4 Functions of Real Variables

This chapter starts a systematic study of properties of functions of real variables, in terms of concepts related to measures. Properties of functions considered in this light are usually referred to as metric properties.

We begin with a characterization of measurable functions due to **N.N. Lusin**. This characterization is an intuitively satisfactory description of measurable functions and has basic and important consequences, in so far as measurable functions are concerned. Riemann integrable functions are then taken up and shown to be Lebesgue integrable and their integrals in either sense are the same.

Push-forward of measures, a natural construct of measures from those given through mappings, is then interposed for the purpose of representation of general integrals as integrals on $\mathbb{R}$, as well as for a transformation formula of the Lebesgue integral of functions on $\mathbb{R}^n$ through change of variables later in the chapter. Then there follows naturally a more detailed study of functions of a real variable, in which considerable emphasis is placed on study of differentiability of functions unfolding from the Lebesgue differentiation theorem for Radon measures on $\mathbb{R}^n$.

Product measures are treated and followed by further studies of functions of several real variables in later sections of the chapter.

A detailed presentation of polar coordinates in $\mathbb{R}^n$ is given in Section 4.11, with applications to integral operators of potential type and integral representation of C^1 functions.

4.1 Lusin theorem

Let μ be a Borel regular measure on $\mathbb{R}^n$, and f a finite-valued function defined on a μ-measurable subset A of $\mathbb{R}^n$. We suppose that $\mu(A) < \infty$. We shall show that f is Σ^{μ}-measurable if and only if it is almost a continuous function; "almost" in the sense given in Theorem 4.1.1. Theorem 4.1.1, is called the Lusin theorem in this book. In the following, μ, A, and f are fixed and specified as previously.

Lemma 4.1.1 *Let h be a simple function defined on A, then for $\varepsilon > 0$, there is a compact set $K \subset A$ such that $h|_K$ is continuous and $\mu(A \backslash K) < \varepsilon$.*

Proof In view of Proposition 3.8.2, we may assume that μ is a Radon measure. The simple function h can be expressed as

$$h = \sum_{j=1}^{k} \alpha_j I_{A_j},$$

where $A_1, \dots, A_k$ are disjoint μ-measurable subsets of A with $A = \bigcup_{j=1}^{k} A_j$. For each $j = 1, \dots, k$, there is a compact set $K_j \subset A_j$ with $\mu(A_j \backslash K_j) < \frac{\varepsilon}{k}$, by Theorem 3.9.1 (ii). Since $K_1, \dots, K_k$ are disjoint compact sets, $\operatorname{dist}(K_i, K_j) > 0$ if $i \neq j$; this, together with the fact that h is constant on each K_j, shows that $h|_K$ is continuous if $K := \bigcup_{j=1}^{k} K_j$. Now, $\mu(A\backslash K) = \sum_{j=1}^{k} \mu(A_j \backslash K_j) < \varepsilon$. The Lemma is proved. ■

Theorem 4.1.1 (Lusin) *Suppose that f is finite-valued and Σ^μ-measurable. Then for $\varepsilon > 0$, there is a compact set $K \subset A$ and a continuous function g defined on $\mathbb{R}^n$ such that $\mu(A\backslash K) < \varepsilon$ and $g = f$ on K.*

Proof There is a sequence $\{f_m\}$ of simple functions defined on A such that $\lim_{m\to\infty} f_m(x) = f(x)$ for $x \in A$. By the Egoroff theorem and Theorem 3.9.1 (ii), there is a compact set $K' \subset A$ such that $\mu(A\backslash K') < \frac{\varepsilon}{2}$ and $f_m(x)$ converges to $f(x)$ uniformly for $x \in K'$. For each m, by Lemma 4.1.1, there is a compact set $K_m \subset A$ such that $f_m|_{K_m}$ is continuous and $\mu(A\backslash K_m) < \frac{\varepsilon}{2^{m+1}}$. Set $K'' = \bigcap_{m=1}^{\infty} K_m$, then $f_m|_{K''}$ is continuous for each m, and

$$\mu(A\backslash K'') = \mu\left(\bigcup_{m=1}^{\infty} (A\backslash K_m)\right) < \sum_{m=1}^{\infty} \frac{\varepsilon}{2^{m+1}} = \frac{\varepsilon}{2}.$$

Now let $K = K' \cap K''$, then $\mu(A\backslash K) < \varepsilon$ and

(a) each $f_m|_K$ is continuous;
(b) $f_m|_K$ converges uniformly to $f|_K$.

From (a) and (b) follows the conclusion that $f|_K$ is continuous. By the Tietze Theorem (Theorem 1.8.1) there is a continuous function g on $\mathbb{R}^n$ such that $g = f|_K$ on K, or $g = f$ on K. ■

Concerning the Lusin theorem, we note first that it still holds if f is finite-valued μ-a.e. on A; and secondly, if f is finite-valued μ-a.e. and satisfies the conclusion of the Lusin theorem, then f is Σ^μ-measurable. To see this, we proceed as follows. For each $m \in \mathbb{N}$ there is a compact set $K_m \subset A$ and a continuous function g_m on $\mathbb{R}^n$ such that $\mu(A\backslash K_m) < \frac{1}{m^2}$ and $g_m = f$ on K_m; now $\sum_m \frac{1}{m^2} < \infty$ implies $\mu(\limsup_{m\to\infty}(A\backslash K_m)) = 0$ (cf. Exercise 2.5.9 (i)), which means that μ-a.e. x in A is in K_m if m is sufficiently large (observe that $A\backslash \limsup_{m\to\infty}(A\backslash K_m) = A\backslash \bigcap_{m=1}^{\infty} \bigcup_{l\geq m}(A\backslash K_l) = \bigcup_{m=1}^{\infty} \bigcap_{l\geq m} K_l = \liminf_{m\to\infty} K_m$), or $f(x) = \lim_{m\to\infty} g_m(x)$ and consequently f is Σ^μ-measurable because each g_m is Σ^μ-measurable due to the fact that μ is a Borel measure. Thus the conclusion of the Lusin theorem is a characterization of Σ^μ-measurable functions on A. We state this explicitly as a theorem for later reference and still call it the Lusin theorem.

Theorem 4.1.2 *Suppose that f is finite-valued μ-a.e. on A. Then f is Σ^{μ}-measurable if and only if for any given $\varepsilon > 0$, there is a compact set $K \subset A$ and a continuous function g on $\mathbb{R}^n$ such that $\mu(A \backslash K) < \varepsilon$ and $f = g$ on K.*

Exercise 4.1.1 Let f be a monotone increasing function defined on a finite open interval (a, b) in $\mathbb{R}$. Show that for any $\varepsilon > 0$, there is a continuous and monotone increasing function g on $\mathbb{R}$ such that the set $\{x \in (a, b) : f(x) \neq g(x)\}$ has Lebesgue measure less than ε. Furthermore, if f is bounded on (a, b), g can also be chosen to be bounded by the same bound as that of f.

Exercise 4.1.2 Suppose that f is integrable on $[a, b]$. Show that for each $\varepsilon > 0$ there is $g \in C[a, b]$ such that $\int_a^b |f - g| d\lambda < \varepsilon$. (Hint: prove first that the conclusion holds for bounded measurable function f.)

To conclude this section, we prove that when μ is the Lebesgue measure λ^n on $\mathbb{R}^n$, a characterization of Lebesgue measurable functions defined on an arbitrary Lebesgue measurable subset A of $\mathbb{R}^n$ similar to Theorem 4.1.2 holds.

Theorem 4.1.3 *Let A be a Lebesgue measurable set in $\mathbb{R}^n$. A function f which is defined and finite almost everywhere on A is measurable if and only if for any $\varepsilon > 0$ there is a closed set $F \subset A$ and a continuous function g on $\mathbb{R}^n$ such that $\lambda^n(A \backslash F) < \varepsilon$ and $f = g$ on F.*

Proof The sufficiency part follows from the same arguments that precede the statement of Theorem 4.1.2. We need only consider the necessity part. So, let f be a measurable function which is defined and finite almost everywhere on A, and let $\varepsilon > 0$ be given. Consider the following sequence $\{A_k\}$ of subsets of $A : A_1 = \{x \in A : |x| < 1\}$ and for $k \geq 2$ let $A_k = \{x \in A : k - 1 < |x| < k\}$. Since each set $\{x \in \mathbb{R}^n : |x| = k\}$ has measure zero (see Exercise 3.4.2), $\bigcup_{k=1}^{\infty} A_k$ consists of almost all points of A. Each A_k is measurable and has finite measure. By Theorem 4.1.1, for each k there is a compact set $F_k \subset A_k$ such that $f|_{F_k}$ is continuous and $\lambda^n(A_k \backslash F_k) < \frac{\varepsilon}{2^k}$. Now let $\{g_k\}$ be a sequence of continuous functions defined as follows: g_1 is a continuous function defined on $\{x \in \mathbb{R}^n : |x| \leq 1\}$ such that $g_1 = f|_{F_1}$ on F_1; suppose $g_1, \ldots, g_k$ have been defined, let g_{k+1} be a continuous function defined on $\{x \in \mathbb{R}^n : |x| \leq k + 1\}$ such that $g_{k+1} = g_k$ on $\{|x| \leq k\}$ and $g_{k+1} = f|F_{k+1}$ on F_{k+1}. That $\{g_k\}$ can be so defined is due to Tietze's extension theorem (Theorem 1.8.1). Then define $g(x) = g_k(x)$ if $|x| \leq k$. Obviously, from our construction of the sequence $\{g_k\}$, g is well defined and is continuous on $\mathbb{R}^n$. If we put $F = \bigcup_k F_k$, F is a closed set, $F \subset A$, and $\lambda^n(A \backslash F) = \sum_k \lambda^n(A_k \backslash F_k) < \sum_k \frac{\varepsilon}{2^k} = \varepsilon$. It is clear that $g = f$ on F. ■

4.2 Riemann and Lebesgue integral

In this section an oriented rectangle in $\mathbb{R}^n$ will be called an oriented interval. We show that a Riemann integrable function defined on a closed oriented interval in $\mathbb{R}^n$ is Lebesgue integrable and its Lebesgue integral coincides with its Riemann integral. First, we recall briefly the Riemann integrability. Fix a finite closed oriented interval

$I = [a_1, b_1] \times \cdots \times [a_n, b_n]$, which is not degenerated i.e. $a_i < b_i$, $i = 1, \ldots, n$. Unless stated otherwise, henceforth in this section, an interval is always a finite, closed, nondegenerate, and oriented interval. Two intervals are said to be nonoverlapping if their interiors are disjoint. A **partition** $\mathcal{P}$ of I is a finite family $\{I_j\}_{j=1}^k$ of nonoverlapping intervals such that $I = \bigcup_{j=1}^k I_j$, where k depends on $\mathcal{P}$; in particular, when $I = [a, b]$ is a finite closed interval in $\mathbb{R}$, a partition $\mathcal{P}$ of I is determined by a sequence $a = x_0 < x_1 < \cdots < x_l = b$ of points in $[a, b]$ and we simply call such a sequence of points in $[a, b]$ a partition of $[a, b]$. For a partition $\mathcal{P} = \{I_j\}_{j=1}^k$ of I, $\|\mathcal{P}\|$ will be used to denote $\max_{1 \leq j \leq k} \operatorname{diam} I_j$, and is called the **mesh** of $\mathcal{P}$.

Consider now a bounded function f defined on I. For an interval $J \subset I$, let $\bar{f}_J = \sup_{x \in J} f(x)$ and $\underline{f}_J = \inf_{x \in J} f(x)$. If $\mathcal{P} = \{I_j\}_{j=1}^k$ is a partition of I, put

$$\bar{S}(f, \mathcal{P}) = \sum_{j=1}^k \bar{f}_{I_j} |I_j|; \quad \underline{S}(f, \mathcal{P}) = \sum_{j=1}^k \underline{f}_{I_j} |I_j|,$$

where $|J|$ denote the volume of the interval J. A partition $\mathcal{P}$ is said to be **finer** than a partition Q if every interval in Q is a union of intervals in $\mathcal{P}$. One verifies easily that if $\mathcal{P}$ is finer than Q, then

$$\bar{S}(f; \mathcal{P}) \leq \bar{S}(f; Q); \quad \underline{S}(f; \mathcal{P}) \geq \underline{S}(f; Q).$$

For partitions $\mathcal{P}$ and Q of I, denote by $\mathcal{P} \vee Q$ the partition of I formed by all the nondegenerate intersections of intervals of $\mathcal{P}$ and those of Q. $\mathcal{P} \vee Q$ is finer than both $\mathcal{P}$ and Q, hence

$$\bar{S}(f; \mathcal{P}) \geq \bar{S}(f; \mathcal{P} \vee Q) \geq \underline{S}(f; P \vee Q) \geq \underline{S}(f; Q),$$

and consequently

$$\inf_{\mathcal{P}} \bar{S}(f; \mathcal{P}) \geq \sup_{\mathcal{P}} \underline{S}(f; \mathcal{P}).$$

$\inf_{\mathcal{P}} \bar{S}(f; \mathcal{P})$ is called the **Darboux upper integral** of f over I and is denoted by $\overline{\int}_I f$, while $\sup_{\mathcal{P}} \underline{S}(f; \mathcal{P})$ is called the **Darboux lower integral** of f over I and is denoted by $\underline{\int}_I f$. We have shown that

$$\underline{\int}_I f \leq \overline{\int}_I f;$$

if $\underline{\int}_I f = \overline{\int}_I f$, then the common value, denoted $\int_I f(x)dx$, is called the **Riemann integral** of f over I, and f is then said to be Riemann integrable over I.

Exercise 4.2.1 Show that a bounded function f defined on I is Riemann integrable if and only if for any $\varepsilon > 0$ there is a partition $\mathcal{P}$ of I such that $\bar{S}(f; \mathcal{P}) - \underline{S}(f; \mathcal{P}) < \varepsilon$. In particular, infer that continuous functions defined on I are Riemann integrable.

For a bounded function f on I, we define related functions $\underline{f}$ and $\bar{f}$ as follows:

$$\underline{f}(x) = \lim_{\delta\to 0+} \inf_{|y-x|<\delta} f(y); \quad \bar{f}(x) = \lim_{\delta\to 0+} \sup_{|y-x|<\delta} f(y).$$

Lemma 4.2.1 *$\underline{f}$ is lower semi-continuous and $\bar{f}$ is upper semi-continuous on I. Hence both are Borel measurable, and therefore are Lebesgue measurable.*

Proof Since $\bar{f} = -(\underline{-f})$, we need only show that $\underline{f}$ is lower semi-continuous.

Let $\lambda \in \mathbb{R}$; we shall show that $E_\lambda := \{\underline{f} > \lambda\}$ is open in I. Let $a \in E_\lambda$, then there is $\delta > 0$, such that

$$\inf_{\substack{|y-a|<2\delta \\ y\in I}} f(y) > \lambda.$$

Now let $x \in I$ and $|x-a| < \delta$; then $|y-x| < \delta$ entails that $|y-a| < 2\delta$ and hence,

$$\inf_{\substack{|y-x|<\delta \\ y\in I}} f(y) \geq \inf_{\substack{|y-a|<2\delta \\ y\in I}} f(y) > \lambda.$$

Consequently, $x \in E_\lambda$ and E_λ is open in I. This shows that $\underline{f}$ is lower semi-continuous on I. ∎

Lemma 4.2.2 $\underline{\int}_I f = \int_I \underline{f} d\lambda^n$, $\overline{\int}_I f = \int_I \bar{f} d\lambda^n$.

Proof Choose a sequence $\{\mathcal{P}_k\}$ of partitions of I such that $\lim_{k\to\infty} \underline{S}(f; \mathcal{P}_k) = \underline{\int}_I f$. Since we still have $\lim_{k\to\infty} \underline{S}(f; Q_k) = \underline{\int}_I f$, if each Q_k is finer than $\mathcal{P}_k$, we may assume that $\|\mathcal{P}_k\| \to 0$ as $k \to \infty$. Let $\mathcal{P}_k = \{I_i^{(k)}\}_{i=1}^{n_k}$ and define $f_k(x) = \underline{f}_{I_i^{(k)}}$ if $x \in [I_i^{(k)})$, $i = 1, \ldots, n_k$ and $f_k(x) = 0$ otherwise, where for an interval $J = [c_1, d_1] \times \cdots \times [c_n, d_n]$, $[J)$ denotes the half-open interval $[c_1, d_1) \times \cdots \times [c_n, d_n)$. We claim now that if $x \in I \setminus \bigcup_{k=1}^{\infty} \bigcup_{i=1}^{n_k} \partial I_i^{(k)}$, then $\lim_{k\to\infty} f_k(x) = \underline{f}(x)$. For each $\delta > 0$, since $\|\mathcal{P}_k\| \to 0$ as $k \to \infty$, $\inf_{|y-x|<\delta} f(y) \leq f_k(x)$, if k is sufficiently large, hence $\inf_{|y-x|<\delta} f(y) \leq \liminf_{k\to\infty} f_k(x)$ and consequently $\underline{f}(x) \leq \liminf_{k\to\infty} f_k(x)$. On the other hand, for each k, $f_k(x) \leq \inf_{|y-x|<\delta} f(y)$ if $\delta > 0$ is small enough, or $f_k(x) \leq \underline{f}(x)$ and hence $\limsup_{k\to\infty} f_k(x) \leq \underline{f}(x)$. Thus,

$$\limsup_{k\to\infty} f_k(x) \leq \underline{f}(x) \leq \liminf_{k\to\infty} f_k(x) \leq \limsup_{k\to\infty} f_k(x),$$

or $\underline{f}(x) = \lim_{k\to\infty} f_k(x)$, as we claim. Now the set $\bigcup_{k=1}^{\infty} \bigcup_{i=1}^{n_k} \partial I_i^{(k)}$ has Lebesgue measure zero and $|f_k(x)| \leq M := \sup_{x\in I} |f(x)|$; we may apply the Lebesgue dominated convergence theorem to obtain the equality $\underline{\int}_I f = \lim_{k\to\infty} \underline{S}(f; \mathcal{P}_k) = \lim_{k\to\infty} \int_I f_k d\lambda^n = \int_I \underline{f} d\lambda^n$. Similarly, $\overline{\int} f = \int_I \bar{f} d\lambda^n$. ∎

Theorem 4.2.1 *A bounded function f on I is Riemann integrable if and only if f is continuous at almost all points of I.*

Proof Since $\underline{f} \leq f \leq \bar{f}$ on I and

$$\underline{\int_I} f = \int_I \underline{f} d\lambda^n \leq \int_I \bar{f} d\lambda^n = \overline{\int_I} f,$$

f is Riemann integrable if and only if $\underline{f} = \bar{f}$ almost everywhere on I. But from Lemma 4.2.1, we know that $\underline{f}$ is lower semi-continuous and $\bar{f}$ is upper semi-continuous; it follows that $\underline{f} = \bar{f}$ almost everywhere on I means, through the inequalities $\underline{f} \leq f \leq \bar{f}$ on I, that f is continuous almost everywhere on I (cf. Exercise 1.5.2 (i) and (ii)). ■

Theorem 4.2.2 *A Riemann integrable function f on I is Lebesgue integrable and $\int_I f(x)dx = \int_I f d\lambda^n$.*

Proof Since $\underline{f} \leq f \leq \bar{f}$ on I, and $\underline{f}(x) = \bar{f}(x)$ for almost all x in I, as we have shown in the proof of Theorem 4.2.1, $f = \underline{f}$ almost everywhere on I and is therefore measurable. As f is bounded and measurable, it is Lebesgue integrable. Now $\int_I f d\lambda^n = \int_I \underline{f} d\lambda^n = \underline{\int_I} f = \int_I f(x)dx$. ■

We note in passing that the function on $[0, 1]$ which takes value 1 on irrational numbers and takes value 0 on rational numbers is not Riemann integrable, but is Lebesgue integrable with Lebesgue integral being 1.

Exercise 4.2.2 Let f be a function defined on $\mathbb{R}$ whose improper integral $\int_{-\infty}^{\infty} f(x)dx$ converges absolutely. Show that f is Lebesgue integrable on $\mathbb{R}$ and $\int_{\mathbb{R}} f d\lambda = \int_{-\infty}^{\infty} f(x)dx$.

Exercise 4.2.3 Give an example to show that the conclusion in Exercise 4.2.2 does not hold if $\int_{-\infty}^{\infty} f(x)dx$ converges, but not absolutely.

We strongly suggest that readers verify that results similar to the conclusion of Exercise 4.2.2 hold for other types of improper integrals.

Notational convention Because of Theorem 4.2.2 and Exercise 4.2.2, we often write $\int_A f d\lambda^n$ as $\int_A f(x)dx$; also, we use $\int_a^b f(x)dx$, $\int_a^{\infty} f(x)dx$, $\int_{-\infty}^b f(x)dx$, and $\int_{-\infty}^{\infty} f(x)dx$ to denote $\int_I f d\lambda$ if I is $[a, b]$, $[a, \infty)$, $(-\infty, b]$, and $(-\infty, \infty)$ in this order. More generally, for a Borel measure μ on $\mathbb{R}$, $\int_a^b f d\mu$, $\int_a^{\infty} f d\mu$, $\int_{-\infty}^b f d\mu$, and $\int_{-\infty}^{\infty} f d\mu$ are similarly connoted.

4.3 Push-forward of measures and distribution of functions

Distribution of a measurable function on a measure space is now considered with its application to representation of the integral of Borel functions of the function as integral on $\mathbb{R}$. For this purpose, a natural method of constructing new measures from one given through mappings will be presented first.

Suppose that μ measures Ω and that t is a map from Ω to a set X; define a set function $t_{\#}\mu$ on 2^X by

$$t_{\#}\mu(A) = \mu(t^{-1}A), \quad A \subset X.$$

Obviously, $t_{\#}\mu$ is a measure on X; it is called the **push-forward** of μ through the map t. Let $A \subset X$ be such that $t^{-1}A$ is μ-measurable, then $t^{-1}A$ is $\mu\lfloor B$-measurable for any subset B of Ω by Exercise 3.1.3 (i). Thus if C is any subset of X, we have, since $(t^{-1}A)^c = t^{-1}A^c$,

$$\mu\lfloor B(t^{-1}C) = \mu\lfloor B(t^{-1}C \cap t^{-1}A) + \mu\lfloor B(t^{-1}C \cap t^{-1}A^c),$$

or,

$$t_{\#}(\mu\lfloor B)(C) = t_{\#}(\mu\lfloor B)(C \cap A) + t_{\#}(\mu\lfloor B)(C \cap A^c).$$

The last equality means that A is $t_{\#}(\mu\lfloor B)$-measurable for any subset B of Ω. Conversely, suppose that a subset A of X is $t_{\#}(\mu\lfloor B)$-measurable for any subset B of Ω; then if we choose $C = X$ in the last equality, it follows that

$$t_{\#}(\mu\lfloor B)(X) = t_{\#}(\mu\lfloor B)(A) + t_{\#}(\mu\lfloor B)(A^c),$$

or,

$$\mu\lfloor B(\Omega) = \mu\lfloor B(t^{-1}A) + \mu\lfloor B(\{t^{-1}A\}^c),$$

and hence,

$$\mu(B) = \mu(B \cap t^{-1}A) + \mu(B \cap \{t^{-1}A\}^c)$$

for any subset B of Ω, implying that $t^{-1}A$ is μ-measurable. We have shown the following proposition.

Proposition 4.3.1 *Let A be a subset of X, then, $t^{-1}A$ is μ-measurable if and only if A is $t_{\#}(\mu\lfloor B)$-measurable for every subset B of Ω.*

Corollary 4.3.1 *A subset A of X is $t_{\#}\mu$-measurable if $t^{-1}A$ is μ-measurable.*

Exercise 4.3.1 Show that if t is injective, then $A \subset X$ is $t_{\#}\mu$-measurable if and only if $t^{-1}A$ is μ-measurable.

Proposition 4.3.2 *If μ is a finite regular measure on Ω, then $A \subset X$ is $t_{\#}\mu$- measurable if and only if $t^{-1}A$ is μ-measurable.*

Proof Because of Corollary 4.3.1, we need only show that if A is $t_{\#}\mu$-measurable, then $t^{-1}A$ is μ-measurable.

Choose $C \in \Sigma^{\mu}$ such that $t^{-1}A \subset C$ and $\mu(t^{-1}A) = \mu(C)$. Using the conclusion of Exercise 3.1.4, we have

$$\begin{aligned}\mu(C \cap t^{-1}A^c) + \mu(C \cup t^{-1}A^c) &= \mu(C) + \mu(t^{-1}A^c)\\ &= \mu(t^{-1}A) + \mu(t^{-1}A^c)\\ &= t_{\#}\mu(A) + t_{\#}\mu(A^c) = t_{\#}\mu(X)\\ &= \mu(\Omega) = \mu(C \cup t^{-1}A^c);\end{aligned}$$

since μ is finite, we may cancel out the term $\mu(C \cup t^{-1}A^c)$ from the far left-hand side and the far right-hand side in the above sequence of equalities to obtain $\mu(C \cap (t^{-1}A)^c) = 0$. Thus $C \cap (t^{-1}A)^c$ is μ-measurable. But from $t^{-1}A \subset C$, we have $t^{-1}A = C \backslash (C \cap (t^{-1}A)^c)$ and hence $t^{-1}A$ is μ-measurable. ■

Suppose now that (Ω, Σ, μ) is a measure space and t is a map from Ω into a set X. Let μ^* be the measure on Ω constructed from μ by Method I; μ^* is the unique Σ-regular measure on Ω such that $\mu^*(A) = \mu(A)$ for $A \in \Sigma$ as asserted by Corollary 3.4.1. Define $t_{\#}\Sigma := \{A \subset X : t^{-1}A \in \Sigma\}$. Since $\Sigma \subset \Sigma^{\mu^*}$ and $\mu^*(A) = \mu(A)$ for $A \in \Sigma$, $t_{\#}\Sigma \subset \Sigma^{t_{\#}\mu^*}$ (by Corollary 4.3.1) and $t_{\#}\mu^*(A) = \mu(t^{-1}A)$ for $A \in t_{\#}\Sigma$. For notational simplicity, denote the restriction of $t_{\#}\mu^*$ to $t_{\#}\Sigma$ by $t_{\#}\mu$; then $(X, t_{\#}\Sigma, t_{\#}\mu)$ is a measure space called the **push-forward** of (Ω, Σ, μ) through the map t. Note that the map t from Ω into X is measure-preserving from (Ω, Σ, μ) to $(X, t_{\#}\Sigma, t_{\#}\mu)$ (cf. Section 2.8.2).

Exercise 4.3.2 Let (Ω, Σ, μ) be a measure space and t a map from Ω into a set X.

(i) Show that a function f on X is $t_{\#}\Sigma$-measurable if and only if $f \circ t$ is Σ-measurable.

(ii) Show that if $f \geq 0$ is $t_{\#}\Sigma$-measurable, then $\int_X f dt_{\#}\mu = \int_\Omega f \circ t d\mu$.

(iii) Show that if f is $t_{\#}\Sigma$-measurable, then $\int_X f dt_{\#}\mu = \int_\Omega f \circ t d\mu$ if one of the integrals is meaningful.

(Hint: start with f as an indicator function of a set.)

Example 4.3.1 (Cf. Exercise 3.4.2 (vi)) Suppose that $\Omega = X = \mathbb{R}^n$, and Σ is the σ-algebra $\mathcal{L}^n$ of all Lebesgue measurable sets in $\mathbb{R}^n$.

(i) For $a \in \mathbb{R}^n$ fixed, let t be the mapping $tx = x + a$, $x \in \mathbb{R}^n$. Then $t_{\#}\mathcal{L}^n = \mathcal{L}^n$, $t_{\#}\lambda^n = \lambda^n$, hence,

$$\int_{\mathbb{R}^n} f(x+a)dx = \int_{\mathbb{R}^n} f(x)dx,$$

if $\int_{\mathbb{R}^n} f(x)dx$ exists, i.e. the Lebesgue integral is translation invariant on $\mathbb{R}^n$.

(ii) For $\alpha \in \mathbb{R}$, $\alpha \neq 0$, consider the mapping $tx = \alpha x$. Then, $t_\# \mathcal{L}^n = \mathcal{L}^n$, $t_\# \lambda^n = \frac{1}{|\alpha|^n} \lambda^n$, hence,

$$\int_{\mathbb{R}^n} f(\alpha x) dx = \frac{1}{|\alpha|^n} \int_{\mathbb{R}^n} f(x) dx,$$

if $\int_{\mathbb{R}^n} f(x) dx$ exists. In particular, take $f = I_{B_1(0)}$, then $\lambda^n(B_r(0)) = r^n \lambda^n(B_1(0))$.

Exercise 4.3.3 Suppose that f is Lebesgue measurable on $\mathbb{R}$ and is periodic with period $l > 0$ i.e. $f(x) = f(x + l)$ for $x \in \mathbb{R}$. Suppose further that f is integrable on $[0, l]$. Show that f is integrable on $[a, a + l]$ and $\int_0^l f d\lambda = \int_a^{a+l} f d\lambda$ for any $a \in \mathbb{R}$.

Exercise 4.3.4 Suppose that t is a continuous and monotone increasing function defined on a finite interval $[a, b]$. Put $c = t(a)$ and $d = t(b)$. Show that for any Borel set $A \subset [c, d]$, $t_\# \mu_t(A) = \lambda(A)$, where μ_t is the Lebesgue–Stieltjes measure generated by t. (Hint: for any interval I open in $[c, d]$, $t_\# \mu_t(I) = |I|$.)

Suppose now that f is a finite-valued measurable function on a measure space (Ω, Σ, μ). Since f is Σ-measurable, $f_\# \Sigma$ contains all Borel subsets of $\mathbb{R}$ and $f_\# \mu$ is a measure on $\mathcal{B}$. Considered as a measure on $\mathcal{B}$, $f_\# \mu$ is called the **distribution** of f. If g is a Borel function on $\mathbb{R}$, then $g \circ f$ is Σ-measurable and

$$\int_{\mathbb{R}} g df_\# \mu = \int_{\Omega} g \circ f d\mu \tag{4.1}$$

if one of the integrals exists. In particular, if g is taken to be $g(t) = |t|^p$, $1 \leq p < \infty$, then

$$\int_{\Omega} |f|^p d\mu = \int_{\mathbb{R}} |t|^p df_\# \mu.$$

Thus, $\int_{\Omega} |f|^p d\mu$ can be expressed as an integral on $\mathbb{R}$ w.r.t. the measure $f_\# \mu$. When $\mu(\{f \leq t\}) < \infty$ for every $t \in \mathbb{R}$, put

$$F(t) = \mu(\{f \leq t\}) = \mu(f^{-1}(-\infty, t]),$$

then F is a monotone increasing function and we might expect $\int_{\mathbb{R}} |t|^p df_\# \mu$ to be the improper Riemann–Stieltjes integral $\int |t|^p dF := \lim_{\substack{b \to \infty \\ a \to -\infty}} \int_a^b |t|^p dF$. We shall see that this is actually true (cf. Exercise 4.5.6).

Exercise 4.3.5 Show that the function F, previously defined, is right-continuous i.e. $F(t) = F(t+)$. Moreover, $\lim_{t \to -\infty} F(t) = 0$, $\lim_{t \to \infty} F(t) = \mu(\Omega)$.

The function F is called the **distribution function** of f. When a function F is mentioned as the distribution function of a measurable function f, it is implicitly assumed that $\mu(\{f \leq t\}) < \infty$ for every $t \in \mathbb{R}$. One sees easily that if f is measurable and finite a.e. on Ω, its distribution $f_\# \mu$ and distribution function can be similarly defined.

As we have seen in Section 3.8, F generates a Lebesgue–Stieltjes measure μ_F on $\mathbb{R}$. It turns out that $\mu_F = f_\# \mu$ on $\mathcal{B}$, as the following theorem claims.

Theorem 4.3.1 *Suppose that F is the distribution function of a finite-valued measurable function f on a measure space (Ω, Σ, μ). Then, $(\mathbb{R}, \mathcal{B}, f_{\#}\mu) = (\mathbb{R}, \mathcal{B}, \mu_F)$, where μ_F is the Lebesgue–Stieltjes measure generated by F, and*

$$\int_{\Omega} g \circ f d\mu = \int_{\mathbb{R}} g d\mu_F, \tag{4.2}$$

for any Borel measurable function g on $\mathbb{R}$ whose μ_F-integral exists.

Proof Since F is right-continuous, $\mu_F((a, b]) = F(b) - F(a)$, from which by letting $a \to -\infty$, we have

$$\mu_F((-\infty, b]) = F(b) = \mu(f^{-1}(-\infty, b]) = f_{\#}\mu((-\infty, b])$$

for $b \in \mathbb{R}$. Now fix $a \in \mathbb{R}$ and consider the family $\mathcal{F}$ of all $B \in \mathcal{B}$ such that $\mu_F((-\infty, a] \cap B) = f_{\#}\mu((-\infty, a] \cap B)$. It is clear that $\mathcal{F}$ is a λ-system and it contains all sets of the form $(-\infty, b]$, $b \in \mathbb{R}$. Since the family of all sets of the form $(-\infty, b]$, $b \in \mathbb{R}$, is a π-system and $\mathcal{B}$ is the smallest σ-algebra containing all sets of the form $(-\infty, b]$, it follows from the (π-λ) theorem that $\mathcal{F} = \mathcal{B}$. Thus,

$$\mu_F((-\infty, a] \cap B) = f_{\#}\mu((-\infty, a] \cap B)$$

for all $B \in \mathcal{B}$. From this, by letting $a \to \infty$, we infer that $\mu_F(B) = f_{\#}(B)$ for all $B \in \mathcal{B}$, or $(\mathbb{R}, \mathcal{B}, \mu_F) = (\mathbb{R}, \mathcal{B}, f_{\#}\mu)$. Then (4.2) follows from (4.1). ■

In the final part of this section we demonstrate using an example the fact that measure spaces, which look very different from one another in appearance, might be the same measure space in different forms.

Example 4.3.2 Let $(\Omega, \sigma(\mathcal{Q}), P)$ be the Bernoulli sequence space of Example 3.4.6. Define a map $t : \Omega \to [0, 1]$ by

$$t(\omega) = \sum_{j=1}^{\infty} \frac{\omega_j}{2^j}, \quad \omega = (\omega_j) \in \Omega.$$

Note that $0.\omega_1\omega_2\omega_3 \cdots$ is a **binary expansion** of $t(\omega)$. For $x \in [0, 1]$, $t^{-1}x$ consists of either two elements or one element, depending on whether x is a **binary rational number** or not, except that $t^{-1}0$ consists of one element; when x is a binary rational number in $(0, 1]$, say $x = \sum_{j=1}^{n} \frac{\varepsilon_j}{2^j}$ with $\varepsilon_n = 1$, then $t^{-1}x$ consists of $(\varepsilon_1, \ldots, \varepsilon_{n-1}, 1, 0, 0, 0, \ldots)$ and $(\varepsilon_1, \ldots, \varepsilon_{n-1}, 0, 1, 1, 1, \ldots)$. Therefore if we put

$$\widehat{\Omega} = \{\omega \in \Omega : \omega_j = 1 \text{ for infinitely many } j\},$$

then $\Omega \setminus \widehat{\Omega}$ is countable and hence $\widehat{\Omega} \in \sigma(\mathcal{Q})$ with $P(\widehat{\Omega}) = 1$. One sees readily that if $\hat{t}$ is the restriction of t to $\widehat{\Omega}$, $\hat{t}$ is bijective from $\widehat{\Omega}$ to $(0, 1]$. As in Section 1.3, for a finite sequence $\varepsilon_1, \ldots, \varepsilon_n$ of 0 and 1, the elementary cylinder $\{\omega \in \Omega : \omega_j = \varepsilon_j, j = 1, \ldots, n\}$ in Ω of rank n is denoted by $E(\varepsilon_1, \ldots, \varepsilon_n)$; and we let $\mathcal{E}$ be the family of empty set $\emptyset$ and all elementary cylinders of all ranks in Ω. $\mathcal{E}$ is a π-system on Ω, and if we let $\widehat{\mathcal{E}} = \{E \cap \widehat{\Omega} : E \in \mathcal{E}\}$, then $\widehat{\mathcal{E}}$ is a π-system on $\widehat{\Omega}$.

(i) Observe first that $\sigma(\widehat{\mathcal{E}}) = \sigma(\mathcal{Q})|\widehat{\Omega}$. Actually, $\Sigma := \{A \in \sigma(\mathcal{Q}) : A \cap \widehat{\Omega} \in \sigma(\widehat{\mathcal{E}})\}$ is a σ-algebra on Ω containing $\mathcal{E}$, implying $\Sigma \supset \sigma(\mathcal{E}) = \sigma(\mathcal{Q}) \supset \Sigma$, or $\Sigma = \sigma(\mathcal{Q}) = \sigma(\mathcal{E})$, and hence $\sigma(\mathcal{Q})|\widehat{\Omega} \subset \sigma(\widehat{\mathcal{E}})$; that $\sigma(\widehat{\mathcal{E}}) \subset \sigma(\mathcal{Q})|\widehat{\Omega}$ follows from the fact that $\sigma(\mathcal{Q})|\widehat{\Omega}$ is a σ-algebra on $\widehat{\Omega}$ containing $\widehat{\mathcal{E}}$.

(ii) For any elementary cylinder $E(\varepsilon_1, \ldots, \varepsilon_n)$ of positive rank n in Ω, put $\widehat{E}(\varepsilon_1, \ldots, \varepsilon_n) = E \cap \widehat{\Omega}$. Observe that $\hat{t}\widehat{E}(\varepsilon_1, \ldots, \varepsilon_n) = (\alpha, \alpha + \frac{1}{2^n}]$, where $\alpha = \sum_{j=1}^{n} \frac{\varepsilon_j}{2^j}$, and since $\hat{t}$ is bijective on $\widehat{\Omega}$ to $(0,1]$, $\hat{t}^{-1}(\alpha, \alpha + \frac{1}{2^n}] = \widehat{E}(\varepsilon_1, \ldots, \varepsilon_n)$, implying that $\hat{t}_\# P((\alpha, \alpha + \frac{1}{2^n}]) = P(\widehat{E}(\varepsilon_1, \ldots, \varepsilon_n)) = \frac{1}{2^n} = \lambda((\alpha, \alpha + \frac{1}{2^n}])$. Now, if we let $\widehat{\mathcal{I}} = \{\hat{t}A : A \in \widehat{\mathcal{E}}\}$, then $\widehat{\mathcal{I}}$ is a π-system on $(0,1]$. Denote temporarily, in this example, by $\mathcal{B}$ and $\widehat{\mathcal{B}}$ the Borel fields on $[0,1]$ and on $(0,1]$ respectively, and let

$$\mathcal{M} = \{B \in \widehat{\mathcal{B}} : \hat{t}^{-1}B \in \sigma(\widehat{\mathcal{E}}) \text{ and } P(\hat{t}^{-1}B) = \lambda(B)\}.$$

As $\hat{t}$ is bijective from $\widehat{\Omega}$ to $(0,1]$, $\mathcal{M}$ is easily seen to be a λ-system on $(0,1]$ containing $\widehat{\mathcal{I}}$; and as $\sigma(\widehat{\mathcal{I}}) = \widehat{\mathcal{B}}$, we conclude by the ($\pi$-$\lambda$) theorem that $\mathcal{M} = \widehat{\mathcal{B}}$, i.e. $\widehat{\mathcal{B}} \subset \hat{t}_\#\sigma(\widehat{\mathcal{E}})$ and $\hat{t}_\# P(B) = \lambda(B)$ for all $B \in \widehat{\mathcal{B}}$.

(iii) We have shown in (ii) that $\widehat{\mathcal{B}} \subset \hat{t}_\#\sigma(\widehat{\mathcal{E}})$ and $\hat{t}_\# P(B) = \lambda(B)$ for all $B \in \widehat{\mathcal{B}}$; now it will be shown that $\widehat{\mathcal{B}} = \hat{t}_\#\sigma(\widehat{\mathcal{E}})$ and thus $((0,1], \widehat{\mathcal{B}}, \lambda)$ is the push-forward of $(\widehat{\Omega}, \sigma(\widehat{\mathcal{E}}), P)$ through the map $\hat{t}$. For this purpose, it is sufficient to claim that $\hat{t}A \in \widehat{\mathcal{B}}$ if $A \in \sigma(\widehat{\mathcal{E}})$. Consider $\mathcal{M} = \{A \in \sigma(\widehat{\mathcal{E}}) : \hat{t}A \in \widehat{\mathcal{B}}\}$. Clearly, $\mathcal{M}$ is a σ-algebra on $\widehat{\Omega}$ containing $\widehat{\mathcal{E}}$ and hence $\mathcal{M} = \sigma(\widehat{\mathcal{E}})$.

From the conclusions in (ii) and (iii) and the fact that $\hat{t}$ is bijective from $\widehat{\Omega}$ to $(0,1]$, we conclude that $B \in \widehat{\mathcal{B}}$ if and only if $\hat{t}^{-1}B \in \sigma(\widehat{\mathcal{E}})$ (equivalently, $A \in \sigma(\widehat{\mathcal{E}})$ if and only if $\hat{t}A \in \widehat{\mathcal{B}}$) and that $\hat{t}_\# P = \lambda$ on $\widehat{\mathcal{B}}$ and $\hat{t}_\#^{-1}\lambda = P$ on $\sigma(\widehat{\mathcal{E}})$. Therefore, $(\widehat{\Omega}, \sigma(\widehat{\mathcal{E}}), P)$ and $((0,1], \widehat{\mathcal{B}}, \lambda)$ are the same measure space labeled differently. Since $\sigma(\widehat{\mathcal{E}}) = \sigma(\mathcal{Q})|\widehat{\Omega}$ and $\Omega \backslash \widehat{\Omega}$ is countable, $([0,1], \mathcal{B}, \lambda)$ is the push-forward of $(\Omega, \sigma(\mathcal{Q}), P)$ through t and $B \in \mathcal{B}$ if and only if $t^{-1}B \in \sigma(\mathcal{Q})$.

Exercise 4.3.6 Let $(\Omega, \sigma(\mathcal{Q}), P)$ and t be as in Example 4.3.2 and P^* be the measure on Ω constructed from P by Method I. Show that $t_\# P^* = \lambda$ on $[0,1]$.

4.4 Functions of bounded variation

This section is devoted to the study of an important class of real-valued functions defined on a finite closed interval $I = [a,b]$. This is the class of **functions of bounded variation**. Functions in this section are all understood to be real-valued and defined on I.

For a real number α, α^+ denotes α or 0 according to whether $\alpha \geq 0$ or $\alpha < 0$, and $\alpha^- := (-\alpha)^+$. It is easily verified that $\alpha = \alpha^+ - \alpha^-$, $(\alpha + \beta)^+ \leq \alpha^+ + \beta^+$, and $(\alpha + \beta)^- \leq \alpha^- + \beta^-$ for any real numbers α and β.

Recall that a finite sequence $a = x_0 < x_1 < \cdots < x_l = b$ of points is called a partition of the interval I, where l varies from partition to partition. A generic partition of an interval will be denoted by $\mathcal{P}$.

Suppose that f is a function and $\mathcal{P} : a = x_0 < x_1 < \cdots < x_l = b$ a partition of I, let

$$P_a^b(f;\mathcal{P}) = \sum_{j=1}^{l}\{f(x_j) - f(x_{j-1})\}^+;$$
$$N_a^b(f;\mathcal{P}) = \sum_{j=1}^{l}\{f(x_j) - f(x_{j-1})\}^-;$$

and

$$V_a^b(f;\mathcal{P}) = \sum_{j=1}^{l}|f(x_j) - f(x_{j-1})|.$$

Observe that

$$V_a^b(f;\mathcal{P}) = P_a^b(f;\mathcal{P}) + N_a^b(f;\mathcal{P}).$$

Now put

$$P_a^b(f) = \sup_{\mathcal{P}} P_a^b(f;\mathcal{P});$$
$$N_a^b(f) = \sup_{\mathcal{P}} N_a^b(f;\mathcal{P});$$

and

$$V_a^b(f) = \sup_{\mathcal{P}} V_a^b(f;\mathcal{P}).$$

$P_a^b(f)$ and $N_a^b(f)$ are called respectively the **positive** and the **negative variation** of f over I, while $V_a^b(f)$ is called the **total variation** of f over I. When $a = b$, $V_a^b(f) = P_a^b(f) = N_a^b(f) = 0$, by definition. A function f is said to be **of bounded variation** on I if $V_a^b(f) < \infty$. Observe that a continuously differentiable function f is of bounded variation over I and $V_a^b(f) \le \int_a^b |f'(x)|dx$, and that a monotone function f is of bounded variation on I with $V_a^b(f) = |f(b) - f(a)|$.

Exercise 4.4.1 Show that $V_a^b(f) = P_a^b(f) + N_a^b(f)$.

Exercise 4.4.2 If $a < c < b$, show that $P_a^b(f) = P_a^c(f) + P_c^b(f)$ and similarly for negative and total variation.

Exercise 4.4.3 Show that if f and g are of bounded variation on I, then $\alpha f + \beta g$ is also of bounded variation on I for any real numbers α and β, and $V_a^b(\alpha f + \beta g) \le |\alpha|V_a^b(f) + |\beta|V_a^b(g)$.

Now suppose that f is a function of bounded variation on I. Let $x \in I$ and $\mathcal{P}$ be a partition of $[a, x]$, then

$$\begin{aligned} f(x) - f(a) &= \sum_{j=1}^{l} \{f(x_j) - f(x_{j-1})\} = P_a^x(f; \mathcal{P}) - N_a^x(f; \mathcal{P}) \\ &\leq P_a^x(f) - N_a^x(f; \mathcal{P}), \end{aligned}$$

or

$$f(x) - f(a) + N_a^x(f; \mathcal{P}) \leq P_a^x(f),$$

from which one infers that

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

Similarly, one has

$$f(x) - f(a) \geq P_a^x(f; \mathcal{P}) - N_a^x(f),$$

and hence

$$f(x) - f(a) + N_a^x(f) \geq P_a^x(f),$$

or

$$f(x) \geq f(a) + P_a^x(f) - N_a^x(f).$$

Consequently,

$$f(x) = f(a) + P_a^x(f) - N_a^x(f), \quad x \in I. \tag{4.3}$$

Since $P_a^x(f)$ and $N_a^x(f)$ are monotone increasing in x, it follows from (4.3) that f is a difference of two monotone increasing functions. Conversely, when f is a difference of two monotone increasing functions, then f is of bounded variation on I. Thus the first part of the following theorem has been shown.

Theorem 4.4.1 *A function f is of bounded variation on I if and only if f is a difference of two monotone increasing functions. Furthermore, if f is of bounded variation on I and $f = f_1 - f_2$, where f_1 and f_2 are monotone increasing and $f_1(a) = f(a)$, then there is a monotone increasing function φ on I with $\varphi(a) = 0$ such that*

$$f_1(x) = f(a) + P_a^x(f) + \varphi(x); \quad f_2(x) = N_a^x(f) + \varphi(x)$$

for $x \in I$.

Proof It remains to show the second part of the theorem. So suppose that f is of bounded variation on I and $f = f_1 - f_2$, where f_1 and f_2 are monotone increasing and $f_1(a) = f(a)$. From monotony of f_1 and f_2, one verifies that for $a \le x' < x'' \le b$,

$$\{f(x'') - f(x')\}^+ = \{f_1(x'') - f_1(x') + f_2(x') - f_2(x'')\}^+ \le f_1(x'') - f_1(x');$$
$$\{f(x'') - f(x')\}^- = \{f_1(x'') - f_1(x') + f_2(x') - f_2(x'')\}^- \le f_2(x'') - f_2(x').$$

From the preceding inequalities it then follows that for $a \le x < y \le b$ and any partition $\mathcal{P}$ of $[x, y]$,

$$P_x^y(f; \mathcal{P}) \le f_1(y) - f_1(x); \quad N_x^y(f; \mathcal{P}) \le f_2(y) - f_2(x),$$

and hence

$$P_x^y(f) \le f_1(y) - f_1(x); \quad N_x^y(f) \le f_2(y) - f_2(x). \tag{4.4}$$

In particular,

$$P_a^x(f) \le f_1(x) - f(a); \quad N_a^x(f) \le f_2(x)$$

for $x \in I$. Let $\varphi(x) = f_1(x) - \{f(a) + P_a^x(f)\}$, then $\varphi \ge 0$ and $\varphi(a) = 0$; from $f(a) + P_a^x(f) - N_a^x(f) = f(x) = f_1(x) - f_2(x)$, it follows that $f_2(x) = N_a^x(f) + \varphi(x)$ for $x \in I$. It remains to see that φ is monotone increasing. For $x < y$ in I we have

$$\varphi(y) - \varphi(x) = f_1(y) - f_1(x) - \{P_a^y(f) - P_a^x(f)\} = f_1(y) - f_1(x) - P_x^y(f) \ge 0,$$

by applying the first inequality in (4.4). This shows that φ is monotone increasing. ■

Henceforth, a function of bounded variation on I will simply be called a **BV function** on I. For a BV function f, let functions f_P, f_N, and f_V be defined by

$$f_P(x) = P_a^x(f); \quad f_N(x) = N_a^x(f); \quad \text{and } f_V(x) = V_a^x(f),$$

then the second part of Theorem 4.4.1 could be interpreted as saying that the decomposition $f = f(a) + f_P - f_N$ is the **minimal decomposition** of f into the difference of monotone increasing functions if a partial order $\prec$ on the family of all monotone increasing functions on I is defined as follows: $f \prec g$ if and only if $g - f$ is nonnegative and monotone increasing on I.

Theorem 4.4.2 *Suppose that f is a BV function on I. If f is right(left)-continuous at $x_0 \in [a, b)$ $(x_0 \in (a, b])$, then so are f_P, f_N, and f_V.*

Proof Since $f(x) - f(a) = f_P(x) - f_N(x)$ and $f_V(x) = f_P(x) + f_N(x)$,

$$f_P(x) = \frac{1}{2}\{f_V(x) + f(x) - f(a)\} \text{ and } f_N(x) = \frac{1}{2}\{f_V(x) - f(x) + f(a)\}$$

for $x \in I$, it is therefore sufficient to show that f_V is right-continuous at x_0. For this, we have to show that $f_V(x_{0^+}) = f_V(x_0)$, or $V_{x_0}^{x_0+h}(f) \to 0$ as $h \to 0+$.

Suppose the contrary, then $\delta_0 = f_V(x_0+) - f_V(x_0) > 0$. Let $\delta = \frac{2}{3}\delta_0$, and choose $h_1 > 0$ small enough so that $x_0 + h_1 \leq b$ and $V_{x_0}^{x_0+h_1}(f) < 2\delta$. Since $V_{x_0}^{x_0+h_1}(f) > \delta$, there is a partition $x_0 < x_1 < \cdots < x_l = x_0 + h_1$ such that

$$\sum_{j=1}^{l} |f(x_j) - f(x_{j-1})| > \delta.$$

As f is right-continuous at x_0, there is $h_2 > 0$ with $x_0 + h_2 < x_1$ such that $|f(x_1) - f(x_0 + h_2)| + \sum_{j=2}^{l} |f(x_j) - f(x_{j-1})| > \delta$; hence $V_{x_0+h_2}^{x_0+h_1}(f) > \delta$. Now repeat the above argument with h_1 replaced by h_2, to obtain $0 < h_3 < h_2$ such that $V_{x_0+h_3}^{x_0+h_2}(f) > \delta$. Then,

$$2\delta > V_{x_0}^{x_0+h_1}(f) \geq V_{x_0+h_3}^{x_0+h_2}(f) + V_{x_0+h_2}^{x_0+h_1}(f) > 2\delta,$$

which is absurd. Thus f_V is right-continuous at x_0. ■

Example 4.4.1 Let f be a Lebesgue integrable function on I and define

$$F(x) = \alpha + \int_a^x f(t)dt, \quad x \in I, \tag{4.5}$$

α being a constant. Then F is a BV function and

$$V_a^b(F) = \int_a^b |f(t)|dt.$$

Actually, for any partition $\mathcal{P} : a = x_0 < x_1 < \cdots < x_l = b$, we have

$$V_a^b(F; \mathcal{P}) = \sum_{j=1}^{l} \left| \int_{x_{j-1}}^{x_j} f(t)dt \right| \leq \int_a^b |f(t)|dt,$$

hence,

$$V_a^b(F) \leq \int_a^b |f(t)|dt < \infty. \tag{4.6}$$

Now, by Exercise 4.1.2, for any $\varepsilon > 0$ there is a step function g such that

$$\int_a^b |f(t) - g(t)|dt < \varepsilon.$$

Choose a partition $\mathcal{P} : a = x_0 < x_1 < \cdots < x_l = b$ of I such that $\{x_0, x_1, \ldots, x_l\}$ contains all the endpoints of the open intervals on which g is constant. We have then,

$$\begin{aligned}\sum_{j=1}^{l} |F(x_j) - F(x_{j-1})| &= \sum_{j=1}^{l} \left| \int_{x_{j-1}}^{x_j} f(t)dt \right| \\ &\geq \sum_{j=1}^{l} \left| \int_{x_{j-1}}^{x_j} g(t)dt \right| - \sum_{j=1}^{l} \left| \int_{x_{j-1}}^{x_j} (f(t) - g(t))dt \right| \\ &= \sum_{j=1}^{l} \int_{x_{j-1}}^{x_j} |g(t)|dt - \sum_{j=1}^{l} \left| \int_{x_{j-1}}^{x_j} (f(t) - g(t))dt \right| \\ &\geq \int_a^b |g(t)|dt - \int_a^b |f(t) - g(t)|dt \\ &\geq \int_a^b |g(t)|dt - \varepsilon \geq \int_a^b |f(t)|dt - 2\varepsilon.\end{aligned}$$

Thus,

$$V_a^b(F) \geq \int_a^b |f(t)|dt - 2\varepsilon.$$

Let $\varepsilon \to 0$; we have $V_a^b(F) \geq \int_a^b |f(t)|dt$, and hence

$$V_a^b(F) = \int_a^b |f(t)|dt,$$

by (4.6).

The function F, defined by (4.5), with f being Lebesgue integrable on I, is called an **indefinite integral** of f.

Exercise 4.4.4 Let F be an indefinite integral of f on I; show that $F_P(x) = \int_a^x f^+(t)dt$ and $F_N(x) = \int_a^x f^-(t)dt$. (Hint: use the fact that $F_P(x) = \frac{1}{2}\{F_V(x) + F(x) - F(a)\}$ and $F_N(x) = \frac{1}{2}\{F_V(x) - F(x) + F(a)\}$.)

4.5 Riemann–Stieltjes integral

The Rieman–Stieltjes integral of bounded functions on I will be defined along the same lines that the Riemann integral is defined. Suppose that g is a monotone increasing function defined on a finite closed interval $I = [a, b]$.

Given a partition $\mathcal{P} : a = x_0 < x_1 < \cdots < x_l = b$ of I and $j = 1, \ldots, l$; put

$$\mathcal{P}_j g = g(x_j) - g(x_{j-1}).$$

For a bounded function f on I, and $\mathcal{P}$ as above, let

$$\underline{f}_j = \inf_{x\in[x_{j-1},x_j]} f(x), \quad \bar{f}_j = \sup_{x\in[x_{j-1},x_j]} f(x);$$

and

$$\underline{S}_g(f,\mathcal{P}) = \sum_{j=1}^{l} \underline{f}_j \mathcal{P}_j g, \quad \bar{S}_g(f,\mathcal{P}) = \sum_{j=1}^{l} \bar{f}_j \mathcal{P}_j g.$$

Observe that for any partitions $\mathcal{P}$ and Q of I, the following sequence of inequalities holds:

$$\underline{S}_g(f,\mathcal{P}) \leq \underline{S}_g(f,\mathcal{P}\vee Q) \leq \bar{S}_g(f,\mathcal{P}\vee Q) \leq \bar{S}_g(f,Q). \tag{4.7}$$

Now let $\underline{\int_a^b} fdg = \sup_{\mathcal{P}} \underline{S}_g(f,\mathcal{P})$ and $\overline{\int_a^b} fdg = \inf_{\mathcal{P}} \bar{S}_g(f,\mathcal{P})$; by (4.7) both $\underline{\int_a^b} fdg$ and $\overline{\int_a^b} fdg$ are finite and $\underline{\int_a^b} fdg \leq \overline{\int_a^b} fdg$. In the case where $\underline{\int_a^b} fdg = \overline{\int_a^b} fdg$, f is said to be **Riemann–Stieltjes integrable** w.r.t. g and the common value, denoted $\int_a^b fdg$, is called the **Riemann–Stieltjes integral** of f w.r.t. g. From (4.7), Theorem 4.5.1 follows directly:

Theorem 4.5.1 *Let g be monotone increasing on $[a,b]$. A bounded function f on $[a,b]$ is Riemann–Stieltjes integrable w.r.t. g if and only if for any $\varepsilon > 0$, there is a partition $\mathcal{P}$ of $[a,b]$ such that*

$$\bar{S}_g(f,\mathcal{P}) - \underline{S}_g(f,\mathcal{P}) < \varepsilon.$$

Example 4.5.1 Let g be a monotone increasing function on $[a,b]$. (i) If f is continuous on $[a,b]$, then $\int_a^b fdg$ exists. (ii) If f is a BV function and g is continuous, then $\int_a^b fdg$ exists.

Clearly, (i) is an easy consequence of Theorem 4.5.1, while (ii) follows also from Theorem 4.5.1 if one notes that for any partition $\mathcal{P} : a = x_0 < x_1 < \cdots < x_n = b$ of $[a,b]$,

$$\begin{aligned}\bar{S}_g(f,\mathcal{P}) - \underline{S}_g(f,\mathcal{P}) &= \sum_{j=1}^{n} (\underline{f}_j - \bar{f}_j)\mathcal{P}_j(g) \\ &\leq \sum_{j=1}^{n} V_{x_{j-1}}^{x_j}(f)\mathcal{P}_j g \leq V_a^b(f) \max_{1\leq j\leq n} \mathcal{P}_j g.\end{aligned}$$

Example 4.5.2 Suppose that w is a nonnegative Lebesgue integrable function on $[a,b]$, and g is an indefinite integral of w (cf. Example 4.4.1), then any Riemann integrable function f on $[a,b]$ is Riemann–Stieltjes integrable w.r.t. g on $[a,b]$ and

$$\int_a^b fdg = \int_a^b f(t)w(t)dt.$$

For a partition $\mathcal{P} : a = x_0 < x_1 < \cdots < x_n = b$, define a function $\bar{f}^{\mathcal{P}}$ by $\bar{f}^{\mathcal{P}}(x) = \bar{f}_j$ if $x \in [x_{j-1}, x_j)$ and $\bar{f}^{\mathcal{P}}(b) = f(b)$; similarly define $\underline{f}^{\mathcal{P}}$ by $\underline{f}^{\mathcal{P}}(x) = \underline{f}_j$ if $x \in [x_{j-1}, x_j)$ and $\underline{f}^{\mathcal{P}}(b) = f(b)$. Now choose a sequence $\{\mathcal{P}^{(k)}\}$ of partitions so that $\|\mathcal{P}^{(k)}\| \to 0$ as $k \to \infty$, and

$$\bar{S}_g(f, \mathcal{P}^{(k)}) \to \overline{\int_a^b} fdg; \quad \underline{S}_g(f, \mathcal{P}^{(k)}) \to \underline{\int_a^b} fdg.$$

Obviously,

$$\bar{S}_g(f, \mathcal{P}^{(k)}) = \int_a^b \bar{f}^{\mathcal{P}^{(k)}}(t)w(t)dt; \quad \underline{S}_g(f, \mathcal{P}^{(k)}) = \int_a^b \underline{f}^{\mathcal{P}^{(k)}}(t)w(t)dt.$$

Since f is Riemann integrable, f is continuous at almost all points of $[a, b]$, and hence

$$\bar{f}^{\mathcal{P}^{(k)}} w \to fw \text{ a.e.}; \quad \underline{f}^{\mathcal{P}^{(k)}} w \to fw \text{ a.e.}$$

If we put $M = \sup_{t \in [a,b]} |f(t)|$, $|\bar{f}^{\mathcal{P}^{(k)}} w| \le Mw$, $|\underline{f}^{\mathcal{P}^{(k)}} w| \le Mw$, hence by LDCT

$$\lim_{k \to \infty} \bar{S}_g(f, \mathcal{P}^{(k)}) = \int_a^b f(t)w(t)dt = \lim_{k \to \infty} \underline{S}_g(f, \mathcal{P}^{(k)}),$$

and thus

$$\overline{\int_a^b} fdg = \underline{\int_a^b} fdg = \int_a^b f(t)w(t)dt,$$

i.e. f is Riemann–Stieltjes integrable w.r.t. g on $[a, b]$ and $\int_a^b fdg = \int_a^b f(t)w(t)dt$.

Exercise 4.5.1 Suppose that f is continuous on $[a, b]$ and g is monotone increasing on $[a, b]$.

(i) Show that $\int_a^b fdg = \overline{\int_a^b} fdg = \inf \bar{S}_g(f, \mathcal{P})$, where the infimum is taken over all those partitions $\mathcal{P}$, the endpoints of whose intervals other than a and b are points of continuity of g.

(ii) Show that $\int_a^b fdg = \int_a^b fd\mu_g$.

The following Lemma is a generalization of Lemma 4.2.2 when $n = 1$.

Lemma 4.5.1 *Suppose that g is a right-continuous and monotone increasing function on $[a, b]$, and f a bounded function on $[a, b]$ which is continuous wherever g is discontinuous, then*

$$\underline{\int_a^b} fdg = \int_a^b \underline{f} d\mu_g; \qquad \overline{\int_a^b} fdg = \int_a^b \overline{f} d\mu_g,$$

where $\underline{f}(x) = \lim_{\delta \to 0+} \inf_{|y-x|<\delta} f(y)$ *and* $\overline{f}(x) = \lim_{\delta \to 0+} \sup_{|y-x|<\delta} f(y)$.

Proof By Lemma 4.2.1, both $\underline{f}$ and $\overline{f}$ are Lebesgue measurable. It is clear that $\underline{f} \le f \le \overline{f}$ on $[a, b]$. Choose a sequence $\{\mathcal{P}^{(k)}\}$ of partitions of $[a, b]$ such that $\|\mathcal{P}^{(k)}\| \to 0$, and

$$\underline{\int_a^b} fdg = \lim_{k\to\infty} \underline{S}_g(f, \mathcal{P}^{(k)}); \qquad \overline{\int_a^b} fdg = \lim_{k\to\infty} \overline{S}_g(f, \mathcal{P}^{(k)}).$$

For each $k \in \mathbb{N}$, let $\mathcal{P}^{(k)}$ be $a = x_0^{(k)} < x_1^{(k)} < \cdots < x_{n_k}^{(k)} = b$, and define $f_k(x) = \inf_{x_{j-1}^{(k)} \le t \le x_j^{(k)}} f(t)$ if $x \in (x_{j-1}^{(k)}, x_j^{(k)}]$ and $f_k(a) = f(a)$. As we have shown in the proof of Lemma 4.2.2, $\lim_{k\to\infty} f_k(x) = \underline{f}(x)$ if $x \in [a, b]$, but is not an endpoint of intervals of the partitions $\mathcal{P}^{(k)}$, $k = 1, 2, \ldots$. Now, since f is continuous wherever g is discontinuous and g is right-continuous, we may assume that all the endpoints of the intervals of the partitions $\mathcal{P}^{(k)}$ are points of continuity of g, except possibly b. Hence $f_k(x) \to \underline{f}(x)$ for μ_g-a.e. x in $[a, b)$; but if b is a point of discontinuity of g, then f is continuous at b and hence $f_k(b) \to f(b) = \underline{f}(b)$. Thus, $\lim_{k\to\infty} f_k(x) = \underline{f}(x)$ for μ_g-a.e. x of $[a, b]$, and $\int_a^b \underline{f} d\mu_g = \lim_{k\to\infty} \int_a^b f_k d\mu_g$ by LDCT, because $|f_k(x)| \le \sup_{a\le t\le b} |f(t)|$. Since g is right-continuous, $\mu_g((c, d]) = g(d) - g(c)$ for $a \le c < d \le b$; we have $\underline{S}_g(f, \mathcal{P}^{(k)}) = \int_a^b f_k d\mu_g$. Consequently, $\int_a^b \underline{f} d\mu_g = \lim_{k\to\infty} \underline{S}_g(f, \mathcal{P}^{(k)}) = \underline{\int_a^b} fdg$. Similarly, $\int_a^b \overline{f} d\mu_g = \overline{\int_a^b} fdg$. ■

Theorem 4.5.2 *Suppose that g is a right-continuous and monotone increasing function on $[a, b]$ and f is a bounded function which is continuous at the μ_g-a.e. point of $[a, b]$, then f is Riemann–Stieltjes integrable w.r.t. g, and*

$$\int_a^b fdg = \int_a^b fd\mu_g.$$

Proof We claim first that f is μ_g-measurable. From $\underline{f} \le f \le \overline{f}$ and the fact that f is continuous μ_g-a.e., it follows that $\underline{f}(x) = f(x) = \overline{f}(x)$ for μ_g-a.e. x in $[a, b]$; hence f differs from $\underline{f}$ only on a set A with $\mu_g(A) = 0$. But $\underline{f}$ is Borel measurable by Lemma 4.2.1, and is therefore μ_g-measurable from the fact that μ_g is a Carathéodory measure.

Thus f is μ_g-measurable as we claim. Now, $\underline{f} = f = \bar{f}$ μ_g-a.e. implies, together with Lemma 4.5.1, that

$$\underline{\int_a^b} f dg = \int_a^b \underline{f} d\mu_g = \int_a^b f d\mu_g = \int_a^b \bar{f} d\mu_g = \overline{\int_a^b} f dg,$$

which entails that f is Riemann–Stieltjes integrable w.r.t. g and $\int_a^b f dg = \int_a^b f d\mu_g$. ■

Theorem 4.5.3 (Integration by parts) *Suppose that f and g are monotone increasing functions on $[a, b]$ and at least one of them is continuous. Then*

$$\int_a^b f dg = f(b)g(b) - f(a)g(a) - \int_a^b g df.$$

Proof Note firstly that $\int_a^b f dg$ and $\int_a^b g df$ exist, from Example 4.5.1. Let $\mathcal{P}$: $a = x_0 < x_1 < \cdots < x_l = b$ be a partition of $[a, b]$, then

$$\begin{aligned}\bar{S}_g(f, \mathcal{P}) &= \sum_{j=1}^{l} f(x_j)[g(x_j) - g(x_{j-1})] \\ &= f(b)g(b) - f(a)g(a) - \sum_{j=1}^{l} g(x_{j-1})[f(x_j) - f(x_{j-1})] \\ &= f(b)g(b) - f(a)g(a) - \underline{S}_f(g, \mathcal{P}),\end{aligned}$$

from which, by taking a sequence $\{\mathcal{P}^{(k)}\}$ of partitions such that

$$\lim_{k\to\infty} \bar{S}_g(f, \mathcal{P}^{(k)}) = \overline{\int_a^b} f dg \quad \text{and} \quad \lim_{k\to\infty} \underline{S}_f(g, \mathcal{P}^{(k)}) = \underline{\int_a^b} g df,$$

we obtain,

$$\begin{aligned}\int_a^b f dg = \overline{\int_a^b} f dg &= f(b)g(b) - f(a)g(a) - \underline{\int_a^b} g df \\ &= f(b)g(b) - f(a)g(a) - \int_a^b g df.\end{aligned}$$ ■

Exercise 4.5.2 Under the same assumptions as in Theorem 4.5.3, show that

$$\int_a^b f d\mu_g = f(b)g(b) - f(a)g(a) - \int_a^b g d\mu_f.$$

(Hint: cf. Exercise 4.5.1.)

Now suppose that g is a BV function on $[a, b]$ and write $g = g_1 - g_2$, where $g_1(x) = g(a) + g_P(x)$ and $g_2(x) = g_N(x)$ for $x \in [a, b]$. Recall that $g_P(x) = P_a^x(g)$ and $g_N(x) = N_a^x(g)$, $x \in [a, b]$. A bounded function f on $[a, b]$ is called Riemann–Stieltjes integrable w.r.t. g if it is Riemann–Stieltjes integrable w.r.t. g_1 and g_2, and in this case the Riemann–Stieltjes integral of f w.r.t. g, denoted $\int_a^b fdg$, is defined by

$$\int_a^b fdg = \int_a^b fdg_1 - \int_a^b fdg_2.$$

With this definition, Corollary 4.5.1 of Theorem 4.5.3 follows, by using Theorem 4.4.2.

Corollary 4.5.1 *Suppose that f and g are BV functions on $[a, b]$ and at least one of them is continuous, then*

$$\int_a^b fdg = f(b)g(b) - f(a)g(a) - \int_a^b gdf.$$

Theorem 4.5.4 (Second mean-value theorem) *Suppose that f is an integrable function on a finite interval $[a, b]$ and φ is a monotone function on $[a, b]$, then there is $c \in [a, b]$ such that*

$$\int_a^b \varphi f d\lambda = \varphi(a) \int_a^c f d\lambda + \varphi(b) \int_c^b f d\lambda.$$

Firstly we prove a lemma.

Lemma 4.5.2 *Let f and φ be as in Theorem 4.5.4 and, further φ is assumed to be nonnegative and monotone decreasing, then there is $c \in [a, b]$ such that*

$$\int_a^b \varphi f d\lambda = \varphi(a) \int_a^c f d\lambda.$$

Proof We may assume that $\varphi(a) > 0$, because otherwise $\varphi \equiv 0$ and the lemma is trivial. Define a function F on $[a, b]$ by

$$F(x) = \int_a^x f d\lambda, \quad x \in [a, b].$$

By Corollary 4.5.1 and Example 4.5.2, we have

$$\int_a^b F d\varphi = F(b)\varphi(b) - F(a)\varphi(a) - \int_a^b \varphi dF = F(b)\varphi(b) - \int_a^b \varphi f d\lambda;$$

hence,

$$\int_a^b \varphi f d\lambda \le M\varphi(b) - M \int_a^b d\varphi = M\varphi(b) + M\{\varphi(a) - \varphi(b)\} = M\varphi(a),$$

or $\frac{1}{\varphi(a)}\int_a^b \varphi f d\lambda \le M$, where $M = \max_{x\in[a,b]} F(x)$. Similarly, if $m = \min_{x\in[a,b]} F(x)$, then $m \le \frac{1}{\varphi(a)}\int_a^b \varphi f d\lambda$; thus,

$$m \le \frac{1}{\varphi(a)}\int_a^b \varphi f d\lambda \le M,$$

from which, by the intermediate-value theorem for continuous functions, there is $c \in [a, b]$ such that $\frac{1}{\varphi(a)}\int_a^b \varphi f d\lambda = F(c) = \int_a^c f d\lambda$. ■

Proof of Theorem 4.5.4 Consider first the case where φ is monotone decreasing. Since $\varphi - \varphi(b)$ is nonnegative and monotone decreasing, by Lemma 4.5.2 there is $c \in [a, b]$ such that $\int_a^b \{\varphi - \varphi(b)\} f d\lambda = \{\varphi(a) - \varphi(b)\}\int_a^c f d\lambda$, or

$$\begin{aligned}\int_a^b \varphi f d\lambda &= \varphi(b)\int_a^b f d\lambda + \{\varphi(a) - \varphi(b)\int_a^c f d\lambda \\ &= \varphi(a)\int_a^c f d\lambda + \varphi(b)\int_c^b f d\lambda.\end{aligned}$$

If φ is monotone increasing, replacing φ by $-\varphi$ in the argument above, we also conclude that there is $c \in [a, b]$ such that

$$\int_a^b \varphi f d\lambda = \varphi(a)\int_a^c f d\lambda + \varphi(b)\int_c^b f d\lambda.$$ ■

Corollary 4.5.2 *Let f be integrable on $[a, b]$ and φ be nonnegative and monotone increasing on $[a, b]$; then there is $c \in [a, b]$ such that*

$$\int_a^b \varphi f d\lambda = \varphi(b)\int_c^b f d\lambda.$$

Proof Replace φ in Theorem 4.5.4 by $\varphi - \varphi(a)$. ■

Remark Lemma 4.5.2, Theorem 4.5.4, and Corollary 4.5.2 will all be referred to as the **second mean-value theorem**.

Exercise 4.5.3 Show that the following improper integrals exist: (i) $\int_0^\infty \frac{\sin x}{x} dx$; (ii) $\int_0^\infty \frac{\sin x}{e^x - 1} dx$.

Exercise 4.5.4 Suppose that h is an integrable function on $[a, b]$ and g is an indefinite integral of h. Show that if f is a Riemann integrable function on $[a, b]$, then f is Riemann–Stieltjes integrable w.r.t. g, and

$$\int_a^b f dg = \int_a^b f h d\lambda.$$

Exercise 4.5.5 Suppose that u and v are integrable functions on $[a, b]$ and that U and V are respectively indefinite integrals of u and v. Show that

$$\int_a^b Uv d\lambda = U(b)V(b) - U(a)V(a) - \int_a^b Vu d\lambda.$$

Exercise 4.5.6 Let f be a measurable and finite a.e. function on a measure space (Ω, Σ, μ). Suppose that $\mu(\{f \leq t\}) < \infty$ for every $t \in \mathbb{R}$ and let $F(t) = \mu(\{f \leq t\})$ for $t \in \mathbb{R}$. Define the improper Riemann–Stieltjes integral $\int_{\mathbb{R}} |t|^p dF$ by

$$\int_{\mathbb{R}} |t|^p dF = \lim_{\substack{b\to\infty \\ a\to-\infty}} \int_a^b |t|^p dF, \quad 1 \leq p < \infty.$$

Show that $\int_\Omega |f|^p d\mu = \int_{\mathbb{R}} |t|^p dF$.

A characterization of functions which are indefinite integrals will be taken up after a treatise on differentiation is given in Section 4.6.

4.6 Covering theorems and differentiation

Our purpose in this section is to establish the Lebesgue differentiation theorem for Radon measures on $\mathbb{R}^n$ and to give some of its relevant applications. To do this, we shall first exhibit a useful procedure of selecting a sequence of disjoint balls from a given collection of balls in $\mathbb{R}^n$, and deduce from it two covering theorems in $\mathbb{R}^n$; one of which is elementary but will be useful when we study the Hardy–Littlewood maximal function in Chapter 6, and the other is a Vitali type covering theorem that is the main tool for the proof of the Lebesgue differentiation theorem.

For convenience, the diameter of a set A is denoted by δA instead of diam A, for the moment, and a ball is either open or closed with positive radius unless, specified explicitly. For a ball B, we shall denote by $\widehat{B}$ the ball concentric with B and with radius five times that of B.

A collection $\mathcal{C}$ of balls in $\mathbb{R}^n$ is said to be **admissible** if $\sup_{B\in\mathcal{C}} \delta B < \infty$. Given an admissible collection $\mathcal{C}$ of balls in $\mathbb{R}^n$, we select a disjoint sequence $\{B_j\}$, finite or infinite, from $\mathcal{C}$ by the following procedure. Let $d_0 = \sup_{B\in\mathcal{C}} \delta B$, then $0 < d_0 < \infty$. Choose a ball B_1 in $\mathcal{C}$ such that $\delta B_1 \geq \frac{1}{2} d_0$. Suppose now that $B_1, \ldots, B_m$ are disjoint balls chosen from $\mathcal{C}$; if $B \cap \bigcup_{j=1}^m B_j \neq \emptyset$ for every $B \in \mathcal{C}$, stop the procedure; otherwise, let

$$d_m = \sup \left\{ \delta B : B \in \mathcal{C},\ B \cap \bigcup_{j=1}^m B_j = \emptyset \right\},$$

and choose a ball B_{m+1} from $\mathcal{C}$ which is disjoint with $\bigcup_{j=1}^m B_j$ and with $\delta B_{m+1} \geq \frac{1}{2} d_m$. Thus a disjoint sequence $\{B_j\}$, finite or infinite, is obtained by this procedure. Such a procedure of selecting $\{B_j\}$ from $\mathcal{C}$ will be referred to as **Procedure**(S).

Lemma 4.6.1 *Suppose that $\mathcal{C}$ is an admissible collection of balls in $\mathbb{R}^n$ and $\{B_j\}$ is a sequence of disjoint balls selected from $\mathcal{C}$ by* Procedure(S). *Then either $\{B_j\}$ is infinite and* $\inf_j \delta B_j > 0$ *or* $\bigcup \mathcal{C} \subset \bigcup_j \widehat{B}_j$ *(recall that* $\bigcup \mathcal{C} := \bigcup_{B \in \mathcal{C}} B$*).*

Proof If $\{B_j\}$ is finite, say $\{B_j\} = \{B_1, \ldots, B_m\}$, meaning that if $B \in \mathcal{C}$, then $B \cap \bigcup_{j=1}^m B_j \neq \emptyset$. Let j_0 be the smallest j, $1 \leq j \leq m$, such that $B \cap B_j \neq \emptyset$. If $j_0 = 1$, then $\delta B \leq d_0 \leq 2B_1$; while if $j_0 \geq 2$, $B \cap \bigcup_{j=1}^{j_0-1} B_j = \emptyset$ and $\delta B \leq d_{j_0-1} \leq 2\delta B_{j_0}$. Hence, $\delta B \leq 2\delta B_{j_0}$ holds; this fact, together with $B \cap B_{j_0} \neq \emptyset$, implies that $B \subset \widehat{B}_{j_0}$. Thus, $\bigcup \mathcal{C} \subset \bigcup_{j=1}^m \widehat{B}_j$.

Now suppose that $\{B_j\}$ is infinite and $\inf_j \delta B_j = 0$. Let again $B \in \mathcal{C}$. Since $\delta B > 0$ and $\inf_j \delta B_j = 0$, there is $j_0 \in \mathbb{N}$ arbitrarily large such that $\delta B > 2\delta B_{j_0}$. But then $B \cap \bigcup_{j=1}^{j_0-1} B_j \neq \emptyset$, because otherwise $\delta B_{j_0} < \frac{1}{2}\delta B \leq \frac{1}{2} d_{j_0-1}$, contradicting the way B_{j_0} is selected by Procedure(S). Since $B \cap \bigcup_{j=1}^{j_0-1} B_j \neq \emptyset$, argue as in the first paragraph of the proof to conclude that B is contained in one of $\widehat{B}_1, \ldots, \widehat{B}_{j_0-1}$, and hence $B \subset \bigcup_j \widehat{B}_j$. Consequently, $\bigcup \mathcal{C} \subset \bigcup_j \widehat{B}_j$. ■

Lemma 4.6.1 leads immediately to the following basic **covering theorem**.

Theorem 4.6.1 *Let $\mathcal{C}$ be an admissible collection of balls in $\mathbb{R}^n$; then there is a disjoint sequence $\{B_j\}$ of balls from $\mathcal{C}$ such that*

$$\lambda^n\left(\bigcup \mathcal{C}\right) \leq 5^n \sum_j \lambda^n(B_j). \tag{4.8}$$

Proof Let $\{B_j\}$ be a sequence of disjoint balls selected from $\mathcal{C}$ by Procedure(S). By Lemma 4.6.1, either $\{B_j\}$ is infinite and $\inf \delta B_j > 0$ or $\bigcup \mathcal{C} \subset \bigcup_j \widehat{B}_j$. If $\{B_j\}$ is infinite and $\inf_j \delta B_j > 0$, then the right-hand side of (4.8) is ∞ and (4.8) holds trivially. Suppose now that $\bigcup \mathcal{C} \subset \bigcup_j \widehat{B}_j$. Then,

$$\lambda^n\left(\bigcup \mathcal{C}\right) \leq \sum_j \lambda^n(\widehat{B}_j) = 5^n \sum_j \lambda^n(B_j),$$

because $\lambda^n(\widehat{B}_j) = 5^n \lambda^n(B_j)$, by Example 4.3.1 (ii). ■

We come now to a **Vitali type** covering theorem. Let E be a subset of $\mathbb{R}^n$; a collection $\mathcal{V}$ of subsets of $\mathbb{R}^n$ is called a **Vitali cover** of E if for every x in E and any positive number ε there is V in $\mathcal{V}$, such that $\delta V < \varepsilon$ and $x \in V$. The following covering theorem is a simple version of the well-known Vitali covering theorem, but it suffices for our purpose.

Theorem 4.6.2 (Vitali) *Let E be a subset of $\mathbb{R}^n$ with $\lambda^n(E) < \infty$, and suppose that $\mathcal{V}$ is a collection of closed balls in $\mathbb{R}^n$ which forms a Vitali cover of E. Then there is a sequence $\{B_j\}$ of disjoint balls from $\mathcal{V}$ such that $\lambda^n(E \setminus \bigcup_j B_j) = 0$.*

Proof Choose an open set $G \supset E$ such that $\lambda^n(G) < \infty$, and let

$$\mathcal{C} = \{V \in \mathcal{V} : V \subset G,\ \delta V \leq 1\}.$$

Then $\mathcal{C}$ is an admissible collection of closed balls and is a Vitali cover of E. Now select a sequence $\{B_j\}$ of disjoint balls from $\mathcal{C}$ by Procedure(S). If $\{B_j\}$ is finite, say $\{B_j\} = \{B_1, \ldots, B_m\}$, then $V \cap \bigcup_{j=1}^m B_j \neq \emptyset$ for every $V \in \mathcal{C}$. Take any $x \in E$ and $\varepsilon > 0$, choose $V \in \mathcal{C}$ such that $x \in V$ and $\delta V < \varepsilon$, then $\operatorname{dist}(x, \bigcup_{j=1}^m B_j) \leq \operatorname{dist}(x, V \cap \bigcup_{j=1}^m B_j) \leq \delta V < \varepsilon$. Since $\varepsilon > 0$ is arbitrary and $\bigcup_{j=1}^m B_j$ is closed, we infer that $x \in \bigcup_{j=1}^m B_j$ or $E \subset \bigcup_{j=1}^m B_j$, and hence $\lambda^n(E \backslash \bigcup_{j=1}^m B_j) = 0$. Suppose now that $\{B_j\}$ is infinite. Since $\sum_j \lambda^n(B_j) = \lambda^n(\bigcup_j B_j) \leq \lambda^n(G) < \infty$, $\inf_{j \geq l} \delta(B_j) = 0$ for any $l \in \mathbb{N}$. Observe then that for any $l \in \mathbb{N}$, $\{B_j\}_{j \geq l+1}$ is a sequence of balls selected from the admissible collection

$$\mathcal{C}^{(l)} := \left\{ V \in \mathcal{C} : V \subset G \backslash \bigcup_{j=1}^{l} B_j \right\},$$

by Procedure(S). Since $\inf_{j \geq l+1} \delta B_j = 0$, it follows from Lemma 4.6.1 that $\bigcup \mathcal{C}^{(l)} \subset \bigcup_{j \geq l+1} \widehat{B}_j$; consequently,

$$\lambda^n \left(E \backslash \bigcup_{j=1}^{l} B_j \right) \leq \lambda^n \left(\bigcup \mathcal{C}^{(l)} \right) \leq \sum_{j \geq l+1} \lambda^n(\widehat{B}_j) = 5^n \sum_{j \geq l+1} \lambda^n(B_j),$$

because $\mathcal{C}^{(l)}$ is a Vitali cover of $E \backslash \bigcup_{j=1}^{l} B_j$. Now from

$$\lambda^n \left(E \backslash \bigcup_j B_j \right) \leq \lambda^n \left(E \backslash \bigcup_{j=1}^{l} B_j \right) \leq 5^n \sum_{j \geq l+1} \lambda^n(B_j)$$

for $l \in \mathbb{N}$, we obtain $\lambda^n(E \backslash \bigcup_j B_j) = 0$ by letting $l \to \infty$. ■

Remark In Theorem 4.6.2, E is not required to be measurable.

Exercise 4.6.1 Show that the union of any family $\mathcal{C}$ of closed balls in $\mathbb{R}^n$ is Lebesgue measurable. (Hint: consider the Vitali cover $\mathcal{V}$ of $\bigcup \mathcal{C}$, which consists of all closed balls each of which is contained in a ball of $\mathcal{C}$).

Exercise 4.6.2 Show that Theorem 4.6.2 still holds if $\mathcal{V}$ is a Vitali cover of E consisting of open balls.

Exercise 4.6.3 Describe in $\mathbb{R}^n$ a procedure for selecting a sequence of disjoint closed cubes, from a collection $\mathcal{C}$ of closed cubes of positive bounded side lengths similar to Procedure(S), when $\mathcal{C}$ is an admissible collection of closed cubes so that Lemma 4.6.1 holds for such a collection $\mathcal{C}$. Then state the Vitali covering theorem for Vitali covers of E consisting of closed (open) cubes, where E is a subset of $\mathbb{R}^n$ with $\lambda^n(E) < \infty$.

Lebesgue differentiation of Radon measures on $\mathbb{R}^n$ is the subject we shall treat in the remaining part of this section. The differentiation is taken w.r.t. Lebesgue measure and with closed balls as base in the sense which will be defined. For the sake of simplicity in expression, a generic closed ball in $\mathbb{R}^n$ is henceforth denoted by B in this section.

Since the expression "λ^n-almost everywhere" appears often, it will hereafter be replaced by "almost everywhere". In other words, a property which holds almost everywhere w.r.t. Lebesgue measure λ^n in $\mathbb{R}^n$ will simply be said to hold almost everywhere. Accordingly, "λ^n-a.e." is often replaced by "a.e.", and λ^n-null sets will simply be called null sets.

Suppose that f is a set function (not necessarily taking only nonnegative values) defined for all closed balls inside an open set $\Omega \subset \mathbb{R}^n$ and $x \in \Omega$, define

$$\liminf_{B\to x} f(B) := \lim_{\sigma\to 0+}\Big\{\inf_{\substack{\delta B<\sigma\\ x\in B}} f(B)\Big\};$$

$$\limsup_{B\to x} f(B) := \lim_{\sigma\to 0+}\Big\{\sup_{\substack{\delta B<\sigma\\ x\in B}} f(B)\Big\}.$$

Clearly, $\liminf_{B\to x} f(B) \le \limsup_{B\to x} f(B)$; in the case $\liminf_{B\to x} f(B) = \limsup_{B\to x} f(B)$, the common value is denoted by $\lim_{B\to x} f(B)$ and we say that $\lim_{B\to x} f(B)$ exists. In the above definitions, B certainly denotes a generic closed ball B in Ω.

Exercise 4.6.4 Show that $\lim_{B\to x} f(B)$ exists and is a finite number l if and only if for any given $\varepsilon > 0$ there is $\sigma > 0$, such that

$$|f(B) - l| < \varepsilon$$

whenever $\delta B < \sigma$ and $x \in B$.

Now let μ be a Radon measure on an open set $\Omega \subset \mathbb{R}^n$; μ is said to be **differentiable** w.r.t. Lebesgue measure λ^n at $x \in \Omega$ with closed balls as base if $\lim_{B\to x} \frac{\mu(B)}{\lambda^n(B)}$ exists. Since the differentiation of Radon measures on $\mathbb{R}^n$ is always taken in this sense in what follows, if $\lim_{B\to x} \frac{\mu(B)}{\lambda^n(B)}$ exists, we simply say that μ is **differentiable at** x with derivate $\frac{d\mu}{d\lambda^n}(x) := \lim_{B\to x} \frac{\mu(B)}{\lambda^n(B)}$. We shall show that μ is differentiable with finite derivate at a.e. x of Ω, and that the function $\frac{d\mu}{d\lambda^n}$ which is defined and finite almost everywhere on Ω is measurable.

Put $\underline{D}\mu(x) = \liminf_{B\to x} \frac{\mu(B)}{\lambda^n(B)}$ and $\bar{D}\mu(x) = \limsup_{B\to x} \frac{\mu(B)}{\lambda^n(B)}$ for $x \in \Omega$. Note that $\underline{D}\mu(x) \le \bar{D}\mu(x)$ for every $x \in \Omega$.

Lemma 4.6.2 *If $\bar{D}\mu \ge \alpha$ on $S \subset \Omega$ for some $\alpha \ge 0$, then $\mu(S) \ge \alpha\lambda^n(S)$.*

Proof Clearly we may assume that $\alpha > 0$. For $l \in \mathbb{N}$, let $S_l = \{x \in S : |x| < l\}$ and let G be any open set which contains S_l and is contained in Ω. Now for any $\varepsilon > 0$ sufficiently small so that $\alpha - \varepsilon > 0$, consider the family $\mathcal{V}$ of all those closed balls $B \subset G$ such that $\mu(B) > (\alpha - \varepsilon)\lambda^n(B)$. Since $\mathcal{V}$ is a Vitali cover of S_l and $\lambda^n(S_l) < \infty$, there is a disjoint sequence $\{B_j\}$ of balls from $\mathcal{V}$ such that $\lambda^n(S_l \setminus \bigcup_j B_j) = 0$, by Vitali the covering theorem (Theorem 4.6.2). Then, $(\alpha - \varepsilon)\lambda^n(S_l) \le (\alpha - \varepsilon)\lambda^n$

$(\bigcup_j B_j) = \sum_j (\alpha - \varepsilon)\lambda^n(B_j) < \sum_j \mu(B_j) = \mu(\bigcup_j B_j) \le \mu(G)$, and since μ is a Radon measure, it follows that $(\alpha - \varepsilon)\lambda^n(S_l) \le \mu(S_l)$ and consequently, by letting $l \to \infty$, $(\alpha - \varepsilon)\lambda^n(S) \le \mu(S)$ follows, as both λ^n and μ are regular measures and S_l increases to S when $l \to \infty$ (cf. Theorem 3.3.2). Finally, let $\varepsilon \searrow 0$ to conclude the proof. ■

Corollary 4.6.1 $\bar{D}\mu < \infty$ *almost everywhere on* Ω.

Proof Since $\Omega = \bigcup_{l\in\mathbb{N}}(\{x \in \Omega : \operatorname{dist}(x, \Omega^c) \ge \frac{1}{l}\} \cap \{x \in \mathbb{R}^n : |x| \le l\})$, Ω is a countable union of compact sets; it is sufficient to show that $\lambda^n(\{x \in K : \bar{D}\mu(x) = \infty\}) = 0$ for any compact set in Ω. For such a compact set K, put $S = \{x \in K : \bar{D}\mu(x) = \infty\}$. Since for any $\alpha > 0$, $\bar{D}\mu \ge \alpha$ on S, by Lemma 4.6.2, $\lambda^n(S) \le \frac{1}{\alpha}\mu(S) \le \frac{1}{\alpha}\mu(K)$, which implies that $\lambda^n(S) = 0$ by letting $\alpha \to \infty$, because $\mu(K) < \infty$. ■

Lemma 4.6.3 *Suppose that* $\underline{D}\mu \le \beta$ *on* $S \subset \Omega$ *for some* $\beta \ge 0$*; then there is a null set* $N \subset S$ *such that* $\mu(S\backslash N) \le \beta\lambda^n(S)$.

Proof Suppose first that $\lambda^n(S) < \infty$. For $l, k \in \mathbb{N}$, take an open set G_k which contains S and is contained in Ω with $\lambda^n(G_k) < \lambda^n(S) + \frac{1}{k}$, and consider the family $\mathcal{V}$ of all those closed balls $B \subset G_k$ such that $\mu(B) < (\beta + \frac{1}{l})\lambda^n(B)$; $\mathcal{V}$ is clearly a Vitali cover of S. Since $\lambda^n(S) < \infty$, by the Vitali covering theorem there is a disjoint sequence $\{B_j\}$ of balls from $\mathcal{V}$ such that $\lambda^n(S\backslash \bigcup_j B_j) = 0$. If we let $N_{l,k} = S\backslash \bigcup_j B_j$ (observe that $\{B_j\}$ depends on l and k), $N_{l,k}$ is a null set contained in S and $(\beta + \frac{1}{l})\lambda^n(G_k) \ge (\beta + \frac{1}{l})\lambda^n(\bigcup_j B_j) > \mu(\bigcup_j B_j) \ge \mu(S\backslash N_{l,k})$. Now let $N = \bigcup_{l,k} N_{l,k}$; N is a null set in S and $(\beta + \frac{1}{l})\lambda^n(S) = (\beta + \frac{1}{l})\inf_k \lambda^n(G_k) \ge \mu(S\backslash N)$ for each l. We simply let $l \to \infty$ to conclude that $\mu(S\backslash N) \le \beta\lambda^n(S)$.

If $\lambda^n(S) = \infty$, for each $l \in \mathbb{N}$, put $S_l = \{x \in S : |x| \le l\}$, then $\lambda^n(S_l) < \infty$. By the first part of the proof, for each $l \in \mathbb{N}$ there is a null set $N_l \subset S_l$ such that $\mu(S_l\backslash N_l) \le \beta\lambda^n(S_l)$; then, $N = \bigcup_l N_l$ is a null set and $\mu(S_l\backslash N) \le \mu(S_l\backslash N_l) \le \beta\lambda^n(S_l) \le \beta\lambda^n(S)$, from which $\mu(S\backslash N) \le \beta\lambda^n(S)$ follows by letting $l \to \infty$. ■

Theorem 4.6.3 (Lebesgue) $\frac{d\mu}{d\lambda^n}$ *exists and is finite almost everywhere on* Ω.

Proof Since $\bar{D}\mu < \infty$ almost everywhere on Ω, by Corollary 4.6.1, it is only necessary to show that $\frac{d\mu}{d\lambda^n}$ exists almost everywhere on Ω. If we put $E = \{x \in \Omega : \bar{D}\mu(x) > \underline{D}\mu(x)\}$, this amounts to showing that $\lambda^n(E) = 0$; but since $\underline{D}\mu \ge 0$, $E = \bigcup_{(\alpha,\beta)} E_{(\alpha,\beta)}$, where $E_{(\alpha,\beta)} = \{x \in \Omega : \bar{D}\mu(x) \ge \alpha > \beta \ge \underline{D}\mu(x)\}$, with (α, β) being a generic pair of rational numbers α, β such that $\alpha > \beta \ge 0$, and since all such pairs (α, β) form a countable set, it suffices to show that $\lambda^n(E_{(\alpha,\beta)}) = 0$ for any such pairs of rational numbers. For such a pair (α, β), put $S = E_{(\alpha,\beta)}$. We now show that $\lambda^n(S) = 0$. Suppose the contrary that $\lambda^n(S) > 0$, then there is $l \in \mathbb{N}$ such that if we put $S_l = \{x \in S : |x| < l\}$ then $\infty > \lambda^n(S_l) > 0$. Now $\underline{D}\mu \le \beta$ on S_l; by Lemma 4.6.3 there is a null set N inside S_l such that $\mu(S_l\backslash N) \le \beta\lambda^n(S_l)$;

on the other hand, the fact that $\bar{D}\mu \geq \alpha$ on $S_l \backslash N$ implies, by Lemma 4.6.2, that $\alpha\lambda^n(S_l) = \alpha\lambda^n(S_l\backslash N) \leq \mu(S_l\backslash N)$. Thus,

$$\mu(S_l\backslash N) \leq \beta\lambda^n(S_l) < \alpha\lambda^n(S_l) = \alpha\lambda^n(S_l\backslash N) \leq \mu(S_l\backslash N),$$

the absurdity of which shows that $\lambda^n(S) = 0$. ■

If we let D denote the set of all $x \in \Omega$ such that $\frac{d\mu}{d\lambda^n}(x)$ exists and is finite, D is measurable because D is the complement in Ω of a null set and null sets are measurable. We shall show in a moment that $\frac{d\mu}{d\lambda^n}$ is measurable as a function defined a.e. on Ω (cf. Section 2.5 for measurability of functions defined a.e. on Ω). For $x \in D$, $\frac{d\mu}{d\lambda^n}(x) = \lim_{B\to x} \frac{\mu(B)}{\lambda^n(B)}$, a fortiori, $\frac{d\mu}{d\lambda^n}(x) = \lim_{r\to 0} \frac{\mu(C_r(x))}{\lambda^n(C_r(x))}$, where $C_r(x)$ is the closed ball centered at x and with radius $r > 0$. Now if, as before, $B_r(x)$ denotes the open ball centered at x and with radius $r > 0$, we claim that

$$\frac{d\mu}{d\lambda^n}(x) = \lim_{r\to 0} \frac{\mu(B_r(x))}{\lambda^n(B_r(x))}. \tag{4.9}$$

To see this one needs only to observe that if $r' = r(1-r)$ for $0 < r < 1$, then

$$(1-r)^n \frac{\mu(C_{r'}(x))}{\lambda^n(C_{r'}(x))} = \frac{\lambda^n(C_{r'}(x))}{\lambda^n(C_r(x))}\frac{\mu(C_{r'}(x))}{\lambda^n(C_{r'}(x))} \leq \frac{\mu(B_r(x))}{\lambda^n(B_r(x))} \leq \frac{\mu(C_r(x))}{\lambda^n(C_r(x))},$$

where the relation $\lambda^n(C_{r'}(x)) = (1-r)^n\lambda^n(C_r(x))$ has been used (cf. Example 4.3.1 (ii)), and (4.9) follows as $r \to 0$.

Lemma 4.6.4 $\frac{d\mu}{d\lambda^n}$ *is measurable.*

Proof For $x \in \Omega$ and $r > 0$, let $\Omega_r(x) = B_r(x) \cap \Omega$. First, we show that as a function of x, $\mu(\Omega_r(x))$ is lower semi-continuous on D (r being fixed). For $x \in D$, let I_x denote the indicator function of the set $\Omega_r(x)$, then $\mu(\Omega_r(x)) = \int_\Omega I_x d\mu$. Suppose now that $\{x_k\}$ is a sequence in D tending to x. Since $I_{x_k} \to I_x$ on $\Omega_r(x)$ and $I_x = 0$ on $\Omega\backslash\Omega_r(x)$, $I_x \leq \liminf_{k\to\infty} I_{x_k}$. It follows from the Fatou Lemma that $\mu(\Omega_r(x)) = \int_\Omega I_x d\mu \leq \liminf_{k\to\infty} \int_\Omega I_{x_k} d\mu = \liminf_{k\to\infty} \mu(\Omega_r(x_k))$. Hence, $\mu(\Omega_r(x))$ is lower semi-continuous as a function of x on D and is therefore measurable on D. Similarly, $\lambda^n(\Omega_r(x))$ is lower semi-continuous on D. By choosing a sequence of r tending to zero, we have

$$\frac{d\mu}{d\lambda^n}(x) = \lim_{r\to 0} \frac{\mu(B_r(x))}{\lambda^n(B_r(x))} = \lim_{r\to 0} \frac{\mu(\Omega_r(x))}{\lambda^n(\Omega_r(x))}$$

for $x \in D$, hence $\frac{d\mu}{d\lambda^n}$ is measurable on D (note that $\Omega_r(x) = B_r(x)$ if r is small). Since $\Omega\backslash D$ is a null set, $\frac{d\mu}{d\lambda^n}$ is measurable. ■

$\frac{d\mu}{d\lambda^n}$ is usually extended from D to Ω by defining it to be zero on $\Omega \backslash D$. In view of Exercise 3.9.1(ii), $\frac{d\mu}{d\lambda^n}$ has a Borel measurable version and we shall henceforth take $\frac{d\mu}{d\lambda^n}$ to be a Borel measurable function on Ω.

Lemma 4.6.5 *For any Borel set $S \subset \Omega$, $\int_S \frac{d\mu}{d\lambda^n} d\lambda^n \leq \mu(S)$. In particular $\int_K \frac{d\mu}{d\lambda^n} d\lambda^n < \infty$ for compact sets $K \subset \Omega$.*

Proof Let g be a generic nonnegative and Borel measurable simple function on Ω satisfying $g \leq \frac{d\mu}{d\lambda^n} \cdot I_S$; there are disjoint Borel sets $A_1, \ldots, A_l$ in S and nonnegative numbers $\alpha_1, \ldots, \alpha_l$ such that $g = \sum_{j=1}^{l} \alpha_j I_{A_j}$. Then,

$$\int_\Omega g d\lambda^n = \sum_{j=1}^{l} \alpha_j \lambda^n(A_j).$$

But $\mu(A_j) \geq \alpha_j \lambda^n(A_j)$, $j = 1, \ldots, l$, by Lemma 4.6.2, consequently,

$$\int_\Omega g d\lambda^n \leq \sum_{j=1}^{l} \mu(A_j) = \mu\left(\sum_{j=1}^{l} A_j\right) \leq \mu(S),$$

and hence,

$$\int_S \frac{d\mu}{d\lambda^n} d\lambda^n = \int_\Omega \frac{d\mu}{d\lambda^n} \cdot I_S d\lambda^n = \sup_g \int_\Omega g d\lambda^n \leq \mu(S). \qquad \blacksquare$$

Lemma 4.6.5 implies that $\{\frac{d\mu}{d\lambda^n}\lambda^n\}^*$ is a Radon measure on Ω (cf. Example 3.8.1). Recall that $\{\frac{d\mu}{d\lambda^n}\lambda^n\}$ denotes the indefinite integral of $\frac{d\mu}{d\lambda^n}$ with respect to λ^n. Since indefinite integrals, considered later in this chapter, are always λ^n-indefinite integrals, the notation $\{\frac{d\mu}{d\lambda^n}\lambda^n\}$ is simplified to $\{\frac{d\mu}{d\lambda^n}\}$ for compactness of expression. Similarly, for a nonnegative measurable function f defined on Ω, $\{f\lambda^n\}$ will be replaced by $\{f\}$. With this notational convention, if f is **locally integrable** on Ω in the sense that f is integrable on compact sets in Ω, then $\{f\}^*$ is a Radon measure on Ω. Thus $\{\frac{d\mu}{d\lambda^n}\}^*$ is a Radon measure on Ω.

Another immediate consequence of Lemma 4.6.5 is the following.

Corollary 4.6.2 *For any $S \subset \Omega$, $\{\frac{d\mu}{d\lambda^n}\}^*(S) \leq \mu(S)$.*

Proof If S is a Borel set, then $\{\frac{d\mu}{d\lambda^n}\}^*(S) = \int_S \frac{d\mu}{d\lambda^n} d\lambda^n \leq \mu(S)$, by Lemma 4.6.5; for general S, the same inequality follows from the fact that both $\{\frac{d\mu}{d\lambda^n}\}^*$ and μ are Borel regular. $\blacksquare$

Remark As shown in Example 3.8.1, $\{f\}^*(S) = \int_S f d\lambda^n$ if S is a measurable subset of Ω and f a nonnegative measurable function. Hence, $\{\frac{d\mu}{d\lambda^n}\}^*(S) = \int_S \frac{d\mu}{d\lambda^n} d\lambda^n$ if S is a measurable subset of Ω.

Corollary 4.6.3 *If f is a nonnegative and locally integrable function on Ω, then $\frac{d}{d\lambda^n}\{f\}^* = f$ a.e. on Ω; i.e. for almost all $x \in \Omega$, $\lim_{B\to x} \frac{\int_B f d\lambda^n}{\lambda^n(B)} = f(x)$; in particular,*

$$f(x) = \lim_{r\to 0} \frac{1}{\lambda^n(B_r(x))} \int_{B_r(x)} f d\lambda^n \tag{4.10}$$

for almost every $x \in \Omega$.

Proof Put $g = \frac{d}{d\lambda^n}\{f\}^*$. By Corollary 4.6.2 and the remark following it, $\int_S g d\lambda^n = \{g\}^*(S) \le \{f\}^*(S) = \int_S f d\lambda^n$ for any measurable set $S \subset \Omega$, hence $g \le f$ a.e. on Ω.

Now, put $E = \{g < f\}$; we will show that $\lambda^n(E) = 0$ to conclude that $f = g$ a.e. For this we need only show that $\lambda^n(E') = 0$, where

$$E' = \left\{ x \in E : \lim_{B\to x} \frac{\{f\}^*(B)}{\lambda^n(B)} \text{ exists} \right\}.$$

Suppose the contrary, that $\lambda^n(E') > 0$, then there are numbers $0 < \beta < \alpha < \infty$ and $R > 0$ such that the set $S = \{x \in E' : |x| < R, f(x) > \alpha > \beta > g(x)\}$ has positive Lebesgue measure. Let G be any open set containing S and contained in Ω, and consider the family $\mathcal{V}$ of all $B \subset G$ satisfying $\beta\lambda^n(B) > \{f\}^*(B)$. $\mathcal{V}$ is a Vitali cover of S; by the Vitali covering theorem, there is a disjoint sequence $\{B_j\}$ of balls from $\mathcal{V}$ such that $\lambda^n(S\setminus \bigcup_j B_j) = 0$ (note that $\lambda^n(S) < \infty$). Then,

$$\begin{aligned}\beta\lambda^n(G) &\ge \beta\lambda^n\left(\bigcup_j B_j\right) = \sum_j \beta\lambda^n(B_j) > \sum_j \{f\}^*(B_j) = \{f\}^*\left(\bigcup_j B_j\right)\\ &= \int_{\bigcup_j B_j} f d\lambda^n \ge \int_S f d\lambda^n,\end{aligned}$$

from which it follows that $\beta\lambda^n(S) \ge \int_S f d\lambda^n$; on the other hand $\int_S f d\lambda^n \ge \alpha\lambda^n(S)$, hence,

$$\beta\lambda^n(S) \ge \int_S f d\lambda^n \ge \alpha\lambda^n(S),$$

the absurdity of which shows that $\lambda^n(E') = 0$. That (4.10) holds for almost all $x \in \Omega$ follows from (4.9). ■

Example 4.6.1 (Density and approximate continuity) Let D be a measurable subset of $\mathbb{R}^n$ with $\lambda^n(D) > 0$. For $x \in \mathbb{R}^n$, if $\lim_{B\to x} \frac{\lambda^n(B\cap D)}{\lambda^n(B)}$ exists, the limit is called the density of D at x. Certainly, the density is nonnegative and ≤ 1. If the density of D at x is 1, x is called a density point of D; while x is called a point of dispersion of D if the density of D at x is 0. A measurable function f on D is said to have **approximate limit** l at x if x is a density point of the set $\{y \in D : |f(y) - l| < \varepsilon\}$ for every $\varepsilon > 0$, and the approximate limit l will be denoted by $\text{aplim}_{y\to x} f(y)$. The function f is called approximately continuous at $x \in D$ if $\operatorname{aplim}_{y\to x} f(y) = f(x)$. We claim that (i) almost every point of D is a density point of D, and almost every point of D^c is a point of dispersion of D, and (ii) a measurable function f on D is approximately continuous a.e. on D. Assertion (i) is a direct consequence of Corollary 4.6.3, by choosing f to be the indicator function of D. Observe that (i) implies that if g is a continuous function on $\mathbb{R}^n$, then f is approximately continuous at almost every point of the set $\{x \in D : f(x) = g(x)\}$. It is clear now that (ii) follows from this observation and the Lusin theorem (Theorem 4.1.1).

Exercise 4.6.5 Suppose that A is a measurable subset of $\mathbb{R}^n$. Show that $\operatorname{dist}(y, A) = o(|y - x|)$ as $y \to x$ for almost every x in A.

For a locally integrable function f on Ω, the set $L(f)$ of all those points $x \in \Omega$ such that $\lim_{B\to x} \frac{1}{\lambda^n B} \int_B |f(y) - f(x)|dy = 0$ is called the **Lebesgue set** of f.

Theorem 4.6.4 *If f is locally integrable on Ω, then $\lambda^n(\Omega \backslash L(f)) = 0$, i.e. $L(f)$ consists of almost all points of Ω.*

Proof Denote by γ the set of all rational numbers in $\mathbb{R}$. For $a \in \gamma$, there is a null set E_a in Ω such that for $x \in \Omega \backslash E_a$, the following holds, by Corollary 4.6.3:

$$\lim_{B\to x} \frac{1}{\lambda^n(B)} \int_B |f(y) - a|dy = |f(x) - a|.$$

Put $E = \bigcup_{a\in\gamma} E_a$, then $\lambda^n(E) = 0$. For $x \in \Omega \backslash E$ and $\varepsilon > 0$, there is $a \in \gamma$ such that $|f(x) - a| < \varepsilon$, hence,

$$\begin{aligned}\limsup_{B\to x} \frac{1}{\lambda^n(B)} \int_B |f(y) - f(x)|dy &\le \limsup_{B\to x} \frac{1}{\lambda^n(B)} \int_B \{|f(y) - a| + |f(x) - a|\}dy \\ &= 2|f(x) - a| < 2\varepsilon.\end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, we have $\limsup_{B\to x} \frac{1}{\lambda^n(B)} \int_B |f(y) - f(x)| = 0$, or $\lim_{B\to x} \frac{1}{\lambda^n(B)} \int_B |f(y) - f(x)|dy = 0$. ■

Theorem 4.6.5 *If f is locally integrable on Ω, then*

$$\lim_{B\to x} \frac{1}{\lambda^n(B)} \int_B f d\lambda^n = f(x)$$

for almost every $x \in \Omega$.

Proof For $x \in L(f)$,

$$\left|\frac{1}{\lambda^n(B)} \int_B f(y)dy - f(x)\right| \le \frac{1}{\lambda^n(B)} \int_B |f(y) - f(x)|dy$$

for any closed ball B containing x; then $\lim_{B\to x} \int_B f d\lambda^n = f(x)$ follows. ■

As an application of Theorem 4.6.5, we shall now prove that the space $C_c(\Omega)$ of all those continuous functions on Ω, each of which vanishes outside a compact subset of Ω, is dense in $L^p(\Omega, \mathcal{L}^n|\Omega, \lambda^n)$:

Proposition 4.6.1 *$C_c(\Omega)$ is dense in $L^p(\Omega, \mathcal{L}^n|\Omega, \lambda^n)$, $1 \le p < \infty$.*

Proof Let $f \in L^p(\Omega, \mathcal{L}^n|\Omega, \lambda^n)$ and $\varepsilon > 0$. For each $k \in \mathbb{N}$, put $F_k = \{x \in \Omega : \operatorname{dist}(x, \Omega^c) \ge \frac{1}{k}\} \cap C_k(0)$; $\{F_k\}$ is an increasing sequence of compact sets in Ω and $\Omega = \bigcup_k F_k$. Set $f_k = I_{F_k} f$, then $\lim_{k\to\infty} f_k(x) = f(x)$ for all $x \in \Omega$ and $|f_k| \le |f|$. LDCT implies that $\lim_{k\to\infty} \|f_k - f\|_p = 0$. There is then k_0 such that $\|f_{k_0} - f\|_p < \frac{\varepsilon}{3}$. Now, for each $l \in \mathbb{N}$, let $g_l(x) = f_{k_0}(x)$ if $|f_{k_0}(x)| \le l$, otherwise let $g_l(x) = 0$. By LDCT again, there is $l_0 \in \mathbb{N}$ such that $\|g_{l_0} - f_{k_0}\|_p < \frac{\varepsilon}{3}$. Put $g = g_{l_0}$; g is a bounded function and $g = 0$ outside F_{k_0}. For $0 < r < \frac{1}{2k_0}$, define $[g]_r(x) = \frac{1}{\lambda^n(B_r(x))} \int_{B_r(x)} g(y)dy$, if $x \in F_{2k_0}$; otherwise let $[g]_r(x) = 0$. Obviously, $[g]_r \in C_c(\Omega)$, $|[g]_r| \le l_{k_0}$ on F_{2k_0} and $[g]_r = 0$ outside F_{2k_0}. $[g]_r$ is therefore in $L^p(\Omega, \mathcal{L}^n|\Omega, \lambda^n)$. Since $\lim_{r\to 0}[g]_r = g$ a.e., by Theorem 4.6.5, LDCT implies $\lim_{r\to 0} \|[g]_r - g\|_p = 0$. Choose $0 < r_0 < \frac{1}{2k_0}$ such that $\|[g]_{r_0} - g\|_p < \frac{\varepsilon}{3}$. Then, $g_{r_0} \in C_c(\Omega)$ and $\|f - [g]_{r_0}\|_p \le \|f - f_{k_0}\|_p + \|f_{k_0} - g\|_p + \|g - [g]_{r_0}\|_p < \varepsilon$. ■

Theorem 4.6.6 *Suppose that D is a measurable set in $\mathbb{R}^n$ with positive measure. Then $L^p(D, \mathcal{L}^n|D, \lambda^n)$ is separable for $1 \le p < \infty$.*

Proof In the proof, we shall denote by $L^p(A)$ the space $L^p(A, \mathcal{L}|A, \lambda^n)$ if $A \in \mathcal{L}^n$. Since if $\{u_k\}_{k\in\mathbb{N}}$ is dense in $L^p(\mathbb{R}^n)$, then $\{u_k|_D\}_{k\in\mathbb{N}}$ is dense in $L^p(D)$, it is sufficient to show that $L^p(\mathbb{R}^n)$ is separable.

We call the indicator function of an oriented interval $I_1 \times \cdots \times I_n$ an elementary unit function of order k, if each I_j, $j = 1, \dots, n$, is of the form $I_j = \left[\frac{l_j}{2^k}, \frac{l_j+1}{2^k}\right)$, $l_j \in \mathbb{Z}$. Consider now the family $\mathcal{E}$ of all finite linear combinations of elementary unit functions of all possible order with rational coefficients. It is clear that $\mathcal{E}$ is a countable set in $L^p(\mathbb{R}^n)$. Let $u \in C_c(\mathbb{R}^n)$ and $\varepsilon > 0$. As u vanishes outside $J = J_1 \times \cdots \times J_n$ with each $J_j = [-n_0, n_0]$ for some $n_0 \in \mathbb{N}$, and u is uniformly continuous on J, for any given $\varepsilon > 0$ there is $g \in \mathcal{E}$ such that $\|u - g\|_p < \varepsilon$; hence the closure of $\mathcal{E}$ in $L^p(\mathbb{R}^n)$ contains $C_c(\mathbb{R}^n)$. Thus the closure of $\mathcal{E}$ in $L^p(\mathbb{R}^n)$ is $L^p(\mathbb{R}^n)$, by Proposition 4.6.1. ■

Lemma 4.6.6 *There is a Radon measure φ on Ω such that $\mu = \{\frac{d\mu}{d\lambda^n}\}^* + \varphi$, i.e. $\mu(S) = \{\frac{d\mu}{d\lambda^n}\}^*(S) + \varphi(S)$ for all $S \subset \Omega$.*

Proof Denote by $\mathcal{K}(\Omega)$ the family of all compact sets in Ω. Both μ and $\{\frac{d\mu}{d\lambda^n}\}^*$ take finite value on $\mathcal{K}(\Omega)$; we define φ_1 on $\mathcal{K}(\Omega)$ by

$$\varphi_1(K) = \mu(K) - \left\{\frac{d\mu}{d\lambda^n}\right\}^*(K)$$

for $K \in \mathcal{K}(\Omega)$. By Corollary 4.6.2, $\varphi_1 \ge 0$. Observe that

(i) φ_1 is monotone on $\mathcal{K}(\Omega)$, i.e. for $K_1 \subset K_2$ in $\mathcal{K}(\Omega)$, $\varphi_1(K_1) \le \varphi_1(K_2)$.

(ii) For any finite number of disjoint sets $K_1, \ldots, K_l$ in $\mathcal{K}(\Omega)$,

$$\varphi_1\left(\bigcup_{j=1}^{l} K_j\right) = \sum_{j=1}^{l} \varphi_1(K_j).$$

Now define φ on $\mathcal{B}(\Omega)$ by

$$\varphi(A) = \sup \varphi_1(K)$$

for $A \in \mathcal{B}(\Omega)$, where the supremum is taken over all $K \in \mathcal{K}(\Omega)$ with $K \subset A$. Then φ is an extension of φ_1 and

$$\mu(A) = \left\{\frac{d\mu}{d\lambda^n}\right\}^* (A) + \varphi(A) \tag{4.11}$$

for $A \in \mathcal{B}(\Omega)$. That φ is an extension of φ_1 follows from (i), while (4.11) holds by taking the limit as $j \to \infty$ on both sides of

$$\mu(K_j) = \left\{\frac{d\mu}{d\lambda^n}\right\}^* (K_j) + \varphi(K_j),$$

for a sequence $\{K_j\} \subset \mathcal{K}(\Omega)$ such that $\lim_{j\to\infty} \mu(K_j) = \mu(A)$, $\lim_{j\to\infty} \{\frac{d\mu}{d\lambda^n}\}^*(K_j) = \{\frac{d\mu}{d\lambda^n}\}^*(A)$, and $\lim_{j\to\infty} \varphi(K_j) = \varphi(A)$. That such a sequence $\{K_j\}$ exists follows by applying Theorem 3.9.1 (ii) to μ and $\{\frac{d\mu}{d\lambda^n}\}^*$, and by definition of φ.

If now $\{A_j\}$ is a disjoint sequence of Borel sets in a given compact set $K \subset \Omega$, then both $\mu(\bigcup_j A_j)$ and $\{\frac{d\mu}{d\lambda^n}\}^*(\bigcup_j A_j)$ are finite, and by (4.11),

$$\begin{aligned}
\varphi\left(\bigcup_j A_j\right) &= \mu\left(\bigcup_j A_j\right) - \left\{\frac{d\mu}{d\lambda^n}\right\}^* \left(\bigcup_j A_j\right) \\
&= \sum_j \left\{\mu(A_j) - \left\{\frac{d\mu}{d\lambda^n}\right\}^* (A_j)\right\} = \sum_j \varphi(A_j),
\end{aligned}$$

hence we have:

(iii) For disjoint sequence $\{A_j\} \subset \mathcal{B}(\Omega)$ with $\bigcup_j A_j \subset K$ for some compact set K in Ω,

$$\varphi\left(\bigcup_j A_j\right) = \sum_j \varphi(A_j).$$

Next, we claim that φ is σ-additive on $\mathcal{B}(\Omega)$. Let $\{A_j\}$ be any disjoint sequence in $\mathcal{B}(\Omega)$. For any compact set $K \subset \bigcup_j A_j$,

$$\varphi(K) = \varphi\left(\bigcup_j \{K \cap A_j\}\right) = \sum_j \varphi(K \cap A_j) \le \sum_j \varphi(A_j),$$

by (iii), and the obvious fact that φ is monotone on $\mathcal{B}(\Omega)$. Consequently,

$$\varphi\left(\bigcup_j A_j\right) \le \sum_j \varphi(A_j). \tag{4.12}$$

On the other hand, fix $l \in \mathbb{N}$ and for each $j = 1, \ldots, l$ take an arbitrary compact set $K_j \subset A_j$, then

$$\varphi\left(\bigcup_j A_j\right) \ge \varphi\left(\bigcup_{j=1}^{l} K_j\right) = \sum_{j=1}^{l} \varphi(K_j), \tag{4.13}$$

by monotony of φ on $\mathcal{B}(\Omega)$ and (ii); since each K_j is an arbitrary compact set in A_j, it follows from (4.13) that

$$\varphi\left(\bigcup_j A_j\right) \ge \sum_{j=1}^{l} \varphi(A_j),$$

and hence,

$$\varphi\left(\bigcup_j A_j\right) \ge \sum_j \varphi(A_j).$$

The last inequality shows, together with (4.12), that $\varphi(\bigcup_j A_j) = \sum_j \varphi(A_j)$. Thus φ is σ-additive on $\mathcal{B}(\Omega)$.

Now construct from φ on $\mathcal{B}(\Omega)$ a measure on Ω by Method I, which is the unique $\mathcal{B}(\Omega)$-regular extension of φ by Corollary 3.4.1 and hence is a Radon measure. The Radon measure so constructed is to be denoted also by φ. That $\mu = \{\frac{d\mu}{d\lambda^n}\}^* + \varphi$ holds follows from (4.11) and Borel regularity of μ, $\{\frac{d\mu}{d\lambda^n}\}^*$, and φ. ■

Theorem 4.6.7 (Lebesgue decomposition theorem) *There is a null set $N \subset \Omega$ such that*

$$\mu = \left\{\frac{d\mu}{d\lambda^n}\right\}^* + \mu\lfloor N.$$

Proof By Lemma 4.6.6, there is a Radon measure φ on Ω such that

$$\mu = \left\{\frac{d\mu}{d\lambda^n}\right\}^* + \varphi. \tag{4.14}$$

Choose a null set $N_1 \subset \Omega$ such that, for $x \in \Omega \backslash N_1$, the derivates $\lim_{B \to x} \frac{\mu(B)}{\lambda^n(B)}$, $\lim_{B \to x} \frac{\{\frac{d\mu}{d\lambda^n}\}^*(B)}{\lambda^n(B)}$, and $\lim_{B \to x} \frac{\varphi(B)}{\lambda^n(B)}$ exist and are finite, and further, $\frac{d}{d\lambda^n}\{\frac{d\mu}{d\lambda^n}\}^*(x) = \frac{d\mu}{d\lambda^n}(x)$. That such a null set N_1 exists is a consequence of Theorem 4.6.3 and Corollary 4.6.3. From the choice of N_1 and (4.14), one concludes that the derivate $\frac{d\varphi}{d\lambda^n}(x) = 0$ for $x \in \Omega \backslash N_1$, and hence, in view of Lemma 4.6.3, there is a null set $N_2 \subset \Omega \backslash N_1$ such that $\varphi(\Omega \backslash (N_1 \cup N_2)) = 0$. Put $N = N_1 \cup N_2$; N is a null set, and for any $S \subset \Omega$,

$$\varphi(S \cap N) \leq \varphi(S) \leq \varphi(S \cap (\Omega \backslash N)) + \varphi(S \cap N) = \varphi(S \cap N),$$

or

$$\varphi(S) = \varphi(S \cap N).$$

Now,

$$\mu(S \cap N) = \left\{\frac{d\mu}{d\lambda^n}\right\}^* (S \cap N) + \varphi(S \cap N) = \varphi(S),$$

consequently,

$$\mu(S) = \left\{\frac{d\mu}{d\lambda^n}\right\}^* (S) + \varphi(S) = \left\{\frac{d\mu}{d\lambda^n}\right\}^* (S) + \mu(S \cap N)$$

for any $S \subset \Omega$; in other words,

$$\mu = \left\{\frac{d\mu}{d\lambda^n}\right\}^* + \mu \lfloor N. \qquad \blacksquare$$

The decomposition of μ into the sum $\{\frac{d\mu}{d\lambda^n}\}^* + \mu\lfloor N$ in Theorem 4.6.7 is called the **Lebesgue decomposition** of μ.

Concepts of absolute continuity and singularity for measures are introduced now for the purpose of singling out a distinguishing feature of the Lebesgue decomposition theorem. Suppose μ and ν are measures on a set Ω. The measure μ is said to be ν-**absolutely continuous** if $\mu(A) = 0$ whenever $\nu(A) = 0$; and μ is said to be ν-**singular** if $\mu = \mu\lfloor N$ where $\nu(N) = 0$. If Ω is a subset of $\mathbb{R}^n$, then a measure μ on Ω being λ^n-absolute continuous or λ^n-singular will simply be said to be absolutely continuous or singular, in this order.

In the decomposition $\mu = \{\frac{d\mu}{d\lambda^n}\}^* + \mu\lfloor N$, where $\lambda^n(N) = 0$, as claimed by Theorem 4.6.7, $\{\frac{d\mu}{d\lambda^n}\}^*$ is absolutely continuous and $\mu\lfloor N$ is singular. Thus, Theorem 4.6.7 claims that any Radon measure on Ω can be decomposed into an absolutely continuous part and a singular part. We shall see presently that such a decomposition is unique.

Lemma 4.6.7 *If μ is an absolutely continuous Radon measure on Ω, then $\mu = \{\frac{d\mu}{d\lambda^n}\}^*$.*

Proof By Theorem 4.6.7,

$$\mu = \left\{\frac{d\mu}{d\lambda^n}\right\}^* + \mu\lfloor N,$$

where $\lambda^n(N) = 0$; but absolute continuity of μ implies $\mu(N) = 0$ and hence $\mu\lfloor N = 0$. ∎

Lemma 4.6.8 *If μ is a singular Radon measure on Ω, then $\frac{d\mu}{d\lambda^n} = 0$ a.e. on Ω.*

Proof There are null sets N and N' in Ω such that

$$\mu = \left\{\frac{d\mu}{d\lambda^n}\right\}^* + \mu\lfloor N$$

and

$$\mu = \mu\lfloor N',$$

by Theorem 4.6.7 and singularity of μ. For any set $S \subset \Omega\backslash N'$, we have

$$\mu(S) = \mu(S \cap N') = \mu(\emptyset) = 0,$$

and hence,

$$0 = \mu(S) = \left\{\frac{d\mu}{d\lambda^n}\right\}^* (S) + \mu(N \cap S),$$

a fortiori, $\{\frac{d\mu}{d\lambda^n}\}^*(S) = 0$. Since $\{\frac{d\mu}{d\lambda^n}\}^*(S) = \int_S \frac{d\mu}{d\lambda^n} d\lambda^n = 0$ for any measurable $S \subset \Omega\backslash N'$, $\frac{d\mu}{d\lambda^n} = 0$ a.e. on $\Omega\backslash N'$, and consequently $\frac{d\mu}{d\lambda^n} = 0$ a.e. on Ω. ∎

Theorem 4.6.8 *For a Radon measure μ on Ω, the Lebesgue decomposition $\mu = \{\frac{d\mu}{d\lambda^n}\}^* + \mu\lfloor N$, where $\lambda^n(N) = 0$, is the unique decomposition of μ into a sum of an absolutely continuous and a singular Radon measure.*

Proof Let $\mu = \mu_a + \mu_s$ be a decomposition of μ into the sum of an absolutely continuous Radon measure μ_a and a singular Radon measure μ_s. Then,

$$\frac{d\mu}{d\lambda^n} = \frac{d\mu_a}{d\lambda^n} + \frac{d\mu_s}{d\lambda^n}$$

almost everywhere on Ω. Since $\frac{d\mu_s}{d\lambda^n} = 0$ a.e. on Ω, by Lemma 4.6.8, $\frac{d\mu}{d\lambda^n} = \frac{d\mu_a}{d\lambda^n}$ a.e. From Lemma 4.6.7, $\mu_a = \{\frac{d\mu_a}{d\lambda^n}\}^* = \{\frac{d\mu}{d\lambda^n}\}^*$. Let $\mu = \{\frac{d\mu}{d\lambda^n}\}^* + \mu\lfloor N$ be the Lebesgue decomposition of μ; then by what has just being proved,

$$\mu(S) = \mu_a(S) + \mu_s(S) = \left\{\frac{d\mu}{d\lambda^n}\right\}^* (S) + \mu\lfloor N(S) = \mu_a(S) + \mu\lfloor N(S);$$

in particular, if $\mu(S) < \infty$, $\mu_s(S) = \mu\lfloor N(S)$, from which $\mu_s = \mu\lfloor N$ follows by Theorem 3.3.2, because both μ_s and $\mu\lfloor N$ are regular. ■

Exercise 4.6.6 Let H^n be the n-dimensional Hausdorff measure on $\mathbb{R}^n$.

(i) Show that H^n is a Radon measure on $\mathbb{R}^n$.

(ii) Show that $\frac{dH^n}{d\lambda^n}(x) = \alpha_n$ for all $x \in \mathbb{R}^n$, where α_n is a constant depending only on the dimension n.

(iii) Show that $H^n = \alpha_n \lambda^n$.

The results in this section will be applied in Section 4.7 to study differentiability of functions of a real variable; while differentiability of measures in a general setting will be taken up in Section 5.7, where a decomposition theorem similar to Theorem 4.6.8 is established.

4.7 Differentiability of functions of a real variable and related functions

Differentiability of functions of a real variable will be studied through differentiation of Lebesgue–Stieltjes measures generated by monotone functions. An important subclass of the class of BV functions will be introduced. This is the class of absolutely continuous functions, which is much larger than the class of continuously differentiable functions, but enjoys many useful properties of the latter; in particular, the formula of integration by parts holds for absolutely continuous functions.

We start with the almost everywhere differentiability for monotone functions.

Lemma 4.7.1 *If g is a finite-valued monotone increasing function on $\mathbb{R}$, then the derivative g' exists and is finite almost everywhere on $\mathbb{R}$ and g' is measurable. Furthermore, $g' = \frac{d\mu_g}{d\lambda}$ a.e.*

Proof Let μ_g be the Lebesgue–Stieltjes measure generated by g. We know from Theorem 4.6.3 that the derivate

$$\frac{d\mu_g}{d\lambda}(x) = \lim_{I \to x} \frac{\mu_g(I)}{|I|}$$

exists and is finite for x in a subset D of $\mathbb{R}$ with $\lambda(\mathbb{R} \backslash D) = 0$, where I denotes a generic finite closed interval in $\mathbb{R}$. We claim that for $x \in D$, $g'(x)$ exists and equals $\frac{d\mu_g}{d\lambda}(x)$. Note first that points in D are necessarily points of continuity of g and $\mu_g([a, b]) = g(b) - g(a)$ if g is continuous at a and b (cf. Lemma 3.7.2). Now for $x \in D$, if $y \to x+$ through points of continuity of g, then $\lim_{y\to x+} \frac{g(y)-g(x)}{y \to x} = \frac{d\mu_g}{d\lambda}(x)$; in general, for any

$y > x$, choose points of continuity y' and y'' such that $x < y' < y < y''$ and such that $\lim_{y\to x+}\frac{y'-x}{y-x} = \lim_{y\to x+}\frac{y''-x}{y-x} = 1$, then,

$$\left(\frac{y'-x}{y-x}\right)\frac{g(y')-g(x)}{y'-x} \le \frac{g(y)-g(x)}{y-x} \le \left(\frac{y''-x}{y-x}\right)\frac{g(y'')-g(x)}{y''-x},$$

from which by taking the limit as $y \to x+$, we obtain $\lim_{y\to x+}\frac{g(y)-g(x)}{y-x} = \frac{d\mu_g}{d\lambda}(x)$. Similarly, $\lim_{y\to x-}\frac{g(y)-g(x)}{y-x} = \lim_{y\to x-}\frac{g(x)-g(y)}{x-y} = \frac{d\mu_g}{d\lambda}(x)$. Thus, $g'(x) = \frac{d\mu_g}{d\lambda}(x)$ for $x \in D$. This means that g' exists almost everywhere on $\mathbb{R}$. That g' is measurable follows from Lemma 4.6.4 and the fact that $g' = \frac{d\mu_g}{d\lambda}$ a.e. ■

Theorem 4.7.1 *If f is a BV function on a finite closed interval $[a, b]$, then f' exists a.e. on $[a, b]$ and is integrable. Furthermore,*

$$V_a^x(f) \ge \int_a^x |f'|d\lambda$$

for $x \in [a, b]$.

Proof Put $f_1(x) = f(a) + P_a^x(f)$ and $f_2(x) = N_a^x(f)$ for $x \in [a, b]$; then f_1 and f_2 are monotone increasing on $[a, b]$ and $f = f_1 - f_2$. That f' exists a.e. on $[a, b]$ and is measurable follows from Lemma 4.7.1 by extending f_1 and f_2 to be defined and monotone increasing on $\mathbb{R}$, as in the last paragraph of Section 3.7 and by extending f by $f = f_1 - f_2$ on $\mathbb{R}$. Then $f' = f_1' - f_2'$ a.e. on $\mathbb{R}$.

If for $i = 1, 2$, we let μ_i be the Lebesgue–Stieltjes measure on $\mathbb{R}$ generated by f_i, then from the Lebesgue decomposition theorem,

$$\mu_i = \left\{\frac{d\mu_i}{d\lambda}\right\}^* + \mu_i\lfloor N_i) \ge \left\{\frac{d\mu_i}{d\lambda}\right\}^* = \{f_i'\}^*,$$

where N_i is a null set in $\mathbb{R}$ and $\frac{d\mu_i}{d\lambda} = f_i'$ a.e., by Lemma 4.7.1. As a consequence, for $x \in [a, b]$, $P_a^x(f) = f_1(x) - f_1(a) \ge \mu_1([a, x)) \ge \int_a^x f_1' d\lambda$; similarly, $N_a^x(f) \ge \int_a^x f_2' d\lambda$. Now, $V_a^x(f) = P_a^x(f) + N_a^x(f) \ge \int_a^x (f_1' + f_2')d\lambda \ge \int_a^x |f'|d\lambda$. That f' is integrable follows from $\int_a^b |f'|d\lambda \le V_a^b(f) < \infty$. ■

Remark Although the measurability of g' in Lemma 4.7.1 follows from that of $\frac{d\mu_g}{d\lambda}$ by Lemma 4.6.4, if a measurable function f is differentiable a.e., the measurability of f' follows from the measurability of the limit of a sequence of measurable functions. Actually,

$$f'(x) = \lim_{k\to\infty} k\left\{f\left(x + \frac{1}{k}\right) - f(x)\right\}$$

if $f'(x)$ exists, and for each $k \in \mathbb{N}$, $k\{f(x + \frac{1}{k}) - f(x)\}$ is a measurable function of x.

Exercise 4.7.1 Let f be a monotone increasing function on a finite closed interval $[a, b]$. Show that $f(x) = f(a) + \int_a^x f' d\lambda$ for all $x \in [a, b]$ if and only if the Lebesgue–Stieltjes measure μ_f generated by f is absolutely continuous.

In the remaining part of this section, functions are finite-valued and defined on a finite closed interval $[a, b]$; and for a function f and an interval I in $[a, b]$ with endpoints $c < d$, the difference $f(d) - f(c)$ will be denoted by $f(I)$.

A monotone increasing function f is said to be **absolutely continuous** if the Lebesgue–Stieltjes measure μ_f generated by f is absolutely continuous. Hence, by Exercise 4.7.1, a monotone increasing function f is absolutely continuous if and only if

$$f(x) = f(a) + \int_a^x f' d\lambda$$

holds for all $x \in [a, b]$. We shall characterize absolute continuity of a monotone increasing function by a property which can be adopted to define absolute continuity for general functions.

Lemma 4.7.2 *For a monotone increasing function f, the following two statements are equivalent:*

(I) *f is absolutely continuous.*

(II) *Given any $\varepsilon > 0$, there is $\delta > 0$ such that if $\{I_j\}$ is a disjoint sequence of intervals open in $[a, b]$ with $\sum_j |I_j| < \delta$, then $\sum_j |f(I_j)| < \varepsilon$.*

Proof For convenience, put $\mu = \mu_f$.

To show the implication (I) $\Rightarrow$ (II), note first that since $\mu(\{x\}) = 0$ for all $x \in [a, b]$, $\mu(\{x\}) = f(x+) - f(x-) = 0$, i.e. f is continuous on $[a, b]$. From Lemma 4.6.7, $\int_a^b \frac{d\mu}{d\lambda} d\lambda = \mu([a, b]) < \infty$, hence $\frac{d\mu}{d\lambda}$ is integrable. Now let $\varepsilon > 0$ be given; by Exercise 2.5.9 (iii) there is $\delta > 0$ such that if A is a measurable set in $[a, b]$ with $\lambda(A) < \delta$, then $\int_A \frac{d\mu}{d\lambda} d\lambda < \varepsilon$; if $\{I_j\}$ is a disjoint sequence of intervals open in $[a, b]$ with $\sum_j |I_j| < \delta$, then $\lambda(\bigcup_j I_j) < \delta$ and $\sum_j |f(I_j)| = \sum_j f(I_j) = \int_{\bigcup_j I_j} \frac{d\mu}{d\lambda} d\lambda < \varepsilon$. Thus (II) holds.

Suppose now that (II) holds. We will show that if N is a null set in $[a, b]$, $\mu(N) = 0$. Given $\varepsilon > 0$, choose $\delta > 0$ according to (II). There is a set G open in $[a, b]$ such that $G \supset N$ and $\lambda(G) < \delta$. But, since $G = \bigcup_j I_j$, where $\{I_j\}$ is a disjoint sequence of intervals open in $[a, b]$, $\sum_j |I_j| = \lambda(G) < \delta$, and consequently,

$$\mu(N) \leq \mu(G) = \sum_j \mu(I_j) = \sum_j f(I_j) < \varepsilon, \tag{4.15}$$

by (II), where the obvious fact that if (II) holds, f is continuous on $[a, b]$ and $\mu(I_j) = f(I_j)$, has been used. Since (4.15) holds for arbitrary $\varepsilon > 0$, $\mu(N) = 0$. ■

Exercise 4.7.2 Show that a monotone increasing and absolutely continuous function maps null sets to null sets.

We take Lemma 4.7.2 as a hint for defining absolute continuity for general functions. A function f is said to be **absolutely continuous** if condition (II) in Lemma 4.7.2 holds for f. Condition (II) in Lemma 4.7.2 will be referred to as **condition (AC)**, and an absolutely continuous function is usually simply called an **AC function**. Immediately following, if $\mathcal{P} : x_0 = c < x_1 < \cdots < x_l = d$ is a partition of $[c, d]$, the intervals (x_{j-1}, x_j), $j = 1, \ldots, l$ are called the intervals of $\mathcal{P}$; and if f is a function defined on $[c, d]$, $\sum_{j=1}^{l} |f(x_j) - f(x_{j-1})|$ will be denoted by $|f(\mathcal{P})|$.

Lemma 4.7.3 *An AC function f is a BV function.*

Proof Since f satisfies condition (AC), there is $\delta > 0$ such that if $\{I_j\}$ is a disjoint sequence of intervals open in $[a, b]$ with $\sum_j |I_j| < \delta$, then $\sum_j |f(I_j)| < 1$. Divide $[a, b]$ into m nonoverlapping closed intervals of equal length $< \delta$, and consider one of these subintervals, say J. Let $\mathcal{P}$ be any partition of J, then $|f(\mathcal{P})| < 1$, because the intervals of $\mathcal{P}$ are in J and the sum of their lengths is smaller than δ. Since $\mathcal{P}$ is an arbitrary partition of J, the total variation of f over J is less than or equal to 1. Hence, $V_a^b(f) \leq m$. ■

Recall that if f is a BV function, the functions f_P, f_N, and f_V are defined by

$$f_P(x) = P_a^x(f); \quad f_N(x) = N_a^x(f); \quad f_V(x) = V_a^x(f)$$

for $x \in [a, b]$.

Lemma 4.7.4 *If f is a BV function, then the following three statements are equivalent:*

(I) *f is an AC function.*

(II) *f_V is an AC function.*

(III) *Both f_P and f_N are AC functions.*

Proof The implication of (II) $\Rightarrow$ (I) and the equivalence of (II) $\Leftrightarrow$ (III) are obvious. It remains to show the implication of (I) $\Rightarrow$ (II).

Suppose (I) holds. For $\varepsilon > 0$ given, choose $\delta > 0$ according to condition (AC). We are going to show that if $\{I_j\}$ is a disjoint sequence of intervals open in $[a, b]$ with $\sum_j |I_j| < \delta$, then $\sum_j f_V(I_j) \leq \varepsilon$. For each j, let $\mathcal{P}_j$ be an arbitrary partition of I_j, and let $\{I_k^{(j)}\}_k$ be the finite family of intervals of $\mathcal{P}_j$, then $\bigcup_j \{I_k^{(j)}\}_k$ is a sequence of disjoint intervals open in $[a, b]$ and $\sum_j \sum_k |I_k^{(j)}| = \sum_j |I_j| < \delta$. From the choice of δ, $\sum_j \sum_k |f(I_k^{(j)})| = \sum_j |f(\mathcal{P}_j)| < \varepsilon$; consequently, $\sum_j f_V(I_j) \leq \varepsilon$, by taking the supremum of $\sum_j |f(\mathcal{P}_j)|$ first for all partitions $\mathcal{P}_1$ of I_1, and then for all partitions $\mathcal{P}_2$ of I_2, and so on. Thus, f_V satisfies condition (AC) and is therefore an AC function. ■

A function f is called an **indefinite integral** if there is an integrable function g such that

$$f(x) = c + \int_a^x g d\lambda \tag{4.16}$$

for some constant c and all $x \in [a, b]$. More precisely, if (4.16) holds, f is called an indefinite integral of g.

Exercise 4.7.3 Show that if f is an indefinite integral of g, then $f' = g$ a.e.

Theorem 4.7.2 *A function f is an AC function if and only if it is an indefinite integral.*

Proof It is obvious that an indefinite integral is an AC function. Suppose now that f is an AC function. Both f_P and f_N are AC functions, by Lemma 4.7.4, hence,

$$f_P(x) = \int_a^x f_P' d\lambda; \quad f_N(x) = \int_a^x f_N' d\lambda$$

for all $x \in [a, b]$, by Exercise 4.7.1. Then,

$$f(x) = f(a) + f_P(x) - f_N(x) = f(a) + \int_a^x (f_P' - f_N') d\lambda$$

for all $x \in [a, b]$. This shows that f is an indefinite integral ($f_P' - f_N'$ is integrable because both f_P' and f_N' are integrable). ■

Exercise 4.7.4 Show that a function f is AC if and only if f' exists a.e., f' is integrable, and $f(x) = f(a) + \int_a^x f' d\lambda$ for all $x \in [a, b]$.

Corollary 4.7.1 *If f is an AC function, then $f_N' = 0$ a.e. on $\{f_P' > 0\}$ and $f_P' = 0$ a.e. on $\{f_N' > 0\}$; in other words, $(f')^+ = f_P'$ a.e. and $(f')^- = f_N'$ a.e.*

Proof Since $f' = f_P' - f_N'$ a.e., by Example 4.4.1, $V_a^b(f) = \int_a^b |f_P' - f_N'| d\lambda$; on the other hand, $V_a^b(f) = P_a^b(f) + N_a^b(f) = \int_a^b f_P' d\lambda + \int_a^b f_N' d\lambda$, since both f_P and f_N are AC, by Lemma 4.7.4. Now, $f_P' + f_N' \geq |f_P' - f_N'|$ and $\int_a^b \{f_P' + f_N' - |f_P' - f_N'|\} d\lambda = 0$ imply that $f_P' + f_N' = |f_P' - f_N'|$ a.e. From the last equality, the conclusion of the corollary follows directly. ■

Exercise 4.7.5 Suppose that f is a BV function.

(i) Show that if $V_a^b(f) = \int_a^b |f'| d\lambda$, then

$$f_V(x) = \int_a^x |f'| d\lambda; \quad f_P(x) = \int_a^x (f')^+ d\lambda; \quad f_N(x) = \int_a^x (f')^- d\lambda$$

for all $x \in [a, b]$.

(ii) Show that a BV function f is AC if and only if $V_a^b(f) = \int_a^b |f'| d\lambda$.

Exercise 4.7.6 A monotone increasing function f is said to be singular if $f' = 0$ a.e. Show that every monotone increasing function is a sum of an AC function and a singular function.

Exercise 4.7.7 Let $\{f_n\}$ be a sequence of AC functions on $[a, b]$ such that $\lim_{n\to\infty} f_n(a)$ exists and is finite, and $\{f_n'\}$ converges in $L^1[a, b]$. Show that $\{f_n\}$ converges uniformly on $[a, b]$ to an AC function.

Example 4.7.1 (Cantor's ternary function) Let $I_0 = [0, 1]$ and let $J_0 = (\frac{1}{3}, \frac{2}{3})$ be the middle third open interval of I_0. Then $I_0 \backslash J_0 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$, and call $[0, \frac{1}{3}]$ and $[\frac{2}{3}, 1]$ I_{11}, I_{12} respectively. The open middle thirds of I_{11} and I_{12} are denoted J_{11}, J_{12} respectively. Continue in this fashion; on the kth step we obtain 2^k open intervals $J_{k,1}, \ldots, J_{k,2^k}$ ordered from left to right, each of length $(\frac{1}{3})^{k+1}$. Put $G = \bigcup_{k=0}^{\infty} \bigcup_{j=1}^{2^k} J_{kj}$, then $\lambda(G) = \sum_{k=0}^{\infty} 2^k (\frac{1}{3})^{k+1} = 1$. The set $P := I_0 \backslash G$ is the intersection of a decreasing sequence of nonempty compact sets, and is therefore a nonempty compact set, called **Cantor's ternary set**. P is small in the sense that $\lambda(P) = 0$; but we shall see that P is large in the sense that cardinality of P is the same as that of $I_0 = [0, 1]$. A function f will now be defined on $[0, 1]$ as follows. For $x \in [0, 1]$, express x in ternary expansion

$$x = \sum_{j=1}^{\infty} \frac{\varepsilon_j}{3^j}, \quad \varepsilon_j \in \{0, 1, 2\},$$

and let $\zeta_j = \frac{1}{2}\varepsilon_j$ for all j. The function f is defined by

$$f(x) = \sum_{j=1}^{n-1} \frac{\zeta_j}{2^j} + \frac{1}{2^n}$$

if $\varepsilon_j \in \{0, 2\}$ for $j = 1, \ldots, n-1$, and $\varepsilon_n = 1$ for some n; otherwise, let $f(x) = \sum_{j=1}^{\infty} \frac{\zeta_j}{2^j}$. Function f is well defined, since the only situation where x has two ternary expansions that might lead to different values of $f(x)$ is when the sequence $\{\varepsilon_j\}$ of one of the expansions is of the form: for some n, $\varepsilon_1, \ldots, \varepsilon_{n-1}$ are in $\{0, 2\}$, $\varepsilon_n = 1$, and either $\varepsilon_j = 0$ for $j \geq n+1$ or $\varepsilon_j = 2$ for $j \geq n+1$; in the first case x can also be expressed as $x = \sum_{j=1}^{n-1} \frac{\varepsilon_j}{3^j} + \frac{0}{3^n} + \sum_{j\geq n+1}^{\infty} \frac{2}{3^j}$, and in either expansion

$$f(x) = \sum_{j=1}^{n-1} \frac{\zeta_j}{2^j} + \frac{1}{2^n},$$

while in the second case x can also be expanded as

$$x = \sum_{j=1}^{n-1} \frac{\varepsilon_j}{3^j} + \frac{2}{3^n} + \sum_{j\geq n+1} \frac{0}{3^j},$$

and $f(x)$ also has the value $\sum_{j=1}^{n-1} \frac{\zeta_j}{2^j} + \frac{1}{2^n}$. The function so defined is called **Cantor's ternary function**.

Exercise 4.7.8 Let f be the Cantor's ternary function.

(i) Show that f is a monotone increasing and continuous function with $f(0) = 0$ and $f(1) = 1$.

(ii) Show that each open interval J_{kj}, $k = 0, 1, 2, \ldots; j = 1, \ldots, 2^k$, defined above is of the form $\left(\sum_{j=1}^{n-1} \frac{\varepsilon_j}{3^j} + \frac{1}{3^n}, \sum_{j=1}^{n-1} \frac{\varepsilon_j}{3^j} + \frac{2}{3^n}\right)$ for some n, where $\varepsilon_1, \ldots, \varepsilon_{n-1}$ are in $\{0, 2\}$. Also show that f is constant on each such interval and find the value.

(iii) Show that if x and y in $[0, 1]$ satisfy $|x - y| \leq \frac{1}{3^n}$, then $|f(x) - f(y)| \leq \frac{1}{2^n}$.

(iv) Show that $\int_P f d\mu_f = \frac{1}{2}$.

Exercise 4.7.9 Let P be the Cantor's ternary set defined previously.

(i) Show that $x \in P$ if and only if x has a ternary expansion $x = \sum_{j=1}^{\infty} \frac{\varepsilon_j}{3^j}$, where each $\varepsilon_j \in \{0, 2\}$.

(ii) Show that Cantor's ternary function maps P onto $[0, 1]$.

(iii) A number x in $[0, 1]$ is called a ternary rational number if $x = \frac{m}{3^n}$, where m and n are nonnegative integers with $0 \leq m \leq 3^n$. Let P_0 be the set obtained by removing all those ternary rational numbers in $(0, 1)$ from P. Show that the Cantor's ternary function is 1-1 on P_0.

(iv) Show that the cardinality of P is the same as that of $[0, 1]$.

Example 4.7.1 (Continued) The Cantor's ternary function f is constant on each open interval J_{kj} and hence $f' = 0$ a.e. on $[0, 1]$. Cantor's ternary function is the most well-known singular function. Observe that $V_a^b(f) = 1$, but $\int_a^b |f'| d\lambda = 0$; hence f is not an AC function. The Cantor's ternary set P is perfect i.e. P is the set of all of its own limit points. Thus P is a perfect compact null set with cardinality that of $\mathbb{R}$.

Example 4.7.2 We now use Cantor's ternary function f on $[0, 1]$ to exhibit the fact that a measurable function of a continuous function may not be measurable.

Define a function g on $[0, 1]$ by $g(x) = f(x) + x$, where f is Cantor's ternary function. Evidently, g is strictly increasing on $[0, 1]$ and maps $[0, 1]$ continuously onto $[0, 2]$. The complement G of Cantor's ternary set P in $[0, 1]$ is an open set which is mapped by g onto an open set in $[0, 2]$ of measure 1 (note that each interval component of G is mapped by f to a point, and is hence mapped by g onto an interval of the same length); as a result, g maps the Cantor's ternary set P onto a compact set K of measure 1. By Proposition 3.11.2, K contains a nonmeasurable set W. Since $g^{-1}W \subset P$ and $\lambda(P) = 0$, $g^{-1}W$ is a null set and is therefore measurable. Put $A = g^{-1}W$ and let $h = I_A$; h is measurable. Because g is a continuous and injective map from the compact set $[0, 1]$ onto $[0, 2]$, g^{-1} is a continuous function from $[0, 2]$ onto $[0, 1]$, by Proposition 1.7.3. Now $h \circ g^{-1}$ is not measurable, because $\{h \circ g^{-1} > 0\} = W$ is nonmeasurable. Thus, a measurable function of a continuous function could be nonmeasurable.

For a right-continuous BV function g on $[a, b]$, let μ_g^+ and μ_g^- be the Lebesgue–Stieltjes measures generated by g_P and g_N respectively, and let $\mu_g = \mu_g^+ - \mu_g^-$. Note that both g_P and g_N are right-continuous, by Theorem 4.4.2. If f is both μ_g^+- and μ_g^--measurable on $[a, b]$ and is integrable w.r.t. μ_g^+ and μ_g^-, we define

$$\int_a^b f d\mu_g = \int_a^b f d\mu_g^+ - \int_a^b f d\mu_g^-.$$

The measure $|\mu_g| := \mu_g^+ + \mu_g^-$ is called the **total variational measure** generated by g, while μ_g^+ and μ_g^- are called respectively the **positive variational measure** and the **negative variational measure** generated by g. If f is a bounded function on $[a, b]$ which is continuous $|\mu_g|$-a.e., then

$$\int_a^b f dg := \int_a^b f dg_P - \int_a^b f dg_N$$

exists and is finite, by Theorem 4.5.2.

Exercise 4.7.10 Suppose that g is an AC function. Show that a Riemann integrable function f is continuous $|\mu_g|$-a.e. Then conclude that $\int_a^b f dg$ is defined and $\int_a^b f dg = \int_a^b f g' d\lambda$. (Hint: cf. Example 4.5.2.)

Theorem 4.7.3 (Integration by parts) *Let f, g be AC functions on $[a, b]$, then*

$$\int_a^b f g' d\lambda = f(b)g(b) - f(a)g(a) - \int_a^b g f' d\lambda.$$

Proof We may assume that both f and g are monotone increasing, then by Theorem 4.5.3,

$$\int_a^b f dg = f(b)g(b) - f(a)g(a) - \int_a^b g df.$$

But by Example 4.5.2,

$$\int_a^b f dg = \int_a^b f g' d\lambda; \quad \int_a^b g df = \int_a^b g f' d\lambda,$$

hence,

$$\int_a^b f g' d\lambda = f(b)g(b) - f(a)g(a) - \int_a^b g f' d\lambda. \qquad \blacksquare$$

Exercise 4.7.11 Let f and g be AC functions. Show that the product fg is AC, and (using integration by parts)

$$\int_c^d (fg)' d\lambda = \int_c^d f' g d\lambda + \int_c^d fg' d\lambda$$

for all $a \leq c < d \leq b$, and conclude that $(fg)' = f'g + fg'$ a.e.

Exercise 4.7.12 Let f be an integrable function on $[a, b]$ with the property that

$$\int_a^b fg' d\lambda = 0$$

for all AC functions g such that $g(a) = g(b) = 0$. Show that f = constant a.e. (Hint: put $c = \int_a^b f d\lambda$, and let

$$g(x) = \int_a^x \left(f - \frac{c}{b-a}\right) d\lambda$$

for $x \in [a, b]$. Observe that $g(a) = g(b) = 0$ and evaluate $\int_a^b (f - \frac{c}{b-a})^2 d\lambda$.)

Exercise 4.7.13 Let f and g be integrable functions on $[a, b]$ and suppose that

$$\int_a^b fh' d\lambda = -\int_a^b gh d\lambda$$

for all AC functions h with $h(a) = h(b) = 0$. Show that f is equivalent to an AC function $\hat{f}$ and $\hat{f}' = g$ a.e.

Theorem 4.7.4 (Change of variable) *Suppose that g is a monotone increasing AC function on $[a, b]$. Put $c = g(a)$ and $d = g(b)$. Then for any nonnegative measurable function f on $[c, d]$, the function $(f \circ g)g'$ is measurable and*

$$\int_c^d f d\lambda = \int_a^b (f \circ g)g' d\lambda.$$

Proof From $|I| = \mu_g(g^{-1}I)$, for any interval I open in $[c, d]$, it follows that $\lambda(G) = \mu_g(g^{-1}G)$ for any set G open in $[c, d]$, and hence for any Borel set B in $[c, d]$ we have (cf. Exercise 4.3.4 and recall that μ_g is absolutely continuous)

$$\lambda(B) = \mu_g(g^{-1}B) = \int_{g^{-1}B} \frac{d\mu_g}{d\lambda} d\lambda = \int_a^b I_{g^{-1}B} g' d\lambda = \int_a^b (I_B \circ g)g' d\lambda,$$

or

$$\lambda(B) = \int_H (I_B \circ g)g' d\lambda,$$

where $H = \{g' > 0\}$. Note that for a Borel set B in $[c, d]$, $I_B \circ g$ is a Borel measurable function and $(I_B \circ g)g'$ is measurable; but in general $I_A \circ g$ may not be measurable for measurable set $A \subset [c, d]$; however, we claim that $(I_A \circ g)g'$ is measurable and

$$\lambda(A) = \int_a^b (I_A \circ g)g' d\lambda.$$

To see this, first consider the case where A is a null set in $[c, d]$. Choose a Borel set B in $[c, d]$ such that $B \supset A$ and $\lambda(B) = \lambda(A) = 0$, then

$$\lambda(B) = \int_H (I_B \circ g)g' d\lambda = 0,$$

which implies that $I_B \circ g = 0$ a.e. on H and, a fortiori, $I_A \circ g = 0$ a.e. on H. Therefore $I_A \circ g$ is measurable on H; as a consequence, $(I_A \circ g)g' = 0$ a.e. on $[a, b]$ and is therefore measurable. Now, let A be any measurable set in $[c, d]$ and choose a Borel set B in $[c, d]$ such that $B \supset A$ and $\lambda(B) = \lambda(A)$; then $S := B \backslash A$ is a null set and $(I_S \circ g)g' = 0$ a.e. as we have just proved. But $(I_B \circ g)g' = (I_A \circ g + I_S \circ g)g' = (I_A \circ g)g'$ a.e. on $[a, b]$, hence $(I_A \circ g)g'$ is measurable and

$$\lambda(A) = \lambda(B) = \int_a^b (I_B \circ g)g' d\lambda = \int_a^b (I_A \circ g)g' d\lambda.$$

If $f \geq 0$ is measurable, $f = \sum_{j=1}^{\infty} \frac{1}{j} I_{A_j}$, where each A_j is a measurable set in $[c, d]$, by Theorem 2.2.1. Then,

$$\begin{aligned}
\int_c^d f d\lambda &= \int_c^d \sum_{j=1}^{\infty} \frac{1}{j} I_{A_j} d\lambda = \sum_{j=1}^{\infty} \frac{1}{j} \lambda(A_j) = \sum_{j=1}^{\infty} \frac{1}{j} \int_a^b (I_{A_j} \circ g)g' d\lambda \\
&= \int_a^b \lim_{l \to \infty} \sum_{j=1}^{l} \frac{1}{j} (I_{A_j} \circ g)g' d\lambda = \int_a^b \lim_{l \to \infty} \left\{ \left(\sum_{j=1}^{l} \frac{1}{j} I_{A_j} \right) \circ g \right\} g' d\lambda \\
&= \int_a^b (f \circ g)g' d\lambda\,,
\end{aligned}$$

where $(f \circ g)g' = \lim_{l \to \infty} \sum_{j=1}^{l} \frac{1}{j} (I_{A_j} \circ g)g'$ is measurable because it is the limit of measurable functions $\sum_{j=1}^{l} \frac{1}{j} (I_{A_j} \circ g)g'$. ■

Remark The change of variable formula in Theorem 4.7.4 is familiar in integral calculus. Here, it is shown under much relaxed conditions on f and g. Note that one of the delicacies in the proof is the measurability of $(f \circ g)g'$, although $f \circ g$ may not be measurable, as we see in Example 4.7.2.

4.8 Product measures and Fubini theorem

We digress in this section from the main theme of the chapter, to the construction and properties of product measures, before going to further studies of functions of several real variables. Consider measure spaces $(\Omega_i, \Sigma_i, \mu_i)$, $i = 1, 2$, and let $R = \{A_1 \times A_2 : A_i \in \Sigma_i, i = 1, 2\}$. R is a π-system. Sets in R are called **measurable rectangles**. The σ-algebra $\sigma(R)$ on $\Omega_1 \times \Omega_2$ generated by R is denoted by $\Sigma_1 \otimes \Sigma_2$. For $E \subset \Omega_1 \times \Omega_2$ and $(w_1, w_2) \in \Omega_1 \times \Omega_2$, we define sets E_{w_1} and E^{w_2} by

$$E_{w_1} = \{y \in \Omega_2 : (w_1, y) \in E\}; \quad E^{w_2} = \{x \in \Omega_1 : (x, w_2) \in E\}.$$

E_{w_1} and E^{w_2} are called respectively the w_1-**section** and w_2-**section** of E.

The lemma that follows is easily verified.

Lemma 4.8.1 *Let Σ be the family of all $E \subset \Omega_1 \times \Omega_2$ such that $E_{w_1} \in \Sigma_2$, $E^{w_2} \in \Sigma_1$ for all $(w_1, w_2) \in \Omega_1 \times \Omega_2$, then Σ is a σ-algebra containing R.*

Corollary 4.8.1 $\Sigma \supset \Sigma_1 \otimes \Sigma_2$.

Corollary 4.8.2 *If f is $\Sigma_1 \otimes \Sigma_2$-measurable, then for $(w_1, w_2) \in \Omega_1 \times \Omega_2$, $x \mapsto f(x, w_2)$ and $y \mapsto f(w_1, y)$ are respectively Σ_1- and Σ_2-measurable.*

Proof Since $I_E(x, w_2) = I_{E^{w_2}}(x)$ and $I_E(w_1, y) = I_{E_{w_1}}(y)$ for $E \subset \Omega_1 \times \Omega_2$, it follows from Lemma 4.8.1 and Corollary 4.8.1 that the corollary holds if f is the indicator function of a set in $\Sigma_1 \otimes \Sigma_2$. Then the corollary holds for $\Sigma_1 \otimes \Sigma_2$-measurable simple functions. For general nonnegative $\Sigma_1 \otimes \Sigma_2$-measurable functions, the corollary follows by Theorem 2.2.1; this is sufficient to conclude that the corollary holds. ■

Lemma 4.8.2 *Suppose that both $(\Omega_1, \Sigma_1, \mu_1)$ and $(\Omega_2, \Sigma_2, \mu_2)$ are σ-finite and $E \in \Sigma_1 \otimes \Sigma_2$, then $w_1 \mapsto \mu_2(E_{w_1})$ is Σ_1-measurable and $w_2 \mapsto \mu_1(E^{w_2})$ is Σ_2-measurable, and*

$$\int_{\Omega_1} \mu_2(E_{w_1}) d\mu_1(w_1) = \int_{\Omega_2} \mu_1(E^{w_2}) d\mu_2(w_2).$$

Proof Ω_1 and Ω_2 can be expressed as

$$\Omega_1 = \bigcup_{n=1}^{\infty} \Omega_n^{(1)}, \quad \Omega_2 = \bigcup_{n=1}^{\infty} \Omega_n^{(2)},$$

where $\{\Omega_n^{(1)}\} \subset \Sigma_1$, $\{\Omega_n^{(2)}\} \subset \Sigma_2$ are both disjoint and $\mu_i(\Omega_n^{(i)}) < \infty$ for $i = 1, 2$ and $n = 1, 2, \ldots$. Consider the family $\mathcal{M}$ of all those $E \in \Sigma_1 \otimes \Sigma_2$, such that the conclusions of the lemma hold if E is replaced by $E \cap (\Omega_n^{(1)} \times \Omega_m^{(2)})$ for all n and m. It is simply routine to verify that $\mathcal{M}$ is a λ-system. But it is to be noted that the only place where $E \cap (\Omega_n^{(1)} \times \Omega_m^{(2)})$ requires considering is when one verifies that if E is in $\mathcal{M}$ then E^c is in $\mathcal{M}$. Since $E \in \mathcal{M}$ is easily seen to satisfy the conclusions of the lemma, and since $\mathcal{M} \supset R$, the lemma follows from the $(\pi$-$\lambda)$ theorem. ■

Now, for $E \in \Sigma_1 \otimes \Sigma_2$, define

$$\mu_1 \times \mu_2(E) = \int_{\Omega_1} \mu_2(E_{w_1})d\mu_1(w_1) = \int_{\Omega_2} \mu_1(E^{w_2})d\mu_2(w_2).$$

Then $\mu_1 \times \mu_2$ is a measure on $\Sigma_1 \otimes \Sigma_2$ and $(\Omega_1 \times \Omega_2, \Sigma_1 \otimes \Sigma_2, \mu_1 \times \mu_2)$ is a measure space, called the **product space** of $(\Omega_1, \Sigma_1, \mu_1)$ and $(\Omega_2, \Sigma_2, \mu_2)$. The measure $\mu_1 \times \mu_2$ is called the **product measure** of μ_1 and μ_2. One notes that $\mu_1 \times \mu_2(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$ if $A_1 \times A_2 \in R$.

Proposition 4.8.1 *Suppose that both $(\Omega_1, \Sigma_1, \mu_1)$ and $(\Omega_2, \Sigma_2, \mu_2)$ are σ-finite, then $\mu_1 \times \mu_2$ is the unique measure on $\Sigma_1 \otimes \Sigma_2$ such that $\mu_1 \times \mu_2(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$ for all $A_1 \in \Sigma_1$ and $A_2 \in \Sigma_2$.*

Proof Let disjoint sequences $\{\Omega_n^{(1)}\} \subset \Sigma_1$ and $\{\Omega_m^{(2)}\} \subset \Sigma_2$ be as in the proof of Lemma 4.8.2, and suppose that μ is a measure on $\Sigma_1 \otimes \Sigma_2$ such that $\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$ for all $A_1 \in \Sigma_1$ and $A_2 \in \Sigma_2$. Consider the family $\mathcal{F}$ of all $E \in \Sigma_1 \otimes \Sigma_2$, such that

$$\mu(E \cap [\Omega_n^{(1)} \times \Omega_m^{(2)}]) = \mu_1 \times \mu_2(E \cap [\Omega_n^{(1)} \times \Omega_m^{(2)}])$$

for all n and m. Then $\mathcal{F}$ is a λ-system containing all measurable rectangles. Since the family R of all measurable rectangles is a π-system, it follows from the $(\pi$-$\lambda)$ theorem that $\mathcal{F} = \Sigma_1 \otimes \Sigma_2$ and thus $\mu = \mu_1 \times \mu_2$. ∎

Theorem 4.8.1 (Simple version of Fubini theorem)

(i) (Tonelli) *If f is $\Sigma_1 \otimes \Sigma_2$-measurable and $f \geq 0$, then $x \mapsto \int_{\Omega_2} f(x, w_2)d\mu_2(w_2)$ is Σ_1-measurable, $y \mapsto \int_{\Omega_1} f(w_1, y)d\mu_1(w_1)$ is Σ_2-measurable, and*

$$\begin{aligned}\int_{\Omega_1 \times \Omega_2} fd\mu_1 \times \mu_2 &= \int_{\Omega_1} \left[\int_{\Omega_2} f(w_1, w_2)d\mu_2(w_2)\right] d\mu_1(w_1)\\ &= \int_{\Omega_2} \left[\int_{\Omega_1} f(w_1, w_2)d\mu_1(w_1)\right] d\mu_2(w_2).\end{aligned}$$

(ii) *If f is $\mu_1 \times \mu_2$-integrable, then conclusions in* (i) *also hold for f.*

Proof Since (ii) is an obvious consequence of (i), it is sufficient to prove (i). If $E \in \Sigma_1 \otimes \Sigma_2$ and $f = I_E$, then (i) follows from Lemma 4.8.2 and hence the lemma holds for nonnegative simple functions. If f is a nonnegative $\Sigma_1 \otimes \Sigma_2$-measurable function, by Theorem 2.2.1,

$$f = \sum_{k=1}^{\infty} \frac{1}{k} I_{A_k} = \lim_{l \to \infty} \sum_{k=1}^{l} \frac{1}{k} I_{A_k},$$

where each $A_k \in \Sigma_1 \otimes \Sigma_2$, then (i) follows from the monotone convergence theorem. ∎

In general, it is not true that the product space of two σ-finite complete measure spaces is complete. For example, consider $(\mathbb{R}^2, \mathcal{L} \otimes \mathcal{L}, \lambda \times \lambda)$, where $\mathcal{L}$ is the σ-algebra of all Lebesgue measurable sets in $\mathbb{R}$ and λ the Lebesgue measure on $\mathbb{R}$. As we have shown in Section 3.11 there is a nonmeasurable set $S \subset \mathbb{R}$. Choose any nonempty null set N in $\mathbb{R}$, and consider the set $N \times S$ in $\mathbb{R}^2$. For $w \in N$, $(N \times S)_w = S$ is not in $\mathcal{L}$; hence $N \times S$ is not in $\mathcal{L} \otimes \mathcal{L}$. But $N \times S \subset N \times \mathbb{R}$ and $\lambda \times \lambda(N \times \mathbb{R}) = \lambda(N)\lambda(\mathbb{R}) = 0$, thus $N \times S$ is a $\lambda \times \lambda$-null set which is not in $\mathcal{L} \otimes \mathcal{L}$. $(\mathbb{R}^2, \mathcal{L} \otimes \mathcal{L}, \lambda \times \lambda)$ is therefore not complete and cannot be $(\mathbb{R}^2, \mathcal{L}^2, \lambda^2)$.

Exercise 4.8.1 Show that $(\mathbb{R}^{k+l}, \mathcal{L}^{k+l}, \lambda^{k+l})$ is the completion of the measure space $(\mathbb{R}^{k+l}, \mathcal{L}^k \otimes \mathcal{L}^l, \lambda^k \times \lambda^l)$ for k, l in $\mathbb{N}$. (Hint: verify first that $\mathcal{B}(\mathbb{R}^{k+l}) \subset \mathcal{L}^k \otimes \mathcal{L}^l$ and $\lambda^{k+l}(B) = \lambda^k \times \lambda^l(B)$ for $B \in \mathcal{B}(\mathbb{R}^{k+l})$.)

Suppose now that both $(\Omega_1, \Sigma_1, \mu_1)$ and $(\Omega_2, \Sigma_2, \mu_2)$ are σ-finite complete measure spaces; then corresponding to Theorem 4.8.1, the following theorem holds.

Theorem 4.8.2 (Fubini) *Let $(\Omega_1 \times \Omega_2, \overline{\Sigma_1 \otimes \Sigma_2}, \overline{\mu_1 \times \mu_2})$ be the completion of $(\Omega_1 \times \Omega_2, \Sigma_1 \otimes \Sigma_2, \mu_1 \times \mu_2)$.*

(i) (Tonelli) *If f is nonnegative $\overline{\Sigma_1 \otimes \Sigma_2}$-measurable, then for μ_1-a.e. w_1 in Ω_1 and μ_2-a.e. w_2 in Ω_2,*

$$v \mapsto f(w_1, v) \text{ is } \Sigma_2\text{-measurable;}$$
$$u \mapsto f(u, w_2) \text{ is } \Sigma_1\text{-measurable.}$$

Furthermore,

$$w_1 \mapsto \int_{\Omega_2} f(w_1, w_2)d\mu_2(w_2) \text{ is } \Sigma_1\text{-measurable;}$$
$$w_2 \mapsto \int_{\Omega_1} f(w_1, w_2)d\mu_1(w_1) \text{ is } \Sigma_2\text{-measurable,}$$

and

$$\int_{\Omega_1 \times \Omega_2} f d\overline{\mu_1 \times \mu_2} = \int_{\Omega_1} \left[\int_{\Omega_2} f(w_1, w_2)d\mu_2(w_2)\right] d\mu_1(w_1)$$
$$= \int_{\Omega_2} \left[\int_{\Omega_1} f(w_1, w_2)d\mu_1(w_1)\right] d\mu_2(w_2).$$

(ii) *If f is $\overline{\mu_1 \times \mu_2}$-integrable, then the same statements in (i) hold for f.*

Lemma 4.8.3 *Suppose that $E \in \Sigma_1 \otimes \Sigma_2$ and $\mu_1 \times \mu_2(E) = 0$. Then for any subset D of E, the following statements hold:*

(1) *$D_{w_1} \in \Sigma_2$ and $\mu_2(D_{w_1}) = 0$ for μ_1-a.e. w_1 in Ω_1.*

(2) *$D^{w_2} \in \Sigma_1$ and $\mu_1(D^{w_2}) = 0$ for μ_2-a.e. w_2 in Ω_2.*

Proof Since $\mu_1 \times \mu_2(E) = \int_{\Omega_1} \mu_2(E_{w_1}) d\mu_1(w_1) = \int_{\Omega_2} \mu_1(E^{w_2}) d\mu_2(w_2) = 0$, and both $\mu_2(E_{w_1})$ and $\mu_1(E^{w_2})$ are nonnegative, $\mu_2(E_{w_1}) = 0$ for μ_1-a.e. w_1 and $\mu_1(E^{w_2}) = 0$ for μ_2-a.e. w_2. For such w_1 and w_2, D_{w_1} and D^{w_2} are in Σ_2 and Σ_1 respectively, because $D_{w_1} \subset E_{w_1}$, $D^{w_2} \subset E^{w_2}$, and both $(\Omega_1, \Sigma_1, \mu_1)$ and $(\Omega_2, \Sigma_2, \mu_2)$ are complete. Trivially, for such w_1 and w_2, $\mu_2(D_{w_1}) = \mu_1(D^{w_2}) = 0$. ■

Proof of Theorem 4.8.2 Since (ii) follows from (i) easily, it suffices to prove (i).

If $f \geq 0$ is $\overline{\Sigma_1 \otimes \Sigma_2}$-measurable, $f = \sum_{j=1}^{\infty} \frac{1}{j} I_{A_j}$, where each A_j is in $\overline{\Sigma_1 \otimes \Sigma_2}$, as claimed by Theorem 2.2.1. It is therefore sufficient to consider the case $f = I_A$ for $A \in \overline{\Sigma_1 \otimes \Sigma_2}$. There are B and C in $\Sigma_1 \otimes \Sigma_2$ such that $B \subset A \subset C$ with $\mu_1 \times \mu_2(C \backslash B) = 0$. This means that $A = B \cup D$ where $D \subset E := C \backslash B$. From Lemma 4.8.3, for μ_1-a.e. w_1 and μ_2-a.e. w_2, $D_{w_1} \in \Sigma_2$ with $\mu_2(D_{w_1}) = 0$ and $D^{w_2} \in \Sigma_1$ with $\mu_1(D^{w_2}) = 0$; for such w_1 and w_2,

$$v \mapsto I_A(w_1, v) = I_{A_{w_1}}(v) = I_{B_{w_1} \cup D_{w_1}}(v) = I_{B_{w_1}}(v) + I_{D_{w_1}}(v)$$

and

$$u \mapsto I_A(u, w_2) = I_{A^{w_2}}(u) = I_{B^{w_2} \cup D^{w_2}}(u) = I_{B^{w_2}}(u) + I_{D^{w_2}}(u)$$

are respectively Σ_2- and Σ_1-measurable. Furthermore,

$$w_1 \mapsto \int_{\Omega_2} I_A(w_1, v) d\mu_2(v) = \mu_2(B_{w_1}),$$

and

$$w_2 \mapsto \int_{\Omega_1} I_A(u, w_2) d\mu_1(u) = \mu_1(B^{w_2})$$

are respectively Σ_1- and Σ_2-measurable by Lemma 4.8.2, and hence,

$$\int_{\Omega_1} \left[\int_{\Omega_2} I_A(w_1, w_2) d\mu_2(w_2) \right] d\mu_1(w_1) = \int_{\Omega_1} \mu_2(B_{w_1}) d\mu_1(w_1);$$

$$\int_{\Omega_2} \left[\int_{\Omega_1} I_A(w_1, w_2) d\mu_2(w_2) \right] d\mu_2(w_2) = \int_{\Omega_2} \mu_1(B^{w_2}) d\mu_2(w_2).$$

Thus (i) holds for $f = I_A$, because by Lemma 4.8.2,

$$\int_{\Omega_1} \mu_2(B_{w_1}) d\mu_1(w_1) = \int_{\Omega_2} \mu_1(B^{w_2}) d\mu_2(w_2) = \mu_1 \times \mu_2(B),$$

and $\mu_1 \times \mu_2(B) = \overline{\mu_1 \times \mu_2}(A) = \int_{\Omega_1 \times \Omega_2} I_A d\overline{\mu_1 \times \mu_2}$. ■

Example 4.8.1 We use the Fubini theorem to evaluate $\int_{-\infty}^{\infty} e^{-x^2} dx$ (cf. Exercise 3.4.7 and Exercise 3.4.8). First note that since $\int_{-\infty}^{\infty} e^{-x^2} dx < \infty$ as an improper integral, $\int_{\mathbb{R}} e^{-x^2} d\lambda(x) = \int_{-\infty}^{\infty} e^{-x^2} dx$ by Exercise 3.4.7 (i). From the Fubini theorem,

$$\int_{\mathbb{R}^2} e^{-(x^2+y^2)} d\lambda^2(x,y) = \int_{\mathbb{R}} \left[\int_{\mathbb{R}} e^{-x^2} d\lambda(x) \right] e^{-y^2} d\lambda(y) = \left[\int_{\mathbb{R}} e^{-x^2} dx \right]^2.$$

Now,

$$\begin{aligned}\int_{\mathbb{R}^2} e^{-(x^2+y^2)} d\lambda^2(x,y) &= \lim_{L\to\infty} \iint_{x^2+y^2\le L^2} e^{-(x^2+y^2)} dxdy \\ &= \lim_{L\to\infty} \int_0^L \rho \int_0^{2\pi} e^{-\rho^2} d\theta d\rho = \lim_{\rho\to\infty} 2\pi \int_0^L \rho e^{-\rho^2} d\rho \\ &= \lim_{L\to\infty} \pi \int_0^L \frac{d}{d\rho}(-e^{-\rho^2}) d\rho = \pi.\end{aligned}$$

Hence $\int_{-\infty}^{\infty} e^{-x^2} dx = \left[\iint_{\mathbb{R}^2} e^{-(x^2+y^2)} dxdy \right]^{\frac{1}{2}} = \sqrt{\pi}$. By the Fubini theorem, again one finds that $\int_{\mathbb{R}^n} e^{-|x|^2} dx = \pi^{\frac{n}{2}}$.

Exercise 4.8.2

(i) Show that $\int_0^{\infty} \frac{|\sin x|}{x} dx = \infty$.

(ii) Show that $\int_0^{\infty} \frac{\sin x}{x} dx = \lim_{b\to\infty} \int_0^b \frac{\sin x}{x} dx = \frac{\pi}{2}$ by integrating $e^{-xy}\sin x$ over a suitable domain in the first quadrant of $\mathbb{R}^2$.

Exercise 4.8.3 Let (Ω, Σ, μ) be a σ-finite measure space and f a nonnegative Σ-measurable function on Ω. Put $G_f = \{(w,y) \in \Omega \times [0,\infty) : 0 < y < f(w)\}$. Show that $G_f \in \Sigma \otimes \mathcal{B}$ and $\mu \times \lambda(G_f) = \int_{\Omega} f d\mu$.

Exercise 4.8.4 Let $f(x,y) = \frac{xy}{(x^2+y^2)^2}$ if $(x,y) \neq (0,0)$, and $f(0,0) = 0$. Verify that $\int_{-1}^{1} \left(\int_{-1}^{1} f(x,y)dx \right) dy = \int_{-1}^{1} \left(\int_{-1}^{1} f(x,y)dy \right) dx = 0$, and decide whether f is Lebesgue integrable on $[-1,1] \times [-1,1]$ or not.

Exercise 4.8.5 Show that $\int_0^{\infty} (\sum_{j=1}^{\infty} e^{-jx} \sin x) dx = \sum_{j=1}^{\infty} \int_0^{\infty} e^{-jx} \sin x dx$ and use this fact to show that $\int_0^{\infty} \frac{\sin x}{e^x - 1} dx = \sum_{j=1}^{\infty} \frac{1}{1+j^2}$.

Exercise 4.8.6

(i) Show that $\int_0^{\infty} \frac{\tan^{-1} t}{t} dt = \infty$ by considering the double integral

$$\int_0^1 \left(\int_0^{\infty} \frac{1}{1+x^2t^2} dt \right) dx.$$

(ii) Show that $\int_0^\infty (\frac{\tan^{-1} t}{t})^2 dt = \pi \ln 2$ by integrating the triple integral

$$\int_0^1 \left(\int_0^1 \left(\int_0^\infty \frac{1}{1+x^2t^2} \cdot \frac{1}{1+y^2t^2} dt \right) dx \right) dy.$$

Example 4.8.2 Let $\{f_n\}$ be a sequence of $L(\Omega_1 \times \Omega_2, \Sigma_1 \otimes \Sigma_2, \mu_1 \times \mu_2)$, in which it converges to f. We claim that there is a subsequence $\{f_{n_k}\}$ of $\{f_n\}$ such that for μ_1-a.e. x $\lim_{k\to\infty} \int_{\Omega_2} |f_{n_k}(x,y) - f(x,y)| d\mu_2(y) = 0$. Define F, F_n on Ω_1 by $F(x) = \int_{\Omega_2} f(x,y) d\mu_2(y)$ and $F_n(x) = \int_{\Omega_2} f_n(x,y) d\mu_2(y)$. Note that F, F_n's are measurable on $(\Omega_1, \Sigma_1, \mu_1)$ and $\int_{\Omega_1} |F_n - F| d\mu_1 = \int_{\Omega_1 \times \Omega_2} |f_n - f| d\mu_1 \times \mu_2$ by the Fubini theorem. Consequently, $\lim_{n\to\infty} \int_{\Omega_1} |F_n - F| d\mu_1 = 0$, and, by Exercise 2.7.9, $\{F_n\}$ has a subsequence $\{F_{n_k}\}$ which converges to F a.e. on Ω_1. Now, the Fatou lemma implies that $\int_{\Omega_1} \lim_{k\to\infty} |F_{n_k} - F| d\mu_1 \le \liminf_{k\to\infty} \int_{\Omega_1} |F_{n_k} - F| d\mu_1 = 0$, which means $\lim_{k\to\infty} |F_{n_k} - F| = 0$ μ_1-a.e. on Ω_1, or

$$\lim_{k\to\infty} \int_{\Omega_2} |f_{n_k}(x,y) - f(x,y)| d\mu_2(y) = 0$$

for μ_1-a.e. x in Ω_1, as we claim.

We conclude this section by applying the Fubini theorem to prove a measurability result which we shall need later. For this purpose, define first a map t from $\mathbb{R}^{2n}$ to $\mathbb{R}^n$ by $t(x,y) = x - y$, where x and y are in $\mathbb{R}^n$. If f is a Borel measurable function on $\mathbb{R}^n$, then $f \circ t$ is Borel measurable on $\mathbb{R}^{2n}$, because $\{f \circ t > \alpha\} = t^{-1}\{f > \alpha\}$, which is a Borel set in $\mathbb{R}^{2n}$. Note that for $A \subset \mathbb{R}^n$, the y-section $(t^{-1}A)^y$ of $t^{-1}A$ is $A + y := \{x + y : x \in A\}$.

Lemma 4.8.4 *If A is a null set in $\mathbb{R}^n$, then $t^{-1}A$ is a null set in $\mathbb{R}^{2n}$.*

Proof There is a Borel set $B \supset A$ with $\lambda^n(B) = 0$. Now $t^{-1}B$ is a Borel set in $\mathbb{R}^{2n}$; by the Fubini theorem,

$$\begin{aligned} \lambda^{2n}(t^{-1}B) &= \int_{\mathbb{R}^{2n}} I_{t^{-1}B} d\lambda^{2n} = \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} I_{t^{-1}B}(x,y) d\lambda^n(x) \right) d\lambda^n(y) \\ &= \int_{\mathbb{R}^n} \lambda^n((t^{-1}B)^y) d\lambda^n(y) = \int_{\mathbb{R}^n} \lambda^n(B + y) d\lambda^n(y) \\ &= \int_{\mathbb{R}^n} \lambda^n(B) d\lambda^n = \int_{\mathbb{R}^n} 0 d\lambda^n = 0, \end{aligned}$$

i.e. $t^{-1}B$ is a null set in $\mathbb{R}^{2n}$. But $t^{-1}A \subset t^{-1}B$ implies that $t^{-1}A$ is a null set. ■

Proposition 4.8.2 *If f is a measurable function on $\mathbb{R}^n$, then $f \circ t$ is a measurable function on $\mathbb{R}^{2n}$.*

Proof There is a Borel function g on $\mathbb{R}^n$ such that $f = g + h$, where $h = 0$ a.e. on $\mathbb{R}^n$. Since $f \circ t = g \circ t + h \circ t$ and $g \circ t$ is Borel measurable, $f \circ t$ is measurable if $h \circ t$ is measurable. We claim that $h \circ t = 0$ a.e. on $\mathbb{R}^{2n}$. There is a null set $A \subset \mathbb{R}^n$ such that $h = 0$ on $\mathbb{R}^n \backslash A$. Then $h \circ t = 0$ on $t^{-1}(\mathbb{R}^n \backslash A) = (t^{-1}\mathbb{R}^n) \backslash t^{-1}A = \mathbb{R}^{2n} \backslash t^{-1}A$. But, by Lemma 4.8.4, $t^{-1}A$ is a null set in $\mathbb{R}^{2n}$, hence $h \circ t = 0$ a.e. on $\mathbb{R}^{2n}$. Since $h \circ t = 0$ a.e. on $\mathbb{R}^{2n}$, it is measurable; consequently, $f \circ t$ is measurable. ■

4.9 Smoothing of functions

Our concern in this section is the smoothing of functions and approximation of functions by smooth ones. The method we shall use is that of the Friederichs mollifier.

We define first some function spaces which will be frequently considered later. Given an open set Ω in $\mathbb{R}^n$ and a positive integer k, we shall denote by $C^k(\Omega)$ the vector space of all functions defined on Ω which have continuous partial derivatives up to order k, and denote by $C^\infty(\Omega)$ the space $\bigcap_k C^k(\Omega)$. The functions considered are either real-valued or complex-valued, as will either be clear from context or explicitly stated. For a function f defined on Ω, recall that the closure in Ω of the set $\{f \neq 0\}$ is called the support of f and is denoted by $\operatorname{supp} f$. If $\operatorname{supp} f$ is a compact set, then f is said to have compact support. The subspace of $C^k(\Omega)$, which consists of all functions in $C^k(\Omega)$ with compact support, is denoted by $C_c^k(\Omega)$; $C_c^\infty(\Omega)$ is similarly defined.

For a measurable subset Ω of $\mathbb{R}^n$, the space $L^p(\Omega, \mathcal{L}^n|\Omega, \lambda^n)$ will be simply denoted by $L^p(\Omega)$, for convenience, and accordingly the space of all those measurable functions which are in $L^p(K)$ for every compact subset K of Ω is denoted by $L^p_{\mathrm{loc}}(\Omega)$. Usually $L^1_{\mathrm{loc}}(\Omega)$ is simply denoted by $L_{\mathrm{loc}}(\Omega)$ and its elements are called locally integrable functions on Ω; correspondingly, functions in $L^p_{\mathrm{loc}}(\Omega)$ are called locally L^p functions on Ω.

Some notations regarding multi-indices are now introduced. By **multi-index**, we mean an ordered n-tuple $\alpha = (\alpha_1, \ldots, \alpha_n)$ of nonnegative integers for some integer $n > 1$ (n will be clearly implied from the context). For a multi-index $\alpha = (\alpha_1, \ldots, \alpha_n)$, the sum $\sum_{j=1}^n \alpha_j$ and the product $\prod_{j=1}^n \alpha_j!$ are denoted respectively by $|\alpha|$ and $\alpha!$; while if $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, x^α will stand for $x_1^{\alpha_1}, \ldots, x_n^{\alpha_n}$. The partial derivative symbol $\frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha^n}}$ will be abbreviated to $\frac{\partial^{|\alpha|}}{\partial x^\alpha}$ or ∂_x^α.

We are now ready to define the Friederichs mollifier. Let $\varphi \in C_c^\infty(\mathbb{R}^n)$ with $\int \varphi d\lambda^n = 1$. For definiteness, assume that $\operatorname{supp} \varphi \subset C_1(0)$, the closed ball in $\mathbb{R}^n$ centered at 0 and with radius 1. Such a function φ is called a **mollifying function**. For $\varepsilon > 0$, define $\varphi_\varepsilon(x) = \varepsilon^{-n}\varphi(\frac{x}{\varepsilon})$ for $x \in \mathbb{R}^n$; then $\operatorname{supp} \varphi_\varepsilon \subset C_\varepsilon(0)$ and $\int \varphi_\varepsilon d\lambda^n = 1$, by Example 4.3.1 (ii).

Corresponding to such a function φ and $\varepsilon > 0$, we define a linear transformation J_ε which maps functions f in $L_{\mathrm{loc}}(\Omega)$ to functions defined on $\Omega_\varepsilon = \{x \in \Omega : \operatorname{dist}(x, \Omega^c) > \varepsilon\}$, by

$$J_\varepsilon f(x) = \int_{C_\varepsilon(x)} f(y)\varphi_\varepsilon(x - y)dy, \quad x \in \Omega_\varepsilon.$$

Note that $C_\varepsilon(x) \subset \Omega$ for $x \in \Omega_\varepsilon$, hence f is integrable on $C_\varepsilon(x)$ and $J_\varepsilon f(x)$ is defined; moreover, since $\varphi_\varepsilon(x-y) = 0$ for y outside $C_\varepsilon(x)$, we may consider the defining integral for $J_\varepsilon f(x)$ as over the whole space $\mathbb{R}^n$, thus

$$J_\varepsilon f(x) = \int f(y)\varphi_\varepsilon(x-y)dy.$$

The family $\{J_\varepsilon\}_{\varepsilon>0}$, which depends on φ, is called a **Friederichs mollifier**. We often consider the case $\varphi \geq 0$, but for the moment, we do not impose this restriction.

The most well-known such nonnegative function φ is that defined as follows:

$$\varphi(x) = \begin{cases} Ce^{-\frac{1}{1-|x|^2}}, & \text{if } |x| < 1; \\ 0, & \text{if } |x| \geq 1, \end{cases}$$

where C is chosen so that $\int \varphi d\lambda^n = 1$.

Exercise 4.9.1

(i) Show that $J_\varepsilon f \in C(\Omega_\varepsilon)$.

(ii) More generally, suppose that h is a continuous function on $\mathbb{R}^n$ with $\operatorname{supp} h \subset C_\varepsilon(0)$; show that $\int f(y)h(x-y)dy$ is a continuous function of $x \in \Omega_\varepsilon$.

Exercise 4.9.2 Show that $\int f(y)\varphi_\varepsilon(x-y)dy = \int f(x-y)\varphi_\varepsilon(y)dy$, for $x \in \Omega_\varepsilon$.

Proposition 4.9.1 *If $f \in C(\Omega)$, $J_\varepsilon f(x) \to f(x)$ uniformly on any compact subset of Ω as $\varepsilon \to 0$.*

Proof Let $K \subset \Omega$ be compact. Fix $0 < \varepsilon_0 < \operatorname{dist}(K, \Omega^c)$ and let $F = \{x \in \Omega : \operatorname{dist}(x, K) \leq \varepsilon_0\}$. F is compact. Since f is uniformly continuous on F, for $\sigma > 0$, there is $\delta > 0$ with $\delta \leq \varepsilon_0$, such that $|f(x) - f(y)| \leq \sigma$ if x, y are in F and $|x-y| < \delta$. For $x \in K$, $0 < \varepsilon < \delta$, we have

$$|J_\varepsilon f(x) - f(x)| = \left| \int (f(y) - f(x))\varphi_\varepsilon(x-y)dy \right| \leq \sigma \int |\varphi_\varepsilon| d\lambda^n \leq \sigma M_\varphi,$$

where $M_\varphi = \int |\varphi| d\lambda^n$. ■

Proposition 4.9.2 *For $f \in L_{\text{loc}}(\Omega)$, $J_\varepsilon f \in C^\infty(\Omega_\varepsilon)$.*

Proof For $h \neq 0$, consider the difference quotient for $x \in \Omega_\varepsilon$,

$$\frac{1}{h}\{J_\varepsilon f(x + he_j) - J_\varepsilon f(x)\} = \int f(y) \frac{\varphi_\varepsilon(x + he_j - y) - \varphi_\varepsilon(x-y)}{h} dy,$$

where $e_j = (\delta_{j1}, \dots, \delta_{jn})$ with δ_{jk} being 1 or zero according to whether or not $k = j$. When h is small, $\operatorname{dist}(x + he_j, \Omega^c) \geq \varepsilon_0 > \varepsilon$, and for all such small enough h, $\varphi_\varepsilon(x + he_j - y) = 0$ for y outside a compact set K in Ω; therefore,

$$\int f(y) \frac{\varphi_\varepsilon(x + he_j - y) - \varphi_\varepsilon(x - y)}{h} dy = \int_K f(y) \frac{\varphi_\varepsilon(x + he_j - y) - \varphi_\varepsilon(x - y)}{h} dy.$$

Now,

$$\left| \frac{\varphi_\varepsilon(x + he_j - y) - \varphi_\varepsilon(x - y)}{h} \right| \leq \max_{z \in \mathbb{R}^n} \left| \frac{\partial \varphi_\varepsilon}{\partial x_j}(z) \right| := M_j,$$

and hence

$$\left| f(y) \frac{\varphi_\varepsilon(x + he_j - y) - \varphi_\varepsilon(x - y)}{h} \right| \leq M_j |f(y)|$$

on K. By LDCT,

$$\frac{\partial}{\partial x_j} J_\varepsilon f(x) = \lim_{h \to 0} \frac{1}{h} \{J_\varepsilon f(x + he_j) - J_\varepsilon f(x)\} = \int f(y) \frac{\partial \varphi_\varepsilon}{\partial x_j}(x - y) dy.$$

So far we have only used the fact that $\varphi_\varepsilon \in C_c^\infty(\mathbb{R}^n)$ with $\operatorname{supp} \varphi_\varepsilon \subset C_\varepsilon(0)$. Hence, we may repeat the argument to obtain

$$\frac{\partial^{|\alpha|}}{\partial x^\alpha} J_\varepsilon f(x) = \int f(y) \frac{\partial^{|\alpha|}}{\partial x^\alpha} \varphi_\varepsilon(x - y) dy.$$

By Exercise 4.9.1 (ii), each $\frac{\partial^{|\alpha|}}{\partial x^\alpha} J_\varepsilon f$ is continuous on Ω_ε. ■

Exercise 4.9.3 If K is a compact set and G is an open set containing K, then there is C^∞ function g with $\operatorname{supp} g \subset G$ and $0 \leq g \leq 1$, such that $g = 1$ on K.

Remark When $f \in L^p(\Omega)$, $1 \leq p \leq \infty$, we may consider f as defined on $\mathbb{R}^n$ by defining f to be zero outside Ω; then $J_\varepsilon f$ is defined for $x \in \mathbb{R}^n$ and hence for $x \in \Omega$.

Theorem 4.9.1 *For $f \in L^p(\Omega)$, $p \geq 1$, we have $\|J_\varepsilon f\|_p \leq L\|f\|_p$, where $L = L(\varphi, p)$.*

Proof By the previous remark, we may assume that $\Omega = \mathbb{R}^n$.

That $\|J_\varepsilon f\|_p \leq L\|f\|_p$ when $p = 1$ or ∞ is obvious. We consider the case $1 < p < \infty$. In this case, let $q > 1$ be the exponent conjugate to p, i.e. $\frac{1}{p} + \frac{1}{q} = 1$, then,

$$
\begin{aligned}
|J_\varepsilon f(x)| &= \left| \int f(y)\varphi_\varepsilon(x-y)dy \right| \\
&\le \int |f(y)||\varphi_\varepsilon(x-y)|dy = \int |f(x-y)||\varphi_\varepsilon(y)|dy \\
&\le \left\{ \int |f(x-y)|^p |\varphi_\varepsilon(y)|dy \right\}^{\frac{1}{p}} \left\{ \int |\varphi_\varepsilon(y)|dy \right\}^{\frac{1}{q}} \\
&= C\left\{ \int |f(x-y)|^p |\varphi_\varepsilon(y)|dy \right\}^{\frac{1}{p}},
\end{aligned}
$$

where $C = \{\int |\varphi_\varepsilon(y)|dy\}^{1/q} = \{\int |\varphi(y)|dy\}^{\frac{1}{q}}$. In one of the steps above, we have used Hölder's inequality w.r.t. the measure ν with $d\nu = |\varphi_\varepsilon| d\lambda^n$ (cf. Exercise 2.5.7). Now the Fubini theorem implies

$$
\begin{aligned}
\|J_\varepsilon f\|_p^p &\le C^p \int \left(\int |f(x-y)|^p |\varphi_\varepsilon(y)|dy \right) dx \\
&= C^p \int \left(\int |f(x-y)|^p |\varphi_\varepsilon(y)|dx \right) dy \\
&= C^p \|f\|_p^p \int |\varphi(y)|dy = C^p C^q \|f\|_p^p,
\end{aligned}
$$

or,

$$
\|J_\varepsilon f\|_p \le L\|f\|_p,
$$

where $L = L(\varphi, p)$. Note that $(x, y) \mapsto |f(x-y)|^p \varphi_\varepsilon(y)$ is measurable by Proposition 4.8.2. ■

Exercise 4.9.4 Show that if $\varphi \ge 0$, the constant L in Theorem 4.9.1 can be taken to be 1.

Theorem 4.9.2 *If $f \in L^p(\Omega)$, $1 \le p < \infty$, then* $\lim_{\varepsilon \to 0} \|J_\varepsilon f - f\|_p = 0$.

Proof We may assume that $\Omega = \mathbb{R}^n$. Let $\sigma > 0$ be given. By Proposition 4.6.1, there is $g \in C_c(\mathbb{R}^n)$ such that $\|f - g\|_p < \frac{\sigma}{2(L+1)}$, where $L = L(\varphi, p)$ is the constant in Theorem 4.9.1. Now,

$$
\begin{aligned}
\|J_\varepsilon f - f\|_p &= \|J_\varepsilon f - J_\varepsilon g + J_\varepsilon g - g + g - f\|_p \\
&\le \|J_\varepsilon (f-g)\|_p + \|J_\varepsilon g - g\|_p + \|g - f\|_p \\
&\le (L+1)\|f-g\|_p + \|J_\varepsilon g - g\|_p \\
&< \frac{\sigma}{2} + \|J_\varepsilon g - g\|_p,
\end{aligned}
$$

where we have used the inequality $\|J_\varepsilon(f-g)\|_p \leq L\|f-g\|_p$ as asserted by Theorem 4.9.1. Let K be the support of g and put $\widehat{K} = \{x \in \mathbb{R}^n : \operatorname{dist}(x, K) \leq 1\}$. $\widehat{K}$ is a compact set, outside which both g and $J_\varepsilon g$ vanish if $0 < \varepsilon \leq 1$. Hence, from Proposition 4.9.1,

$$\|J_\varepsilon g - g\|_p^p = \int_{\widehat{K}} |J_\varepsilon g - g|^p d\lambda^n < \left(\frac{\sigma}{2}\right)^p,$$

or,

$$\|J_\varepsilon g - g\|_p < \frac{\sigma}{2}$$

if ε is sufficiently small, say $\varepsilon < \delta$. This means that $\|J_\varepsilon f - f\|_p < \frac{\sigma}{2} + \|J_\varepsilon g - g\|_p < \sigma$, if $\varepsilon < \delta$. ■

Corollary 4.9.1 *$C_c^\infty(\Omega)$ is dense in $L^p(\Omega)$, $1 \leq p < \infty$.*

Proof Let $f \in L^p(\Omega)$, $1 \leq p < \infty$, and fix $\sigma > 0$. By Proposition 4.6.1, there is $g \in C_c(\Omega)$ such that $\|f - g\|_p < \frac{\sigma}{2}$; while from Theorem 4.9.2, if $\varepsilon > 0$ is small enough, $\|J_\varepsilon g - g\|_p < \frac{\sigma}{2}$. Since g has compact support in Ω, $J_\varepsilon g$ has compact support in Ω if ε is small enough. Hence if ε is small enough, $J_\varepsilon g \in C_c^\infty(\Omega)$ and $\|J_\varepsilon g - g\|_p < \frac{\sigma}{2}$; but then $\|f - J_\varepsilon g\|_p \leq \|f - g\|_p + \|g - J_\varepsilon g\|_p < \sigma$. ■

Exercise 4.9.5 Suppose that $\varphi(x) = \varphi(-x)$ for all x in $\mathbb{R}^n$ and let f, g be in $L^2(\mathbb{R}^n)$. Show that

$$\int_{\mathbb{R}^n} (J_\varepsilon f) g d\lambda^n = \int_{\mathbb{R}^n} f J_\varepsilon g d\lambda^n.$$

4.10 Change of variables for multiple integrals

A transformation formula for multiple integrals under changes of variables will be proved in this section. The changes of variables to be considered are C^1 diffeomorphisms, which we shall now describe. Let Ω be an open set in $\mathbb{R}^n$. A map $t = (t_1, \ldots, t_n)$ from Ω into $\mathbb{R}^n$ is called a **C^1 map** if its component functions t_i are continuously differentiable, i.e. first-order partial derivatives of each t_i exist and are continuous on Ω. For $x \in \Omega$, the linear map from $\mathbb{R}^n$ into $\mathbb{R}^n$ represented by the matrix $\left(\frac{\partial t_i}{\partial x_j}(x)\right)$ in reference to the standard basis of $\mathbb{R}^n$ is called the **differential** of t at x, and is denoted by $d_x t$. By the **standard basis** of $\mathbb{R}^n$ we mean the basis formed by $e_1, \ldots, e_n$, where for each j, $e_j = (\delta_{j1}, \ldots, \delta_{jn})$, with δ_{jk} being 1 or 0 according to whether $k = j$ or $k \neq j$. The symbols δ_{jk} are called **Kronecker symbols**. In this section, linear maps from $\mathbb{R}^n$ to $\mathbb{R}^n$ are represented by matrices with reference to the standard basis. The determinant of $\left(\frac{\partial t_i}{\partial x_j}(x)\right)$, called the **Jacobian** of t at x, is to be denoted by $J(t; x)$. When t is a linear map, $t_i(x) = \sum_{j=1}^n t_{ij} x_j$ for $x = (x_1, \ldots, x_n)$, where (t_{ij}) is the matrix representing t; it follows then that $\left(\frac{\partial t_i}{\partial x_j}(x)\right) = (t_{ij})$, i.e. $d_x t = t$. For a linear map t, the determinant of the matrix representing t is usually denoted by $\det t$, thus $J(t; x) = \det d_x t$ if t is a C^1 map. A C^1 map t from Ω into $\mathbb{R}^n$ is called a

C^1 **diffeomorphism** if it is injective and $d_x t$ is invertible for all $x \in \Omega$. By the **inverse function theorem**, if t is a C^1 diffeomorphism from Ω into $\mathbb{R}^n$, then t^{-1} is a C^1 diffeomorphism from $t\Omega$ onto Ω and $J(t;x)^{-1} = J(t^{-1}; tx)$ for $x \in \Omega$. Note that $J(t;x) \neq 0$ for all x in Ω.

We consider first the transformation formula for integrals when changes of variables are invoked by invertible linear maps. We follow the usual practice of denoting linear maps by capital letters, and, for convenience, the matrix representing a linear map T is also denoted by T. The matrices derived from the unit matrix I by **elementary row operations** are called **elementary matrices**. They are of the following three types:

(i) A type(1) elementary matrix is one obtained from I by multiplying a row of I by a nonzero real number c;

(ii) a type(2) elementary matrix is one obtained from I by multiplying a row of I by a nonzero real number and then adding it to a different row of I;

(iii) a type(3) elementary matrix is one obtained from I by interchanging two rows of I.

Note that if T is an elementary matrix of type(1), then $\det T = c$; while $\det T = 1$ or -1, according to whether T is of type(2) or type(3). If T is an elementary matrix, the corresponding linear map T is called an **elementary linear map** of the same type.

Lemma 4.10.1 *If T is an elementary linear map and $f \geq 0$ is a measurable function on $\mathbb{R}^n$ such that $f \circ T$ is measurable, then*

$$\int_{\mathbb{R}^n} f d\lambda^n = |\det T| \int_{\mathbb{R}^n} f \circ T d\lambda^n. \tag{4.17}$$

Proof Suppose that T is of type(1), then $f \circ T(x_1, \ldots, x_n) = f(x_1, \ldots, cx_j, \ldots, x_n)$ for some $j = 1, \ldots, n$ and $c \neq 0$. By expressing $x = (x_1, \ldots, x_j, \ldots, x_n)$ as $x = (x_j, \hat{x}_j)$ and using the Fubini theorem, we have

$$\begin{aligned}
\int_{\mathbb{R}^n} f \circ T d\lambda^n &= \int_{\mathbb{R}^{n-1}} \left(\int_{\mathbb{R}} f(x_1, \ldots, cx_j, \ldots, x_n) dx_j \right) d\hat{x}_j \\
&= \frac{1}{|c|} \int_{\mathbb{R}^{n-1}} \left(\int_{\mathbb{R}} f(x_1, \ldots, x_j, \ldots, x_n) dx_j \right) d\hat{x}_j \\
&= \frac{1}{|c|} \int_{\mathbb{R}^n} f d\lambda^n,
\end{aligned}$$

where $\int_{\mathbb{R}} f(x_1, \ldots, cx_j, \ldots, x_n) dx_j = \frac{1}{|c|} \int_{\mathbb{R}} f(x_1, \ldots, x_j, \ldots, x_n) dx_j$ follows from the fact stated in Example 4.3.1 (ii). Hence,

$$\int_{\mathbb{R}^n} f d\lambda^n = |c| \int_{\mathbb{R}^n} f \circ T d\lambda^n = |\det T| \int_{\mathbb{R}^n} f \circ T d\lambda^n.$$

Similarly, (4.17) can be verified for the case when T is of type(2) or of type(3). ■

If T is an invertible linear map from $\mathbb{R}^n$ onto $\mathbb{R}^n$ then, as is well known in elementary linear algebra, after a finite number of elementary row operations the corresponding matrix T becomes the unit matrix I, i.e.

$$I = S_1 \cdots S_k \cdot T,$$

where $S_1, \ldots, S_k$ are elementary matrices, or

$$S_k^{-1} \cdots S_1^{-1} = T,$$

where each S_j^{-1} is also elementary and of the some type as S_j; in terms of maps, this means that the invertible linear map T is a composition of a finite number of elementary linear maps, i.e.

$$T = T_1 \circ \cdots \circ T_l, \tag{4.18}$$

with each T_j being elementary.

Theorem 4.10.1 *If T is an invertible linear map from $\mathbb{R}^n$ onto $\mathbb{R}^n$ and f is a measurable function on $\mathbb{R}^n$, then $f \circ T$ is measurable; and if f is either nonnegative or integrable,*

$$\int_{\mathbb{R}^n} f d\lambda^n = |\det T| \int_{\mathbb{R}^n} f \circ T d\lambda^n. \tag{4.19}$$

Proof It is sufficient to prove (4.19) for the case $f \geq 0$. Suppose first that $f \geq 0$ is Borel measurable, then since $f \circ T$ is Borel and T is of the form (4.18), we have from Lemma 4.10.1,

$$\begin{aligned} |\det T| \int_{\mathbb{R}^n} f \circ T d\lambda^n &= \prod_{j=1}^{l} |\det T_j| \int_{\mathbb{R}^n} f \circ T_1 \circ \cdots \circ T_l d\lambda^n \\ &= \left(\prod_{j=1}^{l-1} |\det T_j| \right) \cdot |\det T_l| \int_{\mathbb{R}^n} (f \circ T_1 \circ \cdots \circ T_{l-1}) \circ T_l d\lambda^n \\ &= \prod_{j=1}^{l-1} |\det T_j| \int_{\mathbb{R}^n} f \circ T_1 \circ \cdots \circ T_{l-1} d\lambda^n \\ &= \cdots = \int_{\mathbb{R}^n} f d\lambda^n. \end{aligned}$$

Thus (4.19) holds when f is a nonnegative Borel function on $\mathbb{R}^n$.

Now suppose that f is nonnegative and measurable. We claim first that $f \circ T$ is measurable. Let $B \in \mathcal{B}^n$; we have to show that $(f \circ T)^{-1} B = T^{-1}(f^{-1} B)$ is measurable. As $f^{-1}B$ is measurable, $f^{-1}B = A \cup C$, where A is a Borel set and $\lambda^n(C) = 0$ (cf. Exercise 3.9.1 (i)). There is a Borel set $D \supset C$ such that $\lambda^n(D) = 0$. The indicator function I_D of D is a Borel function; by what we have proved in the first part, $|\det T|\int_{\mathbb{R}^n} I_D \circ T d\lambda^n = \lambda^n(D) = 0$; then $\int_{\mathbb{R}^n} I_D \circ T d\lambda^n = 0$, and consequently $I_D \circ T = 0$ a.e. But $I_D \circ T = I_{T^{-1}D}$ and $I_D \circ T = 0$ a.e. imply $\lambda^n(T^{-1}D) = 0$. Since $T^{-1}C \subset T^{-1}D$, $\lambda^n(T^{-1}C) = 0$. Thus $T^{-1}C$ is measurable. Now, $(f \circ T)^{-1}B = T^{-1}(A \cup C) = T^{-1}A \cup T^{-1}C$ shows that $(f \circ T)^{-1}B$ is measurable. We have shown the claim that $f \circ T$ is measurable. Since $f \circ T$ is measurable, we can repeat the first part of the proof to conclude that (4.19) holds. ■

Corollary 4.10.1 *For a measurable set $A \subset \mathbb{R}^n$, TA is measurable and $\lambda^n(TA) = |detT|\lambda^n(A)$.*

Proof In Theorem 4.10.1, replace T by T^{-1} and consider $f = I_A$. ■

Corollary 4.10.2 *Lebesgue measure is invariant under rotations.*

Proof Let $A \subset \mathbb{R}^n$ and T be a rotation of $\mathbb{R}^n$; we have to show that $\lambda^n(TA) = \lambda^n(A)$. By Corollary 4.10.1, $\lambda^n(TA) = |\det T|\lambda^n(A) = \lambda^n(A)$, because the matrix representing T is an orthogonal matrix and the determinant of an orthogonal matrix is 1 or –1. ■

Now let t be a C^1 diffeomorphism from Ω into $\mathbb{R}^n$. Define a measure $\lambda^n t$ on Ω by

$$\lambda^n t(A) = \lambda^n(tA), \quad A \subset \Omega.$$

That $\lambda^n t$ measures Ω is obvious. Since t is bijective from Ω to $t\Omega$, the measure $\lambda^n t$ on Ω can be considered as a copy of λ^n on the open set $t\Omega$; actually a subset A of Ω is $\lambda^n t$-measurable if and only if tA is λ^n-measurable, and both t and t^{-1} are measure preserving (cf. Section 2.8.2). Furthermore, since a subset B of Ω is Borel if and only if tB is Borel, it follows that $\lambda^n t$ is a Radon measure on Ω.

Proposition 4.10.1 *If $f \geq 0$ is measurable on $t\Omega$, then $f \circ t$ is $\Sigma^{\lambda^n t}$-measurable on Ω and*

$$\int_{t\Omega} f d\lambda^n = \int_{\Omega} f \circ t d\lambda^n t. \tag{4.20}$$

Proof If $f = I_A$ for a measurable set A, then $f \circ t = I_{t^{-1}A}$, where $t^{-1}A$ is $\lambda^n t$-measurable; it follows that (4.20) holds in this case. For the general case, (4.20) follows from Theorem 2.2.1 and what has just been shown. ■

Remark Since $\lambda^n = t_\# \lambda^n t$ on $t\Omega$, Proposition 4.10.1 follows also from Exercise 4.3.2.

Lemma 4.10.2 *$\lambda^n t$ is absolutely continuous on Ω.*

Proof Let $Q \subset \Omega$ be a nondegenerate oriented closed cube, i.e. $Q = I_1 \times \cdots \times I_n$, where $I_1, \ldots, I_n$ are finite closed intervals in $\mathbb{R}$ of the same positive length. Suppose that f is a continuously differentiable function defined on a neighborhood of Q, and consider two points x and y in Q. Let a function g on $[0, 1]$ be defined by $g(s) = f(x + s(y - x))$; then $f(y) - f(x) = g(1) - g(0) =$

$\int_0^1 g'(s)ds = \int_0^1 \left\{\sum_{j=1}^n \frac{\partial f}{\partial x_j}(x + s(y - x)) \cdot (y_j - x_j)\right\} ds = \int_0^1 \nabla f(x + s(y - x)) \cdot (y - x)ds$, where $\nabla f = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\right)$ is the gradient of f. Hence,

$$|f(y) - f(x)| \leq |y - x| \int_0^1 |\nabla f(x + s(y - x))| ds. \tag{4.21}$$

Applying (4.21) to each component function of t, we have

$$|t(y) - t(x)|^2 \leq |y - x|^2 \sum_{i=1}^n \left[\int_0^1 |\nabla t_i(x + s(y - x))| ds\right]^2 \leq |y - x|^2 M(Q)^2,$$

or

$$|t(y) - t(x)| \leq |y - x| M(Q), \tag{4.22}$$

where $M(Q)^2 = \max_{z \in Q} \sum_{i,j=1}^n \left|\frac{\partial t_i}{\partial x_j}(z)\right|^2$.

Suppose now that A is a null set in Ω. Since Ω is a countable union of open sets G, with $\overline{G}$ being a compact subset of Ω (cf. Proposition 3.9.2), to show that $\lambda^n t(A) = 0$, we may assume that A is a null set in an open set G, with $\overline{G}$ a compact set in Ω. Given that $\varepsilon > 0$, there is a sequence $\{Q_k\}$ of nondegenerate closed oriented cubes in G such that $\bigcup Q_k \supset A$ and $\sum_k \lambda^n(Q_k) < \varepsilon$, by Corollary 3.9.1. For each k, let c_k be the center of Q_k, and apply (4.22) for $x = c_k$ and $y \in Q_k$, to obtain

$$|t(y) - t(c_k)| \leq |y - c_k| M(Q_k),$$

which implies that $tQ_k - t(c_k) \subset C_r(t(c_k))$ with $r = (\frac{1}{2} \operatorname{diam} Q_k)M$, where $M^2 = \max_{z \in \overline{G}} \sum_{i,j=1}^n \left|\frac{\partial t_i}{\partial x_j}(z)\right|^2$, and consequently,

$$\lambda^n(tQ_k) = \lambda^n(tQ_k - t(c_k)) \leq \lambda^n(C_r(t(c_k))) = \left(\frac{\sqrt{n}M}{2}\right)^n \lambda^n(C_1(0))\lambda^n(Q_k),$$

by Example 4.3.1. Now,

$$\begin{aligned}\lambda^n t(A) &\leq \lambda^n t\Big(\bigcup_k Q_k\Big) \leq \sum_k \lambda^n t(Q_k) = \sum_k \lambda^n(tQ_k) \\ &\leq \left(\frac{\sqrt{n}M}{2}\right)^n \lambda^n(C_1(0)) \sum_k \lambda^n(Q_k) < \left(\frac{\sqrt{n}M}{2}\right)^n \lambda^n(C_1(0))\varepsilon,\end{aligned}$$

from which, by letting $\varepsilon \to 0$, we conclude that $\lambda^n t(A) = 0$. ■

Corollary 4.10.3 *$A \subset \Omega$ is measurable if and only if tA is measurable. Also, A is measurable if and only if it is $\lambda^n t$-measurable.*

Proof If A is measurable, then $A = B \cup N$, with B a Borel set and N a null set. By Lemma 4.10.2, $\lambda^n(tN) = \lambda^n t(N) = 0$; hence tN is a null set and is therefore measurable. Now, $tA = tB \cup tN$ implies that tA is measurable. Conversely, if tA is measurable, then A is measurable by the same argument, but with Ω replaced by $t\Omega$ and t replaced by t^{-1}.

Since $A \subset \Omega$ is $\lambda^n t$-measurable if and only if tA is measurable, the second part of the corollary follows from the first part. ■

Lemma 4.10.3 *For a.e. x in Ω, $\frac{d\lambda^n t}{d\lambda^n}(x) = \left|\det d_x t\right|$.*

Proof It is sufficient to show that $\lim_{r\to 0} \frac{\lambda^n t(C_r(x))}{\lambda^n(C_r(x))} = \left|\det d_x t\right|$ for $x \in \Omega$, where $C_r(x)$ is the closed ball centered at x and with radius r.

Let $x \in \Omega$ and suppose first that $d_x t = I$, the identity map of $\mathbb{R}^n$. Write

$$t(y) - t(x) = d_x t(y - x) + R(x, y) = (y - x) + R(x, y). \tag{4.23}$$

Since t is differentiable at x, for each $\varepsilon > 0$, there is $\delta > 0$ such that $\left|R(x, y)\right| < \varepsilon\left|y - x\right|$ if $\left|y - x\right| < \delta$. Now if $0 < r < \delta$, we have from (4.23),

$$tC_r(x) - t(x) \subset (1 + \varepsilon)(C_r(x) - x),$$

then $\lambda^n(tC_r(x)) = \lambda^n(tC_r(x) - t(x)) \le (1 + \varepsilon)^n \lambda^n(C_r(x) - x) = (1 + \varepsilon)^n \lambda^n(C_r(x))$, and hence

$$\limsup_{r\to 0} \frac{\lambda^n t(C_r(x))}{\lambda^n(C_r(x))} \le (1 + \varepsilon)^n. \tag{4.24}$$

We show next that $C := C_{r(1-\varepsilon)}(t(x))$ is contained in $tC_r(x)$ if $0 < r < \delta$. Observe first that, by (4.23), $t\Gamma$ is outside C, where Γ is the boundary of $C_r(x)$. To show that $C \subset tC_r(x)$ is to show that the line segment $[t(x), z] := \{t(x) + s(z - t(x)) : 0 \le s \le 1\} \subset tC_r(x)$ for each $z \in \partial C$. Let $z \in \partial C$ be fixed. Define a set L of positive numbers by

$$L = \{0 < \rho \le 1 : t(x) + s(z - t(x)) \in tC_r(x) \text{ for all } 0 \le s \le \rho\}.$$

By the **inverse function theorem**, t maps a neighborhood of x in $C_r(x)$ onto a neighborhood of $t(x)$; hence L is nonempty. Let $\rho_0 = \sup L$. We claim that $\rho_0 \in L$. Note first that $(0, \rho_0) \subset L$. Choose a sequence $\{s_j\}$ in $(0, \rho_0)$ such that $s_j \to \rho_0$ and let $z_j = t(x) + s_j(z - t(x))$. Then $z_j \in tC_r(x)$ for each j. Since $z_j \to z_\infty := t(x) + \rho_0(z - t(x))$ and t^{-1} is continuous, we infer that $t^{-1}z_j \to t^{-1}z_\infty$ and $t^{-1}z_\infty \in C_r(x)$ (note that each $t^{-1}z_j \in C_r(x)$). Now, $t(t^{-1}z_\infty) = z_\infty$ implies that $\rho_0 \in L$. We assert then that $\rho_0 = 1$. If $\rho_0 < 1$, $t^{-1}z_\infty \in B_r(x)$, because $t\Gamma$ is outside C; then by the inverse function theorem again, t maps a neighborhood of $t^{-1}z_\infty$ in $B_r(x)$ onto a neighborhood of z_∞; this would imply that L contains numbers larger than ρ_0, contradicting the definition of ρ_0. Now $\rho_0 = 1$ means the line segment $[t(x), z]$ is contained in

$tC_r(x)$. Thus C is contained in $tC_r(x)$, or $tC_r(x) - t(x) \supset (1-\varepsilon)(C_r(t(x)) - t(x))$. Hence,

$$\lambda^n t(C_r(x)) = \lambda^n(tC_r(x)) = \lambda^n(tC_r(x) - t(x)) \geq (1-\varepsilon)^n \lambda^n(C_r(x)),$$

or

$$\liminf_{r\to 0} \frac{\lambda^n t(C_r(x))}{\lambda^n(C_r(x))} \geq (1-\varepsilon)^n. \tag{4.25}$$

Letting $\varepsilon \to 0$ in (4.24) and (4.25), we have

$$\lim_{r\to 0} \frac{\lambda^n t(C_r(x))}{\lambda^n(C_r(x))} = 1.$$

This shows that $\lim_{r\to 0} \frac{\lambda^n t(C_r(x))}{\lambda^n(C_r(x))} = 1$, if $d_x t = I$. In general, for $x \in \Omega$, consider the map $\hat{t} = (d_x t)^{-1} \circ t$, then $d_x \hat{t} = (d_x t)^{-1} \circ d_x t = I$, hence,

$$\lim_{r\to 0} \frac{\lambda^n \hat{t}(C_r(x))}{\lambda^n(C_r(x))} = 1. \tag{4.26}$$

Now, by Corollary 4.10.1,

$$\begin{aligned} \lambda^n t(C_r(x)) &= \lambda^n(tC_r(x)) = \lambda^n(d_x t \circ (d_x t)^{-1}(tC_r(x))) \\ &= |\det d_x t| \lambda^n(\hat{t}C_r(x)), \end{aligned}$$

from which it follows that

$$\lim_{r\to 0} \frac{\lambda^n t(C_r(x))}{\lambda^n(C_r(x))} = |\det d_x t| \lim_{r\to 0} \frac{\lambda^n \hat{t}(C_r(x))}{\lambda^n(C_r(x))} = |\det d_x t|,$$

by (4.26). ■

Theorem 4.10.2 *Suppose that t is a C^1 diffeomorphism from an open set Ω in $\mathbb{R}^n$ into $\mathbb{R}^n$; then if f is a measurable function on $t\Omega$, $f \circ t$ is measurable on Ω, and if, furthermore, f is nonnegative or integrable, then,*

$$\int_{t\Omega} f d\lambda^n = \int_{\Omega} (f \circ t)(x)|J(t;x)| d\lambda^n(x). \tag{4.27}$$

Proof Since $A \subset t\Omega$ is measurable if and only if $t^{-1}A$ is measurable by Corollary 4.10.3, we infer that if $f = I_A$, then f is measurable if and only if $f \circ t = I_{t^{-1}A}$ is measurable. It follows then from Theorem 2.2.1 that a nonnegative function f is measurable if and only if $f \circ t$ is measurable; from this it follows that f is measurable on $t\Omega$ if and only if

$f \circ t$ is measurable. In particular if f is measurable on $t\Omega$, then $f \circ t$ is measurable on Ω. To verify (4.27), we need only consider the case $f \geq 0$. By Proposition 4.10.1,

$$\int_{t\Omega} f d\lambda^n = \int_{\Omega} f \circ t d\lambda^n t. \tag{4.28}$$

Since $\lambda^n t$ is absolutely continuous,

$$\lambda^n t(A) = \int_A \frac{d\lambda^n t}{d\lambda^n} d\lambda^n = \int_A |\det d_x t| d\lambda^n(x) = \int_A |J(t;x)| d\lambda^n(x)$$

for measurable $A \subset \Omega$ by Lemma 4.10.3; it follows then from Exercise 2.5.7 that $\int_\Omega f \circ t d\lambda^n t = \int_\Omega (f \circ t)(x)|J(t;x)|d\lambda^n(x)$; combining the last equality with (4.28), we conclude that (4.27) holds. ■

We illustrate the way to use Theorem 4.10.2 by an example.

Example 4.10.1 Consider the map t from the open set $\Omega := \{(\rho, \theta) : 0 < \rho < \infty, 0 < \theta < 2\pi\}$ in $\mathbb{R}^2$ into $\mathbb{R}^2$ by

$$(x_1, x_2) = t(\rho, \theta) = (\rho \cos\theta, \rho \sin\theta),$$

then, $\frac{\partial x_1}{\partial \rho} = \cos\theta$, $\frac{\partial x_1}{\partial \theta} = -\rho \sin\theta$; $\frac{\partial x_2}{\partial \rho} = \sin\theta$, $\frac{\partial x_2}{\partial \theta} = \rho\cos\theta$. Hence,

$$d_{(\rho,\theta)} t = \begin{vmatrix} \cos\theta & -\rho\sin\theta \\ \sin\theta & \rho\cos\theta \end{vmatrix} = \rho > 0.$$

t is actually a C^1 diffeomorphism from Ω onto $t\Omega = \{(x_1, x_2) \in \mathbb{R}^2 : x_2 \neq 0, \text{or } x_2 = 0$ but $x_1 < 0\}$, i.e. $t\Omega$ is obtained from $\mathbb{R}^2$ by taking away the positive x_1-axis and the origin. Now if $f \geq 0$ is measurable, then, since $\lambda^2(\mathbb{R}^2 \setminus t\Omega) = 0$, we have

$$\begin{aligned} \int_{\mathbb{R}^2} f d\lambda^2 &= \int_{t\Omega} f d\lambda^n = \int_\Omega (f \circ t)(\rho, \theta) \rho d\lambda^2(\rho, \theta) \\ &= \int_0^\infty \left(\int_0^{2\pi} \rho f(\rho\cos\theta, \rho\sin\theta) d\theta \right) d\rho, \end{aligned}$$

where we have the applied the Fubini theorem in the last step.

Exercise 4.10.1 Suppose that f is a measurable function on $\mathbb{R}^3$ and is either nonnegative or integrable.

(i) Show that

$$\begin{aligned} \int_{\mathbb{R}^3} f(x,y,z) d\lambda^3(x,y,z) &= \int_G f(\rho\cos\varphi, \rho\sin\varphi, z) \rho d\lambda^3(\rho, \varphi, z) \\ &= \int_{-\infty}^{\infty} \int_0^{2\pi} \int_0^\infty f(\rho\cos\varphi, \rho\sin\varphi, z) \rho d\rho d\varphi dz, \end{aligned}$$

where $G = (0, \infty) \times (0, 2\pi) \times \mathbb{R} = \{(\rho, \varphi, z) : 0 < \rho < \infty, 0 < \varphi < 2\pi, z \in \mathbb{R}\}$.

(ii) Show that

$$\begin{aligned}
&\int_{\mathbb{R}^3} f(x,y,z)d\lambda^3(x,y,z) \\
&= \int_H f(\rho\sin\theta\cos\varphi, \rho\sin\theta\sin\varphi, \rho\cos\theta)\rho^2\sin\theta d\lambda^3(\rho,\theta,\varphi) \\
&= \int_0^\infty \int_0^\pi \int_0^{2\pi} f(\rho\sin\theta\cos\varphi, \rho\sin\theta\sin\varphi, \rho\cos\theta)\rho^2\sin\theta d\varphi d\theta d\rho,
\end{aligned}$$

where $H = (0,\infty)\times(0,\pi)\times(0,2\pi) = \{(\rho,\theta,\varphi) : 0<\rho<\infty, 0<\theta<\pi, 0<\varphi<2\pi\}$.

4.11 Polar coordinates and potential integrals

In Example 4.10.1, ρ and θ are the **polar coordinates** of the point $(\rho\cos\theta, \rho\sin\theta)$ in $\mathbb{R}^2$, and $d\theta$ is the line element on the unit circle S^1, described by $(\cos\theta, \sin\theta)$, $0 \le \theta < 2\pi$; while in Exercise 4.10.1 (ii), ρ, φ, and θ are the so-called **spherical coordinates** of the point $(\rho\sin\theta\cos\varphi, \rho\sin\theta\sin\varphi, \rho\cos\theta)$ in $\mathbb{R}^3$, and $\sin\theta d\varphi d\theta$ is the surface element on the unit sphere S^2 in $\mathbb{R}^3$, described by $(\sin\theta\cos\varphi, \sin\theta\cos\varphi, \cos\theta)$, $0 \le \varphi < 2\pi$, $0\le\theta\le\pi$. Therefore, for nonnegative measurable function f on $\mathbb{R}^2$ or $\mathbb{R}^3$, we have

$$\int_{\mathbb{R}^2} f(x)d\lambda^2(x) = \int_0^\infty \left(\int_{S^1} \rho f(\rho x')dl(x')\right)d\rho; \tag{4.29}$$

$$\int_{\mathbb{R}^3} f(x)d\lambda^3(x) = \int_0^\infty \left(\int_{S^2} \rho^2 f(\rho x')d\sigma(x')\right)d\rho, \tag{4.30}$$

where $x = \rho x'$ with $\rho = |x|$ and $x' \in S^1$ or S^2, depending on $x \in \mathbb{R}^2$ or $\mathbb{R}^3$, dl is the line element on S^1, and $d\sigma$ the surface element on S^2. The discussion so far is formal; we shall now put it on a solid basis for $\mathbb{R}^n$ in general.

For $x \in \dot{\mathbb{R}}^n := \mathbb{R}^n\setminus\{0\}$, write $x = \rho x'$, where $\rho = |x|$ and $x' = |x|^{-1}x$ is in $S^{n-1} := \{x \in \mathbb{R}^n : |x| = 1\}$; ρ and x' are called the **polar coordinates** of $x \in \dot{\mathbb{R}}^n$. The polar coordinates of a point $x \in \dot{\mathbb{R}}^n$ will be written as an ordered pair (ρ, x') and hence is represented as a point in $(0,\infty)\times S^{n-1}$. Let p be the map $x \mapsto (\rho, x')$ from $\dot{\mathbb{R}}^n$ to $(0,\infty)\times S^{n-1}$; p is obviously a bijection and both p and p^{-1} are continuous; it follows that a function f on $\dot{\mathbb{R}}^n$ is λ^n-measurable if and only if $f\circ p^{-1}$ is $p_\#\lambda^n$-measurable on $(0,\infty)\times S^{n-1}$, where $p_\#\lambda^n$ is the measure on $(0,\infty)\times S^{n-1}$, defined by $p_\#\lambda^n(A) = \lambda^n(p^{-1}A)$ for subsets A of $(0,\infty)\times S^{n-1}$ (cf. Exercise 4.3.1 and note that $\lambda^n = (p^{-1})_\#(p_\#\lambda^n)$). We then infer from Exercise 4.3.2 that if f is a nonnegative measurable or an integrable function on $\mathbb{R}^n$, then

$$\int_{\mathbb{R}^n} fd\lambda^n = \int_{\dot{\mathbb{R}}^n} fd\lambda^n = \int_{(0,\infty)\times S^{n-1}} f\circ p^{-1}dp_\#\lambda^n. \tag{4.31}$$

We shall presently show that $p_{\#}\lambda^n$ is a product measure. A Borel measure σ on S^{n-1} will be defined first; this measure is interpreted as measuring the surface area of sets in S^{n-1} and is therefore called the surface measure on S^{n-1}. For $E \subset S^{n-1}$ and $r > 0$, let E_r be the set $\bigcup\{\alpha E : 0 < \alpha \le r\}$ in $\mathbb{R}^n$; clearly, $E_r = rE_1$ and E_r is a Borel set in $\mathbb{R}^n$, if $E \in \mathcal{B}(S^{n-1})$. It then follows that

$$\lambda^n(E_r) = r^n \lambda^n(E_1) \tag{4.32}$$

for $E \in \mathcal{B}(S^{n-1})$, by Example 4.3.1 (ii). Observe now that if $h > 0$, $E_{1+h}\backslash E_1$ is a spherically sliced section of the cone $\bigcup\{\alpha E : \alpha > 0\}$ of thickness h, and hence it is natural to define the surface area of $E \in \mathcal{B}(S^{n-1})$, as

$$\lim_{h\to 0+} h^{-1}\lambda^n(E_{1+h}\backslash E_1) = \lim_{h\to 0+} h^{-1}[(1+h)^n - 1]\lambda^n(E_1) = n\lambda^n(E_1),$$

where we have applied (4.32) with $r = 1 + h$. Thus we let $\sigma(E) = n\lambda^n(E_1)$ for $E \in \mathcal{B}(S^{n-1})$. It is readily verified that σ is a finite measure on $\mathcal{B}(S^{n-1})$, and the measure on S^{n-1} constructed from σ by Method I is the unique Radon measure on S^{n-1}, extending σ on $\mathcal{B}(S^{n-1})$ (this measure is also denoted by σ), and $(S^{n-1}, \Sigma^\sigma, \sigma)$ is the completion of $(S^{n-1}, \mathcal{B}(S^{n-1}), \sigma)$ (cf. Exercise 3.4.18).

From (4.32), we have

$$\lambda^n(E_r) = r^n\lambda^n(E_1) = n\lambda^n(E_1)\int_0^r \rho^{n-1}d\rho = \sigma(E)\int_0^r \rho^{n-1}d\rho$$

for $E \in \mathcal{B}(S^{n-1})$ and hence, by Borel regularity of σ, E_r is measurable and

$$\lambda^n(E_r) = \sigma(E)\int_0^r \rho^{n-1}d\rho \tag{4.33}$$

for any σ-measurable set E in S^{n-1} (see Exercise 4.11.1).

Exercise 4.11.1 Let E be a σ-measurable set in S^{n-1}; show that E_r is measurable and (4.33) holds. (Hint: there are Borel sets F and G in S^{n-1} such that $F \subset E \subset G$ and $\sigma(G\backslash F) = 0$.)

Now let γ be the unique Radon measure on $(0, \infty)$ such that $\gamma(B) = \int_B \rho^{n-1}d\rho$ for Borel sets B in $(0, \infty)$. Since $\gamma(A) = 0$ if and only if $\lambda(A) = 0$ for any $A \subset (0, \infty)$, it follows that γ-measurable sets in $(0, \infty)$ are exactly the Lebesgue measurable sets in $(0, \infty)$.

Lemma 4.11.1 *For σ-measurable sets E in S^{n-1} and measurable sets A in $(0, \infty)$,*

$$\gamma \times \sigma(A \times E) = p_{\#}\lambda^n(A \times E).$$

Proof For a fixed σ-measurable set E in S^{n-1}, let $\mathcal{M}$ be the family of all measurable sets A in $(0,\infty)$ such that for every positive integer n,

$$\gamma \times \sigma(A \cap (0,n] \times E) = p_{\#}\lambda^n(A \cap (0,n] \times E),$$

then, $\gamma \times \sigma(A \times E) = p_{\#}\lambda^n(A \times E)$ for $A \in \mathcal{M}$. Since $p_{\#}\lambda^n((0,r] \times E) = \lambda^n(E_r)$, we infer from (4.33) that $\mathcal{M}$ contains $\Pi = \{(0,r] : r > 0\}$, which is a π-system on $(0,\infty)$. It is routine to verify that $\mathcal{M}$ is a λ-system, and the (π-λ) theorem implies that $\mathcal{M}$ contains all Borel sets in $(0,\infty)$. Now if A is a measurable set in $(0,\infty)$, there are Borel sets C and D in $(0,\infty)$ such that $C \subset A \subset D$ and $\lambda(D \backslash C) = \gamma(D \backslash C) = 0$, hence,

$$\begin{aligned}\gamma \times \sigma(C \times E) &= p_{\#}\lambda^n(C \times E) \le p_{\#}\lambda^n(A \times E) \le p_{\#}\lambda^n(D \times E) \\ &= \gamma \times \sigma(D \times E) = \gamma \times \sigma(C \times E),\end{aligned}$$

from which it follows that $\gamma \times \sigma(A \times E) = p_{\#}\lambda^n(A \times E)$. ■

Lemma 4.11.2 $\mathcal{B}((0,\infty) \times S^{n-1}) \subset \Sigma^{\gamma} \otimes \Sigma^{\sigma} \subset \Sigma^{p_{\#}\lambda^n}$.

Proof Since both $(0,\infty)$ and S^{n-1} are separable as metric space, every open set in $(0,\infty) \times S^{n-1}$ is a countable union of sets of the form $A \times B$, where A is open in $(0,\infty)$ and B is open in S^{n-1}; open sets in $(0,\infty) \times S^{n-1}$ are $\Sigma^{\gamma} \otimes \Sigma^{\sigma}$-measurable and hence $\mathcal{B}((0,\infty) \times S^{n-1}) \subset \Sigma^{\gamma} \otimes \Sigma^{\sigma}$. To show that $\Sigma^{\gamma} \otimes \Sigma^{\sigma} \subset \Sigma^{p_{\#}\lambda^n}$, it is sufficient to show that $A \times B \in \Sigma^{p_{\#}\lambda^n}$ if $A \in \Sigma^{\gamma}$ and $B \in \Sigma^{\sigma}$. There are Borel sets C and D in $(0,\infty)$ such that $C \subset A \subset D$ and $\gamma(D \backslash C) = 0$, and there are Borel sets E and F in S^{n-1} such that $E \subset B \subset F$ and $\sigma(F \backslash E) = 0$; then,

$$\gamma \times \sigma(D \times F \backslash C \times E) = 0,$$

and by Lemma 4.11.1,

$$p_{\#}\lambda^n(D \times F \backslash C \times E) = 0,$$

from which we infer that $A \times B = C \times E \cup N$, where $N \subset D \times F \backslash C \times E$ and is therefore a $p_{\#}\lambda^n$-null set. Thus N is $p_{\#}\lambda^n$-measurable and so is $A \times B$, because $C \times E$ is a Borel set in $(0,\infty) \times S^{n-1}$ and is therefore $p_{\#}\lambda$-measurable. ■

Since $\gamma \times \sigma$ is the unique measure on $\Sigma^{\gamma} \otimes \Sigma^{\sigma}$ such that $\gamma \times \sigma(A \times E) = \gamma(A)\sigma(E)$ for measurable set $A \subset (0,\infty)$, and $E \in \Sigma^{\sigma}$ by Proposition 4.8.1, it follows from Lemma 4.11.1 and Lemma 4.11.2 that $p_{\#}\lambda^n = \gamma \times \sigma$ on $\Sigma^{\gamma} \otimes \Sigma^{\sigma}$. Since $\mathcal{B}((0,\infty) \times S^{n-1}) \subset \Sigma^{\gamma} \otimes \Sigma^{\sigma}$, by Lemma 4.11.2, one concludes that the space $((0,\infty) \times S^{n-1}, \Sigma^{p_{\#}\lambda^n}, p_{\#}\lambda^n)$ is the completion of $((0,\infty) \times S^{n-1}, \Sigma^{\gamma} \otimes \Sigma^{\sigma}, \gamma \times \sigma)$, from the fact that $p_{\#}\lambda^n$ is Borel regular (cf. Exercise 3.4.18). That $p_{\#}\lambda^n$ is Borel regular follows from the Borel regularity of λ^n and the fact that $\mathcal{B}(\dot{\mathbb{R}}^n) = p^{-1}\mathcal{B}((0,\infty) \times S^{n-1})$.

Then, on account of (4.31), we infer immediately that if f is a nonnegative measurable function or an integrable function on $\mathbb{R}^n$, then

$$\int_{\mathbb{R}^n} f d\lambda^n = \int_{(0,\infty)\times S^{n-1}} f \circ p^{-1} d\overline{\gamma \times \sigma};$$

consequently, if we put $f(\rho,\theta) = f \circ p^{-1}(\rho,\theta)$, we have from the Fubini theorem the following theorem.

Theorem 4.11.1 (Integral in polar coordinates) *If f is a nonnegative measurable function or an integrable function on $\mathbb{R}^n$, then*

$$\begin{aligned}\int_{\mathbb{R}^n} f d\lambda^n &= \int_0^\infty \left(\int_{S^{n-1}} f(\rho,\theta) d\sigma(\theta)\right) \rho^{n-1} d\rho \\ &= \int_{S^{n-1}} \left(\int_0^\infty \rho^{n-1} f(\rho,\theta) d\rho\right) d\sigma(\theta).\end{aligned}$$

Example 4.11.1 Suppose that $0 \le \alpha < n$ and let $\Gamma_\alpha(x,y) = |x-y|^{-\alpha}$, then for any $r > 0$,

$$\begin{aligned}\int_{B_r(x)} \Gamma_\alpha(x,y) d\lambda^n(y) &= \int_{B_r(0)} \Gamma_\alpha(0,y) d\lambda^n(y) \\ &= \int_0^r \left(\rho^{n-1} \int_{S^{n-1}} \rho^{-\alpha} d\sigma(\theta)\right) d\rho = \frac{\omega_{n-1}}{n-\alpha} r^{n-\alpha},\end{aligned} \tag{4.34}$$

where $\omega_{n-1} = \sigma(S^{n-1})$.

Exercise 4.11.2 Let b_n be the Lebesgue measure of the unit ball in $\mathbb{R}^n$, and let $l_n = \prod_{j=2}^n \int_0^{\frac{\pi}{2}} \cos^j \theta d\theta$ for $n \ge 2$.

(i) Show that $b_n = 2^n l_n$ for $n \ge 2$.

(ii) Show that $b_{2k} = \frac{1}{k!}\pi^k$ and $b_{2k+1} = 2^{2k+1} \frac{k!}{(2k+1)!}\pi^k$.

(Hint: express b_n in terms of b_{n-1} by using the Fubini theorem.)

Exercise 4.11.3

(i) Show that $\int_{\mathbb{R}^n} e^{-|x|^2} dx = \frac{\omega_{n-1}}{2} \int_0^\infty t^{\frac{n}{2}-1} e^{-t} dt$ and $\omega_{n-1} = \frac{n\pi^{\frac{n}{2}}}{\Gamma(\frac{1}{2}n+1)}$, where $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$.

(ii) Compare (i) and Exercise 4.11.2 (ii) to find $\Gamma(\frac{n}{2})$ for $n \in \mathbb{N}$.

In the remaining part of this section, a brief account of integral operators of potential type will be given, with an application to integral representation of C^1 functions.

For $0 < \alpha < n$, let Γ_α be the function on $\mathbb{R}^n \times \mathbb{R}^n$ defined by

$$\Gamma_\alpha(x,\xi) = \frac{1}{|x-\xi|^\alpha}, \quad (x,\xi) \in \mathbb{R}^n \times \mathbb{R}^n.$$

Given a bounded measurable set Ω with positive measure in $\mathbb{R}^n$, we denote by $\widehat{\Omega}$ the smallest closed ball centered at 0 and containing Ω, i.e. $\widehat{\Omega} = C_R(0)$, where $R = \sup_{x\in\Omega} |x|$.

Lemma 4.11.3 *For $u \in L^1(\Omega)$,*

$$\int_\Omega \Gamma_\alpha(x,\xi)|u(\xi)|d\xi < \infty$$

for a.e. x in $\mathbb{R}^n$.

Proof Let R be the radius of the ball $\widehat{\Omega}$, by the Fubini theorem and (4.34),

$$\begin{aligned}\int_{C_{2R}(0)} \left(\int_\Omega \Gamma_\alpha(x,\xi)|u(\xi)|d\xi\right) dx &= \int_\Omega |u(\xi)| \int_{C_{2R}(0)} \Gamma_\alpha(x,\xi)dxd\xi \\ &\le \int_\Omega |u(\xi)| \int_{C_{3R}(\xi)} \Gamma_\alpha(x,\xi)dxd\xi \\ &\le \frac{\omega_{n-1}}{n-\alpha}(3R)^{n-\alpha} \int_\Omega |u(\xi)|d\xi < \infty,\end{aligned}$$

i.e. $\int_\Omega \Gamma_\alpha(x,\xi)|u(\xi)|d\xi$ is an integrable function of x on $C_{2R}(0)$. Hence, $\int_\Omega \Gamma_\alpha(x,\xi)$ $|u(\xi)|d\xi < \infty$ for a.e. x in $C_{2R}(0)$; while if x is outside $C_{2R}(0)$, $\int_\Omega \Gamma_\alpha(x,\xi)|u(\xi)|d\xi$ $\le \int_\Omega |u(\xi)|d\xi < \infty$. ∎

Because of Lemma 4.11.3, for $u \in L^1(\Omega)$, a function $K_\alpha u$ can be defined a.e. on $\mathbb{R}^n$ by

$$(K_\alpha u)(x) = \int_\Omega \Gamma_\alpha(x,\xi)u(\xi)d\xi, \quad x \in \mathbb{R}^n.$$

$K_\alpha u$ is a function measurable by the Fubini theorem; therefore K_α is a linear operator from $L^1(\Omega)$ into the space of measurable functions on $\mathbb{R}^n$. We call K_α an **integral operator of potential type** and Γ_α a **potential kernel**.

Theorem 4.11.2 *Suppose that Ω and D are two bounded measurable sets of positive measure in $\mathbb{R}^n$, then K_α is a bounded linear operator from $L^p(\Omega)$ into $L^p(D)$.*

Proof When $p = 1$ or ∞, the theorem is obvious. We assume that $1 < p < \infty$. Since Ω is bounded, $u \in L^1(\Omega)$ if $u \in L^p(\Omega)$, and hence $(K_\alpha u)(x) = \int_\Omega \Gamma_\alpha(x,\xi)u(\xi)d\xi$ is finite for a.e. x in $\mathbb{R}^n$. Let the radius of the ball $\widehat{\Omega \cup D}$ be R, i.e. $R = \sup_{x\in\Omega\cup D} |x|$, then for $x \in C_R(0)$,

$$
\begin{aligned}
|(K_\alpha u)(x)| &\leq \int_\Omega \Gamma_\alpha(x,\xi)^{\frac{1}{p}} |u(\xi)| \Gamma_\alpha(x,\xi)^{\frac{1}{q}} d\xi \\
&\leq \left(\int_\Omega \Gamma_\alpha(x,\xi) |u(\xi)|^p d\xi \right)^{\frac{1}{p}} \left(\int_\Omega \Gamma_\alpha(x,\xi) d\xi \right)^{\frac{1}{q}} \\
&\leq \left(\int_\Omega \Gamma_\alpha(x,\xi) |u(\xi)|^p d\xi \right)^{\frac{1}{p}} \left(\int_{C_{2R}(x)} \Gamma_\alpha(x,\xi) d\xi \right)^{\frac{1}{q}} \\
&= \left[\frac{\omega_{n-1}(2R)^{n-\alpha}}{n-\alpha} \right]^{\frac{1}{q}} \left(\int_\Omega \Gamma_\alpha(x,\xi) |u(\xi)|^p d\xi \right)^{\frac{1}{p}},
\end{aligned}
$$

where q is the conjugate exponent of p and (4.34) is applied in the last step. Now, denoting $\frac{\omega_{n-1}(2R)^{n-\alpha}}{n-\alpha}$ by M, we have

$$
\begin{aligned}
\|K_\alpha u\|_{p,D}^p &\leq M^{\frac{p}{q}} \int_D \int_\Omega \Gamma_\alpha(x,\xi) |u(\xi)|^p d\xi\, dx \\
&= M^{\frac{p}{q}} \int_\Omega \int_D \Gamma_\alpha(x,\xi) |u(\xi)|^p dx d\xi \\
&\leq M^{\frac{p}{q}} \int_\Omega |u(\xi)|^p \int_{C_{2R}(\xi)} \Gamma_\alpha(x,\xi) dx d\xi \\
&\leq M^{\frac{p}{q}+1} \|u\|_{p,\Omega}^p,
\end{aligned}
$$

where (4.34) is again applied in the last step, and $\|\cdot\|_{p,D}$, $\|\cdot\|_{p,\Omega}$ denote respectively the norms on $L^p(D)$ and $L^p(\Omega)$. Thus $\|K_\alpha\| \leq M = \frac{\omega_{n-1}(2R)^{n-\alpha}}{n-\alpha}$. ■

It is easy to see that, more generally, if b is a bounded measurable function defined on $D \times \Omega$, the function $K_\alpha^b u$ defined for $u \in L^1(\Omega)$ by

$$
(K_\alpha^b u)(x) = \int_\Omega b(x,\xi) \Gamma_\alpha(x,\xi) u(\xi) d\xi
$$

is finite for a.e. x in $\mathbb{R}^n$; furthermore, K_α^b is a bounded linear operator from $L^p(\Omega)$ into $L^p(D)$, $p \geq 1$ with norm $\|K_\alpha^b\| \leq C\frac{\omega_{n-1}(2R)^{n-\alpha}}{n-\alpha}$, where $C = \|b\|_\infty$ and $R = \sup_{x\in\Omega\cup D} |x|$. Of course, we assume as before that Ω and D are bounded measurable sets with positive measure in $\mathbb{R}^n$.

Theorem 4.11.3 *If Ω and D are compact sets in $\mathbb{R}^n$ with positive measure, and b is a continuous function on $D \times \Omega$, then K_α^b maps every bounded measurable function u into a continuous function on D.*

Proof Fix $x \in D$ and for $\delta > 0$, let $h \in \mathbb{R}^n$ be such that $|h| < \delta$ and $x + h \in D$; for such an h,

$$\begin{aligned}
&|(K_\alpha^b u)(x+h) - (K_\alpha^b u)(x)| \\
&= \left| \int_\Omega \{b(x+h,\xi)\Gamma_a(x+h,\xi) - b(x,\xi)\Gamma_\alpha(x,\xi)\}u(\xi)d\xi \right| \\
&\leq \|u\|_\infty \|b\|_\infty \int_{B_{2\delta}(x)} \{\Gamma_\alpha(x+h,\xi) + \Gamma_\alpha(x,\xi)\}d\xi \\
&\quad + \|u\|_\infty \int_{\Omega\setminus B_{2\delta}(x)} |b(x+h,\xi)\Gamma_\alpha(x+h,\xi) - b(x,\xi)\Gamma_\alpha(x,\xi)|d\xi \\
&\leq \|u\|_\infty \|b\|_\infty \frac{\omega_{n-1}}{n-\alpha}\{(3\delta)^{n-\alpha} + (2\delta)^{n-\alpha}\} \\
&\quad + \|u\|_\infty \int_{\Omega\setminus B_{2\delta}(x)} |b(x+h,\xi)\Gamma_\alpha(x+h,\xi) - b(x,\xi)\Gamma_\alpha(x,\xi)|d\xi,
\end{aligned}$$

because by (4.34),

$$\int_{B_{2\delta}(x)} \Gamma_\alpha(x+h,\xi)d\xi \leq \int_{B_{3\delta}(x+h)} \Gamma_\alpha(x+h,\xi)d\xi \leq \frac{\omega_{n-1}(3\delta)^{n-\alpha}}{n-\alpha},$$
$$\int_{B_{2\delta}(x)} \Gamma_\alpha(x,\xi)d\xi \leq \frac{\omega_{n-1}(2\delta)^{n-\alpha}}{n-\alpha}.$$

Now, given $\varepsilon > 0$, choose $\delta > 0$ such that $\|u\|_\infty \|b\|_\infty \{(3\delta)^{n-\alpha} + (2\delta)^{n-\alpha}\} < \frac{\varepsilon}{2}$. Since both $\Gamma_\alpha(x+h,\xi)$ and $\Gamma_\alpha(x,\xi) \leq \delta^{-\alpha}$ for $\xi \in \Omega\setminus B_{2\delta}(x)$, and

$$\begin{aligned}
&|b(x+h,\xi)\Gamma_\alpha(x+h,\xi) - b(x,\xi)\Gamma_\alpha(x,\xi)| \\
&\leq \|b\|_\infty |\Gamma_\alpha(x+h,\xi) - \Gamma_\alpha(x,\xi)| + \Gamma_\alpha(x,\xi)|b(x+h,\xi) - b(x,\xi)|,
\end{aligned}$$

we can then choose $0 < \sigma_0 < \delta$ such that

$$|b(x+h,\xi)\Gamma_\alpha(x+h,\xi) - b(x,\xi)\Gamma_\alpha(x,\xi)| < \{2(\|u\|_\infty \vee 1)\lambda^n(\Omega)\}^{-1}\varepsilon$$

for all $\xi \in \Omega\setminus B_{2\delta}(x)$ whenever $|h| < \sigma_0$ and $x + h \in D$, and consequently $|(K_\alpha^b u)(x+h) - (K_\alpha^b u)(x)| < \varepsilon$ whenever $|h| < \sigma_0$ and $x + h \in D$. Thus, $K_\alpha^b u$ is continuous at $x \in D$. ■

Exercise 4.11.4 Show that if b is a continuous function on $\mathbb{R}^n \times \Omega$, then $K_\alpha^b u$ is continuous on $\mathbb{R}^n$ for $u \in L^\infty(\Omega)$, where Ω is a compact set with positive measure in $\mathbb{R}^n$.

Theorem 4.11.4 (Integral representation of C^1 functions) *Suppose that Ω is a bounded open convex domain in $\mathbb{R}^n$, then there is a bounded map A from $\Omega \times \Omega$ to $\mathbb{R}^n$ which is* *continuous off the diagonal of $\Omega \times \Omega$, such that if u is a C^1 function on Ω with $\nabla u \in L^1(\Omega)$, then*

$$u(x) = \frac{1}{\lambda^n(\Omega)} \int_\Omega u(\xi)d\xi - \int_\Omega A(x,\xi) \cdot \nabla u(\xi)\Gamma_{n-1}(x,\xi)d\xi \tag{4.35}$$

for $x \in \Omega$.

Proof Fix $x \in \Omega$. For $\xi \in \Omega$, let

$$g(t) = u(x + t(\xi - x)), \quad 0 \leq t \leq 1,$$

then, $g'(t) = \nabla u(x + t(\xi - x)) \cdot (\xi - x)$ and

$$u(x) = u(\xi) - \int_0^1 \nabla u(x + t(\xi - x)) \cdot (\xi - x)dt. \tag{4.36}$$

When $0 < t \leq 1$, the map $\xi \mapsto z = x + t(\xi - x)$ is an invertible affine map with Jacobian t^n at all $\xi \in \mathbb{R}^n$; we may use Theorem 4.10.2 to obtain

$$\begin{aligned}\int_\Omega |\nabla u(x + t(\xi - x)) \cdot (\xi - x)|d\xi &= \int_{x+t(\Omega-x)} \left|\nabla u(z) \cdot \frac{z-x}{t}\right| \frac{1}{t^n}dz \\ &= \int_\Omega I_{x+t(\Omega-x)}(z)|\nabla u(z) \cdot (z-x)|t^{-(n+1)}dz;\end{aligned}$$

hence,

$$\begin{aligned}&\int_0^1 \int_\Omega |\nabla u(x + t(z - x)) \cdot (\xi - x)|d\xi\, dt \\ &= \int_\Omega |\nabla u(z) \cdot (z-x)| \int_0^1 I_{x+t(\Omega-x)}(z)t^{-(n+1)}dtdz.\end{aligned}$$

But $I_{x+t(\Omega-x)}(z) = 0$, when $0 < t < \frac{|z-x|}{l(x,z)}$, where $l(x,z)$ is the length of the line segment from x to the boundary of Ω through z, thus,

$$\begin{aligned}&\int_\Omega \int_0^1 |\nabla u(x + t(\xi - x)) \cdot (\xi - x)|dtd\xi \\ &= \frac{1}{n}\int_\Omega |\nabla u(z) \cdot (z-x)| \left(\frac{l(x,z)^n}{|z-x|^n} - 1\right) dz \\ &\leq \frac{1}{n}\int_\Omega |\nabla u(z)|\{l(x,z)^n - |z-x|^n\}\Gamma_{n-1}(x,z)dz;\end{aligned}$$

now, for $0 < \rho < \operatorname{dist}(x, \Omega^c)$, we have

$$\int_{B_\rho(x)} |\nabla u(z)|\{l(x,z)^n - |z-x|^n\}\Gamma_{n-1}(x,z)dz \leq M \int_{B_\rho(x)} \Gamma_{n-1}(x,z)dz < \infty,$$

because $|\nabla u(z)|$ is bounded on $B_\rho(x)$, and consequently

$$\int_\Omega \int_0^1 |\nabla u(x + t(\xi - x)) \cdot (\xi - x)| dt d\xi < \infty.$$

We have shown that $\nabla u(x + t(\xi - x)) \cdot (\xi - x)$ is an integrable function of (ξ, t) on $\Omega \times [0, 1]$ for $x \in \Omega$. Integrate both sides of (4.36) w.r.t. ξ over Ω to obtain (denoting $\lambda^n(\Omega)$ by $|\Omega|$),

$$\begin{aligned} u(x)|\Omega| &= \int_\Omega u(\xi)d\xi - \int_\Omega \int_0^1 \nabla u(x + t(\xi - x)) \cdot (\xi - x) dt d\xi \\ &= \int_\Omega u(\xi)d\xi - \frac{1}{n} \int_\Omega \nabla u(z) \cdot (z - x) \left\{ \frac{l(x, z)^n}{|z - x|^n} - 1 \right\} dz, \end{aligned}$$

by repeating the previous steps with $|\nabla u(x + t(\xi - x)) \cdot (\xi - x)|$ replaced by $\nabla u(x + t(\xi - x)) \cdot (\xi - x)$, as assured by the Fubini theorem. Now let A be the map from $\Omega \times \Omega$ to $\mathbb{R}^n$, defined by

$$A(x, \xi) = \frac{1}{n|\Omega|} \left[\frac{l(x, \xi)^n - |x - \xi|^n}{|\xi - x|} \right] (\xi - x),$$

if $x \neq \xi$ and $A(x, \xi) = 0$; if $x = \xi$, then

$$u(x) = \frac{1}{|\Omega|} \int_\Omega u(\xi)d\xi - \int_\Omega A(x, \xi) \cdot \nabla u(\xi) \Gamma_{n-1}(x, \xi) d\xi$$

for x in Ω. Obviously, A is continuous off the diagonal of $\Omega \times \Omega$ and $|A(x, \xi)| \leq \frac{1}{n} (\operatorname{diam} \Omega)^n |\Omega|^{-1}$, since $l(x, \xi)^n - |x - \xi|^n = l(x, \xi)^n \left(1 - \frac{|x-\xi|^n}{l(x,\xi)^n}\right) \leq l(x, \xi)^n \leq (\operatorname{diam} \Omega)^n$ if $x \neq \xi$. ■

Corollary 4.11.1 *Let $u \in C^1(\mathbb{R}^n)$. Suppose that u and all of its partial derivatives of first order are integrable. Then,*

$$u(x) = \frac{1}{nb_n} \int_{\mathbb{R}^n} \frac{(x - \xi) \cdot \nabla u(\xi)}{|x - \xi|^n} d\xi \tag{4.37}$$

for $x \in \mathbb{R}^n$, where b_n is the Lebesgue measure of the unit ball in $\mathbb{R}^n$.

Proof Observe first that $\frac{(x-\xi)\cdot\nabla u(\xi)}{|x-\xi|^n}$ is integrable on $\mathbb{R}^n$ as a function of ξ; actually for $\rho > 0$, we have

$$\begin{aligned}
&\int_{\mathbb{R}^n} \frac{|(x-\xi)\cdot\nabla u(\xi)|}{|x-\xi|^n} d\xi \\
&\le \int_{B_\rho(x)} \frac{|\nabla u(\xi)|}{|x-\xi|^{n-1}} d\xi + \int_{\mathbb{R}^n\setminus B_\rho(x)} \frac{|\nabla u(\xi)|}{|x-\xi|^{n-1}} d\xi \\
&\le \sup_{\xi\in B_\rho(x)} |\nabla u(\xi)| \int_{B_\rho(x)} \frac{1}{|x-\xi|^{n-1}} d\xi + \frac{1}{\rho^{n-1}} \int_{\mathbb{R}^n} |\nabla u(\xi)| d\xi \\
&< \infty,
\end{aligned}$$

by recalling that $\int_{B_\rho(x)} \frac{1}{|x-\xi|^{n-1}} d\xi = w_{n-1}\rho$.

For $x \in \mathbb{R}^n$ and $R > 0$, apply Theorem 4.11.4 with $\Omega = B_R(x)$, to obtain

$$\begin{aligned}
u(x) &= \frac{1}{R^n b_n} \int_{B_R(x)} u(\xi) d\xi \\
&\quad - \frac{1}{nR^n b_n} \int_{B_R(x)} \frac{R^n - |\xi - x|^n}{|\xi - x|} (\xi - x)\cdot\nabla u(\xi) \Gamma_{n-1}(x,\xi) d\xi \\
&= \frac{1}{R^n b_n} \int_{B_R(x)} u(\xi) d\xi - \frac{1}{nb_n} \int_{B_R(x)} \frac{(\xi - x)\cdot\nabla u(\xi)}{|\xi - x|^n} d\xi \\
&\quad + \frac{1}{nR^n b_n} \int_{B_R(x)} (\xi - x)\cdot\nabla u(\xi) d\xi,
\end{aligned}$$

because $A(x,\xi) = \frac{1}{nR^n b_n}\left[\frac{R^n-|x-\xi|^n}{|\xi-x|}\right](\xi - x)$ in this case. Now, let $R \to \infty$ to conclude that

$$u(x) = -\frac{1}{nb_n} \int_{\mathbb{R}^n} \frac{(\xi - x)\cdot\nabla u(\xi)}{|\xi - x|^n} d\xi = \frac{1}{nb_n} \int_{\mathbb{R}^n} \frac{(x-\xi)\cdot\nabla u(\xi)}{|x-\xi|^n} d\xi,$$

on noting that

$$\left| \frac{1}{R^n b_n} \int_{B_R(x)} u(\xi) d\xi \right| \le \frac{1}{R^n b_n} \int_{\mathbb{R}^n} |u(\xi)| d\xi \to 0$$

and

$$\left| \frac{1}{nR^n b_n} \int_{B_R(x)} (\xi - x)\cdot\nabla u(\xi) d\xi \right| \le \frac{1}{nR^{n-1} b_n} \int_{\mathbb{R}^n} |\nabla u(\xi)| \to 0$$

as $R \to \infty$; while $\int_{B_R(x)} \frac{(\xi-x)\cdot\nabla u(\xi)}{|\xi-x|^n} d\xi \to \int_{\mathbb{R}^n} \frac{(\xi-x)\cdot\nabla u(\xi)}{|\xi-x|^n} d\xi$ as $R \to \infty$ due to the fact that $\frac{(x-\xi)\cdot\nabla u(\xi)}{|x-\xi|^n}$ is integrable on $\mathbb{R}^n$ as a function of ξ. ■

Exercise 4.11.5 Suppose that $u \in C^1(\Omega)$ and $C_r(x) \subset \Omega$. Show that

$$u(x) = \frac{1}{r^n b_n}\left\{\int_{B_r(0)} u(x+\xi)d\xi - \frac{1}{n}\int_{S^{n-1}}\int_0^r (r^n - \rho^n)\frac{\partial u}{\partial \rho}(x+\rho s)d\rho d\sigma(s)\right\}.$$

Example 4.11.2 Let u be a C^1 function on the ball $B_R(x)$ in $\mathbb{R}^n$ such that ∇u is integrable on $B_R(x)$. We establish here the following estimate for the mean of the Lipschitz quotient of u at x:

$$\frac{1}{\lambda^n(B_R(x))}\int_{B_R(x)} \frac{|u(\xi) - u(x)|}{|\xi - x|}d\xi \leq M(\nabla u, x), \tag{4.38}$$

where $M(\nabla u, x) = \sup_{0<r\leq R} \frac{1}{\lambda^n(B_r(x))}\int_{B_r(x)} |\nabla u| d\lambda^n$.

As in the first step of the proof of Theorem 4.11.4, we have

$$\begin{aligned}
\int_{B_R(x)} \frac{|u(\xi) - u(x)|}{|\xi - x|}d\xi &\leq \int_{B_R(x)}\left(\int_0^1 |\nabla u(x + t(\xi - x))|dt\right)d\xi \\
&= \int_0^1\left(\int_{B_R(x)} |\nabla u(x + t(\xi - x))|d\xi\right)dt \\
&= \int_0^1\left(\int_{B_{Rt(x)}} |\nabla u(z)|\frac{1}{t_n}dz\right)dt \\
&= \lambda^n(B_R(x))\int_0^1 \frac{1}{\lambda^n(B_{Rt(x)})}\int_{B_{Rt(x)}} |\nabla u(z)|dzdt \\
&\leq \lambda^n(B_R(x))M(\nabla u, x),
\end{aligned}$$

from which (4.38) follows.