Lecture notes for Math 205A, Version 2014

Lenya Ryzhik*

November 18, 2014

Nothing found here is original except for a few mistakes and misprints here and there. These notes are simply a record of what I cover in class, to spare the students the necessity of taking the lecture notes. The readers should consult the original books for a better presentation and context. We plan to follow the books by Terry Tao on his real analysis class published by the AMS. In addition, the material from the following books will be likely used: H. Royden "Real Analysis", L. Evans and R. Gariepy "Measure Theory and Fine Properties of Functions", J. Duoandikoetxea "Fourier Analysis", and M. Pinsky "Introduction to Fourier Analysis and Wavelets".

1 Basic measure theory

1.1 Definition of the Lebesgue Measure

The Lebesgue measure is a generalization of the notion of the length $l(I)$ of an interval $I = (a, b) \subset \mathbb{R}$, or an area of a polygon in $\mathbb{R}^2$. Restricting ourselves first to one dimension, ideally, we are looking for a non-negative function m defined on the subsets of $\mathbb{R}$ such that:

(i) mE is defined for all subsets of $\mathbb{R}$.

(ii) For an interval I we have $m(I) = l(I)$, the length of I.

(iii) If the sets E_n are disjoint then

$$m(\bigcup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} m(E_n).$$

(iv) m is translationally invariant, that is, $m(E + x) = mE$ for all sets E and $x \in \mathbb{R}$.

The trouble is that such function does not exist, or, rather that for any such function m the measure of any interval is either equal to zero or infinity. More precisely, if we assume that (i), (iii) and (iv) hold then the measure of $[0, 1]$ is either zero or infinite. Let us explain

*Department of Mathematics, Stanford University, Stanford, CA 94305, USA; ryzhik@math.stanford.edu

why this is so. We will do this for the interval $[0,1)$ but generalization to an arbitrary interval is straightforward. Given $x, y \in [0,1)$ define

$$x \oplus y = \begin{cases} x+y, & \text{if } x+y<1, \\ x+y-1, & \text{if } x+y \geq 1, \end{cases}$$

and for a set $E \subseteq [0,1)$ we set $E \oplus y = \{x \in [0,1) : \ x = e \oplus y \text{ for some } e \in E\}$. This is addition on the unit interval identified with the circle.

Lemma 1.1 *Assume (i), (iii), (iv) above. If $E \subseteq [0,1)$ is a set and $y \in [0,1)$, then we have $m(E \oplus y) = m(E)$.*

Proof. Let $E_1 = E \cap [0, 1-y)$ and $E_2 = E \cap [1-y, 1)$, then E_1 and E_2 are and disjoint, and the same is true for $E_1 \oplus y = E_1 + y$, and $E_2 \oplus y = E_2 + (y-1)$. We have

$$E \oplus y = (E_1 \oplus y) \cup (E_2 \oplus y),$$

and, in addition:

$$m(E \oplus y) = m(E_1 \oplus y) + m(E_2 \oplus y) = m(E_1 + y) + m(E_2 + (y-1)) = m(E_1) + m(E_2) = m(E),$$

and we are done. $\square$

Let us introduce an equivalence relation on $[0,1)$: $x \sim y$ if $x - y \in \mathbb{Q}$. Using the axiom of choice, we deduce existence of a set P which contains exactly one element from each equivalence class. Set $P_j = P \oplus q_j$, where q_j is the j-th rational number in $[0,1)$: we write

$$\mathbb{Q} \cap [0,1) = \{q_1, q_2, \ldots\}.$$

Note that the sets P_j are pairwise disjoint: if $x \in P_i \cap P_j$, then

$$x = p_i \oplus q_i = p_j \oplus q_j,$$

and thus $p_j \sim p_i$. It follows that $p_i = p_j$, and $i = j$, since P contains exactly one element from each equivalence class. On the other hand, we have

$$[0,1) = \bigcup_{j=1}^{\infty} P_j,$$

and each P_i is a translation of P by q_i, hence $m(P_i) = m(P)$ for all i, according to (iv). On the other hand, (iii) implies that

$$m([0,1)) = m\left(\bigcup_{n=1}^{\infty} P_n\right) = \sum_{n=1}^{\infty} m(P_n).$$

Thus, we have $m([0,1)) = 0$ if $m(P) = 0$ or $m([0,1)) = +\infty$ if $m(P) > 0$. Therefore, if we want to keep our intended generalization of the length of an interval not totally trivial we have to drop one of the requirements (i)-(iv), and the best candidate to do so is (i) since (ii)-(iv) come from physical considerations.

Let us now define the (outer) Lebesgue measure of a set on the real line – we will first define it for all subsets of $\mathbb{R}$, and then restrict to a smaller class of sets, that we will call measurable, and for which properties (ii)-(iv) will, indeed, hold.

Definition 1.2 *Let A be a subset of $\mathbb{R}$. Its outer Lebesgue measure*

$$m^*(A) = \inf \sum l(I_n),$$

where the infimum is taken over all at most countable collections of open intervals $\{I_n\}$ such that $A \in \bigcup_n I_n$.

Note that $m^*(A)$ is defined for all sets (though it may is possible that $m^*(A) = +\infty$), and we obviously have (i) $m^*(\emptyset) = 0$, and (ii) if $A \subseteq B$ then $m^*(A) \leq m^*(B)$.

Exercise 1.3 Will the notion change if we allow only finite collections of the covering integrals and not countable?

Exercise 1.4 Show that the Lebesgue outer measure of any countable set is zero. Would that still be true if we would allow only finite collections of the intervals I_n in the definution of the outer measure?

Proposition 1.5 *If I is an interval then $m^*(I) = l(I)$.*

Proof. (1) If I is either an open, or a closed, or half-open interval between points a and b then we have

$$m^*(I) \leq l(a - \varepsilon, b + \varepsilon) = b - a + 2\varepsilon$$

for all $\varepsilon > 0$. It follows that $m^*(I) \leq b - a$.

(2) On the other hand, to show that

$$m^*([a, b]) \geq b - a,$$

take a cover $\{I_n\}$ of $[a, b]$ by open intervals. We may choose (using the Heine-Borel lemma) a finite sub-cover $\{J_j\}$, $j = 1, \ldots, N$ which still covers $[a, b]$. As $I \subset \bigcup_{j=1}^{N} J_j$, we have

$$\sum_{j=1}^{N} l(J_j) \geq b - a.$$

Therefore, $m^*[a, b] \geq b - a$ and thus, together with (1) we see that $m^*([a, b]) = b - a$.

(3) For an open interval (a, b) we simply write

$$m^*(a, b) \geq m^*[a + \varepsilon, b - \varepsilon] \geq b - a - 2\varepsilon$$

for all $\varepsilon > 0$, and thus $m^*(a, b) \geq b - a$. □

Proposition 1.6 *(Countable sub-additivity). Let A_n be any collection of subsets of $\mathbb{R}$, then*

$$m^*(\bigcup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} m^*(A_n). \tag{1.1}$$

Proof. If $m^*(A_j) = +\infty$ for some j, then we are done. If $m^*(A_j) < +\infty$ for all $j \in \mathbb{N}$, then for any $\varepsilon > 0$ we may find a countable collection $\{I_k^{(j)}\}$ of intervals such that $A_j \subseteq \bigcup_k I_k^{(j)}$ and

$$\sum_{k=1}^{\infty} l(I_k^{(j)}) - \frac{\varepsilon}{2^j} \leq m^*(A_j) \leq \sum_{k=1}^{\infty} l(I_k^{(j)}).$$

Then we have

$$A := \bigcup_j A_j \subseteq \bigcup_{j,k} I_k^{(j)},$$

and so

$$m^*(A) \leq \sum_{j,k} l(I_k^{(j)}) \leq \sum_j \left(m^*(A_j) + \frac{\varepsilon}{2^j} \right) = \varepsilon + \sum_j m^*(A_j).$$

As this inequality holds for all $\varepsilon > 0$, (1.1) follows. □

This gives an alternative way to look at Exercise 1.4:

Corollary 1.7 *If A is a countable set then $m^*(A) = 0$.*

Indeed, this follows immediately from Proposition 1.6 but, of course, an independent proof is a much better way to see this.

One can not always get countable additivity, as the construction of our "non-measurable" set P above shows. Here is a first step toward getting a form of additivity. We define the distance between two sets A and B as

$$\text{dist}(A, B) = \inf\{|x - y| : \ x \in A, \ y \in B.\}$$

Lemma 1.8 *(Finite additivity for separated sets). Let $E, F \subset \mathbb{R}^n$ be such that dist$(E, F) > 0$, then*

$$m^*(E \cup F) = m^*(E) + m^*(F).$$

Proof. We already know from the countable sub-additivity that

$$m^*(E \cup F) \leq m^*(E) + m^*(F),$$

so all we need to check is that

$$m^*(E) + m^*(F) \leq m^*(E \cup F). \tag{1.2}$$

This is trivial if $m^*(E \cup F) = +\infty$, hence we may assume without loss of generality that

$$m^*(E \cup F) < +\infty.$$

Let I_j be a countable collection of intervals covering $E \cup F$ such that

$$\sum_{j=1}^{\infty} |I_j| \leq m^*(E \cup F) + \varepsilon.$$

Exercise 1.9 Show that, for $\varepsilon > 0$ sufficiently small, we may choose I_j so that

$$|I_j| < \text{dist}(E, F)/24 \text{ for all } j. \tag{1.3}$$

Note that (1.3) means that no I_j can contain both points from E and F. Let us denote by I_j' those intervals from I_j that intersect E and by I_j'' those that intersect F. Then we have

$$E \subset \bigcup_{k=1}^{\infty} I_k', \quad F \subset \bigcup_{k=1}^{\infty} I_k'',$$

and thus

$$m^*(E) \leq \sum_{k=1}^{\infty} |I_k'|, \quad m^*(F) \leq \sum_{k=1}^{\infty} |I_k''|.$$

It follows that

$$m^*(E) + m^*(F) \leq \sum_{k=1}^{\infty} |I_k'| + \sum_{k=1}^{\infty} |I_k''| \leq \sum_{k=1}^{\infty} |I_k| \leq m^*(E \cup F) + \varepsilon.$$

As $\varepsilon > 0$ is arbitrary, (1.2) follows. □

Generalization to higher dimensions

It is a helpful exercise to generalize what we have did so far to dimensions higher than one.

Exercise 1.10 In order to generalize the above discussion to dimensions $n > 1$, define a box in $\mathbb{R}^n$ as the Cartesian product $B = I_1 \times I_2 \times \ldots \times I_n$ of intervals (a box is open if all I_j are open and closed if all I_j are closed). Define the outer Lebesgue measure of a set A in $\mathbb{R}^n$ by replacing the word "interval" in Definition 1.2 by the word "box". Show that all the results we have discussed so far in one dimension hold for higher dimensions as well.

This exercise is much less scary than it may seem: most of the proofs go through verbatim. In particular, if E is a finite union of boxes (either open or closed):

$$E = \bigcup_{j=1}^{N} B_j,$$

then $m^*(E) = |E|$, the usual volume of E.

It will be often helpful to work both with "almost disjoint" boxes – these are boxes whose interior is disjoint.

Exercise 1.11 If $E = \bigcup_{j=1}^{N} E_j$ is a finite union of almost disjoint boxes then

$$m^*(E) = \sum_{j=1}^{N} |B_j|. \tag{1.4}$$

This can be generalized to countable unions.

Lemma 1.12 *Let $E = \bigcup_{j=1}^{\infty} E_j$ be a countable union of almost disjoint boxes, then*

$$m^*(E) = \sum_{j=1}^{\infty} |B_j|. \tag{1.5}$$

Proof. The sub-additivity implies that

$$m^*(B) \leq \sum_{j=1}^{\infty} |B_j|.$$

On the other hand, for any finite N we have the other inclusion

$$\bigcup_{j=1}^{N} B_j \subset E,$$

which, combined with (1.4), implies

$$\sum_{j=1}^{N} |B_j| \leq m^*(E).$$

As this is true for all N, (1.5) follows. □

We will sometimes use a decomposition of an open set into disjoint boxes. In one dimension the result is

Exercise 1.13 *Show that any open set U on the real line is a disjoint union of open intervals.*

One reason for the above property is that the union of two overlapping intervals is also an interval. This is, clearly, false in higher dimensions – a union of two boxes with a non-empty intersection need not be a box. In dimension $n \geq 2$, we consider dyadic cubes: these are the closed cubes obtained by stretching the cubes generated by the integer lattice $\mathbb{Z}^n$ by the factors of $1/2^k$, $k \in \mathbb{Z}$. The convention is that the cubes in the family $\mathcal{Q}_k$ have the sides $1/2^k$. Every cube in $\mathcal{Q}_k$ is contained completely in its "parent" in the level $\mathcal{Q}_{k-1}$. Then, for any open set U we let $\mathcal{C}$ be the collection of all dyadic cubes that are completely contained in U. It is easy to see that every point in U belongs to some cube in $\mathcal{C}$. Moreover, by the construction of the dyadic cubes, if two interiors of dyadic cubes have a non-empty intersection, then one of them lies insider the other. Let us say that a cube in $\mathcal{C}$ is maximal if it does not lie inside any other cube in $\mathcal{C}$. It is an easy exercise to see that the collection of all maximal cubes in $\mathcal{C}$ covers all of U, and that no two maximal cubes have intersecting interiors. Thus, we have proved the following.

Lemma 1.14 *Any open set $U \subset \mathbb{R}^n$ is a countable union of almost disjoint cubes.*

The Lebesgue measurability

As we have seen, we can not define a measure for all sets in a good way. Thus, we have to restrict this notion to a "reasonable" collection of sets. There are several ways to do this. Following Tao's book, we will use the idea of an approximation by open sets, and then show that it is equivalent to the "more traditional" Caratheodory approach.

Definition 1.15 *A set G is said to be $\mathcal{G}_\delta$ if it is an intersection of a countable collection of open sets.*

Proposition 1.16 *(i) Given any set* $A \subset \mathbb{R}^n$ *and any* $\varepsilon > 0$*, there exists an open set* $O \subset \mathbb{R}^n$ *such that* $A \subseteq O$ *and* $m^*(O) \leq m^*(A) + \varepsilon$.
(ii) There exists a set $G \in \mathcal{G}_\delta$ *such that* $A \subseteq G$ *and* $m^*(A) = m^*(G)$.
(iii) We have

$$m^*(A) = \inf\{m^*(O) : \ A \in O, \ O \ open\}.$$

Proof. Part (i) follows immediately from the defition of $m^*(A)$. To show (ii) take open sets O_n which contain A, and such that

$$m^*(A) \geq m^*(O_n) - \frac{1}{n}$$

and define $G = \bigcap_n O_n$. Then $G \in \mathcal{G}_\delta$, $A \subseteq G$, and

$$m^*(A) \leq m^*(G) \leq m^*(O_n) \leq m^*(A) + 1/n \text{ for all } n \in \mathbb{N},$$

hence $m^*(A) = m^*(G)$. Finally, part (iii) follows from part (i). □

The philosophical counter-part of part (i) in Proposition 1.16 is that it does not follow that $m^*(O \setminus A) < \varepsilon$ unless we assume that the set A is "nice enough". This leads to the following definition.

Definition 1.17 *A set* $E \subset \mathbb{R}^n$ *is Lebesgue measurable if for every* $\varepsilon > 0$ *there exists an open set* $U \subset \mathbb{R}^n$ *that contains* E *and such that* $m^*(U \setminus E) < \varepsilon$*. If* E *is Lebesgue measurable, we denote* $m^*(E)$ *simply by* $m(E)$ *(note that it is possible that* $m(E) = +\infty$*).*

Let us now investigate some basic examples of measurable sets. As a first step, we note that all open sets are, obviously, measurable. Another simple example are sets of (outer) measure zero. Indeed, if $m^*(E) = 0$, then, given $\varepsilon > 0$, let us cover E by open boxes B_j such that

$$E \subset U = \bigcup_{j=1}^{\infty} B_j,$$

and

$$m^*(U) < \varepsilon,$$

thus U is an open set such that $m^*(U \setminus E) \leq m^*(U) < \varepsilon$.

The definition also implies immediately that a countable union

$$E = \bigcup_{j=1}^{\infty} E_j$$

of measurable sets E_j is measurable. To see that, given any $\varepsilon > 0$, and for each j take an open set U_j such that $E_j \subset U_j$, and $m^*(U_j \setminus E_j) < \varepsilon/2^j$. Then the union $U = \bigcup_{j=1}^{\infty} U_j$ contains E and, moreover,

$$m^*(U \setminus E) \leq \sum_{j=1}^{\infty} m^*(U_j \setminus E_j) < \varepsilon.$$

Next, we show that every closed set E is measurable. We may write $E = \bigcup_{j=1}^{\infty} E_j$, where each $E_j = \{x \in E : \ |x| \leq j\}$ is a bounded closed set, and by what we have shown above, in order to show that any closed set E is measurable, it suffices to show that any bounded closed set E is measurable. Given $\varepsilon > 0$ we may then find a bounded open set U that contains E such that $m^*(U) < m^*(E) + \varepsilon$. The set $U \setminus E$ is open and is, therefore, a countable union of almost disjoint dyadic closed cubes Q_j. It follows that

$$m^*(U \setminus E) = \sum_{j=1}^{\infty} |Q_j|.$$

The series in the right side converges, as it is bounded from above by $m^*(U)$. In addition, for any finite N, the set $Q_N = \bigcup_{j=1}^{N} Q_j$ is a positive distance away from E, thus

$$m^*(E) + \varepsilon \geq m^*(U) \geq m^*(E \cup Q_N) = m^*(E) + \sum_{j=1}^{N} |Q_j|.$$

We conclude that

$$\sum_{j=1}^{N} |Q_j| < \varepsilon$$

for all N, whence

$$m^*(U \setminus E) = \sum_{j=1}^{\infty} |Q_j| < \varepsilon,$$

as claimed. Thus, any closed set is measurable.

Let us now show that the complement $E^c = \mathbb{R}^n \backslash E$ of a measurable set E is also measurable. To this end, given any $n \in \mathbb{N}$ we can find an open set U_n containing E such that

$$m^*(U_n \setminus E) < 1/n.$$

Then the closed set $F_n = U_n^c$ lies inside E^c and, moreover, $E^c \setminus F_n = U_n \setminus E$, whence

$$m^*(E^c \setminus F_n) < \frac{1}{n}.$$

The set $F = \bigcup_{n=1}^{\infty} F_n \subset E^c$ is measurable by virtue of being a countable union of measurable sets F_n. In addition, we have

$$m^*(E^c \setminus F) = 0,$$

hence $E^c \setminus F$ is a measurable set. It follows that E^c is the union of two measurable sets: F and $E^c \setminus F$, and is thus itself measurable.

Finally, we note that the intersection of countably many measurable sets is measurable: to see that we simply write

$$E = \bigcap_{n=1}^{\infty} E_n, \quad E^c = \bigcup_{n=1}^{\infty} E_n^c.$$

In order to summarize what we shown above, it is convenient to use the notion of a σ-algebra.

Definition 1.18 *A collection $\mathcal{M}$ of sets is a σ-algebra if the following conditions hold:*

(0) The empty set $\emptyset$ is in $\mathcal{M}$.

(i) If $A \in \mathcal{M}$ and $B \in \mathcal{M}$ then $A \cup B \in \mathcal{M}$.

(ii) If $A \in \mathcal{M}$ then its complement $A^c = \mathbb{R} \setminus A$ is also in $\mathcal{M}$.

(iii) If $A_1, A_2, \ldots, A_n, \ldots \in \mathcal{M}$ then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{M}$.

A collection $\mathcal{M}$ which satisfies only (0)-(ii) above is called an algebra of sets.

Note that (i) and (ii) imply that if $A \in \mathcal{M}$ and $B \in \mathcal{M}$ then $A \cap B \in \mathcal{M}$ because of the identity $(A \cap B)^c = A^c \cup B^c$. The same is true for countable intersections.

We have proved above the following result.

Theorem 1.19 *The collection $\mathcal{M}$ of all Lebesgue measurable sets is a σ-algebra.*

We have also shown that the σ-algebra of measurable sets contains all open sets. The smallest σ-algebra that contains all open sets is called the Borel σ-algebra, and its elements are called the Borel sets. We conclude that all Borel sets are Lebesgue measurable. We will later see that not all Lebesgue measurable sets are Borel.

The Caratheodory definition of measurability

Another definition of a measurable set, due to Caratheodory is as follows.

Definition 1.20 *A set $A \subset \mathbb{R}^n$ is (Caratheodory) measurable if for each set $B \subset \mathbb{R}^n$ we have*

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

This definition is less intuitive than the one we have given previously but it is sometimes more convenient and versatile. We will show that the two definitions are equivalent. We will say, for short, that Caratheodory measurable sets are C-measurable (we will drop this designation after we prove the equivalence). As a first remark, it goes without saying that if A is a C-measurable set then so is its complement A^c.

Note that we always have

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

so to check C-measurability of a set A we would need only to verify that

$$m^*(B) \geq m^*(A \cap B) + m^*(A^c \cap B)$$

for all sets $B \subseteq \mathbb{R}^n$.

As a first example of a C-measurable set, we show that if $m^*(E) = 0$ then the set E is C-measurable. Let $B \subset \mathbb{R}^n$ be any set, then $B \cap E \subset E$, so

$$m^*(B \cap E) \leq m^*(E) = 0,$$

while $B \cap E^c \subset B$ and thus

$$m^*(B) \geq m^*(B \cap E^c) = m^*(B \cap E^c) + 0 = m^*(B \cap E^c) + B^*(B \cap E),$$

and thus the set E is C-measurable.

The next step is to show that any open corner $A = \{x_1 > a_1, x_2 > a_2, \dots, x_n > a_n\}$ is C-measurable. Let $B \subset \mathbb{R}^n$ be any set, then we need to verify that

$$m^*(B) \geq m^*(B \cap A) + m^*(B \cap A^c). \tag{1.6}$$

If $m^*(B) = +\infty$ then there is nothing to do. If $m^*(B) < +\infty$ then for any $\varepsilon > 0$ there exists a countable collection of open boxes $\{D_n\}$ so that $B \subseteq \bigcup_n D_n$ and

$$m^*(B) + \varepsilon \geq \sum_n |D_n|.$$

Define a shrunk corner

$$A_n = \{x_1 \geq a_1 + \frac{\varepsilon}{2^n}, x_2 \geq a_2 + \frac{\varepsilon}{2^n}, \dots, x_n \geq a_n + \frac{\varepsilon}{2^n}\},$$

and set $D_n' = D_n \cap A$ and $\tilde{D}_n'' = D_n \cap A_n^c$. Each D_n' is a box, while each $\tilde{D}_n''$ is a finite union of open boxes. We also have

$$B_1 := A \cap B \subseteq \bigcup_n D_n', \quad B_2 := B \cap A^c \subseteq \bigcup_n \tilde{D}_n'',$$

thus

$$m^*(B_1) \leq \sum_n |D_n'|, \quad m^*(B_2) \leq \sum_n |\tilde{D}_n''|.$$

It follows that

$$m^*(B_1) + m^*(B_2) \leq \sum_n (|D_n'| + |\tilde{D}_n''|) \leq \sum_n (|D_n| + \frac{C\varepsilon}{2^n}) \leq m^*(A) + C\varepsilon.$$

As $\varepsilon > 0$ is arbitrary, (1.6) follows. Thus, any open corner is C-measurable.

Next, we go through the steps necessary to show that the C-measurable sets from a σ-algebra. The first step is finite unions.

Lemma 1.21 *If the sets E_1 and E_2 are C-measurable then the set $E_1 \cup E_2$ is also C-measurable.*

Proof. Let A be any set, then

$$m^*(A) = m^*(A \cap E_1) + m^*(A \cap E_1^c) = m^*(A \cap E_1) + m^*((A \cap E_1^c) \cap E_2) + m^*((A \cap E_1^c) \cap E_2^c). \tag{1.7}$$

We used the C-measurability of E_1 in the first equality and of E_2 in the second. Note that

$$(A \cap E_1^c) \cap E_2^c = A \cap (E_1 \cup E_2)^c,$$

and

$$A \cap (E_1 \cup E_2) = (A \cap E_1) \cup (A \cap E_1^c \cap E_2),$$

so that

$$m^*(A \cap (E_1 \cup E_2)) \leq m^*(A \cap E_1) + m^*(A \cap E_1^c \cap E_2). \tag{1.8}$$

Now, going back to (1.7) gives

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

and thus $E_1 \cup E_2$ is measurable. □

As a consequence, the intersection of two measurable sets E_1 and E_2 is measurable because its complement is:

$$(E_1 \cap E_2)^c = E_1^c \cup E_2^c,$$

as well as their difference:

$$E_1 \setminus E_2 = E_1 \cap E_2^c.$$

The next lemma applies to finite unions but will be useful below even when we consider countable unions.

Lemma 1.22 *Let A be any set, and let $E_1, \ldots, E_n$ be a collection of pairwise disjoint C-measurable sets, then*

$$m^*(A \cap (\cup_{i=1}^n E_i)) = \sum_{i=1}^n m^*(A \cap E_i). \tag{1.9}$$

Proof. We prove this by induction. The case $n = 1$ is trivial. Assume that (1.9) holds for $n-1$, then, as E_n is measurable, we have

$$\begin{aligned} m^*(A \bigcap (\bigcup_{i=1}^n E_i)) &= m^*(A \bigcap (\bigcup_{i=1}^n E_i) \bigcap E_n) + m^*(A \bigcap (\bigcup_{i=1}^n E_i) \bigcap E_n^c) \\ &= m^*(A \bigcap E_n) + m^*(A \bigcap (\bigcup_{i=1}^{n-1} E_n) = \sum_{i=1}^n m^*(A \bigcap E_i). \end{aligned}$$

The last equality above follows from the induction assumption while the second one uses pairwise disjointness of E_i. □

Theorem 1.23 *The collection $\mathcal{M}$ of all C-measurable sets is a σ-algebra.*

Proof. Let E be the union of countably many C-measurable sets E_j. Consider the sets $\tilde{E}_j$ defined inductively by $\tilde{E}_1 = E_1$, and

$$\tilde{E}_j = E_j \setminus \bigcup_{i<j} \tilde{E}_i.$$

Then the sets $\tilde{E}_j$ are measurable, disjoint and their union is the same as that of E_j:

$$E = \bigcup_j E_j = \bigcup_j \tilde{E}_j.$$

The set

$$F_n = \bigcup_{j=1}^n \tilde{E}_j \subset E$$

is C-measurable as a finite union of C-measurable sets, and so for any set A we have

$$m^*(A) = m^*(A \cap F_n) + m^*(A \cap F_n^c) \geq m^*(A \cap F_n) + m^*(A \cap E^c).$$

As the sets $\tilde{E}_j$ are disjoint, we may use Lemma 1.22 in the right side above:

$$m^*(A) \geq \sum_{j=1}^{n} m^*(A \cap \tilde{E}_j) + m^*(A \cap E^c).$$

As this is true for all n we may pass to the limit $n \to +\infty$ to obtain

$$m^*(A) \geq \sum_{j=1}^{\infty} m^*(A \cap \tilde{E}_j) + m^*(A \cap E^c). \tag{1.10}$$

However, by sub-additivity we have

$$m^*(A \cap E) = m^*(A \cap (\cup_j \tilde{E}_j)) = m^*(\cup_j (A \cap E_j)) \leq \sum_{j=1}^{\infty} m^*(A \cap \tilde{E}_j).$$

Using this in (1.10) we get

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

Therefore, the set E is measurable. As we already know that if A is a measurable set then so is A^c, it follows that $\mathcal{M}$ is a σ-algebra. □

It is easy to see that the σ-algebra generated by open corners contains all open boxes, and thus all closed boxes. Therefore, any open set is C-measurable (as a countable union of closed boxes).

Proposition 1.24 *A set is C-measurable if and only if it is Lebesgue measurable.*

Proof. Let E be a Lebesgue measurable set, and take any set $A \subset \mathbb{R}^n$. We need to show that

$$m^*(A) \geq m^*(A \cap E) + m^*(A \cap E^c). \tag{1.11}$$

Given any $\varepsilon > 0$, choose an open set O that contains E and such that $m^*(O \setminus E) < \varepsilon$. As the set O is open, it is C-measurable, thus

$$m^*(A) = m^*(A \cap O) + m^*(A \cap O^c) \geq m^*(A \cap E) + m^*(A \cap O^c).$$

Note that

$$m^*(A \cap E^c) \leq m^*(A \cap (O \setminus E)) + m^*(A \cap O^c) \leq m^*(A \cap O^c) + \varepsilon,$$

hence

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

As this is true for any $\varepsilon > 0$, (1.11) follows.

Next, let E be a C-measurable set. We may decompose it as

$$E = \bigcup_{j=1}^{\infty} E_j, \quad E_j = E \cap \{|x| \leq j\},$$

hence it suffices to show that each E_j is measurable, thus we may assume without loss of generality that $m^*(E) < +\infty$. Then tere exists an open set O that contains E and such that

$$m^*(O) \leq m^*(E) + \varepsilon.$$

Let us denote $B = O \setminus E$. As E and B are disjoint and C-measurable, we have

$$m^*(O) = m^*(B) + m^*(E),$$

thus $m^*(B) < \varepsilon$, and thus E is Lebesgue measurable. □

Thus, Lebesgue measurability, as we have first defined it, is equivalent to the Caratheodory definition, and we can use them interchangeably.

1.2 A general definition of a measure

Let us now describe how the construction of the Lebesgue measure may be generalized in a very straightforward way to a more general class of measures.

Definition 1.25 *A mapping $\mu^* : 2^X \to \mathbb{R}$ is an outer measure on a set X if*

(i) $\mu^*(\emptyset) = 0$

(ii) $\mu^*(A) \leq \sum_{k=1}^{\infty} \mu^*(A_k)$ *whenever* $A \subseteq \bigcup_{k=1}^{\infty} A_k$.

The term "outer" in the above definition is not the best since we do not always assume that μ^* comes from some covers by open sets but we will use it anyway, simply to indicate that μ^* is not a true measure.

Definition 1.26 *A measure μ^* defined on a set X is finite if $\mu^*(X) < +\infty$.*

Definition 1.27 *Let μ^* be an outer measure on X and let $A \subset X$ be a set. Then $\mu^*|_A$, a restriction of μ^* to A is the outer measure defined by $\mu^*|_A(B) = \mu^*(A \cap B)$ for $B \subseteq X$.*

Examples. (1) The Lebesgue measure on $\mathbb{R}^n$.
(2) The counting measure: the measure $\mu^{\#}(A)$ is equal to the number of elements in A.
(3) The delta measure on the real line: given a subset $A \subseteq \mathbb{R}$, we set $\mu(A) = 1$ if $0 \in A$, and $\mu(A) = 0$ if $0 \notin A$.
(4) If A is a Lebesgue measurable set, we can take $\mu^*(B) = m(A \cap B)$, a restriction of the Lebesgue measure to to the set A.

Measurable sets

Now, we have to restrict the class of sets for which we will define the notion of a measure (as opposed to the outer measure which is defined for all sets). The Caratheodory definition applies in exactly the same form as for the Lebesgue measure.

Definition 1.28 *A set $A \subset X$ is μ-measurable if for each set $B \subset X$ we have*

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

Once again, if A is a measurable set then so is its complement A^c. If A is a measurable set, we will write $\mu(A)$ instead of $\mu^*(A)$.

The collection of all measurable sets from a σ-algebra – this can be proved verbatim as for the Lebesgue measure.

Theorem 1.29 *Let μ be an outer measure, then the collection $\mathcal{M}$ of all μ-measurable sets is a σ-algebra.*

Remark. The restriction of μ to the σ-algebra of measurable sets is called a measure. In the sequel we will freely use the word "measure" for an outer measure whether this causes confusion or not. We will denote by m the Lebesgue measure on $\mathbb{R}^n$.

Next, we investigate some general properties of measurable sets using the Caratheodory definition – all of them, of course, apply to the Lebesgue measure on $\mathbb{R}^n$.

Countable additivity

Proposition 1.30 *Let μ be a measure and let $\{E_j\}$ be a collection of pairwise disjoint measurable sets, then*

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

Proof. First, sub-additivity implies that

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

Thus, what we need to establish is

$$\mu\left(\bigcup_{j=1}^{\infty} E_j\right) \geq \sum_{j=1}^{\infty} \mu(E_j). \tag{1.12}$$

However, if all E_j are measurable and pairwise disjoint, we have, according to Lemma 1.22 (its proof used nothing about the Lebesgue measure and applied to the general case verbatim), with $A = X$, the whole measure space, for any $n \in \mathbb{N}$

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

As this is true for all $n \in \mathbb{N}$, (1.12) follows. □

Limit of a nested sequence of sets

Proposition 1.31 *Let the sets E_j be measurable, $E_{n+1} \subseteq E_n$ for all $n \geq 1$, and $\mu E_1 < +\infty$, then*

$$\mu\left(\bigcap_{j=1}^{\infty} E_j\right) = \lim_{j\to+\infty} \mu(E_j). \tag{1.13}$$

Proof. Let $E = \bigcap_{j=1}^{\infty} E_j$ and define the annuli $F_i = E_i \setminus E_{i+1}$. Then we have

$$E_1 \setminus E = \bigcup_{j=1}^{\infty} F_j,$$

and all sets F_j are disjoint. It follows from Proposition 1.30 that

$$\mu(E_1 \setminus E) = \sum_{j=1}^{\infty} \mu(F_j). \tag{1.14}$$

On the other hand, as $E \subseteq E_1$ so that $E_1 = (E_1 \setminus E) \cup E$, the same proposition implies that

$$\mu(E_1 \setminus E) = \mu(E_1) - \mu(E),$$

and, similarly,

$$\mu(F_j) = \mu(E_j) - \mu(E_{j+1}).$$

Using this in (1.14) leads to

$$\begin{aligned}\mu(E_1) - \mu(E) &= \sum_{j=1}^{\infty}(\mu(E_j) - \mu(E_{j+1}) = \lim_{n\to+\infty}\sum_{j=1}^{n}(\mu(E_j) - \mu(E_{j+1}) \\ &= \lim_{n\to+\infty}(\mu(E_1) - \mu(E_{n+1}) = \mu(E_1) - \lim_{n\to+\infty}\mu(E_n).\end{aligned}$$

Now, (1.13) follows immedaitely. □

Exercise 1.32 Show that the conclusion of Proposition 1.31 is false without the assumption $\mu(E_1) < +\infty$.

Limit of an increasing sequence of sets

Proposition 1.33 *Let the sets E_j be measurable, $E_n \subseteq E_{n+1}$ for all $n \geq 1$, then*

$$\lim_{j\to+\infty} \mu(E_j) = \mu\left(\bigcup_{j=1}^{\infty} E_j\right). \tag{1.15}$$

Proof. Let us write

$$\mu(E_{k+1}) = \mu(E_1) + \sum_{j=1}^{k}(\mu(E_{j+1}) - \mu(E_j)) = \mu(E_1) + \sum_{j=1}^{k}\mu(E_{j+1} \setminus E_j). \tag{1.16}$$

We used in the last step the fact that $E_j \subseteq E_{j+1}$. Now, let $k \to +\infty$ in (1.16) and use the fact that the sets $E_{j+1} \setminus E_j$ are pairwise disjoint, together with countable additivity of μ from Proposition 1.30

$$\lim_{k\to+\infty} \mu(E_k) = \lim_{k\to+\infty} \left(\mu(E_1) + \sum_{j=1}^{k} \mu(E_{j+1} \setminus E_j)\right)$$
$$= \mu\left(E_1 \bigcup \left(\bigcup_{j=1}^{\infty}(E_{j+1} \setminus E_j)\right)\right) = \mu\left(\bigcup_{j=1}^{\infty} E_j\right),$$

which is (1.15). □

Exercise. The set P defined after Lemma 1.1 is not measurable.

1.3 Borel and Radon measures on $\mathbb{R}^n$

We will now define a special class of measures that is a natural generalization of the notion of the Lebesgue measure on $\mathbb{R}^n$.

Definition 1.34 *(i) A measure μ is Borel if every Borel set is μ-measurable.*
(ii) A measure μ on $\mathbb{R}^n$ is Borel regular if μ is Borel and for each set $A \subset \mathbb{R}^n$ there exists a Borel set B such that $A \subseteq B$ and $\mu^*(A) = \mu(B)$.
(iii) A measure μ is a Radon measure if μ is Borel regular and $\mu(K) < +\infty$ for each compact set $K \subset \mathbb{R}^n$.

Example. 1. The Lebesgue measure is a Radon measure.
2. δ-function is a Radon measure.
3. Lots of example of Borel and Radon measures will be given after we discuss the integration theory.

Restriction of a regular Borel measure

We record a basically bureaucratic fact: the restriction of a regular Borel measure to a measurable set of finite measure is a Radon measure:

Theorem 1.35 *Let μ be a regular Borel measure on $\mathbb{R}^n$. Suppose that the set A is μ-measurable and $\mu(A) < +\infty$. Then the restriction $\mu|_A$ is a Radon measure.*

Proof. We split the proof into (simple) exercises.

Exercise 1.36 Show that $\nu = \mu|_A$ is a Borel measure.

Exercise 1.37 Show that if A is a Borel set then $\mu|_A$ is Borel regular.

Exercise 1.38 Use the result of the previous exercise to show that if μ is Borel regular and A is a measurable set (not necessarily Borel) then $\nu = \mu|_A$ is Borel regular.

Increasing sequences of sets revisited

When the measure μ is Borel regular, we may generalize Proposition 1.33 to not necessarily measurable sets.

Proposition 1.39 *Let μ be a Borel regular measure and the sets E_j satisfy $E_n \subseteq E_{n+1}$ for all $n \geq 1$, then*

$$\lim_{j\to+\infty} \mu(E_j) = \mu\left(\bigcup_{j=1}^{\infty} E_j\right). \tag{1.17}$$

Proof. As the measure μ is Borel regular, we may choose measurable sets B_k so that $E_k \subseteq B_k$ and $\mu^*(E_k) = \mu(B_k)$. The sequence B_k is not necessarily increasing, which can be circumvented by considering the sets $C_k = \bigcap_{j\geq k} B_j$. The sequence C_k is increasing, and, in addition, we have

$$E_k \subseteq E_j \subseteq B_j$$

for all $j \geq k$, whence $E_k \subseteq C_k$. Moreover, we have

$$\mu^*(E_k) = \mu(B_k) \geq \mu(C_k).$$

Proposition 1.33 implies that

$$\liminf_{k\to\infty} \mu^*(E_k) \geq \lim_{k\to\infty} \mu(C_k) = \mu(\bigcup_{k=1}^{\infty} C_k) \geq \mu^*(\bigcup_{k=1}^{\infty} E_k).$$

However, we also have

$$\mu^*(E_k) \leq \mu^*(\bigcup_{k=1}^{\infty} E_k)$$

for each k, and (1.17) follows.

Approximation by open and closed sets

The following result is a generalization of the results on approximation of sets by open and closed sets for the Lebesgue measure. In particular, it shows that any Radon measure is both an "outer" and an "inner" measure in an intuitive sense, and can be constructed as an extension from the open sets.

Theorem 1.40 *Let μ be a Radon measure, then*
(i) for each set $A \subseteq \mathbb{R}^n$ we have

$$\mu^*(A) = \inf\{\mu(U) :\ A \subseteq U,\ U \text{ open}\}. \tag{1.18}$$

(ii) for each μ-measurable set A we have

$$\mu(A) = \sup\{\mu(K) :\ K \subseteq A,\ K \text{ compact.}\} \tag{1.19}$$

Exercise 1.41 Show that the conclusion in part (ii) may fail without the assumption that A is a measurable set.

We begin with the following lemma which proves the statement of the theorem for the Borel sets.

Lemma 1.42 *Let μ be a Borel measure and B be a Borel set.*
(i) If $\mu(B) < +\infty$ then for any $\varepsilon > 0$ there exists a closed set C contained in B, and such that $\mu(B \setminus C) < \varepsilon$.
(ii) If μ is Radon then for any $\varepsilon > 0$ there exists an open set U which contains B, and such that $\mu(U \setminus B) < \varepsilon$.

Proof. (i) Take a Borel set B with $\mu(B) < +\infty$, and define the restriction $\nu = \mu|_B$. As at the beginning of the proof of Theorem 1.35, we deduce that ν is a Borel measure. In addition, the measure ν is finite since $\mu(B) < +\infty$.

Let $\mathcal{F}$ be the collection of all μ-measurable subsets A of $\mathbb{R}^n$ such that for any $\varepsilon > 0$ we can find a closed set C which is contained in A, and such $\nu(A \setminus C) < \varepsilon$. Our goal is to show that $\mathcal{F}$ contains all Borel sets.

Step 1: Closed sets. The first trivial observation is that $\mathcal{F}$ contains all closed sets.

Step 2: Infinite intersections. Let us now show that if the sets $A_j \in \mathcal{F}$ for all $j = 1, 2, \dots,$ then so is their intersection:

$$A = \bigcap_{j=1}^{\infty} A_j \in \mathcal{F}.$$

Indeed, given $\varepsilon > 0$, using the fact that $A_j \in \mathcal{F}$, we choose the closed sets $C_j \subseteq A_j$ so that

$$\nu(A_j \setminus C_j) < \frac{\varepsilon}{2^j}. \tag{1.20}$$

Then the closed set $C = \cap_{j=1}^{\infty} C_j$ is contained in A and, moreover,

$$\nu(A \setminus C) \le \nu\left(\bigcup_{j=1}^{\infty}(A_j \setminus C_j)\right) \le \sum_{j=1}^{\infty} \nu(A_j \setminus C_j) < \varepsilon.$$

Therefore, indeed, $A \in \mathcal{F}$.

Step 3: Infinite unions. Next, we establish that if the sets $A_j \in \mathcal{F}$ for all $j = 1, 2, \dots,$ then so is their union:

$$A = \bigcup_{j=1}^{\infty} A_j \in \mathcal{F}.$$

Choose the sets C_j as in (1.20). As a countable union of closed sets need not be closed, we can not take their union as the approximating set C. However, as $\nu(A) < +\infty$ we have, using Proposition 1.31 (nested sequences of sets),

$$\begin{aligned}\lim_{m \to +\infty} \nu(A \setminus (\bigcup_{j=1}^{m} C_j)) &= \nu(A \setminus \bigcup_{j=1}^{\infty} C_j) = \nu((\bigcup_{j=1}^{\infty} A_j) \setminus (\bigcup_{j=1}^{\infty} C_j)) \\ &\le \nu \bigcup_{j=1}^{\infty}(A_j \setminus C_j) \le \sum_{j=1}^{\infty} \nu(A_j \setminus C_j) < \varepsilon.\end{aligned}$$

Therefore, there exists $m_0 \in \mathbb{N}$ so that

$$\nu(A \setminus (\bigcup_{j=1}^{m_0} C_j)) < \varepsilon,$$

and the set $C = \bigcup_{j=1}^{m_0} C_j$ is closed.

Next, we define $\mathcal{G}$ as the collection of all sets A such that both $A \in \mathcal{F}$ and $A^c \in \mathcal{F}$. It is sufficient to show that

$$\mathcal{G} \text{ contains all open sets and is a } \sigma\text{-algebra}, \tag{1.21}$$

in order to show that $\mathcal{G}$ (and thus $\mathcal{F}$) contains all Borel sets (thus, in particular, the set B with which we have started our construction).

Step 4: collection $\mathcal{G}$ contains all open sets. If O is an open set then O^c is closed and thus $O^c \in \mathcal{F}$ automatically by Step 1. But any open set can be written as a countable union of closed sets, hence by Step 3 collection $\mathcal{F}$ contains all open sets, hence, in particular, our set O. Thus, both O and O^c are in $\mathcal{F}$, so $O \in \mathcal{G}$.

Step 5: G is a σ-algebra. Obviously, if $A \in \mathcal{G}$ then $A^c \in \mathcal{G}$ as well. Therefore, we only need to check that if $A_1, A_2, \dots, A_n, \dots \in \mathcal{G}$ then $A = \cup_{j=1}^{\infty} A_j \in \mathcal{G}$. But $A \in \mathcal{F}$ by Step 3, while Step 2 implies that

$$A^c = \bigcap_{j=1}^{\infty} (A_j^c)$$

is in $\mathcal{F}$ as well, and thus $A \in \mathcal{G}$. We conclude that $\mathcal{G}$ is a σ-algebra, finishing the proof of part (i) of the Lemma.

(ii) Now, we prove the second part of Lemma 1.42. Let B be a Borel set. If $\mu(B^c) < +\infty$, we can use part the already proved part (i) of the present Lemma to find a closed set C contained in B^c such that $\mu(B^c \setminus C) < \varepsilon$. Then we set $U = C^c$ and deduce that

$$\mu(U \setminus B) = \mu(C^c \setminus (B^c)^c) = \mu(B^c \setminus C) < \varepsilon.$$

When $\mu(B^c) + \infty$, we use the usual restriction to balls: let $U_m = U(0, m)$ be the open ball centered at $x = 0$ of radius m. Then $\mu(U_m \cap B^c) < +\infty$ as μ is Radon. We may then apply part (i) to the Borel set $U_m \cap B^c$, and find a closed set $C_m \subseteq U_m \cap B^c$ with

$$\mu\left((U_m \cap B^c) \setminus C_m\right) < \frac{\varepsilon}{2^m}.$$

Then $U_m \cap B \subseteq U_m \setminus C_m$, so that

$$B = \bigcup_{m=1}^{\infty} (U_m \cap B) \subseteq \bigcup_{m=1}^{\infty} (U_m \setminus C_m) := U.$$

The set U is open and

$$\mu(U \setminus B) = \mu\left(\left[\bigcup_{m=1}^{\infty} (U_m \setminus C_m)\right] \setminus B\right) \le \sum_{m=1}^{\infty} \mu((U_m \setminus C_m) \setminus B) < \varepsilon,$$

and we are done $\square$.

Proof of Theorem 1.40

(i) We begin with the first part of the theorem. If $\mu^*(A) = +\infty$ the statement is obvious, just take $U = \mathbb{R}^n$, so we assume that $\mu(A) < +\infty$. If A is a Borel set then (i) holds because of part (ii) of Lemma 1.42. If A is not a Borel set then, as μ is a Borel regular measure, there exists a Borel set B such that $A \subseteq B$ and $\mu^*(A) = \mu(B)$. Then, once again we may apply part (ii) of Lemma 1.42 to see that

$$\begin{aligned}\inf\{\mu(U):\ A \subseteq U,\ U \text{ open}\} \geq \mu^*(A) = \mu(B) &= \inf\{\mu(U):\ B \subseteq U,\ U \text{ open}\}\\ &\geq \inf\{\mu(U):\ A \subseteq U,\ U \text{ open}\},\end{aligned}$$

which implies (1.18).

(ii) Now, we prove (1.19). First, assume that A is a μ-measurable set and $\mu(A) < +\infty$. Then the restriction $\nu = \mu|_A$ is a Radon measure, hence the already proved part (i) of the present Theorem applies to ν. Thus, given $\varepsilon > 0$, we may apply (i) to the set A^c, with $\nu(A^c) = 0$, and find an open set U such that $A^c \subseteq U$ and $\nu(U) < \varepsilon$. The set $C = U^c$ is closed, $C \subseteq A$ and

$$\mu(A \setminus C) = \nu(C^c) = \nu(U) < \varepsilon.$$

It follows that

$$\mu(A) = \sup\{\mu(C):\ C \subseteq A,\ C \text{ closed}\} \text{ if } \mu(A) < +\infty. \tag{1.22}$$

If A is μ-measurable and $\mu(A) = +\infty$, define the annuli $D_k = \{x:\ k-1 \leq |x| < k\}$ and split

$$A = \bigcup_{k=1}^{\infty}(A \cap D_k).$$

Observe that

$$+\infty = \mu(A) = \sum_{k=1}^{\infty}\mu(A \cap D_k),$$

while $\mu(A \cap D_k) < +\infty$ since μ is a Radon measure. We can then use (1.22) to find closed sets $C_k \subseteq A \cap D_k$ such that

$$\mu((A \cap D_k) \setminus C_k) < \frac{1}{2^k}.$$

Once again, the union of all sets C_k need not be closed, so instead we consider the closed sets $G_n = \bigcup_{k=1}^n C_k$. Note that, as all C_k are pairwise disjoint, we have

$$\mu(G_n) = \sum_{k=1}^{n}\mu(C_k) \geq \sum_{k=1}^{n}\left(\mu(A \cap D_k) - \frac{1}{2^k}\right). \tag{1.23}$$

As, by Proposition 1.33 (increasing sequences of sets), we have

$$+\infty = \mu(A) = \lim_{n\to+\infty}\mu\left(\bigcup_{k=1}^{n}(A \cap D_k)\right) = \lim_{n\to+\infty}\sum_{k=1}^{n}\left(\mu(A \cap D_k)\right),$$

we deduce from (1.23) that

$$\lim_{n\to+\infty}\mu(G_n) = +\infty = \mu(A).$$

Therefore, (1.22) actually holds also if $\mu(A) = +\infty$.

What remains is to replace the word "closed" in (1.22) by "compact". However, in the case $\mu(A) = +\infty$ the sets G_n we have constructed are, actually, closed and bounded, hence compact, so we do not need to do anything. If $\mu(A) < +\infty$, we can do the usual "restriction to the balls": given $\varepsilon > 0$ take a closed set $C \subseteq A$ such that

$$\mu(C) > \mu(A) - \varepsilon,$$

and write $C = \bigcup_{k=1}^{\infty} C_k$, with $C_k = C \cap \bar{U}(0, k)$. Then, each C_k is a compact set, $C_k \subset A$, and

$$\mu(C) = \lim_{k\to+\infty} \mu(C_k),$$

because of Proposition 1.33 again. Hence, there exists a positive integer k_0 so that

$$\mu(C_{k_0}) > \mu(A) - \varepsilon,$$

and (1.19) follows. This completes the proof of the theorem.

2 Measurable functions

2.1 Definition and basic properties

Recall that a function is continuous if pre-image of every open set is open. Measurable functions are defined in a similar spirit. We start with the following observation.

Proposition 2.1 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a real-valued function defined on a measurable set. Then the following are equivalent.*
(i) For any $\alpha \in \mathbb{R}$ the set $\{x : f(x) > \alpha\}$ is measurable,
(ii) For any $\alpha \in \mathbb{R}$ the set $\{x : f(x) \geq \alpha\}$ is measurable,
(iii) For any $\alpha \in \mathbb{R}$ the set $\{x : f(x) \leq \alpha\}$ is measurable,
(iv) For any $\alpha \in \mathbb{R}$ the set $\{x : f(x) < \alpha\}$ is measurable.

Proof. First, it is obvious that (i) and (iii) are equivalent, and so are (ii) and (iv). If we write

$$\{x : f(x) > \alpha\} = \bigcup_{n=1}^{\infty} \{x : f(x) \geq \alpha + 1/n\}$$

we see that (ii) implies (i), and, similarly, we get that (iii) implies (iv). □

This leads to the following, somewhat more general definition. Let X be a set, Y a topological space and assume that μ is a measure on X.

Definition 2.2 *A function $f : X \to Y$ is μ-measurable if for each open set $U \subseteq Y$, the pre-image $f^{-1}(U)$ is μ-measurable.*

For real-valued functions it suffices to check that pre-images of the half intervals $(\alpha, +\infty)$ are all open in order to establish measurability of a function.

The next proposition gives some basic properties of measurable functions which are neither surprising nor particularly amusing.

Proposition 2.3 *If $f, g: X \to \mathbb{R}$ are measurable functions and $c \in \mathbb{R}$ is a real number then the following functions are also measurable: cf, $f + c$, $f + g$, $f - g$ and fg.*

Proof. (1) To see that $f + c$ is measurable we simply note that

$$\{x: f(x) + c < \alpha\} = \{x: f(x) < \alpha - c\}.$$

(2) For measurability of cf with $c > 0$ we observe that $\{x: cf(x) < \alpha\} = \{x: f(x) < \alpha/c\}$, and the case $c \leq 0$ is not very different.
(3) To show that $f + g$ is measurable we decompose

$$\{x: f(x) + g(x) < \alpha\} = \bigcup_{r \in \mathbb{Q}} \left[\{x: f(x) < \alpha - r\} \cap \{x: g(x) < r\}\right].$$

(4) The function $f^2(x)$ is measurable because for $\alpha \geq 0$ we have

$$\{x: f^2(x) > \alpha\} = \{x: f(x) > \sqrt{\alpha}\} \cup \{x: f(x) < \sqrt{\alpha}\},$$

and the case $\alpha < 0$ is not that difficult.
(5) Finally, the product fg is measurable because

$$f(x)g(x) = \frac{(f+g)^2 - (f-g)^2}{4},$$

and the right side is measurable by (1)-(4) shown above. $\Box$

Exercise 2.4 Show that if the map $f: Y \to Y$ is continuous and $g: X \to Y$ is measurable, then the composition $f \circ g: X \to Y$ is measurable. Is it true that if $f: \mathbb{R} \to \mathbb{R}$ is a continuous function and $g: \mathbb{R} \to \mathbb{R}$ is measurable function then $g \circ f$ is measurable?

The next theorem is certainly not true in the world of continuous functions: point-wise limits of continuous functions may be quite discontinuous but limits of measurable functions are measurable. In a sense, that is one of the main advantages of the class of measurable functions.

Theorem 2.5 *If the functions $f_1, f_2, \ldots, f_n, \ldots$ are all measurable then so are*

$$g_n(x) = \sup_{1 \leq j \leq n} f_j(x),\ q_n(x) = \inf_{1 \leq j \leq n} f_j(x),\ g(x) = \sup_n f_n(x),\ q(x) = \inf_n f_n(x),$$

as well as

$$s(x) = \limsup_{n \to \infty} f_n(x) \text{ and } w(x) = \liminf_{n \to \infty} f_n(x).$$

Proof. For $g_n(x)$ and $g(x)$ we can write

$$\{g_n(x) > \alpha\} = \bigcup_{j=1}^{n} \{f_j(x) > \alpha\}, \quad \{g(x) > \alpha\} = \bigcup_{j=1}^{\infty} \{f_j(x) > \alpha\},$$

which shows that $g_n(x)$ and $g(x)$ are both measurable, and $q_n(x)$ are $q(x)$ are measurable for a similar reason. Now, for $s(x)$ we use the representation

$$s(x) = \limsup_{n\to\infty} f_n(x) = \inf_n \left(\sup_{k\geq n} f_k(x) \right),$$

and see that $s(x)$ is measurable by what we have just proved. The function $w(x)$ is measurable for a similar reason. □

The next result gives a very convenient representation of a positive function as a sum of simple functions. We denote by χ_A the characteristic function of a set A:

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

Definition 2.6 *A measurable function $f(x)$ is simple if it takes at most countably many values.*

A simple function $f(x)$ that takes values $a_1, a_2, \ldots, a_k, \ldots$ can be written as

$$f(x) = \sum_{k=1}^{\infty} a_k \chi_{A_k}(x).$$

where $A_k = \{x : \ f(x) = a_k\}$ are disjoint sets. It turns out that if we drop the restriction that the sets A_k are disjoint, any measurable function can be written in this way, even with prescribed values of a_k.

Theorem 2.7 *Let a non-negative function f be μ-measurable. Then there exist μ-measurable sets A_k such that*

$$f(x) = \sum_{k=1}^{\infty} \frac{1}{k} \chi_{A_k}(x). \tag{2.1}$$

Proof. Set

$$A_1 = \{x : \ f(x) \geq 1\}$$

and continue inductively by setting

$$A_k = \left\{ x : \ f(x) \geq \frac{1}{k} + \sum_{j=1}^{k-1} \frac{1}{j} \chi_{A_j}(x) \right\}.$$

By construction, we have, for all k:

$$f(x) \geq \sum_{j=1}^{k} \frac{1}{j} \chi_{A_j}(x),$$

and thus

$$f(x) \geq \sum_{k=1}^{\infty} \frac{1}{k} \chi_{A_k}(x). \tag{2.2}$$

If $f(x) = +\infty$ then $x \in A_k$ for all k, hence

$$f(x) = \sum_{k=1}^{\infty} \frac{1}{k} \chi_{A_k}(x) \text{ if } f(x) = +\infty.$$

On the other hand, (2.2) implies that if $f(x) < +\infty$ then $x \notin A_k$ for infinitely many k, which means that

$$\sum_{j=1}^{k-1} \frac{1}{j} \chi_{A_j}(x) \le f(x) \le \frac{1}{k} + \sum_{j=1}^{k-1} \frac{1}{j} \chi_{A_j}(x)$$

for infinitely many k. This implies that (2.1) holds also for the points where $f(x) < +\infty$. □
Remark. Note that this proof works with $1/k$ replaced by any non-negative sequence $a_k \ge 0$ such that both $a_k \to 0$ as $k \to +\infty$ and $\sum_{k=1}^{\infty} a_k = +\infty$.

Exercise 2.8 Show that any measurable function $f(x)$ is a limit of a sequence of simple functions.

2.2 Lusin's and Egorov's theorems

Lusin's theorem says, roughly speaking, that any measurable function coincides with a continuous function on a set of large measure. The catch is that you do not have control of the structure of the set where the two functions coincide. For instance, the Dirichlet function which is equal to one at irrational numbers and to zero at rational ones coincides with the function equal identically to one on the set of irrational numbers which has full measure but lots of holes.

Extension of a continuous function

As a preliminary tool, which is useful in itself, we prove the following extension theorem. Generally, extension theorems deal with extending "good" functions from a set to a larger set preserving "goodness" of the function. The following is just one example of such result.

Theorem 2.9 *Let $K \subseteq \mathbb{R}^n$ be a compact set and $f : K \to \mathbb{R}^m$ be continuous. Then there exists a continuous mapping $\bar{f} : \mathbb{R}^n \to \mathbb{R}^m$ such that $\bar{f}(x) = f(x)$ for all $x \in K$, and $|\bar{f}(x)| \le \sup_{y \in K} |\bar{f}(y)|$ for all $x \in \mathbb{R}^n$.*

Proof. The proof is very explicit. We take $m = 1$ but generalization to $m > 1$ is immediate. Let $U = K^c$ be the complement of K – we need to extend f to the set U. The extension will be a "weighted average" of the values of f on K, with the higher weight given to the points in K which are close to a given point $x \in U$. In order to define an appropriate weight function, given $x \in U$ and $s \in K$, set

$$u_s(x) = \max\left\{2 - \frac{|x-s|}{\text{dist}(x,K)}, 0\right\}.$$

For each $s \in K$ fixed, the function $u_s(x)$ is continuous in $x \in U$, $0 \le u(x) \le 1$ and

$$u_s(x) = 0 \text{ if } |x-s| \ge 2\text{dist}(x,K),$$

which happens when $x \in U$ is "close" to K and s is "sub-optimal" for this x. On the other hand, for a fixed x close to ∂K the function $u_s(x)$ vanishes for s which are far from s_x which realizes the distance from x to K. When x is "far" from K, $u_s(x)$ is close to 1, that is,

$$u_s(x) \to 1 \text{ as } |x| \to +\infty.$$

Now, take a dense set $\{s_j\}$ in K and for $x \in U$ define an averaged cut-off

$$\sigma(x) = \sum_{j=1}^{\infty} \frac{u_{s_j}(x)}{2^j}.$$

Note that for $x \in U$ the function $\sigma(x)$ is continuous, because all $u_{s_j}(x)$ are continuous, and by the Weierstrass test. Moreover, for any $x \in U$ there exists s_{j_0} such that $|x - s_{j_0}| \leq 2\text{dist}(x, K)$, since $\{s_j\}$ are dense. Therefore, $u_{s_{j_0}}(x) > 0$ and thus $\sigma(x) > 0$ for all $x \in U$. Let us also set normalized weights of each point s_j

$$v_j(x) = \frac{2^{-j} u_{s_j}(x)}{\sigma(x)}.$$

Note that

$$\sum_{j=1}^{\infty} v_j(x) \equiv 1 \tag{2.3}$$

for all $x \in U$. Finally, we define the extension of $f(x)$ to all of $\mathbb{R}^n$ by

$$\bar{f}(x) = \begin{cases} f(x), \ x \in K, \\ \sum_{j=1}^{\infty} v_j(x) f(s_j), \ x \in U = \mathbb{R}^n \setminus K. \end{cases} \tag{2.4}$$

The idea is that the function $\bar{f}(x)$ is defined as a weighted average of $f(s_j)$ with the bigger weight going to s_j which are close to x.

Let us check that $\bar{f}(x)$ is continuous. The series in (2.4) for $x \in U$ converges uniformly because $0 \leq u_{s_j}(x) \leq 1$, the function $\sigma(x)$ is continuous and $|f(s_j)| \leq M$ since f is a continuous function, $s_j \in K$, and the set K is compact. As each individual term $f(s_j)v_j(x)$ is a continuous function, the uniform convergence of the series implies that $\bar{f}(x)$ is continuous at $x \in U$.

Now, let us show that for each $x \in K$ we have $f(x) = \lim_{y \to x} \bar{f}(y)$. Given $\varepsilon > 0$, we use the uniform continuity of the function f on the compact set K to choose $\delta > 0$ so that

$$|f(x) - f(x')| < \varepsilon \text{ for all } x, x' \in K \text{ such that } |x - x'| < \delta.$$

Consider $y \in U$ such that $|y - x| < \delta/4$. Then, if $|x - s_k| \geq \delta$ we have

$$\delta \leq |x - s_k| \leq |x - y| + |y - s_k|,$$

thus

$$|y - s_k| \geq \frac{3\delta}{4} \geq 2|x - y|,$$

and hence $u_{s_k}(y) = v_k(y) = 0$ for such s_k. Therefore, we have

$$|f(x) - f(s_k)| < \varepsilon \text{ for all } s_k \text{ such that } v_k(y) \neq 0,$$

and we may simply estimate, using (2.3):

$$|\bar{f}(y) - \bar{f}(x)| = \left| \sum_{k=1}^{\infty} v_k(y) f(s_k) - \sum_{k=1}^{\infty} v_k(y) f(x) \right| \leq \sum_{k=1}^{\infty} v_k(y) |f(x) - f(s_k)| < \varepsilon.$$

Therefore, the function $\bar{f}(x)$ is continuous everywhere. The claim that

$$|\bar{f}(x)| \leq \sup_{y \in K} |\bar{f}(y)|$$

for all $x \in \mathbb{R}^n$ follows immediately from the definition of $\bar{f}(x)$ and (2.3). □

Exercise 2.10 Assume that $f(x)$ is Lipschitz on K. What can you say about the regularity of the extension we have constructed above on $\mathbb{R}^n$?

Lusin's Theorem

Lusin's theorem says that every measurable function coincides with a continuous function on an arbitrarily large set.

Theorem 2.11 *Let μ be a Borel regular measure on $\mathbb{R}^n$ and let $f : \mathbb{R}^n \to \mathbb{R}^m$ be μ-measurable. Assume $A \subset \mathbb{R}^n$ is a μ-measurable set with $\mu(A) < +\infty$. For any $\varepsilon > 0$ there exists a compact set $K_\varepsilon \subseteq A$ such that $\mu(A \setminus K_\varepsilon) < \varepsilon$ and the restriction of the function f to the compact set K_ε is continuous.*

Proof. As usual, it is sufficient to prove this for $m = 1$. We will construct a compact set K_ε on which $f(x)$ is a limit of a uniformly converging sequence of continuous functions and is therefore itself continuous on K_ε. To this end, for each $p \in \mathbb{N}$ take half-open intervals

$$B_{pj} = [j/2^p, (j+1)/2^p), \quad j \in \mathbb{Z},$$

and define the pre-images $A_{pj} = A \cap (f^{-1}(B_{pj}))$. The sets A_{pj} are μ-measurable, and

$$A = \bigcup_{j=1}^{\infty} A_{pj},$$

for each p fixed. Let $\nu = \mu|_A$ be the restriction of μ to the set A, then ν is a Radon measure. Hence, there exists a compact set $K_{pj} \subseteq A_{pj}$ such that

$$\nu\left(A_{pj} \setminus K_{pj}\right) < \frac{\varepsilon}{2^{p+j}},$$

and thus

$$\mu\left(A \setminus \bigcup_{j=1}^{\infty} K_{pj}\right) < \frac{\varepsilon}{2^p}.$$

Now, we choose $N(p)$ so that

$$\mu\left(A \setminus \bigcup_{j=1}^{N(p)} K_{pj}\right) < \frac{\varepsilon}{2^p},$$

and set

$$D_p = \bigcup_{j=1}^{N(p)} K_{pj}.$$

Then, the set D_p is compact. For each p, consider the function

$$g_p(x) = j/2^p \text{ for } x \in K_{pj},\ 1 \le j \le N(p),$$

defined on D_p. As the compact sets K_{pj} are pairwise disjoint, they are all a finite distance apart, thus the function $g_p(x)$ is continuous on the set D_p and, moreover, we have

$$|f(x) - g_p(x)| < \frac{1}{2^p} \text{ for all } x \in D_p. \tag{2.5}$$

Finally, set

$$K_\varepsilon = \bigcap_{p=1}^{\infty} D_p.$$

Then, the set K_ε (which depends on ε through the original choice of the sets K_{pj}) is compact, and

$$\mu(A \setminus K) \le \sum_{p=1}^{\infty} \mu(A \setminus D_p) < \varepsilon.$$

Moreover, (2.5) implies that the sequence $g_p(x)$ converges uniformly to the function $f(x)$ on the set K_ε, and thus f is continuous on the set K_ε. □

A direct consequence of Theorems 2.9 and 2.11 is the following.

Corollary 2.12 *Let μ and A be as in Lusin's theorem. Then there exists a continuous function $\bar{f}: \mathbb{R}^n \to \mathbb{R}^m$ such that $\mu\{x \in A: f(x) \ne \bar{f}(x)\} < \varepsilon$.*

We note that if $f(x)$ is a bounded function: $|f(x)| \le M$, then the function $\bar{f}(x)$ can be chosen so that $|\bar{f}(x)| \le M$ as well – this follows from the last statement in Theorem 2.9.

Egorov's theorem

Egorov's theorem shows that a point-wise converging sequence converges uniformly except possibly on a small set.

Theorem 2.13 *Let μ be a measure on $\mathbb{R}^n$ and let the functions $f_k: \mathbb{R}^n \to \mathbb{R}$ be μ-measurable. Assume that the set A is μ-measurable with $\mu(A) < +\infty$ and $f_k \to g$ almost everywhere on A. Then, for any $\varepsilon > 0$ there exists a μ-measurable set B_ε such that (i) $\mu(A \setminus B_\varepsilon) < \varepsilon$, and (ii) the sequence f_k converges uniformly to g on the set B_ε.*

Proof. Let us define a nested sequence of "not yet good by j" sets

$$C_{ij} = \bigcup_{k=j}^{\infty} \left\{ x \in A : \ |f_k(x) - g(x)| > \frac{1}{2^i} \right\},$$

then $C_{i,j+1} \subset C_{ij}$. The pointwise convergence of f_k to g implies that $\bigcap_{j=1}^{\infty} C_{ij} = \emptyset$, for any $i \in \mathbb{N}$ fixed. Therefore, as $\mu(A) < +\infty$, we have, also for each i fixed:

$$\lim_{j\to\infty} \mu\left(C_{ij}\right) = \mu(\bigcap_{j=1}^{\infty} C_{ij}) = 0.$$

Then, there exists N_i such that

$$\mu\left(C_{i,N_i}\right) < \frac{\varepsilon}{2^i},$$

and for all $x \in A \setminus C_{i,N_i}$ we have

$$|f_k(x) - g(x)| < \frac{1}{2^i} \text{ for all } k \geq i.$$

Set

$$B = A \setminus (\bigcup_{i=1}^{\infty} C_{i,N_i}),$$

then $\mu(A \setminus B) < \varepsilon$, and for each $x \in B$ and for all $n \geq N_i$ we have

$$|f_n(x) - g(x)| \leq \frac{1}{2^i},$$

hence $f_n(x)$ converges uniformly to $g(x)$ on the set B. □

Convergence in probability

We now discuss very briefly another notion of convergence, which arises very often in the probability theory.

Definition 2.14 *A sequence of measurable functions f_n converges in probability to a function f on a set E if for every $\varepsilon > 0$ there exists Nsuch that for all $n \geq N$ we have*

$$\mu\left(x \in E : \ |f_n(x) - f(x)| \geq \varepsilon\right) < \varepsilon.$$

It is quite easy to see that convergence in probability need not imply point-wise convergence anywhere: take a sequence

$$s_n = \left(\sum_{k=1}^{n} \frac{1}{k}\right) (\text{mod} 1)$$

and consider the functions

$$\phi_n(x) = \begin{cases} \chi_{[s_n, s_{n+1}]}(x), & \text{if } 0 \leq s_n < s_{n+1} \leq 1 \\ \chi_{[0,s_{n+1}]}(x) + \chi_{[s_n,1]}(x), & \text{if } 0 \leq s_{n+1} < s_n \leq 1. \end{cases}$$

Then $\phi_n \to 0$ in probability but $\phi_n(x)$ does not go to zero point-wise anywhere on $[0, 1]$. Nevertheless, convergence in probability implies point-wise convergence along a subsequence.

Proposition 2.15 *Assume that f_n converges to f in probability on a set E. Then there exists a subsequence f_{n_k} which converges to $f(x)$ point-wise a.e. on E.*

Proof. For any j we can find a number N_j such that for any $n > N_j$ we have

$$\mu\left(x \in E: \; |f(x) - f_n(x)| \geq \frac{1}{2^j}\right) \leq \frac{1}{2^j}.$$

We will show that the sequence $\phi_j(x) = f_{N_j}(x)$ converges in probability to f as $j \to +\infty$. To this end, define the bad sets

$$E_j = \left\{x \in E: \; |f(x) - f_{N_j}(x)| \geq \frac{1}{2^j}\right\},$$

and the (decreasing sequence of) sets

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

Note that for $x \notin D_k$ we have

$$|f(x) - f_{N_j}(x)| < \frac{1}{2^j}$$

for all $j \geq k$. Thus, $f_{N_j}(x) \to f$ as $j \to \infty$ for all x outside of the "very bad" set

$$D = \bigcap_{k=1}^{\infty} D_k.$$

However, we have

$$\mu(D) \leq \mu(D_k) \leq \frac{1}{2^{k-1}} \text{ for all } k,$$

and thus $\mu(D) = 0$. □

3 Integrals and limit theorems

Definition of the integral

Here, we will define the Lebesgue integral as well as integral with respect to other measures. The main difference with the Riemann integral is that the latter is not very stable under taking limits of functions simply because point-wise limits of continuous functions can be extremely bad and not Riemann integrable. The definition of the Lebesgue integral, on the contrary, makes it very stable under limits.

Recall that a function $f(x)$ is simple if it is measurable and takes countably many values. For a simple, measurable and non-negative function $f(x) \geq 0$ which takes values $y_j \geq 0$:

$$f(x) = \sum_j y_j \chi_{A_j}(x), \tag{3.1}$$

with disjoint μ-measurable sets A_j, we define

$$\int f(x)d\mu = \sum_j y_j \mu(f^{-1}(y_j)) = \sum_j y_j \mu(A_j). \tag{3.2}$$

Compared to the Riemann integral, we allow the sets A_j to be just measurable, and thus have a rather complicated structure which would rule out Riemann integrability of $f(x)$ of the form (3.1).

If $f(x)$ is simple and measurable but is not necessarily positive, we write $f = f^+ - f^-$, where $f^+ = \max(f, 0)$ and $f^- = \max(-f, 0)$. If either

$$\int f^+ d\mu < +\infty,$$

or

$$\int f^- d\mu < +\infty,$$

then we set

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

The next proposition is the key step in the definition of the Lebesgue integral.

Proposition 3.1 *Let f be a bounded function defined on a measurable set E with $\mu E < +\infty$. In order that*

$$\int^* f d\mu := \inf_{f \leq \psi} \int \psi d\mu = \sup_{f \geq \phi} \int \phi d\mu := \int_* f d\mu$$

where the infimum and the supremum are taken over all measurable simple functions $\phi \leq f$ and $\psi \geq f$, respectively, it is necessary and sufficient that f be measurable.

Proof. (1) Let f be a bounded measurable function, with $|f(x)| \leq M$ for all $x \in E$. Choose a mesh step M/n and consider the pre-images

$$E_k = \left\{ x : \frac{(k-1)M}{n} < f(x) \leq \frac{kM}{n} \right\},$$

with $-n \leq k \leq n$, then E_k are disjoint,

$$E = \bigcup_{k=-n}^{n} E_k$$

and each set E_k is measurable. Consider the simple approximants

$$\psi_n(x) = \sum_{k=-n}^{n} \frac{kM}{n} \chi_{E_k}(x), \quad \phi_n(x) = \sum_{k=-n}^{n} \frac{(k-1)M}{n} \chi_{E_k}(x),$$

so that $\phi_n(x) \leq f(x) \leq \psi_n(x)$ for all $x \in E$. Then we have

$$0 \leq \int \psi_n - \int \phi_n = \frac{M}{n} \mu E,$$

and thus

$$\int_* f d\mu = \int^* f d\mu. \tag{3.3}$$

(2) On the other hand, if (3.3) holds then for every n there exist measurable simple functions $\phi_n \leq f$ and $\psi_n \geq f$ such that

$$\int (\psi_n - \phi_n) d\mu \leq \frac{1}{n}. \tag{3.4}$$

Set

$$\psi^*(x) = \liminf_n \psi_n(x), \quad \phi^*(x) = \limsup_n \phi_n(x),$$

then ϕ^* and ψ^* are both measurable and $\phi^* \leq f \leq \psi^*$. Consider the set

$$A = \{x \in E : \ \phi^*(x) < \psi^*(x)\} = \bigcup_{k=1}^{\infty} \{x \in E : \ \phi^*(x) < \psi^*(x) - \frac{1}{k}\} := \bigcup_{k=1}^{\infty} A_k.$$

Given any $k \in \mathbb{N}$ note that for n large enough we have $\phi_n(x) < \psi_n(x) - 1/k$ on the set A_k, and thus, as $\psi_n - \phi_n > 0$, we have

$$\int_E \psi_n d\mu - \int_E \phi_n d\mu = \int_E (\psi_n - \phi_n) d\mu \geq \int_{A_k} (\psi_n - \phi_n) d\mu \geq \frac{\mu(A_k)}{k}.$$

Combining this with (3.4) and letting $n \to +\infty$ we conclude that $\mu(A_k) = 0$ for all k. Thus, we have $\phi^* = \psi^* = f$ except on a set of measure zero, hence the function f is measurable. □

Proposition 3.1 motivates the following definition of the intergal of a measurable function over a set of finite measure.

Definition 3.2 *Let f be a bounded measurable function defined on a measurable set E with $\mu E < +\infty$ then*

$$\int_E f d\mu = \inf \int_E \psi d\mu,$$

with the infimum taken over all simple functions $\psi \geq f$.

The next step in the hierarchy of the definitions is to define the integral of a non-negative function over an arbitrary measurable set.

Definition 3.3 *If $f \geq 0$ is a non-negative measurable function defined on a measurable set E, we define*

$$\int_E f d\mu = \sup_{h \leq f} \int_E h d\mu,$$

with supremum taken over all bounded simple functions h that vanish outside a set of finite measure.

This gives way to the general case.

Definition 3.4 *A non-negative measurable function f defined on a measurable set E is said to be integrable if*

$$\int_E f d\mu < +\infty.$$

A measurable function g defined on a measurable set E is integrable if both $g^+ = \max(g, 0)$ and $g^- = \max(0, -g)$ are integrable.

The Markov and Chebyshev inequalities and convergence in probability

The Markov and Chebyshev inequalities are common computational tools to establish convergence in probability. We first give the Markov inequality.

Theorem 3.5 *Let $f \geq 0$ be a non-negative measurable function, then for any $\lambda > 0$ we have*

$$\mu(\{x: \ f(x) > \lambda\}) \leq \frac{1}{\lambda} \int f(x) d\mu. \tag{3.5}$$

Proof. The simple observation is that

$$f(x) \geq \lambda \chi_{E_\lambda}(x), \quad \text{for all } x \in \mathbb{R}^n, \tag{3.6}$$

where

$$E_\lambda = \{x: \ f(x) > \lambda\}.$$

Integrating (3.6), we obtain (3.5). □

In terms of the probability theory, this can be stated as follows: let X be a random variable, then

$$P(|X| > \lambda) \leq \frac{1}{\lambda} E(|X|),$$

where P is the probability, and $E(|X|)$ is the expected value of $|X|$.

Exercise 3.6 Assume that a sequence of measurable functions f_n satisfies

$$\int |f_n| d\mu \to 0 \text{ as } n \to +\infty.$$

Show that f_n converges to zero in probability.

The Chebyshev inequality allows to estimate how far a random variable deviates from its mean value using the second moment. We say that μ is a probability measure if $\mu(\mathbb{R}^n) = 1$ (or, more generally, $\mu(X) = 1$ where X is the whole measure space).

Theorem 3.7 *Assume that f is an integrable function such that*

$$m_2 := \int (f - \bar{f})^2 d\mu < +\infty,$$

where

$$\bar{f} = \int f d\mu.$$

Then we have

$$\mu(x: \ \ |f - \bar{f}| > \lambda\sqrt{m_2}) \leq \frac{1}{\lambda^2}.$$

Proof. We will apply the Markov inequality to $g(x) = (f(x) - \bar{f})^2$:

$$\mu(x: \ \ |f - \bar{f}|^2 > \lambda^2 m_2) \leq \frac{1}{\lambda^2 m_2} \int (f(x) - \bar{f})^2 d\mu,$$

and we are done. □

Probabilistically, Chebyshev's inequality shows that if the variance of a random variable is small, the probability that it deviates from its mean by a large amount is small.

Exercise 3.8 Let f_n be a sequence of mean-zero functions:

$$\int f_n d\mu = 0,$$

such that

$$\int f_n^2 d\mu \to 0 \text{ as } n \to +\infty.$$

Show that $f_n \to 0$ in probability.

Bounded convergence theorem

We now set to prove several theorems which address the same question: if a sequence $f_n(x)$ converges point-wise to a function $f(x)$, what can we say about the integral of $f(x)$? Let us point out immediately two possible sources of trouble. One example is the sequence of step functions $f_n(x) = \chi_{[n,n+1]}(x)$, and another is the sequence $g_n(x) = n\chi_{[-1/(2n),1/2n]}(x)$. Both $f_n(x)$ and $g_n(x)$ converge point-wise almost everywhere to $f(x) = 0$ but

$$\int_{\mathbb{R}} f_n(x)dx = \int_{\mathbb{R}} g_n(x)dx = 1 \not\to 0 = \int_{\mathbb{R}} f(x)dx.$$

This shows two possible reasons for the integrals of f_n to fail to converge to the integral of $f(x)$: escape to infinity in case of $f_n(x)$ and concentration in the case of $g_n(x)$.

Bounded convergence theorem deals with the situation when neither escape to infinity nor concentration is possible.

Theorem 3.9 *Let f_n be a sequence of measurable functions defined on a measurable set E with $\mu E < +\infty$. Assume that f_n are uniformly bounded: there exists $M > 0$ so that*

$$|f_n(x)| \leq M \text{ for all } n \text{ and all } x \in E.$$

Under these assumptions, if $f_n(x) \to f(x)$ point-wise almost everywhere on E, then

$$\int_E f d\mu = \lim_{n\to\infty} \int_E f_n d\mu. \tag{3.7}$$

Proof. The conclusion is trivial if f_n converges uniformly to f on the set E. In general, given any $\varepsilon > 0$ we may use Egorov's theorem to find a set A_ε such that $\mu(A_\varepsilon) < \varepsilon$, and f_n converges uniformly to f on the set $E \setminus A_\varepsilon$. Then for large enough n we have

$$|f_n(x) - f(x)| < \varepsilon \text{ for all } x \in E \setminus A_\varepsilon,$$

and thus

$$\left|\int_E (f_n - f)d\mu\right| \leq \int_{E\setminus A_\varepsilon} |f_n - f|d\mu + \int_{A_\varepsilon} |f_n - f|d\mu \leq \varepsilon\mu(E) + 2M\mu(A_\varepsilon) \leq (\mu(E) + 2M)\varepsilon,$$

and (3.7) follows. □

Fatou's Lemma

Fatou's lemma tells us that in the limit we may only lose mass, which is exactly what happened in the two examples (concentration and escape to infinity) mentioned at the beginning of this section.

Theorem 3.10 *Let f_n be a sequence of non-negative measurable functions which converges point-wise to a function f on a measurable set E, then*

$$\int_E f d\mu \leq \liminf_{n\to+\infty} \int_E f_n d\mu. \tag{3.8}$$

Proof. Let $h \leq f$ be a bounded non-negative simple function defined on E, which vanishes outside of a set E' with $\mu E' < +\infty$. The sequence $h_n(x) = \min\{h(x), f_n(x)\}$, converges point-wise to $h(x)$: $h_n(x) \to h(x)$ on E. The bounded convergence theorem applies to the sequence h_n on the set E':

$$\int_E h d\mu = \int_{E'} h d\mu = \lim_{n\to\infty} \int_{E'} h_n d\mu \leq \liminf \int_{E'} f_n d\mu \leq \liminf \int_E f_n d\mu.$$

Taking the supremum over all such functions h we arrive to (3.8). □

Exercise 3.11 Show that Fatou's lemma does not generally hold for functions which may take negative values.

The Monotone Convergence Theorem

Fatou's lemma says that you cannot gain mass in the limit. If the sequence f_n is increasing you can hardly lose mass in the limit either.

Theorem 3.12 *Let f_n be a non-decreasing sequence of non-negative measurable functions defined on a measurable set E. Assume that f_n converges point-wise to f almost everywhere on E, then*

$$\int_E f d\mu = \lim_{n\to+\infty} \int_E f_n d\mu.$$

Proof. This is an immediate consequence of Fatou's lemma. □

The monotone convergence theorem has a very simple but useful corollary concerning term-wise Lebesgue integration of a series of non-negative functions.

Corollary 3.13 *Let u_n be a sequence of non-negative measurable functions defined on a measurable set E and let $f(x) = \sum_{n=1}^{\infty} u_n(x)$. Then*

$$\int_E f d\mu = \sum_{n=1}^{\infty} \int_E u_n(x) d\mu.$$

Proof. Apply the monotone convergence theorem to the sequence of partial sums

$$f_n(x) = \sum_{j=1}^{n} u_j(x).$$

The Lebesgue Dominated Convergence Theorem

All the above convergence theorems are part of the Lebesgue dominated convergence theorem.

Theorem 3.14 *Let the functions f_n be measurable and defined on a measurable set E. Assume that $f_n(x) \to f(x)$ almost everywhere on E, and there exists a function $g(x)$ such that*

$$\int_E g(x)d\mu < +\infty,$$

and $|f_n(x)| \leq g(x)$ a.e. on E. Then we have

$$\int_E f d\mu = \lim_{n\to+\infty} \int_E f_n d\mu. \tag{3.9}$$

Proof. As $g - f_n \geq 0$ a.e. on E, Fatous' lemma implies that

$$\int_E (g-f)d\mu \leq \liminf \int_E (g-f_n)d\mu. \tag{3.10}$$

Moreover, as $|f_n| \leq g$, the limit f is integrable, hence it follows from (3.10) that

$$\int_E g d\mu - \int_E f d\mu \leq \int_E g d\mu - \limsup \int_E f_n d\mu,$$

and thus

$$\limsup \int_E f_n d\mu \leq \int_E f d\mu.$$

On the other hand, similarly we know that $g + f_n \geq 0$, which implies

$$\int_E g d\mu + \int_E f d\mu \leq \int_E g d\mu + \liminf \int_E f_n d\mu,$$

and thus

$$\int_E f d\mu \leq \liminf \int_E f_n d\mu.$$

Now, (3.9) follows. □

Absolute continuity of the integral

Proposition 3.15 *Let $f \geq 0$ be a measurable function defined on a measurable set E, and assume that*

$$\int_E f d\mu < +\infty.$$

Then for any $\varepsilon > 0$ there exists $\delta > 0$ so that for any measurable set $A \subseteq E$ with $\mu(A) < \delta$ we have

$$\int_A f d\mu < \varepsilon.$$

Proof. Suppose that this fails. Then there exists $\varepsilon_0 > 0$ and a sequence of sets $A_n \subset E$ so that $\mu(A_n) < 1/2^n$ but

$$\int_{A_n} f d\mu \geq \varepsilon_0.$$

Consider the functions $g_n(x) = f(x)\chi_{A_n}(x)$, then $g_n(x) \to 0$ as $n \to \infty$ except for points x which lie in infinitely many A_n's, that is,

$$x \in A = \bigcap_{n=1}^{\infty} \left(\bigcup_{j=n}^{\infty} A_j \right).$$

However, for any n we have

$$\mu(A) \leq \mu \left(\bigcup_{j=n}^{\infty} A_j \right) \leq \sum_{j=n}^{\infty} \mu(A_j) \leq \frac{1}{2^{n-1}}.$$

It follows that $\mu(A) = 0$ and thus $g_n(x) \to 0$ a.e. on E. Now, set $f_n = f - g_n$, then $f_n \geq 0$ and $f_n \to f$ a.e., so Fatou's lemma can be applied to f_n:

$$\int_E f d\mu \leq \liminf \int_E f_n d\mu \leq \int_E f d\mu - \limsup \int_E g_n d\mu \leq \int_E f d\mu - \varepsilon_0,$$

which is a contradiction. □

4 Differentiation and Integration

We will now address the questions of when the Newton-Leibniz formula

$$\int_a^b f'(x)dx = f(b) - f(a), \tag{4.1}$$

and its twin

$$\frac{d}{dx} \int_a^b f(x)dx = f(x), \tag{4.2}$$

hold. To properly formulate the question, we need the notion of the derivative.

Definition 4.1 *Let f be a real-valued function defined on the real line, then*

$$D^+f(x) = \limsup_{h \downarrow 0} \frac{f(x+h) - f(x)}{h}, \quad D^-f(x) = \limsup_{h \downarrow 0} \frac{f(x) - f(x-h)}{h}$$
$$D_+f(x) = \liminf_{h \downarrow 0} \frac{f(x+h) - f(x)}{h}, \quad D_-f(x) = \liminf_{h \downarrow 0} \frac{f(x) - f(x-h)}{h}.$$

If $D^+f(x) = D^-f(x) = D_+f(x) = D_-f(x) \neq \infty$ then we say that f is differentiable at the point $x \in \mathbb{R}$.

We can now give the answer to the second question.

Theorem 4.2 *Let $f \in L^1[a,b]$ be an integrable function, and*

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

then $F'(x) = f(x)$ a.e.

The answer to the first question requires the notion of absolute continuity.

Definition 4.3 *A function $f : [a,b] \to \mathbb{R}$ is absolutely continuous if for any $\varepsilon > 0$ there exists $\delta > 0$ such that for every finite collection $\{(x_i, x_i')\}$ of non-overlapping intervals with $\sum_{i=1}^n |x_i - x_i'| < \delta$ we have*

$$\sum_{i=1}^n |f(x_i) - f(x_i')| < \varepsilon.$$

With this definition, we can state

Theorem 4.4 *A function $F(x)$ is an indefinite integral, that is, it has the form*

$$F(x) = F(a) + \int_a^x f(t)dt \tag{4.3}$$

with $f \in L^1[a,b]$ if and only if F is absolutely continuous.

The rest of this section contains the proof of these two theorems.

4.1 Differentiation of Monotone Functions

We first show that a monotonic function has a derivative almost everywhere with respect to the Lebesgue measure.

Theorem 4.5 *Let f be an increasing function on an interval $[a,b]$. Then $f'(x)$ exists almost everywhere on $[a,b]$ with respect to the Lebesgue measure and is a measurable function.*

The Vitali lemma

The proof of Theorem 4.5 is not quite as elementary as one would hope, and uses the following covering lemma. Recall that we denote by $m(E)$ the Lebesgue measure of a set $E \subseteq \mathbb{R}$ on the line.

Definition 4.6 *We say that a collection $\mathcal{J}$ of non-trivial closed intervals on the real line forms a fine cover of a set E if for any $\varepsilon > 0$ and any point $x \in E$ there exists an interval I in the collection $\mathcal{J}$ such that $x \in I$ and $m(I) < \varepsilon$.*

Vitali's lemma allows us to find a finite sub-collection of pairwise disjoint balls that covers a very large fraction of a set.

Lemma 4.7 *(Vitali's lemma) Let $E \subset \mathbb{R}$ with $m^*(E) < +\infty$ and let $\mathcal{J}$ be a fine cover of the set E. Then for any $\varepsilon > 0$ there exists a finite subcollection of pairwise disjoint intervals $\{I_1, \ldots, I_N\}$ in $\mathcal{J}$ such that*

$$m^*\left(E \setminus (\bigcup_{j=1}^{N} I_j)\right) < \varepsilon.$$

Proof. Let O be an open set with $m(O) < +\infty$ which contains E: $E \subset O$. Such set exists since $m^*(E) < +\infty$. As O is an open set and $\mathcal{J}$ is a fine cover of E, if we consider the sub-collection $\mathcal{J}'$ of intervals in $\mathcal{J}$ which are contained in O, the new cover $\mathcal{J}'$ is still a fine cover of E. Hence, we may assume from the start that all intervals in $\mathcal{J}$ are contained in O.

Choose any interval I_1 and assume that the intervals $I_1, I_2, \ldots I_n$ have been already chosen. Here is how we choose the interval I_{n+1}. Let k_n be the supremum of the lengths of intervals in $\mathcal{J}$ that do not intersect any of $I_1, I_2, \ldots I_n$. Then $k_n \leq m(O) < +\infty$ and, moreover, if $k_n = 0$ then $E \subset \bigcup_{j=1}^{n} I_j$. Indeed, if $k_n = 0$ and

$$x \in E_n = E \bigcap D_n,\ D_n = (\bigcup_{j=1}^{n} I_j)^c,$$

then, as D_n is open and $\mathcal{J}$ is a fine cover, there exists a non-trivial interval $I \in \mathcal{J}$ such that $I \subset D_n^c$ which contradicts $k_n = 0$. Hence, if $k_n = 0$ for some n then $E \subset \bigcup_{j=1}^{n} I_j$ and we are done.

If $k_n > 0$ for all n, take the interval I_{n+1} disjoint from all of I_j with $1 \leq j \leq n$ such that

$$l(I_{n+1}) \geq k_n/2.$$

This produces a sequence of disjoint intervals I_n such that

$$\sum_n l(I_n) \leq m(O) < +\infty. \tag{4.4}$$

Given $\varepsilon > 0$ find N such that

$$\sum_{j=N+1}^{\infty} l(I_j) < \frac{\varepsilon}{5}$$

and set

$$R = E \setminus \bigcup_{j=1}^{N} I_j.$$

We need to verify that $m^*(R) < \varepsilon$. For any point $x \in R$ there exists an interval $I \in \mathcal{J}$ such that $x \in I$ and I is disjoint from all intervals $\{I_1, I_2, \ldots, I_N\}$. Furthermore, if for some n the interval I is disjoint from intervals $\{I_1, I_2, \ldots I_n\}$ then we have

$$l(I) \leq k_n < 2l(I_{n+1}). \tag{4.5}$$

However, (4.4) implies that $l(I_n) \to 0$ as $n \to +\infty$, thus I must intersect some interval I_n with $n > N$ because of (4.5). Let n_0 be the smallest such n, then

$$l(I) \leq k_{n_0-1} \leq 2l(I_{n_0}).$$

Since $x \in I$ and I intersects I_{n_0}, the distance from x to the midpoint of I_{n_0} is at most

$$l(I) + \frac{l(I_{n_0})}{2} \le \frac{5l(I_{n_0-1})}{2}.$$

Hence, x lies in the interval $\hat{I}_{n_0}$ which has the same midpoint as I_{n_0} and is five times as long as I_{n_0}. Therefore, the set R is covered:

$$R \subseteq \bigcup_{n=N+1}^{\infty} \hat{I}_n,$$

and thus

$$m^*(R) < \sum_{n=N+1}^{\infty} l(\hat{I}_n) \le 5 \sum_{n=N+1}^{\infty} l(I_n) < \varepsilon,$$

and we are done. □

The proof of Vitali's lemma can be easily generalized to dimensions $n > 1$. In particular, we have the following corollary.

Corollary 4.8 *Let $U \subseteq \mathbb{R}^n$ be an open set and $\delta > 0$. There exists a countable collection of disjoint closed balls in U such that $diam B \le \delta$ for all $B \in \mathcal{J}$ and*

$$m\left(U \setminus \bigcup_{B \in \mathcal{J}} B\right) = 0. \tag{4.6}$$

Proof. We first find disjoint closed balls $B_{11}, \dots B_{1,N_1} \subset U$ so that

$$m\left(U \setminus \bigcup_{j=1}^{N_1} B_{1,j}\right) < \frac{m(U)}{3},$$

and set

$$U_1 = U \setminus \bigcup_{j=1}^{N_1} B_j.$$

The set U_1 is also open, and we can find disjoint closed balls $B_{2,1}, \dots, B_{2,N_2} \subset U_1$ so that

$$m\left(U_1 \setminus \bigcup_{j=1}^{N_2} B_{2,j}\right) < \frac{m(U_1)}{3}.$$

Continuing this procedure leads to a sequence of disjoint balls B_n so that (4.6) holds. □

A key point in the proof of Vitali's lemma was the fact that the Lebesgue measure is doubling. This means that there exists a constant $c > 0$ so that for any ball $B(x,r)$ we have a bound $m(B(x,2r)) \le cm(B(x,r))$. Such property is not true in general, for arbitrary measures. A difficult extension of Vitali's lemma and in particular of Corollary 4.8 is the Besikovitch theorem that we will encounter soon which will establish this corollary for non-doubling measures.

Proof of almost everywhere differentiability of monotonic functions

We now prove Theorem 4.5, with the help of Vitali's lemma. We will show that the sets where any pair of derivatives of f are not equal has measure zero. For instance, let

$$E = \{x : \ D^+ f(x) > D^- f(x)\}.$$

We can write E as a countable union:

$$E = \bigcup_{r,s\in\mathbb{Q}} E_{rs}, \quad E_{rs} = \{x : \ D^+ f(x) > r > s > D^- f(x)\},$$

and we will show that

$$m^*(E_{rs}) = 0 \text{ for all } r, s \in \mathbb{Q}.$$

Let $l = m^*(E_{rs})$ and, given $\varepsilon > 0$, enclose E_{rs} in an open set O such that

$$E_{rs} \subseteq O, \text{ with } mO < l + \varepsilon.$$

For each $x \in E_{rs}$ there exists an arbitrary small interval $[x - h, x] \subset O$ such that

$$f(x) - f(x - h) < sh.$$

Using Vitali's lemma, we can choose a finite subcollection $\{I_1, \ldots, I_N\}$ of such disjoint intervals whose interiors cover a set $A = (\bigcup_{n=1}^N I_n^o) \bigcap E_{rs}$, with $l - \varepsilon < m(A) < l + \varepsilon$. It follows that

$$\sum_{n=1}^{N} [f(x_n) - f(x_n - h_n)] < s \sum_{n=1}^{N} h_n < s(l + \varepsilon). \tag{4.7}$$

Next, take any point $y \in A$, then $y \in I_n^o$ for some n, and, as $A \subset E_{rs}$, there exists an arbitrary small interval $[y, y + k] \subset I_n$ such that $f(y + k) - f(y) > rk$. Using Vitali's lemma again, we may choose intervals $\{J_1, \ldots, J_M\}$ such that $J_1, \ldots, J_M \subset \bigcup_{n=1}^N I_n$ and

$$m^*(A \setminus \bigcup_{l=1}^{M} J_l) < \varepsilon.$$

As a consequence,

$$\sum_{n=1}^{M} k_n > m^*(A) - \varepsilon > l - 2\varepsilon,$$

and thus

$$\sum_{n=1}^{M} f(y_n + k_n) - f(y_n) > r \sum_{n=1}^{M} k_n > r(l - 2\varepsilon). \tag{4.8}$$

On the other hand, each interval J_k is contained in some interval I_p and f is increasing so that for each p:

$$\sum_{J_k \subset I_p} (f(y_k + h_k) - f(y_k)) \leq f(x_p) - f(x_p - h_p).$$

Summing over p, and taking into account (4.7) and (4.8), we conclude that $s(l+\varepsilon) \geq r(l-2\varepsilon)$. As this is true for all $\varepsilon > 0$, and $r > s$, it follows that $l = 0$, so that $m^*(E_{rs}) = 0$ for all $r, s \in \mathbb{Q}$, and thus $m^*(E) = 0$.

Now that we know that $f'(x)$ exists a.e., let us show that $f'(x)$ is a measurable function. Let us extend $f(x) = f(b)$ for $x \geq b$ and set

$$g_n(x) = n\left[f(x+\frac{1}{n}) - f(x)\right]. \tag{4.9}$$

Then

$$f'(x) = \lim_{n\to\infty} g_n(x) \tag{4.10}$$

almost everywhere, and thus $f'(x)$ is measurable (as a limit of a sequence of measurable functions). □

Integral of a derivative of a monotone function

We are now ready to establish the Newton-Leibnitz inequality for monotone functions.

Theorem 4.9 *Let $f(x)$ be an increasing function on an interval $[a,b]$, then $f'(x)$ is finite almost everywhere on $[a,b]$, and*

$$\int_a^b f'(x)dx \leq f(b) - f(a). \tag{4.11}$$

Proof. The function $f'(x)$ is measurable according to Theorem 4.5, hence the integral in the left side of (4.11) is well defined. Let us define the approximations $g_n(x)$ by (4.9), once again with the convention $f(x) = f(b)$ for $x > b$, then $g_n(x) \geq 0$, thus $f'(x) \geq 0$ by (4.10), and, moreover, Fatou's lemma implies that

$$\int_a^b f'(x)dx \leq \liminf_{n\to\infty} \int_a^b g_n(x)dx = \liminf_{n\to\infty} \int_a^b n\left[f(x+\frac{1}{n}) - f(x)\right]dx$$
$$= \liminf_{n\to\infty}\left[n\int_b^{b+1/n} f(b)dx - n\int_a^{a+1/n} f(x)dx\right] \leq \liminf_{n\to\infty}\left[n\int_b^{b+1/n} f(b)dx - n\int_a^{a+1/n} f(a)dx\right]$$
$$= f(b) - f(a),$$

and (4.11) follows. As a consequence of (4.11) we also conclude that $f'(x)$ is finite a.e. □

4.2 Differentiability of functions of bounded variation

Let $a = x_0 < x_1 < \ldots < x_{m-1} < x_m = b$ be a partition of an interval $[a,b]$. For a fixed partition we define

$$p = \sum_{k=1}^{m}[f(x_k) - f(x_{k-1})]_+, \quad n = \sum_{k=1}^{m}[f(x_k) - f(x_{k-1})]_-, \quad t = n + p = \sum_{k=1}^{n}|f(x_k) - f(x_{k-1})|.$$

The total variation of a function f over an interval $[a,b]$ is $T_a^b[f] = \sup t$, where supremum is taken over all partitions on $[a,b]$. Similarly, we define $N_a^b[f] = \sup n$ and $P_a^b[f] = \sup p$.

Definition 4.10 *We say that f has a bounded total variation on $[a,b]$ and write $f \in BV[a,b]$ if $T_a^b[f] < +\infty$.*

The simplest example of function of bounded variation is a monotonic function on $[a,b]$, as $T_a^b[f] = |f(b) - f(a)|$ for monotonic functions. It turns out that all functions in $BV[a,b]$ are a difference of two monotonic functions.

Theorem 4.11 *A function f has a bounded variation on an interval $[a,b]$ if and only if f is a difference of two monotonic functions.*

Proof. (1) Assume that $f \in BV[a,b]$. We claim that

$$f(x) - f(a) = P_a^x[f] - N_a^x[f]. \tag{4.12}$$

Indeed, for any partition $a = x_0 < x_1, \ldots < x_m = x$ we have

$$p = n + f(x) - f(a) \leq N_a^x[f] + f(x) - f(a),$$

so that

$$P_a^x[f] \leq N_a^x[f] + f(x) - f(a).$$

Similarly, one shows that

$$N_a^x[f] \leq P_a^x[f] - (f(x) - f(a)),$$

and (4.12) follows. It remains to notice that both functions $u(x) = P_a^x[f]$ and $v(x) = N_a^x[f]$ are non-decreasing to conclude that any BV function is a difference of two monotonic functions.

(2) On the other hand, if $f(x)$ is a difference of two monotonic functions:

$$f(x) = u(x) - v(x),$$

then for any partition of the interval (a,b) we have

$$\sum_{i=1}^{n} |f(x_i) - f(x_{i-1})| \leq \sum_{i=1}^{n} |u(x_i) - u(x_{i-1})| + \sum_{i=1}^{n} |v(x_i) - v(x_{i-1})|$$
$$= \sum_{i=1}^{n} (u(x_i) - u(x_{i-1})) + \sum_{i=1}^{n} (v(x_i) - v(x_{i-1})) = u(b) - u(a) + v(b) - v(a),$$

so that $f \in BV[a,b]$. □

An immediate consequence of Theorem 4.11 is the following observation.

Corollary 4.12 *If a function f has bounded variation on an interval $[a,b]$ then $f'(x)$ exists a.e. on $[a,b]$.*

Differentiation of an integral

Now is the time to prove Theorem 4.2, which is

Theorem 4.13 *Let $f \in L^1[a,b]$ be an integrable function, and*

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

then $F'(x) = f(x)$ a.e.

Proof. First, Proposition 3.15 implies that the function $F(x)$ is continuous. Moreover, F has bounded variation on $[a,b]$ since for any partition of $[a,b]$ we have

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

We need the following basic lemma.

Lemma 4.14 *If $f \in L^1[a,b]$ is integrable and*

$$\int_a^x f(s)ds = 0 \tag{4.13}$$

for all $x \in [a,b]$ then $f(t) = 0$ a.e. on $[a,b]$.

Proof of Lemma 4.14. Suppose that $f(x) > 0$ on a set E with $mE > 0$. Then there exists a compact set $F \subset E$ such that $mF > 0$. Let $O = [a,b] \setminus F$, then

$$0 = \int_a^b f(t)dt = \int_F f(t)dt + \int_O f(t)dt.$$

It follows that

$$\int_O f(t)dt < 0,$$

and thus, as O is a disjoint union of open intervals, there exists an interval $(\alpha, \beta) \subset O$ such that

$$\int_\alpha^\beta f(t)dt < 0,$$

which contradicts (4.13). □

We continue the proof of Theorem 4.13. Let us first assume that the function f is bounded: $|f(x)| \le K$ for all $x \in [a,b]$. As we already know that F has bounded variation, the derivative $F'(x)$ exists a.e. on $[a,b]$ and we only need to show that $F'(x) = f(x)$ a.e. Consider the approximations of $F'(x)$:

$$f_n(x) = \frac{F(x + 1/n) - F(x)}{1/n} = n \int_x^{x+1/n} f(x)dx.$$

These functions are uniformly bounded: $|f_n(x)| \leq K$ and $f_n(x) \to F'(x)$ a.e. The bounded convergence theorem implies that for all $x \in [a, b]$ we have

$$\int_a^x F'(t)dt = \lim_{n\to\infty} \int_a^x f_n(t)dt = \lim_{n\to\infty} \left[n \int_x^{x+1/n} F(t)dt - n \int_a^{a+1/n} F(t)dt \right] = F(x) - F(a).$$

The last step above follows from the continuity of the function $F(t)$. Now, Lemma 4.14 implies that $F'(x) = f(x)$ a.e. on $[a, b]$.

Finally, consider the situation when $f \in L^1[a, b]$ but is maybe unbounded. Without loss of generality we may assume that $f \geq 0$. Consider the cut-offs $g_n(x) = \min\{f(x), n\}$. Then $f - g_n \geq 0$, thus the functions

$$G_n(x) = \int_a^x (f - g_n))dt$$

are increasing, hence $G_n'(x) \geq 0$ a.e. As the functions g_n are bounded for each n fixed, we know from the first part of the proof that

$$\frac{d}{dx} \int_a^x g_n(t)dt = g_n(x)$$

almost everywhere. It follows that $F'(x) = G_n'(x) + g_n(x) \geq g_n(x)$ and, in particular, $F'(x)$ exists almost everywhere. Passing to the limit $n \to \infty$ we deduce that $F'(x) \geq f(x)$ a.e. which, in turn, implies that

$$\int_a^b F'(x)dx \geq \int_a^b f(x)dx = F(b) - F(a).$$

However, as $f \geq 0$, the function F is non-decreasing and thus

$$\int_a^b F'(x)dx \leq F(b) - F(a).$$

Together, the last two inequalities imply that

$$\int_a^b F'(x)dx = F(b) - F(a) = \int_a^b f(t)dt.$$

Since $F'(x) \geq f(x)$ a.e. we conclude that $F'(x) = f(x)$ a.e. □

4.3 The Newton-Leibniz formula for absolutely continuous functions

We will now prove Theorem 4.4, the Newton-Leibniz formula for absolutely continuous functions. A simple observation is that Proposition 3.15 implies that every indefinite integral

$$F(x) = F(a) + \int_a^x f(t)dt \tag{4.14}$$

with $f \in L^1[a, b]$ is absolutely continuous. Our goal is to show that every absolutely continuous function is the indefinite integral of its derivative, that is, the Newton-Leibniz formula holds for absolutely continuous functions. We restate Theorem 4.4.

Theorem 4.15 *A function $F(x)$ is an indefinite integral, that is, it has the form (4.14) with $f \in L^1[a,b]$ if and only if F is absolutely continuous.*

Proof. As we have mentioned, absolute continuity of the indefinite integral follows immediately from Proposition 3.15. Now, let $F(x)$ be absolutely continuous, then, as we have noted above, F has bounded variation on $[a,b]$ and thus can be written as $F(x) = F_1(x) - F_2(x)$, where both of the functions F_1 and F_2 are increasing. Hence, $F'(x)$ exists a.e. and

$$|F'(x)| \leq F_1'(x) + F_2'(x),$$

so that

$$\int_a^b |F'(x)|dx \leq \int_a^b F_1'(x)dx + \int_a^b F_2'(x)dx \leq F_1(b) - F_1(a) + F_2(b) - F_2(a),$$

thus $F'(x)$ is integrable on $[a,b]$. Consider its anti-derivative

$$G(x) = \int_a^x F'(t)dt,$$

then $G(x)$ is absolutely continuous and $G'(x) = F'(x)$ a.e. as follows from Theorem 4.2.

Set $R(x) = F(x) - G(x)$, then $R(x)$ is absolutely continuous and $R'(x) = 0$ a.e. Let us show that $R(x)$ is actually a constant (and thus is equal identically to $F(a)$). This will finish the proof of Theorem 4.4. To this end, we take a point $c \in [a,b]$ and consider the set $A \subseteq [a,c]$ of measure $m(A) = c - a$ such that $R'(x) = 0$ on A. Given $\varepsilon > 0$, for any $x \in A$ and every $n < N(x)$ we choose $h_n(x) < 1/n$ so that

$$|R(x + h_n(x)) - R(x)| < \varepsilon h_n(x). \tag{4.15}$$

This produces a fine covering of A by intervals of the form $I_n(x) = [x, x + h_n(x)]$. Vitali's lemma allows us to find a finite collection $I_k(x_k) = [x_k, y_k]$, $k = 1, \ldots, N$ which covers a set of measure $(c - a - \delta(\varepsilon)/2)$, where $\delta(\varepsilon)$ is δ in the definition of absolute continuity of the function $R(x)$ corresponding to ε, that is, if we set $y_0 = a$ and $x_{N+1} = c$, we have

$$\sum_{k=0}^{N} |x_{k+1} - y_k| < \delta. \tag{4.16}$$

Then, we can estimate, using (4.15) and (4.16):

$$|R(c) - R(a)| \leq \sum_{k=1}^{N} |R(y_k) - R(x_k)| + \sum_{k=1}^{N} |R(x_{k+1}) - R(y_k)| \leq \varepsilon(b-a) + \varepsilon.$$

As $\varepsilon > 0$ is arbitrary, we deduce that $R(x) = R(a)$ for all $c \in [a,b]$. □

A common way to re-phrase Theorem 4.4 is to say that every absolutely continuous function is the integral of its derivative – this identifies functions which satisfy the Newton-Leibniz formula.

5 Product measures and Fubini's theorem

We will now discuss the product measures and the Fubini theorem that relates the iterated and double integrals. Recall the following exercise from multi-variable calculus.

Exercise 5.1 Let $S = [0,1] \times [0,1]$, find a function $f(x,y)$ such that the double integrals

$$\int_0^1 \Big(\int_0^1 f(x,y)dx \Big) dy \text{ and } \int_0^1 \Big(\int_0^1 f(x,y)dy \Big) dx$$

exist but are not equal to each other.

This brings up the question how one can ensure that the equality

$$\int_S f(x,y)dxdy = \int_0^1 \Big(\int_0^1 f(x,y)dx \Big) dy = \int_0^1 \Big(\int_0^1 f(x,y)dy \Big) dx$$

holds, or, more generally, when one can interchange the order of integration for more general measures than the Lebesgue measure. For that, we need first to define the notion of a product measure. The definition is motivated by high school geometry.

Definition 5.2 *Let μ be a measure on a set X and ν a measure on Y, then the outer product measure $\mu \times \nu$ of a set $S \subset X \times Y$ is*

$$(\mu \times \nu)^*(S) = \inf \left(\sum_{j=1}^{\infty} \mu(A_j)\nu(B_j) \right),$$

with the infimum taken over all sets $A_j \subset X$, $B_j \subset Y$ such that $S \subset \bigcup_{j=1}^{\infty}(A_j \times B_j)$.

Our goal in this section is to prove basic statements familiar from the calculus course regarding the connection between the iterated integrals and integrals over the product measure.

Let $\mathcal{F}$ be the collection of sets $S \subseteq X \times Y$ for which the iterated integral can be defined, that is, the characteristic function $\chi_S(x,y)$ is μ-measurable for ν-a.e. $y \in Y$ and the function

$$s(y) = \int_X \chi_S(x,y)d\mu(x)$$

is ν-measurable. For each set $S \in \mathcal{F}$ we define

$$\rho(S) = \int_Y s(y)d\nu(y) = \int_Y \left[\int_X \chi_S(x,y)d\mu(x) \right] d\nu(y).$$

Note that if $U \subseteq V$ and $U, V \in \mathcal{F}$ then $\rho(U) \le \rho(V)$ simply because $\chi_U(x,y) \le \chi_V(x,y)$. Our eventual goal is to show that $\mathcal{F}$ includes all $\mu \times \nu$-measurable sets and that $(\mu \times \nu)(S) = \rho(S)$ for such sets. The first trivial observation in this direction is that all sets of the form $A \times B$, with a μ-measurable set A and a ν-measurable set B, are in $\mathcal{F}$ and

$$\rho(A \times B) = \int_B \mu(A)d\nu(y) = \mu(A)\nu(B).$$

From the way area is defined in elementary geometry we know that the next level of complexity should be countable unions of such sets:

$$\mathcal{P}_1 = \left\{ \bigcup_{j=1}^{\infty} (A_j \times B_j) : \quad A_j \subset X \text{ is } \mu\text{-measurable, and } B_j \subset Y \text{ is } \nu\text{-measurable} \right\}.$$

Note that every set $S = \bigcup_{j=1}^{\infty}(A_j \times B_j) \in \mathcal{P}_1$ is in $\mathcal{F}$. The point is that, using further subdivision of A_j and B_j, such S can be written as a disjoint countable union with

$$(A_j \times B_j) \cap (A_n \times B_n) = \emptyset \text{ for } j \neq n.$$

Then, for each y the cross-section $\{x : \ (x, y) \in S\}$ is an at most countable union of μ-measurable disjoint sets, and

$$\int_X \chi_S(x,y) d\mu_x = \int_X \sum_{j=1}^{\infty} \chi_{A_j}(x)\chi_{B_j}(y) d\mu_x = \sum_{j=1}^{\infty} \mu(A_j)\chi_{B_j}(y)$$

is an ν-integrable function, thus $S \in \mathcal{F}$. Moreover, if $S = \bigcup_{j=1}^{\infty}(A_j \times B_j) \in \mathcal{P}_1$ is a disjoint union then

$$\rho(S) = \sum_{j=1}^{\infty} \mu(A_j)\nu(B_j).$$

Next, we note that for each set $U \subset X \times Y$ its outer measure can be approximated as in elementary geometry:

$$(\mu \times \nu)^*(U) = \inf\{\rho(S) : \ U \subseteq S, \ S \in \mathcal{P}_1\}. \tag{5.1}$$

Indeed, this is somewhat tautological: if $U \subseteq S = \bigcup_{j=1}^{\infty}(A_j \times B_j) \in \mathcal{P}_1$ then (note that the sets $A_j \times B_j$ need not be pairwise disjoint here)

$$\rho(S) = \int_Y \left(\int_X \chi_S(x,y) d\mu_x \right) d\nu_y \leq \int_Y \left(\int_X \sum_{j=1}^{\infty} \chi_{A_j}(x)\chi_{B_j}(y) d\mu_x \right) d\nu_y = \sum_{j=1}^{\infty} \mu(A_j)\nu(B_j).$$

As $(\mu \times \nu)^*(U)$ is the infimum of all possible right sides above, by the definition of the product measure we have

$$\inf \rho(S) \leq (\mu \times \nu)^*(U).$$

On the other hand, any such S can be written as a disjoint union and then

$$\rho(S) = \sum_{j=1}^{\infty} \mu(A_j)\nu(B_j) \geq (\mu \times \nu)^*(U).$$

again by the definition of the product measure. Hence, (5.1) holds. Now, we can show that a product of two measurable sets is measurable.

Proposition 5.3 *Let a set $A \subseteq X$ be μ-measurable and a set $B \subseteq Y$ be ν-measurable. Then the set $A \times B \subset X \times Y$ is $\mu \times \nu$ measurable.*

Proof. Take a set $S = A \times B$ such that A is μ-measurable and B is ν-measurable. Then S is in $\mathcal{P}_0$, thus in $\mathcal{P}_1$ so that

$$(\mu \times \nu)^*(S) \leq \mu(A)\nu(B) = \rho(S) \leq \rho(R)$$

for all $R \in \mathcal{P}_1$ containing S. It follows from (5.1) that $(\mu \times \nu)^*(A \times B) = \mu(A)\nu(B)$.

Let us show that $A \times B$ is $\mu \times \nu$-measurable. Take any set $T \subseteq X \times Y$ and a $\mathcal{P}_1$-set R containing T. Then the sets $R \cap (A \times B)^c$ and $R \cap (A \times B)$ are both disjoint and in $\mathcal{P}_1$. Hence,

$$(\mu \times \nu)^*(T \cap (A \times B)^c) + (\mu \times \nu)^*(T \cap (A \times B)) \leq \rho(R \cap (A \times B)^c) + \rho(R \cap (A \times B)) = \rho(R),$$

because if R and Q are in $\mathcal{P}_1$, $R \cap Q = \emptyset$ then $\rho(R \cup Q) = \rho(R) + \rho(Q)$. Taking infimum over all such R and using (5.1) we arrive to

$$(\mu \times \nu)(T \cap (A \times B)^c) + (\mu \times \nu)(T \cap (A \times B)) \leq (\mu \times \nu)(T),$$

and thus $A \times B$ is a measurable set. $\square$

Once again, following the motivation from approximating areas in elementary geometry we define sets that are countable intersections of those in $\mathcal{P}_1$:

$$\mathcal{P}_2 = \Big\{ \bigcap_{j=1}^{\infty} S_j, \ S_j \in \mathcal{P}_1 \Big\}.$$

Proposition 5.4 *For each set $S \subseteq X \times Y$ there exists a set $R \in \mathcal{P}_2 \cap \mathcal{F}$ such that $S \subseteq R$ and $\rho(R) = (\mu \times \nu)^*(S)$.*

Proof. If $(\mu \times \nu)^*(S) = +\infty$, it suffices to take $R = X \times Y$, so we may assume that

$$(\mu \times \nu)^*(S) < +\infty$$

without loss of generality. Using (5.1), choose the sets $R_j \in \mathcal{P}_1$ such that $S \subseteq R_j$ and

$$\rho(R_j) < (\mu \times \nu)^*(S) + \frac{1}{j}.$$

Consider the sets $R = \bigcap_{j=1}^{\infty} R_j \in \mathcal{P}_2$ and $Q_k = \bigcap_{j=1}^{k} R_j$ and note that

$$\chi_R(x, y) = \lim_{k \to \infty} \chi_{Q_k}(x, y).$$

As each $R_j \in \mathcal{F}$, the functions

$$\chi_{Q_k}(x, y) = \chi_{R_1}(x, y) \ldots \chi_{R_k}(x, y)$$

are μ-measurable functions of x for ν-a.e. y. Therefore, there exists a set $S_0 \subset Y$ of full ν-measure such that $\chi_R(x, y)$ is μ-measurable for each $y \in S_0$ fixed. Moreover, as $\rho(R_1) < +\infty$ (so that for ν-a.e y the function $\chi_{R_1}(x, y)$ is μ-integrable) and $\chi_{Q_k}(x, y) \leq \chi_{R_1}(x, y)$, we have for ν-a.e. y

$$\rho_R(y) = \int_X \chi_R(x, y) d\mu(x) = \lim_{k \to \infty} \rho_k(y), \quad \rho_k(y) = \int_X \chi_{Q_k}(x, y) d\mu(x),$$

and thus $\rho_R(y)$ is ν-integrable and $R \in \mathcal{F}$. As $\rho_k(y) \le \rho_1(y)$, it also follows that

$$\rho(R) = \int_Y \rho_R(y) d\nu(y) = \int_Y \lim_{k\to\infty} \rho_k(y) d\nu(y) = \lim_{k\to\infty} \int_Y \rho_k(y) d\nu(y) = \lim_{k\to\infty} \rho(Q_k). \tag{5.2}$$

However, (5.2) implies that

$$\rho(R) = \lim_{k\to\infty} \rho(Q_k) \le (\mu \times \nu)^*(S).$$

On the other hand, since $S \subseteq Q_k$ we know that $(\mu \times \nu)^*(S) \le \rho(Q_k)$ and thus $\rho(R) = (\mu \times \nu)^*(S)$. □

Corollary 5.5 *The measure $\mu \times \nu$ is regular even if μ and ν are not regular.*

Proof. Proposition 5.3 implies that each set in $\mathcal{P}_2$ is measurable, while Proposition 5.4 implies that for $S \in \mathcal{P}_2$ we have $(\mu \times \nu)(S) = \rho(S)$. The same proposition implies then that the measure $\mu \times \nu$ is regular. □

Definition 5.6 *A set X is σ-finite if $X = \bigcup_{j=1}^\infty B_k$ and the sets B_k are μ-measurable with $\mu(B_k) < +\infty$.*

Theorem 5.7 *(Fubini) Let a set $S \subseteq X \times Y$ be σ-finite with respect to the measure $\mu \times \nu$. Then the cross-section $S_y = \{x : (x, y) \in S\}$ is μ-measurable for ν-a.e. y, the cross-section $S_x = \{x : (x, y) \in S\}$ is ν-measurable for μ-a.e. x, $\mu(S_y)$ is a ν-measurable function of y, and $\nu(S_x)$ is a μ-measurable function of x. Moreover,*

$$(\mu \times \nu)(S) = \int_Y \mu(S_y) d\nu_y = \int_X \nu(S_x) d\mu_x. \tag{5.3}$$

Proof. If $(\mu \times \nu)(S) = 0$ then there exists a set $R \in \mathcal{P}_2$ such that $S \subseteq R$ and $\rho(R) = 0$. Since $\chi_S(x, y) \le \chi_R(x, y)$ it follows that $S \in \mathcal{F}$ and $\rho(S) = 0$.

Now, let $S \subset X \times Y$ be $\mu \times \nu$-measurable and $(\mu \times \nu)(S) < +\infty$. Then there exists $R \in \mathcal{P}_2$, such that $S \subseteq R$ and $(\mu \times \nu)(R \setminus S) = 0$, thus, by the above argument, $\rho(R \setminus S) = 0$. This means that

$$\mu(x : (x, y) \in S) = \mu(x : (x, y) \in R)$$

for ν-a.e. y and thus, as $R \in \mathcal{P}_2$ implies $(\mu \times \nu)(R) = \rho(R)$,

$$(\mu \times \nu)(S) = (\mu \times \nu)(R) = \rho(R) = \int_Y \mu(x : (x, y) \in R) d\nu = \int_Y \mu(x : (x, y) \in S) d\nu,$$

which is (5.3).

Finally, assume that S is a σ-finite set and $(\mu \times \nu)(S) = +\infty$. Then S can be written as a countable union $S = \bigcup_{j=1}^\infty B_j$ of $(\mu \times \nu)$-measurable sets B_j with $(\mu \times \nu)(B_j) < +\infty$. We may assume without loss of generality that all B_j are pairwise disjoint so that by what we have just proved

$$(\mu \times \nu)(S) = \sum_{j=1}^\infty (\mu \times \nu)(B_j) = \sum_{j=1}^\infty \int_Y \mu(x : (x, y) \in B_j) d\nu = \int_Y \sum_{j=1}^\infty \mu(x : (x, y) \in B_j) d\nu$$

$$= \int_Y (\mu : (x, y) \in \bigcup_{j=1}^\infty B_j) d\nu = \int_Y \mu(x : (x, y) \in S) d\nu,$$

so that the claim holds also for such σ-finite sets S. $\square$

Fubini's theorem has a corollary also known as Fubini's theorem.

Corollary 5.8 *Let $X \times Y$ be σ-finite. If $f(x,y)$ is $(\mu \times \nu)$-integrable then the function*

$$p(y) = \int_X f(x,y) d\mu(x)$$

is ν-integrable, the function

$$q(x) = \int_Y f(x,y) d\nu(y)$$

is μ-measurable and

$$\int_{X \times Y} f d(\mu \times \nu) = \int_Y p(y) d\nu(y) = \int_X q(x) d\mu(x). \tag{5.4}$$

Proof. This follows immediately from Theorem 5.7 if $f(x,y) = \chi_S(x,y)$ with a $(\mu \times \nu)$-measurable set S. If $f \geq 0$ use Theorem 2.7 to write

$$f(x,y) = \sum_{k=1}^{\infty} \frac{1}{k} \chi_{A_k}(x,y)$$

and then use Corollary 3.13 to integrate this relation term-wise leading both to

$$\int_Y f(x,y) d\nu(y) = \sum_{k=1}^{\infty} \frac{1}{k} \nu(y : (x,y) \in A_k),$$

if we integrate only in y, and also to

$$\begin{aligned} \int_{X \times Y} f d(\mu \times \nu) &= \sum_{k=1}^{\infty} \frac{1}{k} (\mu \times \nu)(A_k) = \sum_{k=1}^{\infty} \frac{1}{k} \int_X \nu(y : (x,y) \in A_k) d\mu(x) \\ &= \int_X \left(\int_Y f(x,y) d\nu(y) \right) d\mu(x), \end{aligned}$$

which is (5.4). $\square$

6 The Radon-Nikodym theorem

6.1 The Besicovitch theorem

The Besicovtich theorem is a tool to study measures μ on $\mathbb{R}^n$ which do not have the doubling property. The idea is to bypass having to control the measure $\mu(\hat{B})$ (recall that $\hat{B}$ is a ball concentric with B and with a radius that is five times the radius of B) in terms of $\mu(B)$ as in the proof of Vitali's lemma. Here, the doubling property means the following: there exists a constant $C > 0$ so that for any $x \in \mathbb{R}^n$ and $r > 0$ we have

$$\frac{1}{C} \mu(B(x,2r)) \leq \mu(B(x,r)) \leq C \mu(B(x,2r)).$$

In dealing with measures which may not have this property the following theorem is extremely helpful

Theorem 6.1 *(The Besicovitch theorem.) There exists a constant $N(n)$ depending only on the dimension with the following property: if $\mathcal{F}$ is any collection of closed balls in $\mathbb{R}^n$ with*

$$D = \sup\left\{diam\bar{B} |\ \bar{B} \in \mathcal{F}\right\} < +\infty$$

and A is the set of centers of balls $\bar{B} \in \mathcal{F}$ then there exist $\mathcal{J}_1, \mathcal{J}_2, \ldots, \mathcal{J}_{N(n)}$ such that each $\mathcal{J}_k$ is a countable collection of disjoint balls in $\mathcal{F}$ and

$$A \subset \bigcup_{j=1}^{N(n)} \bigcup_{\bar{B} \in \mathcal{J}_j} \bar{B}.$$

The key point here is that we do not have to stretch the balls as in the corollaries of Vitali's lemma – the price to pay is that we have several collections $\mathcal{J}_1, \mathcal{J}_2, \ldots, \mathcal{J}_{N(n)}$, and a ball from a collection $\mathcal{J}_i$ may intersect a ball from another collection $\mathcal{J}_j$ if $i \neq j$. However, this is not that important since the number $N(n)$ is a universal constant depending only on the dimension n.

The following corollary of the Besicovitch theorem is very useful. Note that it does not require the set of centers of the balls to be measurable.

Corollary 6.2 *Let μ be a Borel regular measure on $\mathbb{R}^n$ and $\mathcal{F}$ any collection of non-degenerate closed balls. Let A denote the set of centers of the balls in $\mathcal{F}$. Assume that $\mu^*(A) < +\infty$ and $\inf\{r :\ \bar{B}(a,r) \in \mathcal{F}\} = 0$ for all $a \in A$. Then for each open set $U \subset \mathbb{R}^n$ there exists a countable collection $\mathcal{J}$ of pairwise disjoint balls in $\mathcal{F}$ such that $\bigcup_{\bar{B} \in \mathcal{J}} \bar{B} \subseteq U$ and*

$$\mu((A \cap U) \setminus \bigcup_{\bar{B} \in \mathcal{J}} \bar{B}) = 0. \tag{6.1}$$

Proof. Let $N(n)$ be the number of required collections in the Besicovitch theorem and take $\theta = 1 - 1/(2N(n))$. Then, using this theorem, we may find a countable collection $\mathcal{J}$ of disjoint balls in $\mathcal{F}_1 = \{\bar{B} :\ \bar{B} \in \mathcal{F},\ \bar{B} \subset U,\ \mathrm{diam}\bar{B} \leq 1\}$ such that

$$\mu^*\left((A \cap U) \cap (\bigcup_{\bar{B} \in \mathcal{J}} \bar{B})\right) \geq \frac{1}{N(n)} \mu^*(A \cap U).$$

Therefore, using the increasing sets theorem in the form of Proposition 1.39 (which holds for sets that may be not measurable) we may choose a finite sub-collection $\bar{B}_1, \ldots, \bar{B}_{M_1}$ of $\mathcal{J}$ such that

$$\mu^*\left((A \cap U) \cap (\bigcup_{j=1}^{M_1} \bar{B}_j)\right) \geq \frac{1}{2N(n)} \mu^*(A \cap U).$$

It follows that (since $\bigcup_{j=1}^{M_1} \bar{B}_j$ is a measurable set)

$$\mu\left((A \cap U) \setminus (\bigcup_{j=1}^{M_1} \bar{B}_j)\right) \leq \left(1 - \frac{1}{2N(n)}\right) \mu^*(A \cap U).$$

Applying the same reasoning to the set $U_2 = U \setminus \left(\bigcup_{j=1}^{M_1} \bar{B}_j\right)$ and the collection

$$\mathcal{F}_2 = \{\bar{B} : \ \bar{B} \in \mathcal{F}, \ \bar{B} \subset U_2, \ \text{diam}\bar{B} \le 1\}$$

we get a finite set of balls $\bar{B}_{M_1+1}, \ldots, \bar{B}_{M_2}$ such that

$$\mu^* \left((A \cap U_2) \setminus (\bigcup_{j=M_1+1}^{M_2} \bar{B}_j) \right) \le \left(1 - \frac{1}{2N(n)}\right) \mu^*(A \cap U_2).$$

It follows that

$$\mu^* \left((A \cap U) \setminus (\bigcup_{j=1}^{M_2} \bar{B}_j) \right) = \mu^* \left((A \cap U_2) \setminus (\bigcup_{j=M_1+1}^{M_2} \bar{B}_j) \right) \le \left(1 - \frac{1}{2N(n)}\right) \mu(A \cap U_2)$$
$$\le \left(1 - \frac{1}{2N(n)}\right)^2 \mu^*(A \cap U).$$

Continuing this procedure, for each k we obtain a finite collection of balls $\bar{B}_1, \ldots, \bar{B}_{M_k}$ so that

$$\mu^* \left((A \cap U) \setminus (\bigcup_{j=1}^{M_k} \bar{B}_j) \right) \le \left(1 - \frac{1}{2N(n)}\right)^k \mu^*(A \cap U).$$

Then the collection $\mathcal{J} = \{\bar{B}_1, \bar{B}_2, \ldots, \bar{B}_k, \ldots\}$ satisfies (6.1). $\square$

6.2 The proof of the Besicovitch theorem

The proof of this theorem proceeds in several technical steps. Step 1 is to reduce the problem to the situation when the set A of the centers is bounded. Step 2 is to choose the balls $\bar{B}_1, \bar{B}_2, \ldots, \bar{B}_n, \ldots$– this procedure is quite similar to that in Vitali's lemma but not identical. Step 3 is to show that the balls we have chosen cover the set A. The last step is to to show that the balls $\bar{B}_j$ can be split into $N(n)$ separate sub-collections $\mathcal{J}_k$, $k = 1, \ldots, N(n)$ such that each $\mathcal{J}_k$ itself is a collection of pair-wise disjoint balls. For that one has to estimate how many of the balls $\bar{B}_1, \ldots, \bar{B}_{k-1}$ the ball $\bar{B}_k$ intersects – it turns out that this number depends only on the dimension (and not on k, the set A or anything else) and that is the number $N(n)$ we are looking for. The crux of the matter is in this estimate and it is not trivial.

Reduction to counting the number of balls a given ball $\bar{B}_k$ may intersect

Let us first explain why our main interest is in estimating how many of the "preceding" balls $\bar{B}_1, \ldots, \bar{B}_{k-1}$ the ball $\bar{B}_k$ intersects.

Lemma 6.3 *Let $\bar{B}_1, \bar{B}_2, \ldots, \bar{B}_n, \ldots$ be a countable collection $\mathcal{F}$ of balls in $\mathbb{R}^n$. Assume that there exists $M > 0$ so that each ball $\bar{B}_n$ intersects at most M balls out of $\{\bar{B}_1, \bar{B}_2, \ldots, \bar{B}_{n-1}\}$.*

Then the collection $\mathcal{F}$ can be split into $(M+1)$ sub-collections $\mathcal{J}_1, \mathcal{J}_2, \ldots, \mathcal{J}_{M+1}$ so that each $\mathcal{J}_m$ is a collection of pair-wise disjoint balls and

$$\bigcup_{\bar{B}\in\mathcal{F}} \bar{B} = \bigcup_{j=1}^{M+1} \bigcup_{\bar{B}\in\mathcal{J}_j} \bar{B}.$$

Proof. Let us prepare $M+1$ "baskets" $\mathcal{J}_1, \mathcal{J}_2, \ldots, \mathcal{J}_{M+1}$. We put $\bar{B}_k$ into these baskets in the following way: $\bar{B}_1$ goes into the basket $\mathcal{J}_1$, $\bar{B}_2$ into $\mathcal{J}_2$, and so on until $\bar{B}_{M+1}$ which goes into $\mathcal{J}_{M+1}$. After that we proceed as follows: assume the balls $\bar{B}_1, \ldots, \bar{B}_{k-1}$ were already put into baskets. Take the ball $\bar{B}_k$ – by assumption only M out the $M+1$ baskets may contain a ball $\bar{B}_j$, $j = 1, \ldots, k-1$ which intersects $\bar{B}_k$. Hence at least one basket contains no balls which intersect $\bar{B}_k$ – this is the basket that $\bar{B}_k$ is put in (if there are several such baskets we just put $\bar{B}_k$ into one of such baskets, it does not matter which one). Then we go to the next ball $\bar{B}_{k+1}$, and so on. □

Reduction to a bounded set of centers A

Assume that we have proved the Besicovitch theorem for the situation when the set of centers A of all balls $\bar{B} \in \mathcal{F}$ is bounded. If A is an unbounded set, let $D = \sup\left\{\text{diam}\bar{B} \mid \bar{B} \in \mathcal{F}\right\}$ and

$$A_l = A \cap \{x: \ 3D(l-1) \le |x| < 3Dl\}, \ l \ge 1,$$

be the sets of centers in annuli of width $3D$. Then cover each A_l by disjoint collections $\{\mathcal{J}_1^{(l)}, \ldots, \mathcal{J}_{N(n)}^{(l)}\}$ of balls in $\mathcal{F}$ – this is possible since all A_l are bounded sets. The point is that if a ball $\bar{B}_1$ is in one of the collections $\mathcal{J}_p^{(l)}$ covering the set A_l, and a ball $\bar{B}_2$ is in one of the collections $\mathcal{J}_r^{(m)}$ covering the set A_m with $|m-l| \ge 2$, then $\bar{B}_1$ and $\bar{B}_2$ do not intersect. The reason is that if $\bar{B}_1 = \bar{B}(x_1, R_1)$ and $\bar{B}_2 = \bar{B}(x_2, R_2)$ then

$$x_1 \in \{x: \ 3D(l-1) - D/2 \le |x| < 3Dl + D/2\}$$

while

$$x_2 \in \{x: \ 3D(m-1) - D/2 \le |x| < 3Dm + D/2\},$$

thus, $|x_1 - x_2| \ge 2D > \text{diam}\bar{B}_1 + \text{diam}\bar{B}_2$. Therefore, if we double the number $N(n)$ needed to cover a bounded set we can set up the baskets as in the proof of Lemma 6.3 and cover an unbounded set A by $2N(n)$ countable collections of disjoint balls.

Remark. From now on we assume that the set A is bounded.

Choosing the balls

Recall that $D = \sup\left\{\text{diam}\bar{B} \mid \bar{B} \in \mathcal{F}\right\} < +\infty$ – so we may choose a ball $\bar{B}_1 \in \mathcal{F}$ with radius

$$r_1 \ge \frac{3}{4} \cdot \frac{D}{2}.$$

After that, if the balls $\bar{B}_k$, $k = 1, \ldots, j-1$ have been chosen, choose $\bar{B}_j$ as follows. Let

$$A_j = A \setminus \bigcup_{i=1}^{j-1} \bar{B}_i$$

be the subset of A not covered by the first $(j-1)$ balls. If $A_j = \emptyset$, stop and set the counter $J = j$ (note that even in that case we are not done yet – the balls $\bar{B}_j$ may intersect each other and we still have to distribute them into $N(n)$ baskets so that balls inside each basket do not intersect). If $A_j \neq \emptyset$ choose $\bar{B}_j = \bar{B}(a_j, r_j)$ such that $a_j \in A_j$ and

$$r_j \geq \frac{3}{4} \sup \left\{ r : \ \bar{B}(a,r) \in \mathcal{F}, \ a \in A_j \right\}.$$

Note that we do not care whether the ball $\bar{B}(a_j, r_j)$ is contained in the set A_j, but only ensure that $a_j \in A_j$. If $A_j \neq \emptyset$ for any j we set the counter $J = \infty$.

Facts about the balls

We now prove some simple properties of the balls $\bar{B}_k$ that we have chosen. First, we show that a ball $\bar{B}_j$ chosen after a ball $\bar{B}_i$ can not be "much larger" than $\bar{B}_i$.

Lemma 6.4 *If $j > i$ then $r_j \leq 4r_i/3$.*

Proof. Note that if $j > i$ then $A_j \subset A_i$ – hence, the ball $\bar{B}_j$ was "a candidate ball" when $\bar{B}_i$ was chosen. Thus,

$$r_j \leq \sup \left\{ r : \ \bar{B}(a,r) \in \mathcal{F}, \ a \in A_i \right\},$$

and so

$$r_i \geq \frac{3}{4} \sup \left\{ r : \ \bar{B}(a,r) \in \mathcal{F}, \ a \in A_i \right\} \geq \frac{3}{4} r_j,$$

as claimed. □

The next lemma shows that if we shrink the balls $\bar{B}_j$ by a factor of three, the resulting balls are disjoint – without having to put them into any kind of separate sub-collections.

Lemma 6.5 *The balls $\bar{B}(a, r_j/3)$ are all disjoint.*

Proof. Let $j > i$, then the center a_j is not inside the ball B_i by construction as $A_j \cap B_i = \emptyset$. Therefore, we have $|a_j - a_i| > r_i$, and using Lemma 6.4 this leads to

$$|a_j - a_i| > r_i = \frac{r_i}{3} + \frac{2r_i}{3} \geq \frac{r_i}{3} + \frac{2}{3} \cdot \frac{3}{4} r_j \geq \frac{r_i}{3} + \frac{r_j}{2} > \frac{r_i}{3} + \frac{r_j}{3}.$$

This implies that the balls $\bar{B}(a_i, r_i/3)$ and $\bar{B}(a_j, r_j/3)$ do not intersect. □

Next, we prove that if we have chosen infinitely many balls in our construction then their radius tends to zero.

Lemma 6.6 *If $J = \infty$ then $\lim_{j \to +\infty} r_j = 0$.*

Proof. Since A is a bounded set, all $a_j \in A$ and $D < +\infty$, the set

$$Q = \bigcup_{j=1}^{\infty} \bar{B}(a, r_j/3)$$

is bounded. However, all the balls $\bar{B}\left(a_j, r_j/3\right)$ are disjoint by Lemma 6.5 and thus

$$\sum_{j=1}^{\infty} |r_j|^n < +\infty.$$

Therefore, $r_j \to 0$ and we are done. □

The next lemma shows that the balls $\bar{B}_j$ cover the whole set A of centers of all balls in the collection $\mathcal{F}$.

Lemma 6.7 *We have*

$$A \subset \bigcup_{j=1}^{J} \bar{B}(a_j, r_j).$$

Proof. If $J < \infty$ this is obvious – the only reason we can stop at a finite J is if the whole set A is covered by $\bigcup_{j=1}^{J} \bar{B}(a_j, r_j)$. Suppose $J = \infty$ and let $a \in A$ be a center of a ball $\bar{B}(a, r) \in \mathcal{F}$. Assume that a is not in the union $\bigcup_{j=1}^{\infty} \bar{B}(a_j, r_j)$. Lemma 6.6 implies that there exists j such that $r_j < 3r/4$. This is a contradiction: the point a is not in the set $\bigcup_{i=1}^{j-1} \bar{B}(a_i, r_i)$, hence the ball $\bar{B}(a, r)$ was "a candidate ball" at stage j and its radius r satisfies $r > 4r_j/3$ – this is impossible. Hence, no point in A can fail to be in the set $\bigcup_{j=1}^{\infty} \bar{B}(a_j, r_j)$, and we are done. □

Estimating the ball intersections

The rest of the proof is devoted to the following proposition.

Proposition 6.8 *There exists a number M_n which depends only on dimension n so that each ball $\bar{B}_k$ intersects at most M_n balls $\bar{B}_j$ with indices j less than k.*

This proposition together with Lemma 6.3 completes the proof of the Besicovitch Theorem. Hence, all we need to do is to prove Proposition 6.8. The proof is rather technical. We will do it in two steps. Given $m \in \mathbb{N}$ we will split the set of preceding balls $\bar{B}_j$, $j = 1, \ldots, m-1$, into the "good" ones which do not intersect $\bar{B}_m$ and the "bad" ones that do. Further, we split the "bad" ones into "small" (relative to $\bar{B}_m$) and "large" balls. Next, we will estimate the number of small bad balls by 20^n. Estimating the number of "large" balls is the final and more daunting task.

To begin we fix a positive integer m and define the set of bad preceding indices

$$I_m = \{j : \ 1 \le j \le m-1, , \ \bar{B}_j \cap \bar{B}_m \neq \emptyset\}.$$

Out of these we first consider the "small bad balls":

$$K_m = I_m \cap \{j : \ r_j \le 3r_m\}.$$

Intersecting small balls

An estimate for the cardinality of K_m is as follows.

Lemma 6.9 *The number of elements in K_m is bounded above as $|K_m| \le 20^n$.*

The main point of this lemma is of course than the number 20^n depends only on the dimension n and not on m or the collection $\mathcal{F}$.

Proof. Let $j \in K_m$ – we will show that then the smaller ball $\bar{B}(a_j, r_j/3)$ is contained in the stretched ball $\bar{B}(a_m, 5r_m)$. As Lemma 6.5 tells us that all the balls of the form $\bar{B}(a_j, r_j/3)$ are disjoint, it will follow that

$$5^n r_m^n \geq \sum_{j \in K_m} \frac{r_j^n}{3^n}. \tag{6.2}$$

However, as $j < k$, we know from Lemma 6.4 that $r_j \geq 3r_m/4$, and thus (6.2) implies that

$$5^n r_m^n \geq \sum_{j \in K_m} \frac{r_j^n}{3^n} \geq |K_m| \frac{3^n r_m^n}{4^n 3^n} = \frac{|K_m| r_m^n}{4^n},$$

and thus $|K_m| \leq 20^n$.

Thus, we need to show only that if $j \in K_m$ then $\bar{B}(a_j, r_j/3) \subset \bar{B}(a_m, 5r_m)$. To see that take a point $x \in \bar{B}(a_j, r_j/3)$, then, as $\bar{B}_j$ and $\bar{B}_m$ intersect, and $r_j \leq 3r_m$, we have

$$|x - a_m| \leq |x - a_j| + |a_j - a_m| \leq \frac{r_j}{3} + r_j + r_m = \frac{4}{3} r_j + r_m \leq 4r_m + r_m \leq 5r_m.$$

Therefore, $x \in \bar{B}(a_m, 5r_m)$ and we are done. □

Intersecting large balls

Now we come to the hardest part in the proof – estimating the cardinality of the set $P_m = I_m \setminus K_m$, that is, the number of balls $\bar{B}_j$ with indices j smaller than m which intersect the ball $\bar{B}_m = \bar{B}(a_m, r_m)$ and have a radius $r_j > 3r_m$.

Proposition 6.10 *There exists a number L_n which depends only on dimension n such that the cardinality of the set P_m satisfies $|P_m| \leq L_n$.*

We will assume without loss of generality that the center $a_m = 0$. The key to the proof of Proposition 6.10 is the following lemma which shows that the balls in the set P_m are sparsely distributed in space.

Lemma 6.11 *Let $i, j \in P_m$ with $i \neq j$, and let θ be the angle between the two lines $(a_i, 0)$ and $(a_j, 0)$ that connect the centers a_i and a_j to $a_m = 0$. Then $\theta \geq \cos^{-1} \frac{61}{64} = \theta_0 > 0$.*

Before proving this technical lemma let us finish the proof of Proposition 6.10 assuming the statement of Lemma 6.11 holds. To this end, let $r_0 > 0$ be such that if a point $x \in \mathbb{R}^n$ lies on the unit sphere in $\mathbb{R}^n$, $|x| = 1$, and $y, z \in \bar{B}(x, r_0)$ are two points in a (small) ball of radius r_0 around x then the angle between the lines connecting the points y and z to zero is less than the angle θ_0 from Lemma 6.11. Let L_n be a number (one may take the smallest such number if desired) so that the unit sphere $\{|x| = 1\} \in \mathbb{R}^n$ can be covered by L_n balls of radius r_0 (some of these balls will overlap). Then Lemma 6.11 implies that $|P_m| \leq L_n$. Indeed, if $i, j \in P_m$ then, according to this lemma, the rays connecting a_j and a_i to $a_m = 0$ have an angle larger than θ_0 between them and thus they may not intersect the same ball of radius r_0 with the center on the unit sphere. Therefore, their total number is at most L_n. □

The proof of Lemma 6.11

By now the whole proof of the Besicovitch theorem was reduced to the proof of Lemma 6.11. Let i and j be as in that lemma and assume without loss of generality that $|a_i| \leq |a_j|$. Let us denote by θ the angle between the lines $(a_j, 0)$ and $(a_i, 0)$. Lemma 6.11 is a consequence of the following two lemmas. Recall that we need to prove that θ can not be too small – it is bounded from below by $\cos^{-1}(61/64)$. The first lemma says that if θ is smaller than $\cos^{-1}(5/6)$ then the point a_i is in the ball $\bar{B}_j$ (recall that we are under the assumption that $|a_i| \leq |a_j|$), and thus $j > i$.

Lemma 6.12 *If* $\cos\theta > 5/6$ *then* $a_i \in \bar{B}_j$.

The second lemma says that if $a_i \in \bar{B}_j$ then the angle θ is at least $\cos^{-1}(61/64)$ – this finishes the proof of Lemma 6.11.

Lemma 6.13 *If* $a_i \in \bar{B}_j$ *then* $\cos\theta \leq 61/64$.

Proof of Lemma 6.12. First, we know that $i, j < m$ – hence, $a_m \notin \bar{B}_i \cup \bar{B}_j$ – this follows from how we choose the balls $\bar{B}_m$. As $a_m = 0$ this means that $r_i < |a_i|$ and $r_j < |a_j|$. In addition, the balls $\bar{B}_m$ and $\bar{B}_i$ intersect, and so do the balls $\bar{B}_m$ and $\bar{B}_j$, hence $|a_i| < r_m + r_i$, and $|a_j| < r_m + r_j$. Moreover, as $i, j \in P_m$, we have $r_i > 3r_m$ and $r_j > 3r_m$. Let us put these facts together:

$$\begin{aligned}
&3r_m < r_i < |a_i| \leq r_i + r_m,\\
&3r_m < r_j < |a_j| \leq r_j + r_m,\\
&|a_i| \leq |a_j|.
\end{aligned}$$

We claim that

$$|a_i - a_j| \leq |a_j| \text{ if } \cos\theta > 5/6. \tag{6.3}$$

Indeed, assume that $|a_i - a_j| \geq |a_j|$. Then we have

$$\cos\theta = \frac{|a_i|^2 + |a_j|^2 - |a_i - a_j|^2}{2|a_i||a_j|} \leq \frac{|a_i|^2}{2|a_i||a_j|} \leq \frac{|a_i|}{2|a_j|} \leq \frac{1}{2} < \frac{5}{6},$$

which contradicts assumptions of Lemma 6.12. Therefore, $|a_i - a_j| \geq |a_j|$ is impossible and thus $|a_i - a_j| \leq |a_j|$. This already implies that $a_i \in \bar{B}(a_j, |a_j|)$ but we need a stronger condition $a_i \in \bar{B}(a_j, r_j)$ (recall that $r_j < |a_j|$ so the ball $\bar{B}(a_j, r_j)$ is smaller than $\bar{B}(a_j, |a_j|)$).

Assume that $a_i \notin B_j$ – we will show that this would imply that $\cos\theta \leq 5/6$, which would be a contradiction. As $a_i \notin \bar{B}_j$, we have $r_j < |a_i - a_j|$, which, together with (6.3) gives

$$\begin{aligned}
\cos\theta &= \frac{|a_i|^2 + |a_j|^2 - |a_i - a_j|^2}{2|a_i||a_j|} = \frac{|a_i|}{2|a_j|} + \frac{(|a_j| - |a_i - a_j|)(|a_j| + |a_i - a_j|)}{2|a_i||a_j|}\\
&\leq \frac{1}{2} + \frac{(|a_j| - |a_i - a_j|)2|a_j|}{2|a_i||a_j|} \leq \frac{1}{2} + \frac{|a_j| - |a_i - a_j|}{|a_i|} \leq \frac{1}{2} + \frac{|a_j| - r_j}{r_i} \leq \frac{1}{2} + \frac{r_j + r_m - r_j}{r_i}\\
&\leq \frac{1}{2} + \frac{r_m}{r_i} \leq \frac{1}{2} + \frac{1}{3} = \frac{5}{6}.
\end{aligned}$$

This contradicts the assumption that $\cos\theta > 5/6$, hence $a_i \notin \bar{B}_j$ is impossible and the proof of Lemma 6.12 is complete. □

The last remaining step in the proof of the Besicovitch theorem is the proof of Lemma 6.13.

Proof of Lemma 6.13. The triangle inequality implies, obviously that

$$0 \le |a_i - a_j| + |a_i| - |a_j|.$$

Let us show that if $a_i \in \bar{B}_j$ and $|a_i| \le |a_j|$ then we have a stronger estimate:

$$|a_i - a_j| + |a_i| - |a_j| \ge \frac{|a_j|}{8}, \tag{6.4}$$

which is a measure of the notion that a_j and a_i are not collinear – which is what we need to quantify. Indeed, as $a_i \in \bar{B}_j$ we have $i < j$ and $a_j \notin \bar{B}_i$ – this follows from the way we chose the balls $\bar{B}_j$, so

$$r_i < |a_i - a_j| \le r_j.$$

Moreover, as $i < j$, we have $r_j \le 4r_i/3$. Recall also that $r_i < |a_i|$ – this is because $a_m = 0$ and $i < m$, and also $|a_j| < r_j + r_m$ – this is because $\bar{B}_j$ and $\bar{B}_m$ intersect, and, finally, that $r_j \ge 3r_m$ because $j \in P_m$. Together, these imply

$$\begin{aligned}|a_i - a_j| + |a_i| - |a_j| &\ge r_i + r_i - (r_j + r_m) = 2r_i - r_j - r_m \ge 2\cdot\frac{3}{4}r_j - r_j - \frac{r_j}{3} = \frac{r_j}{6} \\ &= \frac{1}{6}\cdot\frac{3}{4}\left(r_j + \frac{r_j}{3}\right) \ge \frac{1}{8}(r_j + r_m) \ge \frac{1}{8}|a_j|,\end{aligned}$$

which is what we claimed.

The final step is to involve the angle θ: show that

$$|a_i - a_j| + |a_i| - |a_j| \le \frac{8|a_j|}{3}(1-\cos\theta). \tag{6.5}$$

Once again, as by the assumptions of Lemma 6.13 we have $a_i \in B_j$, we must have $i < j$. Since $i < j$, we also have $a_j \notin B_i$, and thus $|a_i - a_j| > r_i$, which implies (we also use our assumption that $|a_i| \le |a_j|$ in the computation below)

$$\begin{aligned}0 &\le \frac{|a_i - a_j| + |a_i| - |a_j|}{|a_j|} \le \frac{|a_i - a_j| + |a_i| - |a_j|}{|a_j|}\cdot\frac{|a_i - a_j| + |a_j| - |a_i|}{|a_j - a_i|} \\ &= \frac{|a_i - a_j|^2 - (|a_i| - |a_j|)^2}{|a_j||a_i - a_j|} = \frac{2|a_i||a_j|(1-\cos\theta)}{|a_j||a_i - a_j|} = \frac{2|a_i|(1-\cos\theta)}{|a_i - a_j|} \\ &\le \frac{2(r_i + r_m)(1-\cos\theta)}{r_i} \le \frac{2\cdot 4r_i(1-\cos\theta)}{3r_i} = \frac{8(1-\cos\theta)}{3},\end{aligned}$$

so (6.5) holds.

It follows from (6.5) that

$$\frac{1}{8}|a_j| \le \frac{8|a_j|}{3}(1-\cos\theta),$$

and thus $\cos\theta \le 61/64$. This finishes the proof of Lemma 6.13 and hence that of the Besicovitch theorem! □

Exercise 6.14 *(i) Simplify the above prove in dimension $n = 1$. (ii) Find the best $N(n)$ in dimensions $n = 1$ and $n = 2$. Warning: it is not very difficult in one dimension but not at all simple in two dimensions.*

6.3 Differentiation of measures

Let μ and ν be two Radon measures defined on $\mathbb{R}^n$. The density of one measure with respect to another is defined as follows.

Definition 6.15 *We define*

$$\overline{D}_\mu\nu(x) = \begin{cases} \limsup\limits_{r\to 0} \dfrac{\nu(\bar{B}(x,r))}{\mu(\bar{B}(x,r))}, & \text{if } \mu(\bar{B}(x,r)) > 0 \text{ for all } r > 0, \\ +\infty, & \text{if } \mu(\bar{B}(x,r_0)) = 0 \text{ for some } r_0 > 0, \end{cases}$$

and

$$\underline{D}_\mu\nu(x) = \begin{cases} \liminf\limits_{r\to 0} \dfrac{\nu(\bar{B}(x,r))}{\mu(\bar{B}(x,r))}, & \text{if } \mu(\bar{B}(x,r)) > 0 \text{ for all } r > 0, \\ +\infty, & \text{if } \mu(\bar{B}(x,r_0)) = 0 \text{ for some } r_0 > 0, \end{cases}$$

If $\overline{D}_\mu\nu(x) = \underline{D}_\mu\nu(x) < +\infty$ then we say that ν is differentiable with respect to μ at the point x, and $D_\mu\nu$ is the density of ν with respect to μ.

Our task is to find out when $D_\mu\nu$ exists and when ν can be recovered by integrating $D_\mu\nu$, as with functions. To guide our expectations, consider an example when $\mu(x)$ is the Lebesgue measure $m(x)$ on $\mathbb{R}^n$, and $d\nu(x) = f(x)d\mu(x)$, that is, for a Lebesgue measurable set A we set

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

with a given continuous non-negative function $f(x)$. Then, clearly, we have

$$D_\mu\nu(x) = f(x),$$

and thus

$$\nu(A) = \int_A D_\mu\nu(x)d\mu(x),$$

for any Lebesgue measurable set A. On the other hand, if ν is a δ-function at $x = 0$, that is, $\nu(A) = 1$ if $0 \in A$ and $\nu(A) = 0$ otherwise, then $D_\mu\nu(x) = 0$ for all $x \neq 0$, which is μ-a.e. (though not ν-a.e), thus

$$\int_A D_\mu\nu(x)d\mu(x) = 0$$

for all Lebesgue measurable sets A. Therefore, for this pair of measures it is not true that

$$\nu(A) = \int_A D_\mu\nu(x)d\mu(x) \tag{6.6}$$

for all Lebesgue measurable sets A.

The first step in investigating when (6.6) holds is to show that the function $D_\mu\nu(x)$ is integrable with respect to the measure μ.

Theorem 6.16 *Let μ and ν be Radon measures on $\mathbb{R}^n$. Then $D_\mu\nu$ exists and is finite a.e. Moreover, $D_\mu\nu$ is a μ-measurable function.*

Proof. First, it is clear that $D_\mu\nu(x)$ in a ball $B(0,R)$ would not change if we restrict the measures μ and ν to the ball $B(0,2R)$. Hence, we may assume without loss of generality that the measures μ and ν are both finite: $\mu(\mathbb{R}^n), \nu(\mathbb{R}^n) < +\infty$.

Lemma 6.17 *Let ν and μ be two finite Radon measures on $\mathbb{R}^n$ and let $0 < s < +\infty$, then (i) $A \subseteq \{x \in \mathbb{R}^n : \underline{D}_\mu\nu \le s\}$ implies $\nu^*(A) \le s\mu^*(A)$, and (ii) $A \subseteq \{x \in \mathbb{R}^n : \overline{D}_\mu\nu \ge s\}$ implies $\nu^*(A) \ge s\mu^*(A)$.*

Proof of Lemma 6.17. Let A be as in (i) and let U be an open set containing the set A. Then for any $\varepsilon > 0$ and any $x \in A$ we may find a sequence $r_n(x) \to 0$, as $n \to +\infty$, such that

$$\nu(\bar{B}(x, r_n(x)) \le (s+\varepsilon)\mu(\bar{B}(x, r_n(x)),$$

and $\bar{B}(x, r_n(x)) \subset U$. The balls $\bar{B}(x, r_n(x))$, $x \in A$, $n \in \mathbb{N}$, form a collection $\mathcal{F}$ satisfying the assumptions of Corollary 6.2 since ν is a finite measure. Note that this corollary does not require A to be a ν-measurable set. Hence, we may choose a countable sub-collection $\mathcal{J}$ of pairwise disjoint balls such that

$$\nu\left(A \setminus \bigcup_{B \in \mathcal{J}} \bar{B}\right) = 0.$$

It follows that

$$\nu^*(A) \le \sum_{B \in \mathcal{J}} \nu(\bar{B}) \le (s+\varepsilon)\sum_{B \in \mathcal{J}} \mu(\bar{B}) \le (s+\varepsilon)\mu(U).$$

Taking infimum over all open sets U containing the set A we obtain that $\nu^*(A) \le s\mu^*(A)$. The proof of part (ii) is almost identical. □

Returning to the proof of Theorem 6.16 consider the set $I = \{x : \overline{D}_\mu\nu(x) = +\infty\}$. Then for all $s > 0$ we have $s\mu^*(I) \le \nu^*(I)$, which means that $\mu(I) = 0$, as $\nu^*(I) \le \nu^*(\mathbb{R}^n) < +\infty$. Moreover, for any $b > a$ if we set $R_{ab} = \{x : \underline{D}_\mu\nu < a < b < \overline{D}_\mu\nu\}$, we have, using Lemma 6.17:

$$b\mu^*(R_{ab}) \le \nu^*(R_{ab}) \le a\mu^*(R_{ab}),$$

thus $\mu(R_{ab}) = 0$. It follows that $D_\mu\nu(x)$ exists and is finite μ-a.e. It remains to show that the function $D_\mu\nu(x)$ is μ-measurable.

Lemma 6.18 *For each $x \in \mathbb{R}^n$ and $r > 0$ we have $\limsup_{y \to x} \mu(\bar{B}(y,r)) \le \mu(\bar{B}(x,r))$ and $\limsup_{y \to x} \nu(\bar{B}(y,r)) \le \nu(\bar{B}(x,r))$.*

Proof of Lemma 6.18. Let $y_k \to x$ and set $f_k(z) = \chi_{\bar{B}(y_k,r)}(z)$. We claim that

$$\limsup_{k\to\infty} f_k(z) \le \chi_{\bar{B}(x,r)}(z). \tag{6.7}$$

Indeed, all we need to verify is that if $z \notin \bar{B}(x,r)$ then

$$\limsup_{k\to\infty} f_k(z) = 0.$$

However, as $U = (\bar{B}(x,r))^c$ is an open set, and $y_k \to x$ it follows that for k large enough we have $z \notin \bar{B}(y_k, r)$, and thus (6.7) holds. It follows that

$$\liminf_{k\to\infty}(1 - f_k(z)) \geq 1 - \chi_{\bar{B}(x,r)}(z),$$

and thus, by Fatou's lemma, we have

$$\int_{\bar{B}(x,2r)} (1 - \chi_{\bar{B}(x,r)}(z))d\mu \leq \liminf_{k\to\infty} \int_{\bar{B}(x,2r)} (1 - f_k(z))d\mu.$$

This is nothing but

$$\mu(\bar{B}(x,2r)) - \mu(\bar{B}(x,r)) \leq \liminf_{k\to\infty}[\mu(\bar{B}(x,2r)) - \mu(\bar{B}(y_k,r))] = \mu(\bar{B}(x,2r)) - \limsup_{k\to\infty} \mu(\bar{B}(y_k,r)),$$

and thus

$$\mu(\bar{B}(x,r)) \geq \limsup_{k\to\infty} \mu(\bar{B}(y_k,r)),$$

and we are done. □

All that remains to finish the proof of Theorem 6.16 is to notice that Lemma 6.18 implies that the functions $f_\mu(x) = \mu(\bar{B}(x,r))$ and $f_\nu(x) = \nu(\bar{B}(x,r))$ are upper semi-continuous and thus μ-measurable for all $r > 0$ fixed. Therefore, the derivative

$$D_\mu\nu(x) = \lim_{r\to 0} \frac{f_\mu(x;r)}{f_\nu(x;r)}$$

is also μ-measurable. □

6.4 The Radon-Nikodym theorem

Definition 6.19 *We say that a measure ν is absolutely continuous with respect to a measure μ and write $\nu \ll \mu$ if for any set A such that $\mu(A) = 0$ we have $\nu(A) = 0$.*

Theorem 6.20 *Let μ and ν be Radon measures on $\mathbb{R}^n$ and assume that ν is absolutely continuous with respect to μ. Then for any μ-measurable set A we have*

$$\nu(A) = \int_A D_\mu\nu(x)d\mu. \tag{6.8}$$

Proof. Let A be a μ-measurable set. We claim that A then is ν-measurable. Indeed, there exists a Borel set B such that $A \subseteq B$ and $\mu(B \setminus A) = 0$. As $\nu \ll \mu$, it follows that we have $\nu(B \setminus A) = 0$ as well, so that the set $B \setminus A$ is ν-measurable. In addition, as B is a Borel set, B is also ν-measurable. Writing $A = B \cap (B \setminus A)^c$ we see that A is, indeed, ν-measurable.

Set now $Z = \{x : \ D_\mu\nu(x) = 0\}$ and $I = \{x : \ D_\mu\nu(x) = +\infty\}$. Then $\mu(I) = 0$ by Theorem 6.16 and thus $\nu(I) = 0$. Moreover, for any $R > 0$ we have

$$\nu(Z \cap B(0,R)) \leq s\mu(Z \cap B(0,R))$$

for all $s > 0$ by Lemma 6.17. It follows that $\nu(Z \cap B(0,R)) = 0$ for all $R > 0$, thus $\nu(Z) = 0$. Summarizing, we have

$$\nu(Z) = \int_Z (D_\mu \nu) d\mu = 0, \quad \text{and } \nu(I) = \int_I (D_\mu \nu) d\mu = 0. \tag{6.9}$$

The rest is done with the help of Lemma 6.17. Consider a μ-measurable set A, fix $t > 1$ and decompose A as

$$A = \bigcup_{m=-\infty}^{+\infty} A_m \bigcup Z \bigcup I,$$

with

$$A_m = \left\{ x : \; t^m \le D_\mu \nu(x) < t^{m+1} \right\}.$$

The sets A_m are μ-measurable, hence they are ν-measurable as well. Furthermore, as

$$\nu(Z) = \nu(I) = 0,$$

we have

$$\nu(A) = \sum_{m=-\infty}^{+\infty} \nu(A_m) \le \sum_{m=-\infty}^{+\infty} t^{m+1} \mu(A_m) \le t \sum_{m=-\infty}^{+\infty} t^m \mu(A_m) \le t \int_{\tilde{A}} (D_\mu \nu) d\mu,$$

and

$$\nu(A) = \sum_{m=-\infty}^{+\infty} \nu(A_m) \ge \sum_{m=-\infty}^{+\infty} t^m \mu(A_m) = \frac{1}{t} \sum_{m=-\infty}^{+\infty} t^{m+1} \mu(A_m) \ge \frac{1}{t} \int_{\tilde{A}} (D_\mu \nu) d\mu,$$

where

$$\tilde{A} = \bigcup_{m=-\infty}^{+\infty} A_m = A \setminus (Z \cup I).$$

Passing to the limit $t \to 1$ and using (6.9) to replace $\tilde{A}$ by A as the domain of integration we obtain (6.8). □

6.5 The Lebesgue decomposition

Definition 6.21 *We say that two Radon measures μ and ν are mutually singular and write $\mu \perp \nu$ if there exists a Borel set B such that $\mu(\mathbb{R}^n \setminus B) = \nu(B) = 0$.*

Theorem 6.22 *(The Lebesgue Decomposition) Let μ and ν be Radon measures on $\mathbb{R}^n$. Then (i) there exist measures $\nu_{ac} \ll \mu$ and $\nu_s \perp \mu$ so that $\nu = \nu_{ac} + \nu_s$, and (ii) $D_\mu \nu(x) = D_\mu \nu_{ac}(x)$ and $D_\mu \nu_s = 0$, both for μ-a.e. x so that for each Borel set A we have*

$$\nu(A) = \int_A (D_\mu \nu) d\mu + \nu_s(A). \tag{6.10}$$

Proof. As before, since both μ and ν are Radon measures we may assume that $\mu(\mathbb{R}^n) < \infty$ and $\nu(\mathbb{R}^n) < +\infty$. If one or both of these measures is not finite we would simply restrict both μ and ν to balls $B(0, R)$ and let $R \to +\infty$ at the end of the proof.

We will define ν_{ac} and ν_s as $\nu_{ac} = \nu|_B$ and $\nu_s = \nu|_{B^c}$ with an appropriately chosen Borel set B. Consider the collection

$$\mathcal{F} = \{A \subset \mathbb{R}^n,\ A \text{ Borel},\ \mu(\mathbb{R}^n \setminus A) = 0.\}$$

The set B should be, in the measure-theoretical sense, the smallest element of $\mathcal{F}$. To this end choose $B_k \in \mathcal{F}$ such that

$$\nu(B_k) \le \inf_{A\in\mathcal{F}} \nu(A) + \frac{1}{k},$$

and set $B = \bigcap_{k=1}^{\infty} B_k$. Then

$$\mu(\mathbb{R}^n \setminus B) \le \sum_{k=1}^{\infty} \mu(\mathbb{R}^n \setminus B_k) = 0, \tag{6.11}$$

and thus $B \in \mathcal{F}$ and B is the smallest element of $\mathcal{F}$ in the sense that $\nu(B) = \inf_{A\in\mathcal{F}} \nu(A)$. Note that (6.11) implies that the measure $\nu_s = \nu|_{B^c}$ is mutually singular with μ.

Let us show that $\nu_{ac} = \nu|_B$ is absolutely continuous with respect to μ. Let $A \subset \mathbb{R}^n$ and assume that $\mu(A) = 0$ but $\nu_{ac}^*(A) > 0$. Take a Borel set A' such that $A \subset A'$, and $\mu(A') = 0$, while $\nu_{ac}(A') \ge \nu_{ac}^*(A) > 0$ and consider $\tilde{A} = B \cap A'$. For $\tilde{A}$ we still have, using (6.11),

$$\mu(\tilde{A}) = \mu(A') - \mu(A' \cap B^c) = 0, \tag{6.12}$$

and

$$\nu_{ac}(\tilde{A}) = \nu_{ac}(A') > 0. \tag{6.13}$$

Now, (6.12) implies that $B' = B \setminus \tilde{A} \in \mathcal{F}$ but (6.13) means that

$$\nu(B') = \nu(B) - \nu(\tilde{A}) < \nu(B),$$

which is a contradiction. Therefore, ν_{ac} is absolutely continuous with respect to μ.

Finally, to show that $D_\mu\nu(x) = D_\mu\nu_{ac}(x)$ for μ-a.e. x, let $z > 0$, consider the set

$$C_z = \{x :\ D_\mu\nu_s \ge z\},$$

and write $C_z = C_z' \cup C_z''$ with $C_z' = C_z \cap B$, $C_z'' = C_z \cap B^c$. Then $\mu(C_z'') = 0$ since $B \in \mathcal{F}$, while Lemma 6.17 implies that

$$z\mu(C_z') \le \nu_s(C_z') \le \nu_s(B) = 0.$$

It follows that $D_\mu\nu_s = 0$ μ-a.e., which, in turn, means that $D_\mu\nu = D_\mu\nu_{ac}$ μ-a.e. Now, Theorem 6.20 implies that (6.10) holds. □

The Lebesgue-Besicovitch theorem

Given a function f, we define its average over a measurable set E with $\mu(E) > 0$ as

$$\fint_E f d\mu = \frac{1}{\mu(E)} \int_E f d\mu.$$

A trivial observation is that for a continuous function $f(x)$ we have

$$\lim_{r \downarrow 0} \fint_{\bar{B}(x,r)} f dy = f(x).$$

The following generalization is much less immediately obvious.

Theorem 6.23 *Let μ be a Radon measure and assume that $f \in L^1_{loc}(\mathbb{R}^n, d\mu)$, then*

$$\lim_{r \to 0} \fint_{\bar{B}(x,r)} f d\mu = f(x) \text{ for } \mu\text{-a.e. } x \in \mathbb{R}^n. \tag{6.14}$$

Proof. The proof is surprisingly simple and is based on the Radon-Nikodym theorem. Let us define the measures $\nu_\pm$ as follows. For a Borel set B we set

$$\nu_\pm(B) = \int_B f_\pm d\mu, \tag{6.15}$$

with $f_+ = \max(f, 0)$ and $f_- = \max(-f, 0)$, and for an arbitrary set A define

$$\nu^*_\pm(A) = \inf(\nu_\pm(B) : \ A \subseteq B, \ B \text{ Borel}).$$

Then, ν_+ and ν_- are Radon measures, absolutely continuous with respect to μ, thus

$$\nu_+(A) = \int_A D_\mu \nu_+ d\mu, \quad \nu_-(A) = \int_A D_\mu \nu_- d\mu \tag{6.16}$$

for all μ-measurable sets A. Together, (6.15) and (6.16) imply that

$$D_\mu \nu_\pm = f_\pm \ \mu\text{-a.e.} \tag{6.17}$$

Indeed, consider, for instance, the set $S = \{x : \ f_+(x) > D_\mu \nu_+(x)\} = \bigcup_{q \in \mathbb{Q}} S_q$, with

$$S_q = \{x : \ f_+(x) - D_\mu \nu_+(x) > q\}.$$

The set S_q is μ-measurable, and

$$\int_{S_q} (f_+ - D_\mu \nu_+) d\mu \geq q\mu(S_q),$$

thus $\mu(S_q) = 0$ so that $\mu(S) = 0$ as well. Using (6.17), we get

$$\lim_{r \to 0} \fint_{\bar{B}(x,r)} f d\mu = \lim_{r \to 0} \frac{1}{\mu(\bar{B}(x,r)} [\nu^+(\bar{B}(x,r)) - \nu^-(\bar{B}(x,r))] = D_\mu \nu_+ - D_\mu \nu_- = f_+ - f_- = f,$$

for μ-a.e. x. □

The Lebesgue-Besicovitch theorem has several interesting corollaries.

Definition 6.24 *Let $f \in L^p_{loc}(\mathbb{R}^n, d\mu)$ with $1 \le p < +\infty$. A point x is a Lebesgue point of f*

$$\lim_{r\to 0} \fint_{\bar{B}(x,r)} |f(y) - f(x)|^p d\mu_y = 0.$$

Corollary 6.25 *Let μ be a Radon measure on $\mathbb{R}^n$, $1 \le p < +\infty$ and let $f \in L^p_{loc}(\mathbb{R}^n, d\mu)$ with $1 \le p < +\infty$, then*

$$\lim_{r\to 0} \fint_{\bar{B}(x,r)} |f(y) - f(x)|^p d\mu_y = 0 \tag{6.18}$$

for μ-a.e. $x \in \mathbb{R}^n$.

Proof. Let ξ_j be a countable dense subset of $\mathbb{R}$, then for each j fixed we have

$$\lim_{r\to 0} \fint_{\bar{B}(x,r)} |f(y) - \xi_j|^p d\mu_y = |f(x) - \xi_j|^p \tag{6.19}$$

for μ-a.e. $x \in \mathbb{R}^n$. Hence, there exists a set S of full measure, $\mu(\mathbb{R}^n \setminus S) = 0$ so that (6.19) holds for all j for $x \in S$. Next, given $x \in S$ and $\varepsilon > 0$ choose ξ_j so that $|f(x) - \xi_j|^p < \varepsilon/2^p$, then we have

$$\begin{aligned}
&\limsup_{r\to 0} \fint_{\bar{B}(x,r)} |f(y) - f(x)|^p d\mu_y \\
&\le 2^{p-1} \limsup_{r\to 0} \fint_{\bar{B}(x,r)} |f(y) - \xi_j|^p d\mu_y + 2^{p-1} \limsup_{r\to 0} \fint_{\bar{B}(x,r)} |\xi_j - f(x)|^p d\mu_y \\
&= 2^{p-1}|f(x) - \xi_j|^p + 2^{p-1}|f(x) - \xi_j|^p \le \varepsilon,
\end{aligned}$$

and, as $\varepsilon > 0$ is arbitrary, (6.18) holds. □

The next corollary describes the "density" of measurable sets.

Corollary 6.26 *Let $E \subseteq \mathbb{R}^n$ be Lebesgue measurable, then*

$$\lim_{r\to 0} \frac{|B(x,r) \cap E|}{|B(x,r)|} = 1 \text{ for a.e. } x \in E,$$

and

$$\lim_{r\to 0} \frac{|B(x,r) \cap E|}{|B(x,r)|} = 0 \text{ for a.e. } x \notin E.$$

Proof. This follows immediately from the Lebesgue-Besicovitch theorem applied to the function $f(x) = \chi_E(x)$. □

7 Signed measures and the Riesz representation theorem

7.1 The Hahn decomposition

Definition 7.1 *A signed measure ν on a σ-algebra $\mathcal{B}$ is a function defined on sets from $\mathcal{B}$ that satisfies* *(i)* ν *assumes only one of the values* $+\infty$ *and* $-\infty$.

(ii) $\nu(\emptyset) = 0$.

(iii) $\nu\left(\bigcup_{j=1}^{\infty} E_j\right) = \sum_{j=1}^{\infty} \nu(E_j)$ *for any sequence* E_j *of disjoint sets in* $\mathcal{B}$ *and the series converges absolutely.*

Definition 7.2 *A set* A *is positive with respect to a signed measure* ν *if* $A \in \mathcal{B}$ *and* $\nu(E) \geq 0$ *for all* $E \subseteq A$.

Proposition 7.3 *Let* E *be a measurable set,* $0 < \nu(E) < +\infty$*, then there exists a positive set* $A \subseteq E$ *with* $\nu(A) > 0$.

Proof. If E is not positive, we construct a sequence of sets $A_1, \ldots, A_k, \ldots$ as follows. Let n_1 be the smallest integer so that E contains a subset A_1 with $\nu(A_1) < -1/n_1$. Then, inductively, having chosen $A_1, \ldots, A_{k-1}$ choose A_k as follows. Set

$$E_{k-1} = E \setminus (\bigcup_{j=1}^{k-1} A_j),$$

and let n_k be the smallest integer so that E_{k-1} contains a subset Q with $\nu(Q) < -1/n_k$. Finally, take $A_k \subseteq E_{k-1}$ with $\nu(A_k) < -1/n_k$. This procedure can be continued unless at some step k_0 the set E_{k_0} is positive. In that case we are done, as

$$\nu(E_{k_0}) = \nu(E) - \sum_{j=1}^{k_0-1} \nu(A_j) \geq \nu(E) > 0.$$

On the other hand, if we never stop, we set $A = E \setminus \bigcup_{j=1}^{\infty} A_j$. Note that, since $\nu(E) > 0$, we have

$$\sum_{j=1}^{\infty} |\nu(A_j)| < +\infty,$$

and thus $n_j \to +\infty$ as $j \to +\infty$. Moreover, A can not contain a subset S of negative measure because in that case we would have $\nu(S) < -1/(n_k - 1)$ for a large enough k which would give a contradiction. □

Theorem 7.4 *Let* ν *be a signed measure on* X. *Then there exists a positive set* A *and a negative set* B *so that* $X = A \bigcup B$.

Proof. Assume that ν omits the value $+\infty$ and set $\lambda = \sup\{\nu(A) : A \text{ is a positive set}\}$. Choose positive sets A_j such that $\nu(A_j) > \lambda - 1/j$ and set $A = \bigcup_{j=1}^{\infty} A_j$. Since A is a union of positive sets, A is positive itself. Therefore,

$$\nu(A) = \nu(A_j) + \nu(A \setminus A_j) \geq \lambda - 1/j,$$

for all $j \in \mathbb{N}$, and thus $\nu(A) = \lambda$. No subset S of the set $B = A^c$ can have positive measure for if $\nu(S) > 0$, S contains a positive subset S' with $\nu(S') > 0$ by Proposition 7.3. Then the set $A' = A \cup S'$ would be positive with $\nu(A') > \lambda$ which would contradict the definition of λ. Hence, the set B is negative. □

Corollary 7.5 *Let ν be a signed measure on X. There exists a pair of mutually singular measures ν^+ and ν^- such that $\nu = \nu^+ - \nu^-$.*

Proof. Simply decompose $X = A \cup B$ as in Theorem 7.4, set $\nu^+ = \nu|_A$ and $\nu^- = \nu|_B$, and observe that both ν^+ and ν^- are measures (and not signed measures). □

We will denote by $|\nu| = \nu^+ + \nu^-$ the total variation of the measure ν. The decomposition

$$\nu = \nu^+ - \nu^-$$

shows that Radon-Nikodym theorem applies to signed measures as well, in the following sense. We say that $\nu \ll \mu$ if $\mu(A) = 0$ implies that $\nu^+(A) = \nu^-(A) = 0$. In that case, we may use the Radon-Nikodym theorem to write

$$\nu^+(S) = \int_S f^+ d\mu, \quad \nu^-(S) = \int_S f^- d\mu,$$

and

$$\nu(S) = \int_S f d\mu,$$

with $f = f^+ - f^-$.

7.2 The Riesz Representation Theorem in L^p

Recall that a linear functional $F : X \to \mathbb{R}$ acting on a normed linear space X is bounded if there exists a constant $C > 0$ so that $|F(x)| \le C\|x\|_X$ for all $x \in X$, and

$$\|F\| = \sup_{\|x\|_X = 1} |F(x)|.$$

An example of a bounded linear functional on $L^p(\mathbb{R}^n)$ is

$$F(f) = \int_{\mathbb{R}^n} fg dx,$$

where $g \in L^q(\mathbb{R}^n)$ and $\|F\| \le \|g\|_{L^q}$ – this follows from the Hölder inequality. It turns out that for $1 \le p < +\infty$ all bounded linear functionals on L^p have this form.

Theorem 7.6 *Let μ be a Radon measure, $1 \le p < +\infty$, and $F : L^p(\mathbb{R}^n, d\mu) \to \mathbb{R}$ be a bounded linear functional. Then there exists a unique function $g \in L^q(\mathbb{R}^n, d\mu)$, where $1/p + 1/q = 1$, such that*

$$F(f) = \int_{\mathbb{R}^n} f(x)g(x)d\mu$$

for any function $f \in L^p(\mathbb{R}^n, d\mu)$, and $\|F\| = \|g\|_{L^q}$.

Proof. The proof is long but straightforward. First, we construct the only candidate for the function g rather explicitly in terms of the functional F. Then we check that the candidate g lies in $L^q(\mathbb{R}^n, d\mu)$, and, finally, we verify that, indeed, both

$$F(f) = \int fg d\mu,$$

and $\|F\| = \|g\|_{L^q(\mathbb{R}^n, d\mu)}$.

First, we assume that μ is a finite measure: $\mu(\mathbb{R}^n) < +\infty$ so that the function $f \equiv 1$ lies in all $L^p(\mathbb{R}^n, d\mu)$. For a μ-measurable set E let us set $\nu(E) = F(\chi_E)$. The linearity and boundedness of F, and finiteness of μ imply that ν is a signed measure with

$$|\nu(E)| \leq \|F\|\|\chi_E\|_{L^p} \leq \|F\|[\mu(E)]^{1/p} \leq \|F\|[\mu(\mathbb{R}^n)]^{1/p}. \tag{7.1}$$

Let us decompose $\nu = \nu^+ - \nu^-$ as in Corollary 7.5, and introduce the Hahn decomposition of $\mathbb{R}^n$ relative to ν: $\mathbb{R}^n = A \cup B$, so that ν^+ supported in A, and ν^- supported in B. Then (7.1) implies that

$$\nu^+(E) = \nu(A \cap E) = |\nu(A \cap E)| \leq \|F\|[\mu(A \cap E)]^{1/p} \leq \|F\|[\mu(E)]^{1/p}, \tag{7.2}$$

and thus ν^+ (and also ν^- by the same argument) is absolutely continuous with respect to μ. Therefore, ν has the Radon-Nikodym derivative $g(x)$:

$$\nu(E) = \int_E g d\mu,$$

with $g^+ = D_\mu \nu^+$ and $g^- = D_\mu \nu^-$.

Using (7.2), we conclude that

$$\|g\|_{L^1(\mathbb{R}^n, d\mu)} = \nu^+(\mathbb{R}^n) + \nu^-(\mathbb{R}^n) \leq 2\|F\|(\mu(\mathbb{R}^n))^{1/p},$$

thus $g \in L^1(\mathbb{R}^n, d\mu)$.

Let us now show that $g \in L^q(\mathbb{R}^n, d\mu)$, where $1/p + 1/q = 1$. It follows from the definition of g that for any simple function ϕ which takes only finitely many values we have

$$F(\phi) = \int \phi g d\mu. \tag{7.3}$$

For $1 < p < +\infty$, let ψ_n be a point-wise non-decreasing sequence of simple functions taking finitely many values such that $\psi_n^{1/q} \to |g|$. For example, we may take

$$\psi_n(x) = \frac{j}{2^n} \quad \text{if } |x| \leq n \text{ and } \quad \frac{j}{2^n} \leq |g(x)|^q < \frac{j+1}{2^n}, \quad 0 \leq j \leq 2^{2n} - 1,$$

and $\psi_n(x) = 0$ if $|x| > n$ or $|g(x)|^q \geq 2^n$. The function ψ_n lies then in all spaces $L^r(\mathbb{R}^n; d\mu)$, with $1 \leq r \leq +\infty$, for each fixed n. Define also $\phi_n = (\psi_n)^{1/p} \text{sgn } g$,

$$\|\phi_n\|_{L^p} = \left(\int \psi_n d\mu \right)^{1/p},$$

therefore, as ϕ_n is also a simple function taking finitely many values, we get

$$\int \psi_n d\mu = \int \psi_n^{1/p+1/q} d\mu = \int |\psi_n|^{1/q} |\phi_n| d\mu \leq \int |g||\phi_n| d\mu = \int g\phi_n d\mu$$
$$= F(\phi_n) \leq \|F\|\|\phi_n\|_{L^p} \leq \|F\| \left(\int \psi_n d\mu \right)^{1/p}.$$

It follows that

$$\left(\int \psi_n d\mu\right)^{1/q} \leq \|F\|$$

and thus

$$\int |g|^q d\mu \leq \|F\|^q \tag{7.4}$$

by the Monotone Convergence Theorem, hence $g \in L^q(\mathbb{R}^n, d\mu)$ and $\|g\|_{L^q(\mathbb{R}^n,d\mu)} \leq \|F\|$.

In order to finish the proof, note that, as $g \in L^q(\mathbb{R}^n, d\mu)$, the linear functional

$$G(f) = \int fg d\mu$$

is bounded: $\|G\| \leq \|g\|_{L^q}$. Moreover, $G(\phi) = F(\phi)$ for any simple function in $L^p(\mathbb{R}^n, d\mu)$ which takes finitely many values. As such simple functions are dense in this space, and both G and F are bounded functionals, it follows that $G(f) = F(f)$ for all $f \in L^p(\mathbb{R}^n, d\mu)$, thus

$$F(f) = \int fg d\mu$$

for all $f \in L^p(\mathbb{R}^n, d\mu)$. Hence, we have $\|F\| \leq \|g\|_{L^q}$, which, together with the bound (7.4) implies that $\|F\| = \|g\|_{L^q}$.

When the measure μ is not finite: $\mu(\mathbb{R}^n) = +\infty$, and still in the case $1 < p < +\infty$, consider the balls $B_R = B(0, R)$ and the corresponding restrictions $\mu_R = \mu|_{B_R}$. Define also the bounded linear functionals $F_R(f) = F(f\chi_R)$, with $\chi_R(x) = \chi_{B_R}(x)$. Note that for any function $f \in L^p(\mathbb{R}^n; d\mu)$ we have

$$\|f - f\chi_R\|_{L^p} \to 0 \text{ as } R \to +\infty,$$

whence

$$|F(f) - F_R(f)| = |F(f - f\chi_R)| \leq \|F\|\|f - f\chi_R\|_{L^p} \to 0 \text{ as } R \to +\infty.$$

Moreover, we have

$$|F_R(f)| \leq \|F\|\|f\chi_R\|_{L^p(\mathbb{R},d\mu)} = \|F\|\|f\|_{L^p(\mathbb{R},d\mu_R)},$$

so that F_R is a bounded linear functional on $L^p(\mathbb{R}, d\mu_R)$ with $\|F_R\| \leq \|F\|$. It follows that there exists a unique function $g_R \in L^q(\mathbb{R}, d\mu_R)$ such that

$$F_R(f) = \int fg_R d\mu_R,$$

and

$$\|g_R\|_{L^q(\mathbb{R},d\mu_R)} = \|F_R\| \leq \|F\|.$$

We may assume without loss of generality that g_R vanishes outside of the ball $B(0, R)$. Given any $R' > R''$, the natural restriction of $F_{R'}$ to $L^p(\mathbb{R}, d\mu_{R''})$ coincides with $F_{R''}$. The uniqueness of the kernel $g_{R''}$ implies that $g_R(x)$ is of the form

$$g_R(x) = g(x)\chi_R(x),$$

with a fixed function $g(x)$:

$$F_R(f) = \int f(x)\chi_R(x)g(x)d\mu.$$

Hence, we may pass to the limit $R \to \infty$ and Fatou's lemma implies that the limit $g(x)$ is in $L^q(\mathbb{R}^n, d\mu)$ with $\|g\|_{L^q(\mathbb{R}^n,d\mu)} \leq \|F\|$. It follows that, for any $f \in L^p(\mathbb{R}^n; d\mu)$:

$$F(f) = \lim_{R\to+\infty} F(f\chi_R) = \lim_{R\to+\infty} \int f(x)\chi_R(x)g(x)d\mu = \int f(x)g(x)dx,$$

with the last equality following from the Lebesgue dominated convergence theorem, which can be applied because $fg \in L^1(\mathbb{R}^n; d\mu)$ since $f \in L^p(\mathbb{R}^n; d\mu)$ and $g \in L^q(\mathbb{R}^n; d\mu)$. As a consequence, we have $\|F\| \leq \|g\|_{L^q}$, and thus $\|F\| = \|g\|_{L^q}$.

It remains only to consider the case $p = 1$. First, if $\mu(\mathbb{R}^n) < +\infty$, we know that

$$F(f) = \int fgd\mu$$

for any simple function that takes finitely many values. Hence, for any measurable set E we have

$$\Big|\int_E g(x)d\mu\Big| \leq \|F\|\mu(E).$$

It follows that for any $\varepsilon > 0$

$$\mu(x:\ g(x) > \|F\| + \varepsilon) = 0,$$

and

$$\mu(x:\ g(x) < -\|F\| - \varepsilon) = 0.$$

Thus, $g \in L^\infty(\mathbb{R}^n; d\mu)$ and $\|g\|_{L^\infty} \leq \|F\|$. The density in $L^1(\mathbb{R}^n; d\mu)$ of simple functions taking finitely many values shows then that

$$F(f) = \int fgd\mu$$

for any $f \in L^1(\mathbb{R}^n; d\mu)$, which, in turn, implies that $\|F\| \leq \|g\|_{L^\infty}$, whence $\|F\| = \|g\|_{L^\infty}$, and we are done. When $\mu(\mathbb{R}^n) = +\infty$, we proceed by the same cut-off argument as for $1 < p < +\infty$. □

7.3 The Riesz representation theorem for $C_c(\mathbb{R}^n)$

The Riesz representation theorem for the L^p spaces we have discussed above holds for $1 \leq p < +\infty$. The nature of the dual space to L^∞ is beyond the scope of these notes but it has a meaningful counterpart which describes bounded linear functionals on $C_c(\mathbb{R}^n; \mathbb{R}^m)$ – vector valued continuous functions of compact support.

Theorem 7.7 *Let $L : C_c(\mathbb{R}^n; \mathbb{R}^m) \to \mathbb{R}$ be a linear functional such that for each compact set K we have*

$$\sup\{L(f):\ f \in C_c(\mathbb{R}^n; \mathbb{R}^m),\ |f| \leq 1,\ suppf \subseteq K\} < +\infty. \tag{7.5}$$

Then there exists a Radon measure μ on $\mathbb{R}^n$ and a μ-measurable function $\sigma : \mathbb{R}^n \to \mathbb{R}^m$ such that (i) $|\sigma(x)| = 1$ for μ-a.e. $x \in \mathbb{R}^n$, and (ii) $L(f) = \int_{\mathbb{R}^n} (f \cdot \sigma)d\mu$ for all $f \in C_c(\mathbb{R}^n; \mathbb{R}^m)$.

In the scalar case when $m = 1$, the function $\sigma(x) = \pm 1$ μ-a.e., thus we may define a signed measure $d\nu(x) = \sigma(x) d\mu(x)$ and get the representation

$$L(f) = \int_{\mathbb{R}^n} f d\nu.$$

That is, the space dual to $C_c(\mathbb{R}^n)$ consists of signed measures.

Proof of Theorem 7.7

Let us momentarily consider the case $m = 1$, when we need to show that there exists a signed measure ν so that

$$L(f) = \int_{\mathbb{R}^n} f d\nu,$$

for every $f \in C_c(\mathbb{R}^n)$. What we really would like to do then is define the signed measure ν

$$\nu(A) = L(\chi_A),$$

which is essentially what we did for the L^p spaces, $1 \le p < +\infty$. Unfortunately, the characteristic function χ_A is not continuous, forcing us to resort to approximations. To this end (returning to the general situation $m \ge 1$), define the variation measure by

$$\mu^*(V) = \sup\{L(f) : \ f \in C_c(\mathbb{R}^n; \mathbb{R}^m), \ |f| \le 1, \ \text{supp} f \subseteq V\}$$

for open sets V, and for an arbitrary set $A \subset \mathbb{R}^n$ define

$$\mu^*(A) = \inf\{\mu(V) : \ A \subset V, \ V \text{ is open}\}.$$

Our task is to show that μ and an appropriately defined function σ will satisfy (i) and (ii). We will proceed gingerly in several steps. First, we need to show that μ is actually a Radon measure. Next, for $f \in C_c^+ = \{f \in C_c(\mathbb{R}^n) : \ f \ge 0\}$ we will define a functional

$$\lambda(f) = \sup\{L(g) : \ g \in C_c(\mathbb{R}^n; \mathbb{R}^m), \ |g| \le f\}. \tag{7.6}$$

It turns out that λ is actually a linear functional on $C_c^+(\mathbb{R}^n)$. Moreover, we will show that λ has an explicit form

$$\lambda(f) = \int_{\mathbb{R}^n} f d\mu. \tag{7.7}$$

The function σ will come about as follows: for every unit vector $e \in \mathbb{R}^m$, $|e| = 1$, we define a linear functional λ_e on $C_c(\mathbb{R}^n)$ by

$$\lambda_e(f) = L(fe). \tag{7.8}$$

We will extend λ_e to a bounded linear functional on $L^1(\mathbb{R}^n, d\mu)$ and use the Riesz representation theorem for $L^1(\mathbb{R}^n, d\mu)$ to find a function $\sigma_e \in L^\infty(\mathbb{R}^n)$ so that

$$\lambda_e(f) = \int f \sigma_e d\mu$$

for all $f \in L^1(\mathbb{R}^n, d\mu)$. Finally we will set $\sigma(x) = \sum_{j=1}^m \sigma_{e_j}(x) e_j$, where e_j is the standard basis for $\mathbb{R}^m$. Then for any $f \in C_c(\mathbb{R}^n; \mathbb{R}^m$ we have

$$L(f) = \sum_{j=1}^m L((f \cdot e_j)e_j) = \sum_{j=1}^m \lambda_{e_j}(f \cdot e_j) = \sum_{j=1}^m \int (f \cdot e_j)\sigma_{e_j} d\mu = \int (f \cdot \sigma) d\mu,$$

and we would be done, except for showing that $|\sigma(x)| = 1$ for μ-a.e. $x \in \mathbb{R}^n$.

Step 1. As promised, we first show that μ is a Radon measure. Let us check that μ is a measure: we take open sets V_j, $j \geq 1$, and an open set $V \subset \bigcup_{j=1}^\infty V_j$. Next, choose a function $g \in C_c(\mathbb{R})$ with $|g(x)| \leq 1$ and $K_g = \text{supp } g \subset V$. Since K_g is a compact set, there exists k so that $K_g \subset \bigcup_{j=1}^k V_j$. Consider smooth functions ζ_j such that supp $\zeta_j \subset V_j$ and

$$\sum_{j=1}^k \zeta_j(x) \equiv 1 \text{ on } K_g.$$

Then we can write

$$g = \sum_{j=1}^k g\zeta_j,$$

so, as $|g(\zeta_j)| \leq 1$ on V_j and supp $\zeta_j \subset V_j$, we get

$$|L(g)| \leq \sum_{j=1}^k |L(g\zeta_j)| \leq \sum_{j=1}^k \mu^*(V_j).$$

Since this is true for all functions g supported in V with $|g| \leq 1$, we have

$$\mu^*(V) \leq \sum_{j=1}^\infty \mu^*(V_j).$$

Next, let A and A_j, $j \geq 1$ be arbitrary sets with $A \subseteq \bigcup_{j=1}^\infty A_j$. Given $\varepsilon > 0$ choose open sets V_j such that $A_j \subset V_j$ and

$$\mu^*(A_j) \geq \mu^*(V_j) - \varepsilon/2^j.$$

Then we have $A \subset V := \bigcup_{j=1}^\infty V_j$ and thus

$$\mu^*(A) \leq \mu^*(V) \leq \sum_{j=1}^\infty \mu^*(V_j) \leq \sum_{j=1}^\infty \left(\mu^*(A_j) + \frac{\varepsilon}{2^j}\right) = \varepsilon + \sum_{j=1}^\infty \mu^*(A_j).$$

As this is true for all $\varepsilon > 0$ we conclude that μ is measure.

To see that μ is a Borel measure we use the following criterion due to Caratheodory.

Lemma 7.8 *Let μ be a measure on $\mathbb{R}^n$. If $\mu^*(A \bigcup B) = \mu^*(A) + \mu^*(B)$ for all sets $A, B \subseteq \mathbb{R}^n$ with dist$(A, B) > 0$ then μ is a Borel measure.*

We postpone the proof of the Caratheodory criterion for the moment, as it is not directly related to the crux of the matter in the proof of the Riesz representation theorem.

Now, if U_1 and U_2 are two open sets such that $\text{dist}(U_1, U_2) > 0$ then

$$\mu^*(U_1 \cup U_2) = \mu^*(U_1) + \mu^*(U_2) \tag{7.9}$$

simply be the definition of μ. Then, for any pair of sets A_1 and A_2 with $\text{dist}(A_1, A_2) > 0$ we can find open sets V_1 and V_2 with $\text{dist}(V_1, V_2) > 0$ which contain A_1 and A_2, respectively. Then, for any open set V containing $A_1 \cup A_2$ we can set $U_1 = V \cap V_1$, $U_2 = V \cap V_2$, then (7.9) implies that

$$\mu^*(V) = \mu^*(U_1) + \mu^*(U_2) \geq \mu^*(A_1) + \mu^*(A_2),$$

thus

$$\mu^*(A_1 \cup A_2) \geq \mu^*(A_1) + \mu^*(A_2),$$

and the measure μ is Borel. The definition of μ as an outer measure immediately implies that μ is Borel regular: for any set A we can choose open sets V_k containing A such that $\mu(V_k) \leq \mu^*(A_k) + 1/k$, then the Borel set $V = \bigcap_{k=1}^{\infty} V_k$ contains A and $\mu(V) = \mu^*(A)$. Finally, (7.5) and the definition of μ imply that $\mu(K) < +\infty$ for any compact set K and thus μ is a Radon measure.

Step 2. Next, in order to show that λ_e introduced in (7.8) is a bounded linear functional, consider first the functional λ defined by (7.6) on $C_c^+(\mathbb{R}^n)$. Let us show that λ is linear, that is,

$$\lambda(f_1 + f_2) = \lambda(f_1) + \lambda(f_2). \tag{7.10}$$

Let $f_1, f_2 \in C_c^+(\mathbb{R}^n)$, take arbitrary functions $g_1, g_2 \in C_c(\mathbb{R}^n; \mathbb{R}^m)$ such that $|g_1| \leq f_1$, $|g_2| \leq f_2$ and consider $g_1' = g_1 \text{sgn}(L(g_1))$, $g_2' = g_2 \text{sgn}(L(g_2))$. Then $|g_1' + g_2'| \leq f_1 + f_2$, and thus

$$|L(g_1)| + |L(g_2)| = L(g_1') + L(g_2') = L(g_1' + g_2') \leq \lambda(f_1 + f_2).$$

It follows that

$$\lambda(f_1) + \lambda(f_2) \leq \lambda(f_1 + f_2), \tag{7.11}$$

so that λ is super-linear. On the other hand, given $g \in C_c(\mathbb{R}^n; \mathbb{R}^m)$ such that $|g| \leq f_1 + f_2$ we may set, for $j = 1, 2$:

$$g_j(x) = \begin{cases} \dfrac{f_j(x) g(x)}{f_1(x) + f_2(x)}, & \text{if } f_1(x) + f_2(x) > 0, \\ 0, & \text{if } f_1(x) + f_2(x) = 0. \end{cases}$$

It is easy to check that g_1 and g_2 are continuous functions with compact support. Then, as $g = 0$ where $f_1 + f_2 = 0$, we have $g = g_1 + g_2$, and $|g_j(x)| \leq f_j(x)$, $j = 1, 2$, for all $x \in \mathbb{R}$. It follows that

$$|L(g)| \leq |L(g_1)| + |L(g_2)| \leq \lambda(f_1) + \lambda(f_2),$$

thus $\lambda(f_1 + f_2) \leq \lambda(f_1) + \lambda(f_2)$, which, together with (7.11) implies (7.10).

Step 3. The next step is to show that λ has the explicit form (7.7).

Lemma 7.9 *For any function $f \in C_c^+(\mathbb{R}^n)$ we have*

$$\lambda(f) = \int_{\mathbb{R}^n} f d\mu. \tag{7.12}$$

Proof. Once again, we would like to approximate f by piecewise constant functions – this, however, is impossible if we want to keep the approximants in $C_c^+(\mathbb{R}^n)$. Instead, given a function $f \in C_c^+(\mathbb{R}^n)$ choose a partition

$$0 = t_0 < t_1 < \ldots < t_N = 2\|f\|_{L^\infty},$$

with $0 < t_i - t_{i-1} < \varepsilon$ and so that $\mu(f^{-1}\{t_j\}) = 0$ for $j = 1, \ldots, N$. The sets $U_j = f^{-1}(t_{j-1}, t_j)$, are bounded and open, hence $\mu(U_j) < \infty$. We will now construct functions h_j such that $0 \le h_j \le 1$, $\mathrm{sup} h_j \subset U_j$, and

$$\mu(U_j) - \frac{\varepsilon}{N} \le \lambda(h_j) \le \mu(U_j). \tag{7.13}$$

As μ is a Radon measure, there exist compact sets $K_j \subseteq U_j$ with

$$\mu(U_j \setminus K_j) < \varepsilon/N.$$

There also exist functions $g_j \in C_c(\mathbb{R}^n; \mathbb{R}^m)$ with $|g_j| \le 1$, supp $g_j \subseteq U_j$, and

$$|L(g_j)| \ge \mu(U_j) - \varepsilon/N,$$

We choose $h_j \in C_c^+(\mathbb{R}^n)$ so that

$$\text{supp } h_j \subseteq U_j,\ 0 \le h_j \le 1 \text{ and } h_j \equiv 1 \text{ on the compact set } K_j \cup \text{supp } g_j.$$

Then $h_j \ge |g_j|$ and thus

$$\lambda(h_j) \ge |L(g_j)| \ge \mu(U_j) - \varepsilon/N,$$

while $\lambda(h_j) \le \mu(U_j)$ since supp $h_j \subseteq U_j$ and $0 \le h_j \le 1$. Summarizing, we have (7.13). Consider the open set

$$A = \{x : \ f(x) > 0 \text{ and } 0 \le h_j(x) < 1 \text{ for all } 1 \le j \le N\},$$

then

$$\mu(A) = \mu\left(\bigcup_{j=1}^{N} (U_j \setminus \{h_j = 1\}\right) \le \sum_{j=1}^{N} \mu(U_j \setminus K_j) < \varepsilon.$$

This gives an estimate

$$\lambda(f - f\sum_{j=1}^{N} h_j) = \sup\left\{|L(g)| : \ g \in C_c(\mathbb{R}^n; \mathbb{R}^m), |g| \le f(1 - \sum_{j=1}^{N} h_j)\right\}$$
$$\le \sup\left\{|L(g)| : \ g \in C_c(\mathbb{R}^n; \mathbb{R}^m), |g| \le \|f\|_{L^\infty}\chi_A\right\} = \|f\|_{L^\infty}\mu(A) \le \varepsilon\|f\|_{L^\infty}.$$

It follows that

$$\lambda(f) \le \sum_{j=1}^{N} \lambda(fh_j) + \varepsilon\|f\|_{|L^\infty} \le \sum_{j=1}^{N} t_j\mu(U_j) + \varepsilon\|f\|_{|L^\infty},$$

and

$$\lambda(f) \ge \sum_{j=1}^{N} \lambda(fh_j) \ge \sum_{j=1}^{N} t_{j-1}(\mu(U_j) - \frac{\varepsilon}{N}) \ge \sum_{j=1}^{N} t_{j-1}\mu(U_j) - 2\varepsilon\|f\|_{L^\infty}.$$

As a consequence,

$$\left|\lambda(f) - \int_{\mathbb{R}^n} f d\mu\right| \le \sum_{j=1}^{N} (t_j - t_{j-1})\mu(U_j) + 3\varepsilon\|f\|_{|L^\infty} \le \varepsilon\mu(\mathrm{supp} f) + 3\varepsilon\|f\|_{|L^\infty},$$

and thus (7.12) holds. □

Step 4. We now construct the function σ.

Lemma 7.10 *There exists a μ-measurable function $\sigma : \mathbb{R}^n \to \mathbb{R}^m$ such that*

$$L(f) = \int_{\mathbb{R}^n} (f \cdot \sigma) d\mu. \tag{7.14}$$

Proof. For a fixed vector $e \in \mathbb{R}^n$ with $|e| = 1$ and $f \in C_c(\mathbb{R}^n)$ define $\lambda_e(f) = L(fe)$. Then, λ_e is a linear functional on $C_c(\mathbb{R}^n)$ and

$$|\lambda_e(f)| \le \sup\{|L(g)| : \; g \in C_c(\mathbb{R}^n; \mathbb{R}^m), \; |g| \le |f|\} \le \lambda(|f|) = \int_{\mathbb{R}^n} |f| d\mu. \tag{7.15}$$

Thus, λ_e can be extended to a bounded linear functional on $L^1(\mathbb{R}^n, d\mu)$, hence by the Riesz representation theorem for L^p-spaces there exists $\sigma_e \in L^\infty(\mathbb{R}^n, d\mu)$ such that

$$\lambda_e(f) = \int_{\mathbb{R}^n} f \sigma_e d\mu. \tag{7.16}$$

Moreover, (7.15) implies that, as a bounded linear functional on $L^1(\mathbb{R}^n, d\mu)$, λ_e has the norm $\|\lambda_e\| \le 1$. Therefore, $\|\sigma_e\|_{L^\infty(\mathbb{R}^n, d\mu)} \le 1$ as well. Setting

$$\sigma = \sum_{j=1}^{m} \sigma_{e_j} e_j,$$

where e_j is the standard basis in $\mathbb{R}^n$ we obtain

$$L(f) = \sum_{j=1}^{m} L((f \cdot e_j)e_j) = \sum_{j=1}^{m} \int_{\mathbb{R}^n} (f \cdot e_j)\sigma_{e_j} d\mu = \int_{\mathbb{R}^n} (f \cdot \sigma) d\mu,$$

which is (7.14). □

Step 5. The last step is

Lemma 7.11 *The function σ defined above satisfies $|\sigma| = 1$ μ-a.e.*

Proof. Let U be an open set, $\mu(U) < +\infty$ and set $\sigma'(x) = \sigma(x)/|\sigma(x)|$ where $\sigma(x) \ne 0$, and $\sigma'(x) = 0$ where $\sigma(x) = 0$. Using Theorem 2.3 and Corollary 2.12, we may find a compact set $K_j \subset U$ such that $\mu(U \setminus K_j) < 1/j$ and σ' is continuous on K_j. We can extend σ' to a continuous function f_j on all of $\mathbb{R}^n$ so that $|f_j| \le 1$. Next, since K_j is a proper compact subset of an open set U we can find a cut-off function $h_j \in C_c(\mathbb{R}^n)$ such that

$$0 \le h_j \le 1, \; h_j \equiv 1 \text{ on } K_j \subseteq U, \text{ and } h_j = 0 \text{ outside of } U.$$

This produces a sequence of functions $g_j = f_j h_j$ such that

$$|g_j| \le 1, \text{ supp } g_j \in U \text{ and } g_j \cdot \sigma \to |\sigma| \text{ in probability on } U.$$

Using Proposition 2.15, we may pass to a subsequence $j_k \to +\infty$ so that $g_{j_k} \cdot \sigma \to |\sigma|$ μ-a.e. in U. Then, as $|g_j| \le 1$, $|\sigma| \le \sqrt{m}$ and $\mu(U) < +\infty$, bounded convergence theorem implies that

$$\int_U |\sigma| d\mu = \lim_{k\to+\infty} \int_U (g_{j_k} \cdot \sigma) d\mu = \lim_{k\to\infty} L(g_{j_k}) \le \mu(U), \tag{7.17}$$

by the definition of the measure μ. On the other hand, for any function $f \in C_c(\mathbb{R}^n; \mathbb{R}^m)$ supported inside U with $|f| \le 1$ we have

$$L(f) = \int_U (f \cdot \sigma) d\mu \le \int_U |\sigma| d\mu,$$

thus

$$\mu(U) \le \int_U |\sigma| d\mu. \tag{7.18}$$

Putting (7.17) and (7.18) together we conclude that $|\sigma| = 1$ μ-a.e. in U.

Step 6. Finally, we prove the Caratheodory criterion, Lemma 7.8. Let μ satisfy the assumptions of this lemma and let C be a closed set. We need to show that for any set A

$$\mu^*(A) \ge \mu^*(A \cap C) + \mu^*(A \setminus C). \tag{7.19}$$

If $\mu^*(A) = +\infty$ this is trivial, so we assume that $\mu^*(A) < +\infty$. Define the sets

$$C_n = \{x \in \mathbb{R}^n : \ \text{dist}(x, C) \le 1/n\}.$$

Then $\text{dist}(A \setminus C_n, A \cap C) \ge 1/n$, thus, by the assumption of Lemma 7.8,

$$\mu^*(A \setminus C_n) + \mu^*(A \cap C) = \mu^*((A \setminus C_n) \cup (A \cap C)) \le \mu^*(A). \tag{7.20}$$

We claim that

$$\lim_{n\to\infty} \mu^*(A \setminus C_n) = \mu^*(A \setminus C). \tag{7.21}$$

Indeed, consider the annuli

$$R_k = \Big\{x \in A : \ \frac{1}{k+1} < \text{dist}(x, C) \le \frac{1}{k}\Big\}$$

As C is closed, we have $\mathbb{R}^n \setminus C = \bigcup_{k=1}^\infty R_k$. Moreover, $\text{dist}(R_k, R_j) > 0$ if $|k - j| \ge 2$, hence

$$\sum_{k=1}^m \mu^*(R_{2k}) = \mu^*\Big(\bigcup_{k=1}^m R_{2k}\Big) \le \mu^*(A),$$

and

$$\sum_{k=1}^m \mu^*(R_{2k-1}) = \mu^*\Big(\bigcup_{k=1}^m R_{2k-1}\Big) \le \mu^*(A),$$

both for all $m \geq 1$. It follows that $\sum_{k=1}^{\infty} \mu^*(R_k) < +\infty$. In that case

$$(A \setminus C) = (A \setminus C_n) \bigcup \Big(\bigcup_{k=n}^{\infty} R_k \Big),$$

thus

$$\mu^*(A \setminus C_n) \leq \mu^*(A \setminus C) \leq \mu^*(A \setminus C_n) + \sum_{k=n}^{\infty} \mu^*(R_k),$$

and (7.21) follows. Passing to the limit $n \to +\infty$ in (7.20) with the help of (7.21) we obtain (7.19). Therefore, all closed sets are μ-measurable, thus the measure μ is Borel. □

8 The Fourier transform on the circle

8.1 Pointwise convergence on $\mathbb{S}^1$

Given a function $f \in L^1(\mathbb{S}^1)$ (here $\mathbb{S}^1$ is the unit circle), or equivalently, a periodic function $f \in L^1[0,1]$, we define the Fourier coefficients, for $k \in \mathbb{Z}$:

$$\hat{f}(k) = \int_0^1 f(x) e^{-2\pi i k x} dx.$$

Trivially, we have $|\hat{f}(k)| \leq \|f\|_{L^1}$ for all $k \in \mathbb{Z}$. The Riemann-Lebesgue lemma shows that an L^1-signal can not have too much high-frequency content and $\hat{f}(k)$ have to decay for large k.

Lemma 8.1 *(The Riemann-Lebesgue lemma) If $f \in L^1(\mathbb{S}^1)$ then $\hat{f}(k) \to 0$ as $k \to +\infty$.*

Proof. Note that

$$\hat{f}(k) = \int_0^1 f(x) e^{-2\pi i k x} dx = -\int_0^1 f(x) e^{-2\pi i k (x + 1/(2k))} dx = -\int_0^1 f(x - \frac{1}{2k}) e^{-2\pi i k x} dx,$$

and thus

$$\hat{f}(k) = \frac{1}{2} \int_0^1 \left[f(x) - f(x - \frac{1}{2k}) \right] e^{-2\pi i k x} dx.$$

As a consequence, we have

$$|\hat{f}(k)| \leq \frac{1}{2} \int_0^1 \left| f(x) - f(x - \frac{1}{2k}) \right| dx,$$

hence $\hat{f}(k) \to 0$ as $k \to +\infty$. □

A simple implication of the Riemann-Lebesgue lemma is that

$$\int_0^1 f(x) \sin(mx) dx \to 0$$

as $m \to \infty$ for any $f \in L^1(\mathbb{S}^1)$. Indeed, for $m = 2k$ this is an immediate corollary of Lemma 8.1, while for an odd $m = 2k+1$ we would simply write

$$\int_0^1 f(x)e^{\pi i(2k+1)x}dx = \int_0^1 f(x)e^{\pi ix}e^{2\pi ikx}dx,$$

and apply this lemma to $\tilde{f}(x) = f(x)e^{i\pi x}$.

In order to investigate convergence of the Fourier series

$$\sum_{k=-\infty}^{\infty} \hat{f}(k)e^{2\pi ikx}$$

let us introduce the partial sums

$$S_N f(x) = \sum_{k=-N}^{N} \hat{f}(k)e^{2\pi ikx}.$$

A convenient way to represent $S_N f$ is by writing it as a convolution:

$$S_n f(x) = \int_0^1 f(t) \sum_{k=-N}^{N} e^{2\pi ik(x-t)}dt = \int_0^1 f(x-t)D_N(t)dt.$$

Here the Dini kernel is

$$\begin{aligned} D_N(t) &= \sum_{k=-N}^{N} e^{2\pi ikt} = e^{-2\pi iNt}(1 + e^{2\pi it} + e^{4\pi it} + \ldots + e^{4\pi iNt}) = e^{-2\pi iNt}\frac{e^{2\pi i(2N+1)t} - 1}{e^{2\pi it} - 1} \\ &= \frac{e^{2\pi i(N+1/2)t} - e^{-2\pi i(N+1/2)t}}{e^{\pi it} - e^{-\pi it}} = \frac{\sin((2N+1)\pi t)}{\sin(\pi t)}. \end{aligned}$$

The definition of the Dini kernel as a sum of exponentials implies immediately that

$$\int_0^1 D_N(t)dt = 1 \tag{8.1}$$

for all N, while the expression in terms of sines shows that

$$|D_N(t)| \le \frac{1}{\sin(\pi\delta)}, \quad \delta \le |t| \le 1/2.$$

The "problem" with the Dini kernel is that its L^1-norm is not uniformly bounded in N. Indeed, consider

$$L_N = \int_{-1/2}^{1/2} |D_N(t)|dt. \tag{8.2}$$

Let us show that

$$\lim_{N\to+\infty} L_N = +\infty. \tag{8.3}$$

We compute:

$$
\begin{aligned}
L_N &= 2\int_0^{1/2}\frac{|\sin((2N+1)\pi t)|}{|\sin\pi t|}dt \geq 2\int_0^{1/2}\frac{|\sin((2N+1)\pi t)|}{|\pi t|}dt \\
&- 2\int_0^{1/2}|\sin((2N+1)\pi t)|\left|\frac{1}{\sin\pi t}-\frac{1}{\pi t}\right|dt = 2\int_0^{N+1/2}\frac{|\sin(\pi t)|}{\pi t}dt + O(1) \\
&\geq \frac{2}{\pi}\sum_{k=0}^{N-1}\int_0^1\frac{|\sin\pi t|}{t+k}dt + O(1) \geq C\log N + O(1),
\end{aligned}
$$

which implies (8.3). This means that, first, the sequence D_N does not form an approximation of the delta function in the usual sense, that is D_N does not behave like a kernel of the form $\phi_N(t) = N\phi(Nt)$, with $\phi \in L^1(\mathbb{S}^1)$, and, second, that (8.1) holds because of cancellation of many oscillatory terms, and not because D_N is uniformly bounded in $L^1(\mathbb{S}^1)$. These oscillations may cause difficulties in the convergence of the Fourier series.

Convergence of the Fourier series for regular functions

Nevertheless, for "reasonably regular" functions the Fourier series converges and Dini's criterion for the convergence of the Fourier series is as follows.

Theorem 8.2 *(Dini's criterion) Let $f \in L^1(\mathbb{S}^1)$ satisfy the following condition at the point x: there exists $\delta > 0$ so that*

$$
\int_{|t|<\delta}\left|\frac{f(x+t)-f(x)}{t}\right|dt < +\infty, \tag{8.4}
$$

then $\lim_{N\to\infty} S_N f(x) = f(x)$.

Proof. Let $\delta > 0$ be as in (8.4). It follows from the normalization (8.1) that

$$
S_N f(x) - f(x) = \int_{-1/2}^{1/2}[f(x-t)-f(x)]D_N(t)dt \tag{8.5}
$$

$$
= \int\limits_{|t|\leq\delta}[f(x-t)-f(x)]\frac{\sin((2N+1)\pi t)}{\sin(\pi t)}dt + \int\limits_{\delta\leq|t|\leq 1/2}[f(x-t)-f(x)]\frac{\sin((2N+1)\pi t)}{\sin(\pi t)}dt.
$$

Consider the first term above (with the change of variables $t \to (-t)$):

$$
I_1 = \int_{|t|\leq\delta}[f(x-t)-f(x)]\frac{\sin((2N+1)\pi t)}{\sin(\pi t)}dt = \int_{-1/2}^{1/2} g_x(t)\sin((2N+1)\pi t)dt,
$$

with

$$
g_x(t) = \frac{f(x+t)-f(x)}{\sin(\pi t)}\chi_{[-\delta,\delta]}(t).
$$

Assumption (8.4) means that, as a function of the variable t, and for x fixed, $g_x \in L^1(\mathbb{S}^1)$. The Riemann-Lebesgue lemma implies then that $I_1 \to 0$ as $N \to +\infty$. The second term in (8.5) is treated similarly: the function

$$
r_x(t) = \frac{f(x+t)-f(x)}{\sin(\pi t)}\chi_{[\delta\leq|t|\leq 1/2]}(t)
$$

is uniformly bounded by a constant $C(\delta)$ which depends on δ, thus the Riemann-Lebesgue lemma, once again, implies that

$$I_2 = \int_{|t|\geq\delta} g_x(t)\sin((2N+1)\pi t)dt,$$

vanishes as $N \to 0$ with $\delta > 0$ fixed. □

Another criterion for the convergence of the Fourier series was given by Jordan:

Theorem 8.3 *(Jordan's criterion) If f has bounded variation on some interval $(x-\delta, x+\delta)$ around the point x then*

$$\lim_{N\to+\infty} S_N f(x) = \frac{1}{2}[f(x^+) + f(x^-)]. \tag{8.6}$$

Proof. Let us set $x = 0$ for convenience. As f has a bounded variation on the interval $(-\delta, \delta)$, it is equal to the difference of two monotonic functions, and we can assume without loss of generality that f is monotonically increasing on $(-\delta, \delta)$, and also that $f(0^+) = 0$. Let us write

$$S_N f(0) = \int_{-1/2}^{1/2} f(-t)D_N(t)dt = \int_0^{1/2} [f(t) + f(-t)]D_N(t)dt.$$

Given $\varepsilon > 0$, choose $\delta > 0$ so that $0 \leq f(t) < \varepsilon$ for all $t \in (0, \delta)$. The first term above may be split as

$$\int_0^{1/2} f(t)D_N(t)dt = \int_0^{\delta} f(t)D_N(t)dt + \int_{\delta}^{1/2} f(t)D_N(t)dt = II_1 + II_2,$$

and

$$II_2 = \int_{\delta}^{1/2} f(t)D_N(t)dt \to 0,$$

exactly for the same reason as in the corresponding term I_2 in the proof of Theorem 8.2, since the function $g(t) = f(t)/\sin(\pi t)$ is uniformly bounded on the interval $[\delta, 1/2]$.

In order to treat I_1 we recall the following basic fact: if h is an increasing function on $[a, b]$ and ϕ is continuous on $[a, b]$ then there exists a point $c \in (a, b)$ such that

$$\int_a^b h(x)\phi(x)dx = h(b_-)\int_c^b \phi(x)dx + h(a_+)\int_a^c \phi(x)dx. \tag{8.7}$$

To see that such $c \in (a, b)$ exists, define the function

$$\eta(y) = h(b_-)\int_y^b \phi(x)dx + h(a_+)\int_a^y \phi(x)dx,$$

then η is continuous and

$$\eta(a) = h(b_-)\int_a^b \phi(x)dx \geq \int_a^b h(x)\phi(x)dx,$$

while

$$\eta(b) = h(a_+)\int_a^b \phi(x)dx \leq \int_a^b h(x)\phi(x)dx,$$

thus there exists $c \in [a,b]$ as in (8.7). Therefore, as $f(0^+) = 0$, we have, with some $c \in (0,\delta)$:

$$II_1 = \int_0^\delta f(t) D_N(t) dt = f(\delta^-) \int_c^\delta D_N(t) dt,$$

and

$$\left| \int_c^\delta D_N(t) dt \right| \leq \left| \int_c^\delta \sin(\pi(2N+1)t) \left[\frac{1}{\sin \pi t} - \frac{1}{\pi t} \right] dt \right| + \left| \int_c^\delta \frac{\sin(\pi(2N+1)t)}{\pi t} dt \right|$$
$$\leq \int_0^1 \left| \frac{1}{\sin \pi t} - \frac{1}{\pi t} \right| dt + \sup_{M>0} \left| \int_0^M \frac{\sin(\pi t)}{\pi t} dt \right| = C < +\infty,$$

with the constant $C > 0$ independent of δ. It follows that $|II_1| \leq C\varepsilon$ for all $N \in \mathbb{N}$. This shows that for a monotonically increasing function f:

$$\int_0^{1/2} f(t) D_N(t) dt \to f(0^+), \quad \text{as } N \to +\infty.$$

A change of variables $t \to (-t)$ shows that then for a monotonically decreasing function f we have

$$\int_0^{1/2} f(-t) D_N(t) dt \to f(0^-),$$

and (8.6) follows. $\Box$

The localization principle

The Fourier coefficients are defined non-locally, nevertheless it turns out that if two functions coincide in an interval $(x-\delta, x+\delta)$ then the sums of the corresponding Fourier series coincide at the point x. More precisely, we have the following.

Theorem 8.4 *(Localization theorem) Let $f \in L^1(\mathbb{S}^1)$ and assume that $f \equiv 0$ on an interval $(x-\delta, x+\delta)$. Then*

$$\lim_{N \to \infty} S_N(x) = 0.$$

Proof. Under the assumptions of Theorem 8.4 we have

$$S_N f(x) = \int_{\delta \leq |t| \leq 1} f(x-t) D_N(t) dt = \int g_x(t) \sin((2N+1)\pi t) dt,$$

where the function

$$g_x(t) = \frac{f(x-t)}{\sin(\pi t)} \chi_{\delta \leq |t| \leq 1}(t)$$

is in $L^1(\mathbb{S}^1)$ as a function of t for each x fixed, because of the cut-off around $t = 0$. It follows from the Riemann-Lebesgue lemma that $S_N f(x) \to 0$ as $N \to +\infty$. $\Box$

The du Bois-Raymond example

In 1873, surprisingly, du Bois-Raymond proved that the Fourier series of a continuous function may diverge at a point. We will give an indirect proof of the existence of such function. First, we need first a result from the functional analysis.

Theorem 8.5 *(Banach-Steinhaus theorem) Let X be a Banach space, Y a normed vector space and let $\{T_\alpha\}$ be a family of bounded linear operators $X \to Y$. Then either* $\sup_\alpha \|T_\alpha\| < +\infty$*, or there exists $x \in X$ so that* $\sup_\alpha \|T_\alpha x\|_Y = +\infty$.

Proof. Let us define $\phi_\alpha(x) = \|T_\alpha(x)\|_Y$ – this is a continuous function on X, as

$$|T_\alpha(x) - T_\alpha(x')| = |\|T_\alpha(x)\|_Y - \|T_\alpha(x')\|_Y| \le \|T_\alpha(x - x')\|_Y \le C_\alpha \|x - x'\|_X.$$

Consider also $\phi(x) = \sup_\alpha \phi_\alpha(x)$, and let

$$V_n = \{x \in X : \ \phi(x) > n\} = \bigcup_\alpha \{x \in X : \ \phi_\alpha(x) > n\}.$$

Continuity of $\phi_\alpha(x)$ means that the set V_n is a union of open sets, hence V_n itself is open. Let us assume that V_N is not dense in X for some N. Then there exists $x_0 \in X$ and $r > 0$ such that $\|x\| < r$ implies that $x_0 + x \notin V_N$. Therefore, $\phi(x_0 + x) \le N$ for all such x, thus

$$\|T_\alpha(x_0 + x)\|_Y \le N \text{ for all } x \in B(0, r) \text{ and all } \alpha,$$

and, in particular, $\|T_\alpha(x_0)\| \le N$. As a consequence, we obtain

$$\|T_\alpha(x)\|_Y \le N + \|T_\alpha(x_0)\| \le 2N \text{ for all } x \in B(0, r) \text{ and all } \alpha.$$

It follows that

$$\|T_\alpha\| \le \frac{2N}{r} \text{ for all } \alpha.$$

On the other hand, if all sets V_N are dense, the Baire Category Theorem implies that the intersection $\bar{V} = \bigcap_{n=1}^\infty V_n$ is not empty. In that case, taking $x_0 \in \bar{V}$ we observe that for any n there exists α_n with

$$\|T_{\alpha_n} x_0\|_Y \ge n = (n/\|x_0\|_X)\|x_0\|_X,$$

thus $\|T_{\alpha_n}\| \ge n/\|x_0\|_X$ and thus $\sup_\alpha \|T_\alpha\| = +\infty$. □

We now prove the du Bois-Raymond theorem.

Theorem 8.6 *There exists a continuous function $f \in C(\mathbb{S}^1)$ so that its Fourier series diverges at $x = 0$.*

Proof. Let $X = C(\mathbb{S}^1)$ and $Y = \mathbb{C}$, and define $T_N : X \to Y$ by

$$T_N f = S_N f(0) = \int_{-1/2}^{1/2} f(t) D_N(t) dt.$$

Then

$$\|T_N\| \le L_N = \int_{-1/2}^{1/2} |D_N(t)| dt,$$

and, moreover, as D_N changes sign only finitely many times, we may construct a sequence of continuous functions f_j^N such that $f_j^N \to |D_N|$ pointwise as $j \to +\infty$, $|f_j^N| \le 1$ and $|T_N f_j^N| \ge L_N - 1/j$. It follows that $\|T_N\| = L_N$. Recall (see (8.3)) that

$$\lim_{N\to+\infty} L_N = +\infty. \tag{8.8}$$

With (8.8) in hand we may use the Banach-Steinhaus theorem to conclude that there exists a function $f \in C(\mathbb{S}^1)$ such that $|S_N f(0)| \to +\infty$ as $N \to +\infty$. □

Kolmogorov showed in 1926 that an L^1-function may have a Fourier series that diverges at every point. Then Carelson in 1965 proved that the Fourier series of an L^2-function converges almost everywhere and then Hunt improved this result to an arbitrary L^p for $p > 1$.

8.2 Approximation by trigonometric polynomials

The Cesaro sums

In order to "improve" the convergence of the Fourier series consider the corresponding Cesaro sums

$$\sigma_N f(x) = \frac{1}{N+1}\sum_{k=0}^{N} S_k f(x) = \int_0^1 f(t) F_N(x-t)dt,$$

where F_N is the Fejér kernel

$$\begin{aligned} F_N(t) &= \frac{1}{N+1}\sum_{k=0}^{N} D_k(t) = \frac{1}{(N+1)\sin^2(\pi t)}\sum_{k=0}^{N} \sin(\pi(2k+1)t)\sin(\pi t) \\ &= \frac{1}{2(N+1)\sin^2(\pi t)}\sum_{k=0}^{N} [\cos(2\pi k t) - \cos(2\pi(k+1)t] \\ &= \frac{1}{2(N+1)\sin^2(\pi t)}[1 - \cos(2\pi(N+1)t] = \frac{1}{N+1}\frac{\sin^2(\pi(N+1)t)}{\sin^2(\pi t)}. \end{aligned}$$

The definition and explicit form of F_N show that, unlike the Dini kernel, F_N is non-negative and has L^1-norm

$$\int_0^1 |F_N(t)|dt = 1. \tag{8.9}$$

Moreover, its mass outside of any finite interval around zero vanishes as $N \to +\infty$:

$$\lim_{N\to\infty} \int_{\delta<|t|<1/2} F_N(t)dt = 0 \quad \text{for any } \delta > 0. \tag{8.10}$$

This improvement is reflected in the following approxtmation theorem.

Theorem 8.7 *Let $f \in L^p(\mathbb{S}^1)$, $1 \le p < \infty$, then*

$$\lim_{N\to\infty} \|\sigma_N f - f\|_p = 0. \tag{8.11}$$

Moreover, if $f \in C(\mathbb{S}^1)$, then

$$\lim_{N\to\infty} \|\sigma_N f - f\|_{C(\mathbb{S}^1)} = 0. \tag{8.12}$$

Proof. We use the Minkowski inequality, with the notation $f_t(x) = f(x-t)$:

$$\sigma_N f(x) - f(x) = \int_{-1/2}^{1/2} [f(x-t) - f(x)] F_N(t) dt,$$

thus

$$\|\sigma_N f - f\|_p \leq \int_{-1/2}^{1/2} \|f_t - f\|_p F_N(t) dt = \int\limits_{|t|<\delta} \|f_t - f\|_p F_N(t) dt + \int\limits_{\delta \leq |t| \leq 1/2} \|f_t - f\|_p F_N(t) dt$$
$$= I_N^\delta + II_N^\delta. \tag{8.13}$$

Recall that, for $f \in L^p(\mathbb{S}^1)$, with $1 \leq p < +\infty$ we have

$$\|f_t - f\|_p \to 0 \text{ as } t \to 0.$$

Hence, in order to estimate the first term above, given $\varepsilon > 0$, we may choose δ so small that

$$\|f_t - f\|_p < \varepsilon \text{ for all } t \in (-\delta, \delta), \tag{8.14}$$

then

$$|I_N^\delta| < \varepsilon \int_{-1/2}^{1/2} F_N(t) dt = \varepsilon.$$

Given such δ we choose N_ε so large that for all $N > N_\varepsilon$ we have

$$\int_{\delta \leq |t| \leq 1/2} F_N(t) dt < \varepsilon.$$

This is possible because of (8.10). The second term in (8.13) may then be estimated as

$$|II_N^\delta| \leq 2\|f\|_p \int_{\delta \leq |t| \leq 1/2} F_N(t) dt < 2\varepsilon \|f\|_p.$$

Now, (8.11) follows. In order to prove (8.12) all we need to do is replace (8.14) with the corresponding estimate in $C(\mathbb{S}^1)$ and repeat the above argument. □

Theorem 8.7 has a couple of useful corollaries. First, it shows that the trigonometric polynomials are dense in $L^p(\mathbb{S}^1)$:

Corollary 8.8 *Trigonometric polynomials are dense in $L^p(\mathbb{S}^1)$ for any $p \in [1, +\infty)$.*

Proof. This follows immediately from (8.11) since each $\sigma_N f$ is a trigonometric polynomial, for all N and f. □

Corollary 8.9 *(The Parceval identity) The map $f \to \hat{f}$ is an isometry between $L^2(\mathbb{S}^1)$ and l^2.*

Proof. The trigonometric exponentials $\{e^{2\pi ikx}\}$, $k \in \mathbb{Z}$, form an orthonormal set in $L^2(\mathbb{S}^1)$, which is complete by Corollary 8.8, hence they form a basis of $L^2(\mathbb{S}^1)$ and the Fourier coefficients of $f \in L^2(\mathbb{S}^1)$ are the coefficients in the expansion

$$f(x) = \sum_{k \in \mathbb{Z}} \hat{f}(k) e^{2\pi ikx},$$

so that

$$\sum_{k\in\mathbb{Z}}|\hat{f}(k)|^2=\int_0^1|f(x)|^2dx, \tag{8.15}$$

by the standard Hilbrt space theory used here for $L^2(\mathbb{S}^1)$. □

Corollary 8.10 *Let $f\in L^2(\mathbb{S}^1)$, then $\|S_Nf-f\|_2\to 0$ as $N\to+\infty$.*

Proof. This is a consequence of Corollary 8.9:

$$\begin{aligned}\|S_Nf-f\|_2^2&=\|S_Nf\|_2^2+\|f\|^2-2\langle S_Nf,f\rangle=\|S_Nf\|_2^2+\|f\|^2-2\langle S_Nf,S_Nf\rangle\\&=\|f\|^2-\|S_Nf\|_2^2\to 0,\end{aligned}$$

as $N\to+\infty$ by (8.15). □

Another useful immediate consequence if Theorem 8.7 is

Corollary 8.11 *Let $f\in L^1(\mathbb{S}^1)$ be such that $\hat{f}(k)=0$ for all $k\in\mathbb{Z}$. Then $f=0$.*

Ergodicity of irrational rotations

Corollary 8.11 itself has an interesting implication. Consider the shift $T_\alpha:\mathbb{S}^1\to\mathbb{S}^1$ on the unit circle by a number α: $T_\alpha(x)=(x+\alpha)\text{mod } 1$. The map T_α is invertible and preserves the Lebesgue measure: $m(R)=m(T_\alpha(R))$ for any Lebesgue measurable set $R\subseteq\mathbb{S}^1$. It turns out that for $\alpha\notin\mathbb{Q}$ this map is ergodic. This means that if T_α leaves a measurable set $R\subseteq\mathbb{S}^1$ invariant: $T_\alpha(R)=R$, then either $m(R)=1$ or $m(R)=0$. Indeed, let α be irrational and R be a T_α-invariant set. Then the characteristic function χ_R of the set R satisfies

$$\chi_R^\alpha(x):=\chi_R(x+\alpha)=\chi_R(x),\quad \text{for all } x\in\mathbb{S}^1. \tag{8.16}$$

On the other hand, the Fourier transform of χ_R^α is

$$\hat{\chi}_R^\alpha(k)=\int_0^1\chi_R(x+\alpha)e^{-2\pi ik\cdot x}dx=\int_0^1\chi_R(x)e^{-2\pi ik\cdot(x-\alpha)}dx=\hat{\chi}_R(k)e^{2\pi ik\alpha}.$$

Comparing this to (8.16), we see that

$$\hat{\chi}_R(k)e^{2\pi ik\alpha}=\hat{\chi}_R(k)$$

for all $k\in\mathbb{Z}$, which, as α is irrational, implies that $\hat{\chi}_R(k)=0$ for all $k\neq 0$, hence χ_R is equal almost everywhere to a constant. This constant can be equal only to zero or one. In the former case R has measure zero, in the latter its measure is equal to one.

9 The Fourier transform in $\mathbb{R}^n$

We now look at the basic properties of the Fourier transform in the whole space. Given a function $f\in L^1(\mathbb{R}^n)$, its Fourier transform is

$$\hat{f}(\xi)=\int f(x)e^{-2\pi ix\cdot\xi}dx.$$

Then, obviously, $\hat{f} \in L^\infty(\mathbb{R}^n)$, with $\|\hat{f}\|_{L^\infty} \le \|f\|_{L^1}$. Moreover, the function $\hat{f}(\xi)$ is continuous:

$$\hat{f}(\xi) - \hat{f}(\xi') = \int f(x)\left(e^{-2\pi i x\cdot\xi} - e^{-2\pi i x\cdot\xi'}\right)dx \to 0$$

as $\xi' \to \xi$, by the Lebesgue dominated convergence theorem since $f \in L^1(\mathbb{R}^n)$. The Riemann-Lebesegue lemma is easily generalized to the Fourier transform on $\mathbb{R}^n$, and

$$\lim_{\xi\to\infty} \hat{f}(\xi) = 0.$$

9.1 The Schwartz class $\mathcal{S}(\mathbb{R}^n)$

For a smooth compactly supported function $f \in C_c^\infty(\mathbb{R}^n)$, we have the following remarkable algebraic relations between taking derivatives and multiplying by polynomials:

$$\widehat{\frac{\partial f}{\partial x_i}}(\xi) = 2\pi i \xi \hat{f}(\xi), \tag{9.1}$$

and

$$(-2\pi i)(\widehat{x_j f})(\xi) = \frac{\partial \hat{f}}{\partial \xi_j}(\xi). \tag{9.2}$$

This motivates the following definition.

Definition 9.1 *The Schwartz class $\mathcal{S}(\mathbb{R}^n)$ consists of functions f such that for any pair of multi-indices α and β*

$$p_{\alpha\beta}(f) := \sup_x |x^\alpha D^\beta f(x)| < +\infty.$$

As $C_c^\infty(\mathbb{R}^n)$ lies inside the Schwartz class, the Schwartz functions are dense in $L^1(\mathbb{R}^n)$.

Convergence in $\mathcal{S}(\mathbb{R}^n)$ is defined as follows: a sequence $\phi_k \to 0$ in $\mathcal{S}(\mathbb{R}^n)$ if

$$\lim_{k\to\infty} p_{\alpha\beta}(\phi_k) = 0 \tag{9.3}$$

for all multi-indices α, β. Note that if $\phi_k \to 0$ in $\mathcal{S}(\mathbb{R}^n)$, then all functions

$$\psi_k^{\alpha\beta}(x) = x^\alpha D^\beta \phi_k(x)$$

converge to zero as $k \to +\infty$ not only in $L^\infty(\mathbb{R}^n)$ (which is directly implied by (9.3)) but also in any $L^p(\mathbb{R}^n)$, $1 \le p \le +\infty$ because

$$\int |\psi_k^{\alpha\beta}|^p dx \le \int_{|x|\le 1} |\psi_n^{\alpha\beta}|^p dx + 2\int_{|x|\ge 1} \frac{|x|^{n+1}|\psi_n^{\alpha\beta}|^p}{1+|x|^{n+1}} dx \le C_n |p_{\alpha\beta}|^p + 2C_n' |p_{\alpha'\beta}|^p,$$

with $\alpha' = \alpha + (n+1)/p$ and the constants C_n and C_n' that depend only on the dimension n.

The main reason to introduce the Schwartz class is the following theorem.

Theorem 9.2 *The Fourier transform is a continuous map $\mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n)$ such that*

$$\int_{\mathbb{R}^n} f(x)\hat{g}(x)dx = \int_{\mathbb{R}^n} \hat{f}(x)g(x)dx, \tag{9.4}$$

and

$$f(x) = \int \hat{f}(\xi)e^{2\pi i x\cdot\xi}d\xi \tag{9.5}$$

for all $f, g \in \mathcal{S}(\mathbb{R}^n)$.

Proof. We begin with a lemma that is one of the cornerstones of the probability theory.

Lemma 9.3 *Let $f(x) = e^{-\pi|x|^2}$, then $\hat{f}(x) = f(x)$.*

Proof. First, as

$$f(x) = e^{-\pi|x_1|^2}e^{-\pi|x_2|^2}\dots e^{-\pi|x_n|^2},$$

so that both f and $\hat{f}$ factor into a product of functions of one variable, it suffices to consider the case $n = 1$. The proof is a glimpse of how useful the Fourier transform is for differential equations and vice versa: the function $f(x)$ satisfies an ordinary differential equation

$$u' + 2xu = 0, \tag{9.6}$$

with the boundary condition $u(0) = 1$. However, relations (9.1) and (9.2) together with (9.6) imply that $\hat{f}$ satisfies the same differential equation (9.6), with the same boundary condition $\hat{f}(0) = 0$. It follows that $f(x) = \hat{f}(x)$ for all $x \in \mathbb{R}$. □

We continue with the proof of Theorem 9.2. Relations (9.1) and (9.2) imply that if $f_k \to 0$ in $\mathcal{S}(\mathbb{R}^n)$ then for any pair of multi-indices α, β:

$$\sup_{x\in\mathbb{R}^n} |\xi^\alpha D^\beta \hat{f}_k(\xi)| \le C\|D^\alpha(x^\beta f_k)\|_{L^1} \to 0,$$

thus $\hat{f}_k \to 0$ in $\mathcal{S}(\mathbb{R}^n)$ as well, hence the Fourier transform is a continuous map $\mathcal{S} \to \mathcal{S}$.

The Parceval identity can be verified directly using Fubini's theorem:

$$\int_{\mathbb{R}^n} f(x)\hat{g}(x)dx = \int_{\mathbb{R}^{2n}} f(x)g(\xi)e^{-2\pi i\xi\cdot x}dxd\xi = \int_{\mathbb{R}^n} \hat{f}(\xi)g(\xi)d\xi.$$

Finally, we prove the inversion formula using a rescaling argument. Let $f, g \in \mathcal{S}(\mathbb{R}^n)$ then for any $\lambda > 0$ we have

$$\int_{\mathbb{R}^n} f(x)\hat{g}(\lambda x)dx = \int_{\mathbb{R}^{2n}} f(x)g(\xi)e^{-2\pi i\lambda\xi\cdot x}dx = \int \hat{f}(\lambda\xi)g(\xi)d\xi = \frac{1}{\lambda^n}\int_{\mathbb{R}^n} \hat{f}(\xi)g\left(\frac{\xi}{\lambda}\right)d\xi.$$

Multiplying by λ^n and changing variables on the left side we obtain

$$\int_{\mathbb{R}^n} f\left(\frac{x}{\lambda}\right)\hat{g}(x)dx = \int_{\mathbb{R}^n} \hat{f}(\xi)g\left(\frac{\xi}{\lambda}\right)d\xi.$$

Letting now $\lambda \to \infty$ using the Lebesgue dominated convergence theorem gives

$$f(0)\int_{\mathbb{R}^n} \hat{g}(x)dx = g(0)\int_{\mathbb{R}^n} \hat{f}(\xi)d\xi, \tag{9.7}$$

for all functions f and g in $\mathcal{S}(\mathbb{R}^n)$. Taking $g(x) = e^{-\pi|x|^2}$ in (9.7) and using Lemma 9.3 leads to

$$f(0) = \int_{\mathbb{R}^n} f(\xi)d\xi. \tag{9.8}$$

The inversion formula (9.5) now follows if we apply (9.8) to a shifted function $f_y(x) = f(x+y)$, because

$$\hat{f}_y(\xi) = \int_{\mathbb{R}^n} f(x+y)e^{-2\pi i\xi\cdot x}dx = e^{2\pi i\xi\cdot y}\hat{f}(\xi),$$

so that

$$f(y) = f_y(0) = \int_{\mathbb{R}^n} \hat{f}_y(\xi)d\xi = \int_{\mathbb{R}^n} e^{2\pi i\xi\cdot y}\hat{f}(\xi)d\xi,$$

which is (9.5). □

The Schwartz distributions

Definition 9.4 *The space $\mathcal{S}'(\mathbb{R}^n)$ of Schwartz distirbutions is the space of linear functionals T on $\mathcal{S}(\mathbb{R}^n)$ such that $T(\phi_k) \to 0$ for all sequences $\phi_k \to 0$ in $\mathcal{S}(\mathbb{R}^n)$.*

Theorem 9.2 allows us to extend the Fourier transform to distributions in $\mathcal{S}'(\mathbb{R}^n)$ by setting $\hat{T}(f) = T(\hat{f})$ for $T \in \mathcal{S}'(\mathbb{R}^n)$ and $f \in \mathcal{S}(\mathbb{R}^n)$. The fact that $\hat{T}(f_k) \to 0$ for all sequences $f_k \to 0$ in $\mathcal{S}(\mathbb{R}^n)$ follows from the continuity of the Fourier transform as a map $\mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n)$, hence $\hat{T}$ is a Schwartz distribution for all $T \in \mathcal{S}'(\mathbb{R}^n)$. For example, if δ_0 is the Schwartz distribution such that $\delta_0(f) = f(0)$, $f \in \mathcal{S}(\mathbb{R}^n)$, then

$$\hat{\delta}_0(f) = \hat{f}(0) = \int_{\mathbb{R}^n} f(x)dx,$$

so that $\hat{\delta}(\xi) \equiv 1$ for all $\xi \in \mathbb{R}^n$.

Similarly, since differentiation is a continuous map $\mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n)$, we may define the distributional derivative as

$$\frac{\partial T}{\partial x_j}(f) = -T\left(\frac{\partial f}{\partial x_j}\right),$$

for all $T \in \mathcal{S}'(\mathbb{R}^n)$ and $f \in \mathcal{S}(\mathbb{R}^n)$ – the minus sign here comes from the integration by parts formula, for if T happens to have the form

$$T_g(f) = \int_{\mathbb{R}^n} f(x)g(x)dx,$$

with a given $g \in \mathcal{S}(\mathbb{R}^n)$, then

$$T\left(\frac{\partial f}{\partial x_j}\right) = \int_{\mathbb{R}^n} \frac{\partial f}{\partial x_j}(x)g(x)dx = -\int_{\mathbb{R}^n} f(x)\frac{\partial g}{\partial x_j}(x)dx.$$

For instance, in one dimension $\delta_0(x) = 1/2(\mathrm{sgn}(x))'$ in the distributional sense because for any function $f \in \mathcal{S}(\mathbb{R})$ we have

$$\langle(\mathrm{sgn})', f\rangle = -\langle\mathrm{sgn}, f'\rangle = -\int_{-\infty}^{\infty} \mathrm{sgn}(x)f'(x)dx = \int_{-\infty}^{0} f'(x)dx - \int_{0}^{\infty} f'(x)dx = 2f(0) = 2\langle\delta_0, f\rangle.$$

9.2 The law of large numbers and the central limit theorem

The law of large numbers and the central limit theorem deal with the question of how a sum of the large number of identically distributed random variables behaves. We will not discuss them here in great detail but simply explain how the Fourier transform is useful in this problem. Let X_j be a sequence of real-valued independent, identically distributed random variables with mean zero and finite variance:

$$\mathbb{E}(X_n) = 0, \quad \mathbb{E}(X_n^2) = D < +\infty. \tag{9.9}$$

Let us define

$$Z_n = \frac{X_1 + X_2 + \ldots + X_n}{n}. \tag{9.10}$$

Recall that if X and Y are two independent random variables with probability densities p_X and p_Y, that is,

$$\mathbb{E}(f(X)) = \int f(x)p_X(x)dx, \quad \mathbb{E}(f(X)) = \int f(x)p_Y(x)dx,$$

then the sum $Z = X + Y$ has the probability density

$$p_Z(x) = (p_X \star p_Y)(x) = \int_{\mathbb{R}} p_X(x-y)p_Y(y)dy.$$

On the other hand, if X has a probability density p_X, the variable $X_\lambda = X/\lambda$ satisfies

$$P(X_\lambda \in A) = P(X \in \lambda A),$$

so that

$$\int_A p_{X_\lambda}(x)dx = \int_{\lambda A} p(x)dx,$$

which means that $p_{X_\lambda}(x) = \lambda p(\lambda x)$.

Going back to the averaged sum Z_n in (9.10), it follows that its probability density is

$$p_n(x) = n\left[p_X \star p_X \star \ldots p_X\right](nx),$$

with the convolution above taken n times. The Fourier transform of a convolution has a simple form

$$\widehat{(f \star g)}(\xi) = \int f(y)g(x-y)e^{-2\pi i\xi\cdot x}dxdy = \int f(y)g(z)e^{-2\pi i\xi\cdot(z+y)}dzdy = \hat{f}(\xi)\hat{g}(\xi). \tag{9.11}$$

Hence, the Fourier transform of p_n is

$$\hat{p}_n(\xi) = \left[\hat{p}_X\left(\frac{\xi}{n}\right)\right]^n.$$

As

$$\int_{\mathbb{R}} p_X(x)dx = 1,$$

we have $\hat{p}_X(0) = 1$. Since X has mean zero,

$$\hat{p}'_X(0) = -2\pi i \int_{\mathbb{R}} x p_X(x) dx = 0, \tag{9.12}$$

and the second derivative at zero is

$$\hat{p}''_X(0) = (-2\pi i)^2 \int_{\mathbb{R}} x^2 p_X(x) dx = -4\pi^2 D. \tag{9.13}$$

We can now compute, with the help of (9.12) and (9.13), for any $\xi \in \mathbb{R}$:

$$\lim_{n\to\infty} \hat{p}_n(\xi) = \lim_{n\to\infty} \left(1 - \frac{2\pi^2 D|\xi|^2}{n^2}\right)^n = 1.$$

As a consequence, for any test function $f \in \mathcal{S}(\mathbb{R})$ we have

$$\mathbb{E}(f(Z_n)) = \int f(x) p_n(x) dx = \int_{\mathbb{R}} \hat{f}(\xi)\hat{p}_n(\xi) d\xi \to \int_{\mathbb{R}} \hat{f}(\xi) d\xi = f(0).$$

Thus, the random variable Z_n converges in law to a non-random value $Z = 0$. This is the weak law of large numbers.

In order to get a non-trivial limit for a sum of random variables we consider "the central limit scaling":

$$R_n = \frac{X_1 + X_2 + \ldots + X_n}{\sqrt{n}}.$$

As we did for Z_n, we may compute the probability density q_n for R_n:

$$q_n(x) = \sqrt{n}\,[p_X \star p_X \star \ldots p_X]\,(\sqrt{n}x),$$

and its Fourier transform is

$$\hat{q}_n(\xi) = \left[\hat{p}_X\left(\frac{\xi}{\sqrt{n}}\right)\right]^n.$$

We may also compute, point-wise in $\xi \in \mathbb{R}^n$ the limit

$$\lim_{n\to\infty} \hat{q}_n(\xi) = \lim_{n\to\infty} \left(1 - \frac{2\pi^2 D|\xi|^2}{n}\right)^n = e^{-2\pi^2 D|\xi|^2},$$

which is now non-trivial. This means that, say, for any function $f(x) \in C_c(\mathbb{R})$ we have

$$\mathbb{E}(f(R_n)) \to \int \hat{f}(\xi) e^{-2\pi^2 D|\xi|^2} d\xi,$$

thus R_n converges in law to a random variable with the Gaussian probability density

$$q(x) = \int e^{2\pi i\xi\cdot x} e^{-2\pi^2 D|\xi|^2} d\xi = \int e^{2\pi i\xi\cdot x/\sqrt{2\pi D}} e^{-\pi|\xi|^2} \frac{d\xi}{\sqrt{2\pi D}} = \frac{e^{-|x|^2/(2D)}}{\sqrt{2\pi D}}.$$

This is the central limit theorem.

9.3 Interpolation in L^p-spaces

A simple example of an interpolation inequality is a bound that tells us that a function f which lies in two spaces $L^{p_0}(\mathbb{R}^n, d\mu)$ and $L^{p_1}(\mathbb{R}^n, d\mu)$ has to lie also in all intermediate spaces $L^p(\mathbb{R}^n, d\mu)$ with $p_0 \leq p \leq p_1$. Indeed, if $p = (1-\alpha)p_0 + \alpha p_1$, $0 < \alpha < 1$, then, by Hölder's inequality,

$$\int |f|^{(1-\alpha)p_0+\alpha p_1} d\mu \leq \left(\int |f|^{p_0} d\mu\right)^{1-\alpha} \left(\int |f|^{p_1} d\mu\right)^{\alpha}.$$

The Riesz-Thorin interpolation theorem

The Riesz-Thorin interpolation theorem deals with the following question, somewhat motivated by above. Let (M,μ) and (N,ν) be two measure spaces and consider an operator A which maps $L^{p_0}(M)$ to a space $L^{q_0}(N)$, and also $L^{p_1}(M)$ to a space $L^{q_1}(N)$. More precisely, there exist operators $A_0 : L^{p_0}(M) \to L^{q_0}(N)$ and $A_1 : L^{p_1}(M) \to L^{q_1}(N)$ so that $A = A_0 = A_1$ on $L^{p_0}(M) \cap L^{p_1}(N)$. The question is whether A can be defined on $L^p(M)$ with $p_0 < p < p_1$, and what is its target space. Let us define $p_t \in (p_0, p_1)$ and $q_t \in (q_0, q_1)$ by

$$\frac{1}{p_t} = \frac{t}{p_1} + \frac{1-t}{p_0}, \quad \frac{1}{q_t} = \frac{t}{q_1} + \frac{1-t}{q_0}, \quad 0 \leq t \leq 1, \tag{9.14}$$

as well as

$$k_0 = \|A\|_{L^{p_0}(M)\to L^{q_0}(N)}, \quad k_1 = \|A\|_{L^{p_1}(M)\to L^{q_1}(N)}.$$

Theorem 9.5 *(The Riesz-Thorin interpolation theorem) For any $t \in [0,1]$ there exists a bounded linear operator $A_t : L^{p_t}(M) \to L^{q_t}(N)$ that coincides with A on $L^{p_0}(M) \cap L^{p_1}(M)$ and whose operator norm satisfies*

$$\|A_t\|_{L^{p_t}(M)\to L^{q_t}(N)} \leq k_0^{1-t} k_1^t. \tag{9.15}$$

Before proving the Riesz-Thorin interpolation theorem we mention some of its implications. We already know that the Fourier transform maps $L^1(\mathbb{R}^n)$ to $L^\infty(\mathbb{R}^n)$ and $L^2(\mathbb{R}^n)$ to itself. This allows us to extend the Fourier transform to all intermediate spaces $L^p(\mathbb{R}^n)$ with $1 \leq p \leq 2$.

Corollary 9.6 *(The Hausdorff-Young inequality) If $f \in L^p(\mathbb{R}^n)$ then its Fourier transform $\hat{f} \in L^{p'}(\mathbb{R}^n)$ with $\dfrac{1}{p} + \dfrac{1}{p'} = 1$ and $\|\hat{f}\|_{L^{p'}} \leq \|f\|_{L^p}$.*

Proof. We take $p_0 = 1$, $p_1 = 2$, $q_0 = \infty$, $q_1 = 2$. Then for any $t \in [0,1]$ the corresponding p_t and q_t are given by

$$\frac{1}{p_t} = \frac{1-t}{1} + \frac{t}{2} = 1 - \frac{t}{2}, \quad \frac{1}{q_t} = \frac{t}{2},$$

which means that $1/p_t + 1/q_t = 1$, as claimed. Furthermore, as $\|\hat{f}\|_{L^2} = \|f\|_{L^2}$ by the Parceval identity and $\|\hat{f}\|_{L^\infty} \leq \|f\|_{L^1}$, it follows that $\|\hat{f}\|_{L^{p_t}\to L^{q_t}} \leq 1$. $\square$

The next corollary allows to estimate convolutions.

Corollary 9.7 *Let $f \in L^p(\mathbb{R}^n)$ and $g \in L^q(\mathbb{R}^n)$, then $f \star g \in L^r(\mathbb{R}^n)$, and*

$$\|f \star g\|_{L^r} \leq \|f\|_{L^p} \|g\|_{L^q}, \tag{9.16}$$

with

$$\frac{1}{r} + 1 = \frac{1}{p} + \frac{1}{q}. \tag{9.17}$$

Proof. We do this in two steps. First, fix $g \in L^1(\mathbb{R}^n)$. Obviously, we have

$$\|f \star g\|_{L^1} \leq \int |f(x-y)||g(y)|dydx = \|f\|_{L^1} \|g\|_{L^1}, \tag{9.18}$$

and

$$\|f \star g\|_{L^\infty} \leq \|f\|_{L^\infty} \|g\|_{L^1}. \tag{9.19}$$

The Riesz-Thorin theorem applied to the map $f \to f \star g$ implies then that

$$\|f \star g\|_{L^p} \leq \|g\|_{L^1} \|f\|_{L^p}, \tag{9.20}$$

which is a special case of (9.16) with $q = 1$ and $r = p$. On the other hand, Hölder's inequality implies that

$$\|f \star g\|_{L^\infty} \leq \|f\|_{L^p} \|g\|_{L^{p'}}, \quad \frac{1}{p} + \frac{1}{p'} = 1. \tag{9.21}$$

Let us take $p_0 = 1$, $q_0 = p$, $p_1 = p'$ and $q_1 = \infty$ in the Riesz-Thorin interpolation theorem applied to the mapping $g \to f \star g$, with f fixed. Then (9.20) and (9.21) imply that, for all $t \in [0, 1]$,

$$\|f \star g\|_{L^r} \leq \|f\|_{L^p} \|g\|_{L^q},$$

with

$$\frac{1}{q} = \frac{1}{p_t} = \frac{1-t}{1} + \frac{t}{p'},$$

and

$$\frac{1}{r} = \frac{1}{q_t} = \frac{1-t}{p} + \frac{t}{\infty}.$$

It follows that $t = 1 - p/r$, thus

$$\frac{1}{q} = 1 - (1 - \frac{p}{r}) + \frac{1}{p'}(1 - \frac{p}{r}) = \frac{p}{r} + (1 - \frac{1}{p})(1 - \frac{p}{r}) = 1 - \frac{1}{p} + \frac{1}{r},$$

which is (9.17). □

The next example arises in microlocal analysis. Given a function $a(x, \xi) \in \mathcal{S}(\mathbb{R}^{2n})$ we define a semiclassical operator

$$A(x, \varepsilon D)f = \int e^{2\pi i \xi \cdot x} a(x, \varepsilon \xi) \hat{f}(\xi) d\xi.$$

Corollary 9.8 *The family of operators $A(x, \varepsilon D)$, $0 < \varepsilon \leq 1$, is uniformly bounded from any $L^p(\mathbb{R}^n)$, $1 \leq p \leq +\infty$, to itself.*

Proof. Let us write

$$A(x,\varepsilon D)f=\int e^{2\pi i\xi\cdot x}a(x,\varepsilon\xi)\hat f(\xi)d\xi=\int e^{2\pi i\xi\cdot x+2\pi i\varepsilon\xi\cdot y}\tilde a(x,y)\hat f(\xi)d\xi dy=\int\tilde a(x,y)f(x+\varepsilon y)dy,$$

where $\tilde a(x,y)$ is the Fourier transform of the function $a(x,\xi)$ in the variable ξ. It follows that

$$\|A(x,\varepsilon D)f\|_{L^\infty}\le\|f\|_{L^\infty}\sup_{x\in\mathbb{R}^n}\int|\tilde a(x,y)|dy=C_1(a)\|f\|_{L^\infty},$$

and

$$\|A(x,\varepsilon D)\|_{L^1}\le\int|\tilde a(x,y)||f(x+\varepsilon y)|dydx\le\int(\sup_{z\in\mathbb{R}^n}|\tilde a(z,y)|)|f(x+\varepsilon y)|dydx$$
$$=\|f\|_{L^1}\int(\sup_{z\in\mathbb{R}^n}|\tilde a(z,y)|)dy=C_2(a)\|f\|_{L^1}.$$

The Riesz-Thorin interpolation theorem implies that for any $p\in[1,+\infty]$ there exists $C_p(a)$ which does not depend on $\varepsilon\in(0,1]$ so that $\|A(x,\varepsilon D)\|_{L^p\to L^p}\le C_p$. □

The three lines theorem

A key ingredient in the proof of the Riesz representation theorem is the following basic result from complex analysis.

Theorem 9.9 *Let $F(z)$ be a bounded analytic function in the strip $S=\{z:\ 0\le Re z\le 1\}$, such that $|F(iy)|\le m_0$, $|F(1+iy)|\le m_1$, with $m_0,m_1>0$ for all $y\in\mathbb{R}$. Then*

$$|F(x+iy)|\le m_0^{1-x}m_1^x \text{ for all } 0\le x\le 1,\ y\in\mathbb{R}. \tag{9.22}$$

Proof. It is convenient to set

$$F_1(z)=\frac{F(z)}{m_0^{1-z}m_1^z},$$

so that $|F_1(iy)|\le 1$, $|F_1(1+iy)|\le 1$ and F_1 is uniformly bounded in S. It suffices to show that $|F(x+iy)|\le 1$ for all $(x,y)\in S$ under these assumptions. If the strip S were a bounded domain, this would follow immediately from the maximum modulus principle.

Assume first that $F_1(x+iy)\to 0$ as $|y|\to+\infty$, uniformly in $x\in[0,1]$. Then $|F_1(x\pm iM)|\le 1/2$ for all y with $|y|\ge M$, and $M>0$ large enough. The maximum modulus principle implies that $|F_1(x+iy)|\le 1$ for $|y|\le M$, and, since, $|F_1(x+iy)|\le 1/2$ for all y with $|y|\ge M$, it follows that $|F_1(x+iy)|\le 1$ for all $(x,y)\in S$.

In general, set

$$G_n(z)=F_1(z)e^{(z^2-1)/n},$$

then

$$|G_n(iy)|\le|F_1(iy)|e^{(-y^2-1)/n}\le 1,$$

and

$$|G_n(1+iy)|\le F_1(1+iy)|e^{-y^2}\le 1,$$

but in addition, G_n goes to zero as $|y| \to +\infty$, uniformly in $x \in [0,1]$:

$$|G_n(x+iy)| \le |F_1(z)| e^{(x^2-y^2-1)/n} \le C_0 e^{-y^2/n},$$

with a constant C_0 such that $|F_1(z)| \le C_0$ for all $z \in S$. It follows from the previous part of the proof that $|G_n(z)| \le 1$, hence

$$|F_1(z)| \le e^{(1+y^2)/n},$$

for all $z \in S$ and all $n \in \mathbb{N}$. Letting $n \to +\infty$ we deduce that $|F_1(z)| \le 1$ for all $z \in S$. $\square$

The proof of the Riesz-Thorin interpolation theorem

First, let us define the operator A on $L^{p_t}(M)$ with p_t as in (9.14). Given $f \in L^{p_t}(M)$ we decompose it as

$$f(x) = f_1(x) + f_2(x), \quad f_1(x) = f(x)\chi_{|f|\le 1}(x), \quad f_2(x) = f(x)\chi_{|f|\ge 1}(x).$$

Then, as $p_t \le p_1$:

$$\int_M |f_1|^{p_1} d\mu = \int_M |f|^{p_1}\chi_{|f|\le 1} d\mu \le \int_M |f|^{p_t}\chi_{|f|\le 1} d\mu \le \int_M |f|^{p_t} d\mu = \|f\|_{L^{p_t}}^{p_t},$$

and, as $p_0 \le p_t$:

$$\int_M |f_2|^{p_0} d\mu = \int_M |f|^{p_t}\chi_{|f|\ge 1} d\mu \le \int_M |f|^{p_t}\chi_{|f|\ge 1} d\mu \le \int_M |f|^{p_t} d\mu = \|f\|_{L^{p_t}}^{p_t},$$

so that $f_1 \in L^{p_1}(M)$ and $f_2 \in L^{p_0}(M)$. As A is defined both on $L^{p_0}(M)$ and $L^{p_1}(M)$, we can set

$$Af = Af_1 + Af_2.$$

We need to verify that A maps $L^{p_t}(M)$ to $L^{q_t}(N)$ continuously. Note that the norm of a bounded linear functional $L_f : L^{p'}(M) \to \mathbb{R}$,

$$L_f(g) = \int_M fg d\mu, \quad f \in L^p(M),$$

is $\|L_f\| = \|f\|_{L^p}$, for all $p \in [1, +\infty]$, with

$$\frac{1}{p} + \frac{1}{p'} = 1.$$

To see that, for $f(x) = |f(x)|e^{i\alpha(x)}$ simply take $g(x) = |f(x)|^{p/p'} \exp\{-i\alpha(x)\}$ for $1 < p < +\infty$, $g(x) = \exp\{-i\alpha(x)\}$ for $p = 1$, and $g(x) = \chi_{A_\varepsilon}(x)\exp\{-i\alpha(x)\}$, where A_ε is a set of a finite measure such that $|f(x)| > (1-\varepsilon)\|f|_{L^\infty}$ on A_ε for $p = +\infty$. We conclude that

$$\|f\|_{L^p} = \sup_{\|g\|_{L^{p'}}=1} \int_M fg d\mu, \quad \frac{1}{p} + \frac{1}{p'} = 1.$$

For an operator mapping L^p to L^q we have the corresponding representation for its norm:

$$\|A\|_{L^p(M)\to L^q(N)} = \sup_{\|f\|_{L^p(M)}=1} \|Af\|_{L^q(N)} = \sup_{\substack{\|f\|_{L^p(M)}=1 \\ \|g\|_{L^{q'}(N)}=1}} \int_N (Af) g d\nu. \tag{9.23}$$

We will base our estimate of the norm of $A : L^{p_t}(M) \to L^{q_t}(N)$ on (9.23) . Moreover, as simple functions are dense in $L^{p_t}(M)$ and $L^{q_t'}(N)$, it suffices to use in (9.23) only simple functions f and g with $\|f\|_{L^{p_t}(M)} = \|g\|_{L^{q_t'}(N)} = 1$, of the form

$$f(x) = \sum_{j=1}^{n} a_j e^{i\alpha_j(x)} \chi_{A_j}(x), \quad g(y) = \sum_{j=1}^{m} b_j e^{i\beta_j(y)} \chi_{B_j}(y), \quad x \in M, \ y \in N, \tag{9.24}$$

with $a_j, b_j > 0$, μ-measurable sets A_j and ν-measurable sets B_j. Since $0 < t < 1$, neither p_t nor q_t' can be equal to $+\infty$, hence $\mu(A_j), \nu(B_j) < +\infty$.

Let us now extend the definition of p_t and q_t to all complex numbers ζ with $0 \le \mathrm{Re}\ \zeta \le 1$:

$$\frac{1}{p(\zeta)} = \frac{1-\zeta}{p_0} + \frac{\zeta}{p_1}, \quad \frac{1}{q(\zeta)} = \frac{1-\zeta}{q_0} + \frac{\zeta}{q_1}, \quad \frac{1}{q'(\zeta)} = \frac{1-\zeta}{q_0'} + \frac{\zeta}{q_1'}.$$

Fix $t \in (0,1)$ and a pair of (complex-valued) functions $f \in L^{p_t}(M)$ and $g \in L^{q_t'}(M)$ of the form (9.24). Consider a family of functions

$$u(x,\zeta) = \sum_{j=1}^{n} a_j^{p_t/p(\zeta)} e^{i\alpha_j(x)} \chi_{A_j}(x), \quad v(y,\zeta) = \sum_{j=1}^{m} b_j^{q_t'/q'(\zeta)} e^{i\beta_j(y)} \chi_{B_j}(y),$$

with $x \in M$, $y \in N$ and $0 \le \mathrm{Re}\ \zeta \le 1$. Note that, when $\zeta = t$,

$$u(x,t) = f(x) \text{ and } v(y,t) = g(y). \tag{9.25}$$

As both $1/p(\zeta)$ and $1/q'(\zeta)$ are linear in ζ, the functions $u(x,\zeta)$ and $v(x,\zeta)$ are analytic in ζ in the strip $S = \{\zeta : \ 0 \le \mathrm{Re}\ \zeta \le 1\}$. Since $u(x,\zeta)$ and $v(y,\zeta)$ are simple functions of x and y, respectively, vanishing outside of a set of finite measure for each $\zeta \in S$ fixed, they lie in $L^{p_0}(M) \cap L^{p_1}(M)$, and $L^{q_0'}(M) \cap L^{q_1'}(M)$, respectively. Therefore, we can define

$$F(\zeta) = \int_N (Au)(y,\zeta) v(y,\zeta) d\nu = \sum_{j=1}^{n} \sum_{k=1}^{m} a_j^{p_t/p(\zeta)} b_k^{q_t'/q'(\zeta)} \int_N (A\Psi_j)(y) e^{i\beta_k(y)} \chi_{B_k}(y) d\nu,$$

with $\Psi_j(x) = e^{i\alpha_j(x)} \chi_{A_j}(x)$. According to (9.23) and (9.25) , in order to prove that

$$\|A_t\|_{L^{p_t}(M)\to L^{q_t}(N)} \le k_0^{1-t} k_1^t, \tag{9.26}$$

it suffices to show that

$$|F(t)| \le k_0^{1-t} k_1^t. \tag{9.27}$$

The function $F(\zeta)$ is analytic and bounded in the strip S, as, for instance, for $\zeta = \eta + i\xi$, $0 \le \eta \le 1$:

$$\left| a_j^{p_t/p(\zeta)} \right| = \left| a_j^{p_t\zeta/p_1 + p_t(1-\zeta)/p_0} \right| = \left| a_j^{p_t\eta/p_1 + p_t(1-\eta)/p_0} \right| \le C_j < +\infty.$$

On the boundary of the strip S we have the following bounds: along the line $\eta = 0$, for $z = i\xi$,

$$\|u(x,i\xi)\|_{L^{p_0}(M)} = \left(\int_M \sum_{j=1}^n \left|a_j^{[p_t(i\xi)/p_1+p_t(1-i\xi)/p_0]p_0}\right| \chi_{A_j}(x)d\mu\right)^{1/p_0}$$
$$= \left(\int_M \sum_{j=1}^n |a_j|^{p_t} \chi_{A_j}(x)d\mu\right)^{1/p_0} = \|f\|_{L^{p_t}(M)}^{p_t/p_0} = 1,$$

and

$$\|v(y,i\xi)\|_{L^{q_0'}(N)} = \left(\int_N \sum_{j=1}^m \left|b_j^{[q_t'(i\xi)/q_1'+q_t'(1-i\xi)/q_0']q_0'}\right| \chi_{B_j}(y)d\nu\right)^{1/q_0'}$$
$$= \left(\int_N \sum_{j=1}^m |b_j|^{q_t'} \chi_{B_j}(y)d\nu\right)^{1/p_0} = \|g\|_{L^{q_t'}(N)}^{q_t'/q_0'} = 1.$$

It follows that

$$|F(i\xi)| \le \|(Au)(i\xi)\|_{L^{q_0}(N)}\|v(i\xi)\|_{L^{q_0'}(N)} \le \|A\|_{L^{p_0}(M)\to L^{q_0}(N)}\|u(i\xi)\|_{L^{p_0}(N)}\|v(i\xi)\|_{L^{q_0'}(N)} \le k_0.$$

Similarly, along the line $\zeta = 1+i\xi$ we have $\|u(x,1+i\xi)\|_{L^{p_1}(M)} \le 1$ and $\|v(x,1+i\xi)\|_{L^{q_1'}(N)} \le 1$, which implies that $|F(1+i\xi)| \le k_1$. The three lines theorem implies now that $|F(\eta+i\xi)| \le k_0^{1-\eta}k_1^{\eta}$, hence (9.27) holds. □

9.4 The Hilbert transform

The Poisson kernel

Given a Schwartz class function $f(x) \in \mathcal{S}(\mathbb{R}^n)$ define a function

$$u(x,t) = \int_{\mathbb{R}^n} e^{-2\pi t|\xi|}\hat{f}(\xi)e^{2\pi i x\xi}d\xi, \quad t \ge 0, \quad x \in \mathbb{R}^n.$$

The function $u(x,t)$ is harmonic:

$$\Delta_{x,t}u = 0 \text{ in } \mathbb{R}^{n+1}_+ = \mathbb{R}^n \times (0,+\infty),$$

and satisfies the boundary condition on the hyper-plane $t = 0$:

$$u(x,0) = f(x), \quad x \in \mathbb{R}^n.$$

We can write $u(x,t)$ as a convolution

$$u(x,t) = P_t \star f = \int P_t(x-y)f(y),$$

with

$$\hat{P}_t(\xi) = e^{-2\pi t|\xi|},$$

and

$$P_t(x) = C_n \frac{t}{(t^2+|x|^2)^{(n+1)/2}}.$$

Here the constant n depends only on the spatial dimension.

The conjugate Poisson kernel

In the same spirit, for $f \in \mathcal{S}(\mathbb{R})$ define $u(x,t) = P_t \star f$, set $z = x + it$ and write

$$u(z) = \int_{\mathbb{R}} e^{-2\pi t|\xi|} \hat{f}(\xi) e^{2\pi i x\xi} d\xi = \int_0^\infty \hat{f}(\xi) e^{2\pi i z\xi} d\xi + \int_{-\infty}^0 \hat{f}(\xi) e^{2\pi i \bar{z}\xi} d\xi.$$

Consider the function $v(z)$ given by

$$iv(z) = \int_0^\infty \hat{f}(\xi) e^{2\pi i z\xi} d\xi - \int_{-\infty}^0 \hat{f}(\xi) e^{2\pi i \bar{z}\xi} d\xi.$$

As the function

$$u(z) + iv(z) = \int_0^\infty \hat{f}(\xi) e^{2\pi i z\xi} d\xi$$

is analytic in the upper half-plane $\{\mathrm{Im} z > 0\}$, the function v is the harmonic conjugate of u. It can be written as

$$v(z) = \int_{\mathbb{R}} (-i\mathrm{sgn}(\xi)) e^{-2\pi t|\xi|} \hat{f}(\xi) e^{2\pi i x\xi} d\xi = Q_t \star f,$$

with

$$\hat{Q}_t(\xi) = -i\mathrm{sgn}(\xi) e^{-2\pi t|\xi|}, \tag{9.28}$$

and

$$Q_t(x) = \frac{1}{\pi} \frac{x}{t^2 + x^2}.$$

The Poisson kernel and its conjugate are related by

$$P_t(x) + iQ_t(x) = \frac{i}{\pi(x + iy)},$$

which is analytic in $\{\mathrm{Im} z \geq 0\}$. The main problem with the conjugate Poisson kernel is that it does not decay fast enough at infinity to be in $L^1(\mathbb{R})$ nor is regular at $x = 0$ as $t \to 0$.

The principle value of $1/x$

In order to consider the limit of Q_t as $t \to 0$ let us define the principal value of $1/x$ which is an element of $\mathcal{S}'(\mathbb{R})$ defined by

$$\mathrm{P.V.}\frac{1}{x}(\phi) = \lim_{\varepsilon \to 0} \int_{|x|>\varepsilon} \frac{\phi(x)}{x} dx, \quad \phi \in \mathcal{S}(\mathbb{R}).$$

This is well-defined because

$$\mathrm{P.V.}\frac{1}{x}(\phi) = \int_{|x|<1} \frac{\phi(x) - \phi(0)}{x} dx + \int_{|x|>1} \frac{\phi(x)}{x} dx,$$

thus

$$\left| \mathrm{P.V.}\frac{1}{x}(\phi) \right| \leq C(\|\phi'\|_{L^\infty} + \|x\phi\|_{L^\infty}),$$

and therefore P.V.$(1/x)$ is, indeed, a distribution in $\mathcal{S}'(\mathbb{R})$. The conjugate Poisson kernel Q_t and the principal value of $1/x$ are related as follows.

Proposition 9.10 *Let* $Q_t = \frac{1}{\pi}\frac{x}{t^2+x^2}$*, then for any function* $\phi \in \mathcal{S}(\mathbb{R})$

$$\frac{1}{\pi}P.V.\frac{1}{x}(\phi) = \lim_{t\to 0}\int_{\mathbb{R}} Q_t(x)\phi(x)dx.$$

Proof. Let

$$\psi_t(x) = \frac{1}{x}\chi_{t<|x|}(x)$$

so that

$$\text{P.V.}\frac{1}{x}(\phi) = \lim_{t\to 0}\int_{\mathbb{R}} \psi_t(x)\phi(x)dx.$$

Note, however, that

$$\begin{aligned}
&\int (\pi Q_t(x) - \psi_t(x))\phi(x)dx = \int_{\mathbb{R}} \frac{x\phi(x)}{x^2+t^2}dx - \int_{|x|>t}\frac{\phi(x)}{x}dx \\
&= \int_{|x|<t}\frac{x\phi(x)}{x^2+t^2}dx + \int_{|x|>t}\left[\frac{x}{x^2+t^2} - \frac{1}{x}\right]\phi(x)dx \qquad (9.29)\\
&= \int_{|x|<1}\frac{x\phi(tx)}{x^2+1}dx - \int_{|x|>t}\frac{t^2\phi(x)}{x(x^2+t^2)}dx = \int_{|x|<1}\frac{x\phi(tx)}{x^2+1}dx - \int_{|x|>1}\frac{\phi(tx)}{x(x^2+1)}dx.
\end{aligned}$$

The dominated convergence theorem implies that both integrals on the utmost right side above tend to zero as $t \to 0$. □

It is important to note that the computation in (9.29) worked only because the kernel $1/x$ is odd – this produces the cancellation that saves the day. This would not happen, for instance, for a kernel behaving as $1/|x|$ near $x = 0$.

The Hilbert transform

Motivated by the previous discussion, for a function $f \in \mathcal{S}(\mathbb{R})$, we define the Hilbert transform as

$$Hf(x) = \lim_{t\to 0} Q_t \star f(x) = \frac{1}{\pi}\lim_{\varepsilon\to 0}\int_{|y|>\varepsilon}\frac{f(x-y)}{y}dy.$$

It follows from (9.28) that

$$\widehat{Hf}(\xi) = \lim_{\varepsilon\to 0}\hat{Q}_t(\xi)\hat{f}(\xi) = -i\text{sgn}(\xi)\hat{f}(\xi). \qquad (9.30)$$

Therefore, the Hilbert transform may be extended to an isometry $L^2(\mathbb{R}) \to L^2(\mathbb{R})$, with $\|Hf\|_{L^2} = \|f\|_{L^2}$, $H(Hf) = -f$ and

$$\int (Hf)(x)g(x)dx = -\int f(x)(Hg)(x)dx. \qquad (9.31)$$

The following extension of the Hilbert transform to L^p-spaces for $1 < p < \infty$ is due to M. Riesz.

Theorem 9.11 *Given* $1 < p < \infty$ *there exists* $C_p > 0$ *so that*

$$\|Hf\|_{L^p} \le C_p\|f\|_{L^p} \text{ for all } f \in L^p(\mathbb{R}^n). \qquad (9.32)$$

Proof. We first consider $p \geq 2$. It suffices to establish (9.32) for $f \in \mathcal{S}(\mathbb{R})$. Consider a smaller set

$$\mathcal{S}_0 = \{f \in \mathcal{S} : \ \exists \varepsilon > 0 \text{ such that } \hat{f}(\xi) = 0 \text{ for } |\xi| < \varepsilon\}.$$

Let us show that $\mathcal{S}_0$ is dense in $L^p(\mathbb{R})$. Given any $f \in \mathcal{S}$ we'll find a sequence $g_n \in \mathcal{S}_0$ such that $\|f - g_n\|_{L^p} \to 0$ as $n \to +\infty$. For $p = 2$ this is trivial: take a smooth function $\chi(\xi)$ such that $0 \leq \chi(\xi) \leq 1$, $\chi(\xi) = 0$ for $|\xi| \leq 1$, $\chi(\xi) = 1$ for $|\xi| > 2$, and set

$$g_n(x) = \int e^{2\pi i \xi x} \hat{f}(\xi) \chi(n\xi)\, d\xi,$$

so that

$$\|f - g_n\|_{L^2}^2 \leq \int_{-2/n}^{2/n} |\hat{f}(\xi)|^2 d\xi \to 0 \text{ as } n \to +\infty. \tag{9.33}$$

On the other hand, for $p = +\infty$ we have

$$\|f - g_n\|_{L^\infty} \leq \int_{-2/n}^{2/n} |\hat{f}(\xi)| d\xi \to 0 \text{ as } n \to +\infty. \tag{9.34}$$

Interpolating between $p = 2$ and $p = +\infty$ we conclude that

$$\|f - g_n\|_{L^p} \to 0 \text{ as } n \to +\infty \tag{9.35}$$

for all $p \geq 2$, hence $\mathcal{S}_0$ is dense in $L^p(\mathbb{R})$ for $2 \leq p < +\infty$

Given $f \in \mathcal{S}_0$, $\widehat{Hf}(\xi) = -i(\mathrm{sgn}\xi)\hat{f}(\xi)$ is a Schwartz class function (there is no discontinuity at $\xi = 0$), thus Hf is also in $\mathcal{S}(\mathbb{R})$. We may then write

$$p(x) = (f + iHf)(x) = \int_{\mathbb{R}} (1 + \mathrm{sgn}(\xi))\hat{f}(\xi) e^{2\pi i \xi x} d\xi = 2\int_0^\infty \hat{f}(\xi) e^{2\pi i \xi x} d\xi,$$

and consider its extension to the complex plane:

$$p(z) = 2\int_0^\infty \hat{f}(\xi) e^{2\pi i \xi z} d\xi.$$

The function $p(z)$ is holomorphic in the upper half-plane $\{\mathrm{Im} z > 0\}$ and is continuous up to the boundary $y = 0$. Since $f \in \mathcal{S}_0$ there exists $\varepsilon > 0$ so that $\hat{f}(\xi) = 0$ for $|\xi| \leq \varepsilon$. Thus, $p(z)$ satisfies an exponential decay bound

$$|p(z)| \leq 2e^{-2\pi\varepsilon y} \|\hat{f}\|_{L^1}, \quad z = x + iy. \tag{9.36}$$

Integrating $p^4(z)$ along the contour C_R which consists of the interval $[-R, R]$ along the real axis and the semicircle $\{x^2 + y^2 = R^2, \ y > 0\}$, and passing to the limit $R \to 0$ with the help of (9.36) leads to

$$\lim_{R \to +\infty} \int_{-R}^{R} (f(x) + iHf(x))^4 dx = 0.$$

As both f and Hf are in $\mathcal{S}_0$, the integral above converges absolutely, hence

$$\int_{\mathbb{R}} (f(x) + iHf(x))^4 dx = 0.$$

The real part above gives

$$\int_{\mathbb{R}} (Hf(x))^4 dx = \int_{\mathbb{R}} [-f^4(x) + 2f^2(x)(Hf)^2(x)]dx \leq 2\int f^2(x)(Hf)^2(x)dx$$
$$\leq \int (2f^4(x) + \frac{1}{2}(Hf)^4(x))dx,$$

hence

$$\int_{\mathbb{R}} (Hf(x))^4 dx \leq 4\int f^4(x)dx, \tag{9.37}$$

for any function $f \in \mathcal{S}_0$. As we have shown that $\mathcal{S}_0$ is dense in any $L^p(\mathbb{R})$, $2 \leq p < \infty$, (9.37) holds for all $f \in L^4(\mathbb{R})$. Therefore, the Hilbert transform is a bounded operator $L^4(\mathbb{R}) \to L^4(\mathbb{R})$. As we know that it is also bounded from $L^2(\mathbb{R})$ to $L^2(\mathbb{R})$, the Riesz-Thorin interpolation theorem implies that $\|Hf\|_{L^p} \leq C_p\|f\|_{L^p}$ for all $2 \leq p \leq 4$.

An argument identical to the above, integrating the function $p^{2k}(z)$ over the same contour, shows that H is bounded from $L^{2k}(\mathbb{R})$ to $L^{2k}(\mathbb{R})$ for all integers k. It follows then from Riesz-Thorin interpolation theorem that $\|Hf\|_{L^p} \leq C_p\|f\|_{L^p}$ for all $2 \leq p < +\infty$.

It remains to consider $1 < p < 2$ – this is done using the duality argument. Let $q > 2$ be the dual exponent, $1/p+1/q = 1$. As the operator $H : L^q(\mathbb{R}) \to L^q(\mathbb{R})$ is bounded, so is its adjoint $H^* : L^p(\mathbb{R}) \to L^p(\mathbb{R})$ defined by $\langle H^*f, g\rangle = \langle f, Hg\rangle$, with $f \in L^p(\mathbb{R})$, $g \in L^q(\mathbb{R})$. However, (9.31) says that $H^* = -H$, hence the boundedness of H^* implies that $H : L^p(\mathbb{R}) \to L^p(\mathbb{R})$ is also bounded. □

The Hilbert transform does not map $L^1(\mathbb{R}) \to L^1(\mathbb{R})$ but we have the following result due to Kolmogorov.

Theorem 9.12 *Let $f \in L^1(\mathbb{R})$, then there exists $C > 0$ so that for any $\lambda > 0$ the following estimate holds:*

$$m\{x : \ |Hf(x)| \geq \lambda\} \leq \frac{C}{\lambda}\int_{\mathbb{R}} |f(x)dx.$$

We will not prove this theorem here.

10 The Haar functions and the Brownian motion

10.1 The Haar functions and their completeness

The Haar functions

The basic Haar function is

$$\psi(x) = \begin{cases} 1 & \text{if } 0 \leq x < 1/2, \\ -1 & \text{if } 1/2 \leq x < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{10.1}$$

It has mean zero

$$\int_0^1 \psi(x)dx = 0,$$

and is normalized so that

$$\int_0^1 \psi^2(x)dx = 1.$$

The rescaled and shifted Haar functions are

$$\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k), \quad j,k \in \mathbb{Z}.$$

These functions form an orthonormal set in $L^2(\mathbb{R})$ because if $j = j'$ and $k \neq k'$ then

$$\int_{\mathbb{R}} \psi_{jk}(x)\psi_{jk'}(x)dx = 2^j \int_{\mathbb{R}} \psi(2^j x - k)\psi(2^j x - k')dx = 0$$

because $\psi(y-k)\psi(y-k') = 0$ for any $y \in \mathbb{R}$ and $k \neq k'$. On the other hand, if $j \neq j'$, say, $j < j'$, then

$$\begin{aligned}
&\int_{\mathbb{R}} \psi_{jk}(x)\psi_{j'k'}(x)dx = 2^{j/2+j'/2} \int_{\mathbb{R}} \psi(2^j x - k)\psi(2^{j'} x - k')dx \\
&= 2^{j'/2-j/2} \int_{\mathbb{R}} \psi(y)\psi(2^{j'-j}y + 2^{j'-j}k - k')dy \\
&= 2^{j'/2-j/2} \int_0^{1/2} \psi(2^{j'-j}y + 2^{j'-j}k - k')dy - 2^{j'/2-j/2} \int_{1/2}^1 \psi(2^{j'-j}y + 2^{j'-j}k - k')dy.
\end{aligned}$$

Both of the integrals above equal to zero. Indeed, $2^{j'-j} \geq 2$, hence, for instance,

$$\int_0^{1/2} \psi(2^{j'-j}y + 2^{j'-j}k - k')dy = 2^{j-j'} \int_0^{2^{j'-j-1}} \psi(y + 2^{j'-j}k - k')dy = 0,$$

because

$$\int_m^n \psi(y)dy = 0,$$

for all $m, n \in \mathbb{Z}$, and $j' > j$. Finally, when $j = j'$, $k = k'$ we have

$$\int_{\mathbb{R}} |\psi_{jk}(x)|^2 = 2^j \int_{\mathbb{R}} |\psi(2^j x - k)|^2 dx = \int_{\mathbb{R}} |\psi(x-k)|^2 dx = 1.$$

The Haar coefficients of a function $f \in L^2(\mathbb{R})$ are defined as the inner products

$$c_{jk} = \int f(x)\psi_{jk}(x)dx, \tag{10.2}$$

and the Haar series of f is

$$\sum_{j,k \in \mathbb{Z}} c_{jk}\psi_{jk}(x). \tag{10.3}$$

Orthonormality of the family $\{\psi_{jk}\}$ ensures that

$$\sum_{j,k} |c_{jk}|^2 \leq \|f\|_{L^2}^2 < +\infty,$$

and the series (10.3) converges in $L^2(\mathbb{R})$. In order to show that it actually converges to the function f itself we need to prove that the Haar functions form a basis for $L^2(\mathbb{R})$.

Completeness of the Haar functions

To show that Haar functions form a basis in $L^2(\mathbb{R})$ we consider the dyadic projections P_n defined as follows. Given $f \in L^2(\mathbb{R})$, and $n, k \in \mathbb{Z}$, consider the intervals

$$I_{nk} = ((k-1)/2^n, k/2^n],$$

then

$$P_n f(x) = \fint_{I_{nk}} f dx = 2^n \int_{I_{nk}} f dx, \quad \text{for } x \in I_{nk}.$$

The function $P_n f$ is constant on each of the dyadic intervals I_{nk}. In particular, each Haar function ψ_{jk} satisfies $P_n \psi_{jk}(x) = 0$ for $j \geq n$, while $P_n \psi_{jk}(x) = \psi_{jk}(x)$ for $j < n$. We claim that, actually, for any $f \in L^2(\mathbb{R})$ we have

$$P_{n+1} f - P_n f = \sum_{k \in \mathbb{Z}} c_{nk} \psi_{nk}(x), \tag{10.4}$$

with the Haar coefficients c_{nk} given by (10.2). Indeed, let $x \in I_{nk}$ and write

$$I_{nk} = \Big(\frac{(k-1)}{2^n}, \frac{k}{2^n}\Big] = \Big(\frac{2(k-1)}{2^{n+1}}, \frac{(2k-1)}{2^{n+1}}\Big] \bigcup \Big(\frac{(2k-1)}{2^{n+1}}, \frac{2k}{2^{n+1}}\Big] = I_{n+1,2k-1} \bigcup I_{n+1,2k}.$$

The function $P_n f$ is constant on the whole interval I_{nk} while $P_{n+1} f$ is constant on each of the sub-intervals $I_{n+1,2k-1}$ and $I_{n+1,2k}$. In addition,

$$\int_{I_{nk}} (P_n f) dx = \int_{I_{nk}} (P_{n+1} f) dx.$$

This means exactly that

$$P_{n+1}(x) = P_n f(x) + c_{nk} \psi_{nk}(x) \text{ for } x \in I_{nk},$$

which is (10.4).

As a consequence of (10.4) we deduce that

$$P_{n+1} f(x) - P_{-m} f(x) = \sum_{j=-m}^{n} \sum_{k \in \mathbb{Z}} c_{jk} \psi_{jk}(x), \tag{10.5}$$

for all $m, n \in \mathbb{Z}$ with $n > m$. It remains to show that for any $f \in L^2(\mathbb{R})$ we have

$$\lim_{m \to +\infty} P_{-m} f(x) = 0, \qquad \lim_{n \to +\infty} P_n f(x) = f(x), \tag{10.6}$$

both in the L^2-sense. The operators $P_n f$ are uniformly bounded because for all $n, k \in \mathbb{Z}$ we have

$$\int_{I_{nk}} |(P_n f)(x)|^2 dx = 2^{-n} 2^{2n} \left| \int_{I_{nk}} f(y) dy \right|^2 \leq \int_{I_{nk}} |f(y)|^2 dy.$$

Summing over $k \in \mathbb{Z}$ for a fixed n we get

$$\int_{\mathbb{R}} |P_n f(x)|^2 \leq \int_{\mathbb{R}} |f(x)|^2,$$

thus $\|P_n f\|_{L^2} \leq \|f\|_{L^2}$. Uniform boundedness of P_n implies that it is sufficient to establish both limits in (10.6) for functions $f \in C_c(\mathbb{R})$. However, for such f we have, on one hand,

$$|P_{-m} f(x)| \leq \frac{1}{2^m} \int_{\mathbb{R}} |f(x)| dx \to 0 \text{ as } m \to +\infty,$$

and, on the other, f is uniformly continuous on $\mathbb{R}$, so that $\|P_n f(x) - f(x)\|_{L^\infty} \to 0$ as $n \to +\infty$, which, as both $P_n f$ and f are compactly supported, implies the second limit in (10.6). Therefore, ψ_{jk} form an orthonormal basis in $L^2(\mathbb{R})$ and every function $f \in L^2(\mathbb{R})$ has the reperesentation

$$f(x) = \sum_{j,k=-\infty}^{\infty} c_{jk} \psi_{jk}(x), \quad c_{jk} = \int_{\mathbb{R}} f(y) \psi_{jk}(y) dy. \tag{10.7}$$

10.2 The Brownian motion

Brownian motion is a random process $X_t(\omega)$, $t \geq 0$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ which has the following properties:

(i) The function $X_t(\omega)$ is continuous in t for a.e. realization ω.

(ii) For all $0 \leq s < t < +\infty$ the random variable $X_t(\omega) - X_s(\omega)$ is Gaussian with mean zero and variance $t - s$:

$$\mathbb{E}(X(t) - X(s)) = 0, \quad \mathbb{E}(X(t) - X(s))^2 = t - s.$$

(iii) For any subdivision $0 = t_0 < t_1 < \ldots < t_N = t$ of the interval $[0, t]$, the random variables $X_{t_1} - X_{t_0}, \ldots, X_{t_N} - X_{t_{N-1}}$ are independent.

Construction of the Brownian motion

We will construct the Brownian motion on the interval $0 \leq t \leq 1$ – the restriction to a finite interval is a simple convenience but by no means a necessity. The Haar functions $\psi_{jk}(x)$, with $j \geq 0$, $0 \leq k \leq 2^j - 1$, form a basis for the space $L^2[0, 1]$. Let us denote accordingly $\phi_n(x) = \psi_{jk}(x)$ for $n = 2^j + k$, $0 \leq k \leq 2^j - 1$, and $\phi_0(x) = 1$ so that $\{\phi_n\}$ form an orthonormal basis for $L^2[0, 1]$. Let $Z_n(\omega)$, $n \geq 0$, be a collection of independent Gaussian random variables of mean zero and variance one, that is,

$$P(Z_n < y) = \int_{-\infty}^{y} e^{-y^2} \frac{dy}{\sqrt{2\pi}}.$$

We will show that the process

$$X_t(\omega) = \sum_{n=0}^{\infty} Z_n(\omega) \int_0^t \phi_n(s) ds \tag{10.8}$$

is a Brownian motion.

First, we need to verify that the series (10.8) converges in $L^2(\Omega)$ for a fixed $t \in [0,1]$. Note that

$$\mathbb{E}\left(\sum_{k=n}^{m} Z_k(\omega)\int_0^t \phi_k(s)ds\right)^2 = \sum_{k=n}^{m}\left(\int_0^t \phi_k(s)ds\right)^2 = \sum_{k=n}^{m}\langle \chi_{[0,t]}, \phi_k\rangle^2.$$

As ϕ_k form a basis for $L^2[0,1]$, the series (10.8) satisfies the Cauchy criterion and thus converges in $L^2(\Omega)$. Moreover, for any $0 \le s < t \le 1$ we have

$$\begin{aligned}\mathbb{E}\left(X_t - X_s\right)^2 &= \mathbb{E}\left(\sum_{k=0}^{\infty} Z_k(\omega)\int_s^t \phi_k(u)du\right)^2 = \sum_{k=0}^{\infty}\left(\int_s^t \phi_k(u)du\right)^2 = \sum_{k=0}^{\infty}\langle \chi_{[s,t]}, \phi_k\rangle^2\\ &= \|\chi_{[s,t]}\|_{L^2}^2 = t - s,\end{aligned}$$

hence the increments $X_t - X_s$ have the correct variance. Let us show that they are independent: for $0 \le t_0 < t_1 \le t_2 < t_3 \le 1$:

$$\begin{aligned}\mathbb{E}\left((X_{t_3} - X_{t_2})(X_{t_1} - X_{t_0})\right) &= \mathbb{E}\left(\sum_{k=0}^{\infty}\int_{t_2}^{t_3}\phi_k(u)du\int_{t_0}^{t_1}\phi_k(u')du'\right)\\ &= \sum_{k=0}^{\infty}\langle \chi_{[t_2t_3]}, \phi_k\rangle\langle \chi_{[t_0t_1]}, \phi_k\rangle = \langle \chi_{[t_2t_3]}, \chi_{[t_0t_1]}\rangle = 0.\end{aligned}$$

As the variables $X_t - X_s$ are jointly Gaussian, independence of the increments follows.

Continuity of the Brownian motion

In order to prove continuity of the process $X_t(\omega)$ defined by the series (10.8) we show that the series converges uniformly in t almost surely in ω. To this end let us show that

$$M(\omega) = \sup_n \frac{|Z_n(\omega)|}{\sqrt{\log n}} < +\infty \text{ almost surely in } \omega. \tag{10.9}$$

Note that, for each $n \ge 0$:

$$\mathbb{P}\left(|Z_n(\omega)| \ge 2\sqrt{\log n}\right) \le e^{-(2\sqrt{\log n})^2/2} = \frac{1}{n^2},$$

thus

$$\sum_{n=0}^{\infty}\mathbb{P}\left(|Z_n(\omega)| \ge 2\sqrt{\log n}\right) < +\infty.$$

The Borel-Cantelli lemma implies that almost surely the event $\left\{|Z_n(\omega)| \ge 2\sqrt{\log n}\right\}$ happens only finitely many times, so that $|Z_n(\omega)| < 2\sqrt{\log n}$ for al $n \ge n_0(\omega)$ almost surely, and (10.9) follows.

Another useful observation is that for each fixed $t \ge 0$ and $j \in \mathbb{N}$ there exists only one k so that

$$\int_0^t \phi_{2^j+k}(s)ds \ne 0,$$

and for that k we have

$$\left|\int_0^t \phi_{2^j+k}(s)ds\right| \leq 2^{j/2}2^{-j} = \frac{1}{2^{j/2}}.$$

Hence, we may estimate the dyadic blocs, using (10.9):

$$\left|\sum_{k=0}^{2^j-1} Z_{2^j+k}(\omega)\int_0^t \phi_{2^j+k}(s)ds\right| \leq M(\omega)\sqrt{(j+1)\log 2}\sum_{k=0}^{2^j-1}\left|\int_0^t \psi_{jk}(s)ds\right| \leq \frac{\sqrt{j}M_1(\omega)}{2^{j/2}}.$$

Therefore, the dyadic blocs are bounded by a convergent series which does not depend on $t \in [0,1]$, hence the sum $X_t(\omega)$ of the series is a continuous function for a.e. ω.

Nowhere differentiability of the Brownian motion

Theorem 10.1 *The Brownian path $X_t(\omega)$ is nowhere differentiable for almost every ω.*

Proof. Let us fix $\beta > 0$. Then if $\dot{X}_s$ exists at some $s \in [0,1]$ and $|\dot{X}_s| < \beta$ then there exists n_0 so that

$$|X_t - X_s| \leq 2\beta|t-s| \text{ if } |t-s| \leq \frac{2}{n} \tag{10.10}$$

for all $n > n_0$. Let A_n be the set of functions $x(t) \in C[0,1]$ for which (10.10) holds for some $s \in [0,1]$. Then $A_n \subset A_{n+1}$ and the set $A = \bigcup_{n=1}^\infty A_n$ includes all functions $x(t) \in C[0,1]$ such that $|\dot{x}(s)| \leq \beta$ at some point $s \in [0,1]$.

The next step is to replace (10.10) by a discrete set of conditions – this is a standard trick in such situations. Assume that (10.10) holds for a function $x(t) \in C[0,1]$ and let $k = \sup\{j : \ j/n \leq s\}$, then

$$y_k = \max\left(\left|x\left(\frac{k+2}{n}\right) - x\left(\frac{k+1}{n}\right)\right|, \left|x\left(\frac{k+1}{n}\right) - x\left(\frac{k}{n}\right)\right|, \left|x\left(\frac{k}{n}\right) - x\left(\frac{k-1}{n}\right)\right|\right) \leq \frac{8\beta}{n}.$$

Therefore, if we denote by B_n the set of all functions $x(t) \in C[0,1]$ for which $y_k \leq 8\beta/n$ for some k, then $A_n \subseteq B_n$. Therefore, in order to show that $\mathbb{P}(A) = 0$ it suffices to check that

$$\lim_{n\to\infty} \mathbb{P}(B_n) = 0. \tag{10.11}$$

This, however, can be estimated directly, using translation invariance of the Brownian motion:

$$\begin{aligned}
\mathbb{P}(B_n) &\leq \sum_{k=1}^{n-2}\mathbb{P}\left[\max\left[\left|X(\frac{k+2}{n}) - X(\frac{k+1}{n})\right|, \left|X(\frac{k+1}{n}) - X(\frac{k}{n})\right|, \left|X(\frac{k}{n}) - X(\frac{k-1}{n})\right|\right] \leq \frac{8\beta}{n}\right] \\
&\leq n\mathbb{P}\left[\max\left[\left|X\left(\frac{3}{n}\right) - X\left(\frac{2}{n}\right)\right|, \left|X\left(\frac{2}{n}\right) - X\left(\frac{1}{n}\right)\right|, \left|X\left(\frac{1}{n}\right)\right|\right] \leq \frac{8\beta}{n}\right] \\
&= n\mathbb{P}\left[\left|X\left(\frac{1}{n}\right)\right| \leq \frac{8\beta}{n}\right]^3 = n\left(\sqrt{\frac{n}{2\pi}}\int_{-8\beta/n}^{8\beta/n} e^{-nx^2/2}dx\right)^3 \leq n\left(\sqrt{\frac{n}{2\pi}}\frac{16\beta}{n}\right)^3 \leq \frac{C}{\sqrt{n}},
\end{aligned}$$

which implies (10.11). It follows that $\mathbb{P}(A) = 0$ as well, hence Brownian motion is nowhere differentiable with probability one. □

Corollary 10.2 *Brownian motion does not have bounded variation with probability one.*