * \varphi_A \to g$ in L^1.

We start by writing

$$\begin{aligned}g * \varphi_A(x) - g(x) &= \int g(x-y)\varphi_A(y)dy - g(x) = \int g(x - Ay)\varphi(y)dy - g(x)\\ &= \int [g(x - Ay) - g(x)]\varphi(y)dy.\end{aligned}$$

We used a change of variables and the fact that $\int \varphi(y)dy = 1$. Because g is continuous with compact support, then g is bounded, and the integral on the right goes to 0 by dominated convergence, the dominating function being $\|g\|_\infty|\varphi(y)|$. Therefore $g * \varphi_A(x)$ converges to $g(x)$ pointwise.

To show the convergence in L^1, we have

$$\int |g * \varphi_A(x) - g(x)| dx \le \int\int |g(x - Ay) - g(x)|\, |\varphi(y)| dy\, dx$$
$$= \int\int |g(x - Ay) - g(x)|\, |\varphi(y)| dx\, dy.$$

Since g is continuous with compact support and hence bounded, for each y

$$G_A(y) = \int |g(x - Ay) - g(x)| dx$$

converges to 0 as $A \to 0$ by dominated convergence. Also

$$G_A(y) \le \int |g(x - Ay)| dx + \int |g(x)| dx \le 2\|g\|_1 < \infty.$$

Then

$$\int G_A(y) |\varphi(y)| dy$$

converges to 0 as $A \to 0$ by dominated convergence, the dominating function being $2\|g\|_1 |\varphi(y)|$.

(b) This follows from the argument we used for g above. □

Now we are ready to give the inversion formula. The proof seems longer than it might be, but there is no avoiding the introduction of the function H_a or some similar function.

Theorem 14.6 *Suppose $f, \widehat{f} \in L^1$. Then*

$$f(y) = \frac{1}{(2\pi)^n} \int e^{-iu\cdot y} \widehat{f}(u) du, \qquad a.e.$$

Proof. If $g(x) = a^{-n} f(x/a)$, then its Fourier transform is $\widehat{f}(au)$. So the Fourier transform of

$$\frac{1}{a^n} \frac{1}{(2\pi)^{n/2}} e^{-x^2/2a^2}$$

is $e^{-a^2u^2/2}$. Therefore if we let

$$H_a(x) = \frac{1}{(2\pi)^n} e^{-|x|^2/2a^2},$$

we have

$$\widehat{H}_a(u) = (2\pi)^{-n/2} a^n e^{-a^2|u|^2/2}.$$

We have

$$\begin{aligned}
\int \widehat{f}(u) e^{-iu\cdot y} H_a(u) du& \\
&= \int\int e^{iu\cdot x} f(x) e^{-iu\cdot y} H_a(u) dx\, du \\
&= \int\int e^{iu\cdot(x-y)} H_a(u) du\, f(x)\, dx \\
&= \int \widehat{H}_a(x-y) f(x) dx. \qquad (14.3)
\end{aligned}$$

We can interchange the order of integration because

$$\int\int |f(x)|\,|H_a(u)| dx\, du < \infty.$$

The left hand side of the first line of (14.3) converges to $(2\pi)^{-n}\int \widehat{f}(u) e^{-iu\cdot y} dy$ as $a \to \infty$ by dominated convergence and the fact that $\widehat{f} \in L^1$. The last line of (14.3) is equal to

$$\int \widehat{H}_a(y-x) f(x) dx = f * \widehat{H}_a(y) \qquad (14.4)$$

since $\widehat{H}_a$ is symmetric. But by Proposition 14.5, $f * \widehat{H}_a$ converges to f in L^1 as $a \to \infty$. □

The last topic that we consider is the Plancherel theorem.

Theorem 14.7 *(a) Suppose f is continuous with compact support. Then $\widehat{f} \in L^2$ and*

$$\|f\|_2 = (2\pi)^{-n/2} \|\widehat{f}\|_2. \qquad (14.5)$$

(b) We can use the result in (a) to define $\widehat{f}$ when $f \in L^2$ and so that (14.5) holds.

Proof. (a) Let $g(x) = \overline{f(-x)}$. Note

$$\widehat{g}(u) = \int e^{iu\cdot x}\overline{f(-x)}dx = \overline{\int e^{-iu\cdot x} f(-x)dx} = \overline{\int e^{iu\cdot x} f(x)dx} = \overline{\widehat{f}(u)}.$$

By (14.3) and (14.4) with $y = 0$

$$f * g * \widehat{H}_a(0) = \int \widehat{f * g}(u) H_a(u)du. \tag{14.6}$$

Since $\widehat{f * g}(u) = \widehat{f}(u)\widehat{g}(u) = |\widehat{f}(u)|^2$, the right hand side of (14.6) converges by monotone convergence to $(2\pi)^{-n} \int |\widehat{f}(u)|^2 du$ as $a \to \infty$. Since f and g are continuous with compact support, then it is easy to see that $f * g$ is also, and so the left hand side of (14.6) converges to $f * g(0) = \int f(y)g(-y)dy = \int |f(y)|^2 dy$ by Proposition 14.5(b).

(b) The set of continuous functions with compact support is dense in L^2. Given a function f in L^2, choose a sequence of continuous functions with compact support $\{f_m\}$ such that $f_m \to f$ in L^2. By the result in (a), $\{\widehat{f_m}\}$ is a Cauchy sequence in L^2, and therefore converges to a function in L^2, which we call $\widehat{f}$. If $\{f'_m\}$ is another sequence of continuous functions with compact support converging to f in L^2, then $\{f_m - f'_m\}$ is a sequence of continuous functions with compact support converging to 0 in L^2; by the result in (a), $\widehat{f}_m - \widehat{f}'_m$ converges to 0 in L^2, and therefore $\widehat{f}$ is defined uniquely up to almost everywhere equivalence. By passing to the limit in L^2 on both sides of (14.5), we see that (14.5) holds for $f \in L^2$. □

Richard F. Bass
Department of Mathematics
University of Connecticut
Storrs, CT 06269-3009, USA
bass@math.uconn.edu