Lemma 6.1.2. *Let d be a positive integer, and let M_d be the set of all d by d matrices with real entries. Then*

(a) $\det(AB) = \det(A)\det(B)$ *holds for all A, B in M_d,*
(b) $\det(A)$ *is nonzero if and only if A is invertible,*
(c) $\det(A)$ *is a polynomial in the components of A, and*
(d) $\det(A^t) = \det(A)$*, where A^t is the transpose of A.*

Proofs of Lemmas 6.1.1 and 6.1.2 can be found in Halmos [53] and Hoffman and Kunze [61].

Recall that if $T: \mathbb{R}^d \to \mathbb{R}^d$ is linear, if A is the matrix of T with respect to some ordered basis of $\mathbb{R}^d$, and if B is the matrix of T with respect to some possibly different ordered basis of $\mathbb{R}^d$, then there is an invertible matrix U such that $A = UBU^{-1}$. It follows that $\det(A) = \det(U)\det(B)\det(U^{-1}) = \det(B)$. Thus $\det(T)$, the *determinant* of the linear operator T, can be defined to be the determinant of a matrix representing T; it does not matter which ordered basis is used to compute the matrix.

Let us prove the following special case of Theorem 6.1.7.

Proposition 6.1.3. *Let $T: \mathbb{R}^d \to \mathbb{R}^d$ be an invertible linear map. Then*

$$\lambda(T(B)) = |\det(T)|\lambda(B)$$

holds for each Borel subset B of $\mathbb{R}^d$.

Proof. The maps T and T^{-1} are continuous (check this) and hence measurable[1] with respect to $\mathscr{B}(\mathbb{R}^d)$ and $\mathscr{B}(\mathbb{R}^d)$; thus $T(B)$ is a Borel set if and only if B is a Borel set.

Since T is invertible, there exist linear maps $T_1, T_2, \ldots, T_n$ such that $T = T_1 \circ T_2 \circ \cdots \circ T_n$ and such that each T_k operates on a vector x in one of the following ways:

(a) one component of x is multiplied by a nonzero number, and the other components are left unchanged;
(b) two components of x are interchanged, and the other components are left unchanged;
(c) for some i and j the component x_i is replaced with $x_i + x_j$, while the other components of x are left unchanged

(see Exercise 1). In view of the relation $\det(T) = \det(T_1)\det(T_2)\ldots\det(T_n)$, it suffices to show that

$$\lambda(T_k(B)) = |\det(T_k)|\lambda(B) \tag{1}$$

holds for each k and each Borel set B.

[1] Since $T^{-1}(U)$ and $T(U)$ are open and hence Borel whenever U is a open subset of $\mathbb{R}^d$, the measurability of T and T^{-1} follows from Proposition 2.6.2.

First suppose that T_k arises through case (a) or case (b) above. Then it is easy to check that (1) holds if B is a cube with edges parallel to the coordinate axes and hence if B is an open set (use Lemma 1.4.2) or an arbitrary Borel set (use the regularity of λ).

Next suppose that T_k arises through case (c). Then there exist indices i and j such that if $x = (x_1, \dots, x_d)$, then the ith component of $T_k(x)$ is $x_i + x_j$, while the other components of $T_k(x)$ agree with the corresponding components of x. Let us view $\mathbb{R}^d$ as the product of $\mathbb{R}$ (corresponding to the ith coordinate in $\mathbb{R}^d$) with $\mathbb{R}^{d-1}$ (corresponding to the remaining coordinates). Let B be a Borel subset of $\mathbb{R}^d$. It is easy to check that for each u in $\mathbb{R}^{d-1}$ the sections at u of B and of $T_k(B)$ are translates of one another and hence have the same Lebesgue measure. Thus it follows from the theory of product measures (in particular, from Theorem 5.1.4, an extension of Example 5.1.5 to $\mathbb{R}^d$, and the remarks at the end of Sect. 5.2) that $\lambda(B) = \lambda(T_k(B))$. Since $\det(T_k) = 1$ holds whenever T_k arises through case (c), the proof is complete. □

We will need the following standard facts about derivatives of vector-valued functions; proofs can be found in a number of advanced calculus or basic analysis texts,[2] and are sketched in Exercises 2, 3, and 4.

Let X and Y be Banach spaces, let U be an open subset of X, and let x_0 belong to U. A function $F: U \to Y$ is *differentiable* at x_0 if there is a continuous linear map $T: X \to Y$ such that

$$\lim_{x \to x_0} \frac{\|F(x) - F(x_0) - T(x - x_0)\|}{\|x - x_0\|} = 0. \tag{2}$$

It is easy to check that given x_0 and F, there is at most one such map T; it is called the *derivative* of F at x_0 and is denoted by $F'(x_0)$. It is also easy to check that if F is differentiable at x_0, then F is continuous at x_0. Furthermore, if $T: X \to Y$ is continuous and linear, then it is differentiable, with derivative T, at each point in X.

The *chain rule* now takes the following form.

Proposition 6.1.4. *Let X, Y, and Z be Banach spaces, and let U and V be open subsets of X and Y. If $x_0 \in U$, if $G: U \to Y$ is differentiable at x_0 and satisfies $G(U) \subseteq V$, and if $F: V \to Z$ is differentiable at $G(x_0)$, then $F \circ G$ is differentiable at x_0, and*

$$(F \circ G)'(x_0) = F'(G(x_0)) \circ G'(x_0).$$

A method for proving Proposition 6.1.4 is suggested in Exercise 5.

Let us now restrict our attention to the special case of the space $\mathbb{R}^d$. It will prove convenient to endow $\mathbb{R}^d$ with the norm $\|\cdot\|_\infty$ defined by

$$\|x\|_\infty = \max(|x_1|, |x_2|, \dots, |x_d|)$$

---

[2]See Bartle [4], Hoffman [60], Loomis and Sternberg [85], Rudin [104], or Thomson, Bruckner, and Bruckner [117].

(here $x_1, x_2, \ldots, x_d$ are the components of the vector x). It is easy to check that the open sets and the continuous functions determined by $\|\cdot\|_\infty$ are the same as those determined by the usual norm $\|\cdot\|_2$ (see Exercise 6). If $\mathbb{R}^d$ is given the norm $\|\cdot\|_\infty$, if $T\colon \mathbb{R}^d \to \mathbb{R}^d$ is linear, and if (a_{ij}) is the matrix of T with respect to the usual ordered basis of $\mathbb{R}^d$, then T is continuous and its norm (see Sect. 3.5) is given by

$$\|T\| = \max_i \sum_{j=1}^{d} |a_{ij}| \tag{3}$$

(see Exercise 7).

Now let U be an open subset of $\mathbb{R}^d$, let F be a function from U to $\mathbb{R}^d$, and let $f_1, \ldots, f_d$ be the components of F; thus $F(x) = (f_1(x), \ldots, f_d(x))$ holds at each x in U. Then F is said to be a C^1 *function* (or to be *of class* C^1) if the partial derivatives $\partial f_i/\partial x_j$, $i, j = 1, \ldots, d$ exist and are continuous at each point in U.

We will need the following facts.

Lemma 6.1.5. *Let U be an open subset of $\mathbb{R}^d$, and let $F\colon U \to \mathbb{R}^d$ be a C^1 function. Then F is differentiable at each point in U, and the matrix of $F'(x)$ (with respect to the usual ordered basis of $\mathbb{R}^d$) is $(\partial f_i(x)/\partial x_j)$.*

Lemma 6.1.6. *Let U be an open subset of $\mathbb{R}^d$, and let $F\colon U \to \mathbb{R}^d$ be differentiable at each point in U. If x_0 and x_1, together with all the points on the line segment connecting them, belong to U and if $\|F'(x)\| \le C$ holds at each point x on this line segment, then*

$$\|F(x_1) - F(x_0)\|_\infty \le C\|x_1 - x_0\|_\infty.$$

See Exercises 8 and 9 for sketches of proofs of these lemmas.

The *Jacobian* J_F of the C^1 function F is defined by $J_F(x) = \det(F'(x))$. In view of Lemmas 6.1.2 and 6.1.5, the Jacobian of such a function is continuous and hence Borel measurable.

We turn to the main result of this section.

Theorem 6.1.7. *Let U and V be open subsets of $\mathbb{R}^d$, and let T be a bijection of U onto V such that T and T^{-1} are both of class C^1. Then each Borel subset B of U satisfies*

$$\lambda(T(B)) = \int_B |J_T(x)|\,\lambda(dx), \tag{4}$$

and each Borel measurable function $f\colon V \to \mathbb{R}$ satisfies

$$\int_V f\,d\lambda = \int_U f(T(x))|J_T(x)|\,\lambda(dx), \tag{5}$$

in the sense that if either of the integrals in (5) exists, then both exist and (5) holds.

Note that, in view of the identity $T^{-1}(T(x)) = x$, the chain rule implies that $(T^{-1})'(T(x)) \circ T'(x) = I$ holds at each x in U. Thus $T'(x)$ is invertible, and so $J_T(x)$ is nonzero, for each such x.

Note also that T and T^{-1} are Borel measurable (since they are continuous); thus a subset B of U is a Borel set if and only if $T(B)$ is Borel.

We need the following two lemmas for the proof of Theorem 6.1.7.

Lemma 6.1.8. *Let U be an open subset of $\mathbb{R}^d$, let $G\colon U \to \mathbb{R}^d$ be a differentiable function, let ε be a positive number, and let C be a cube that is a Borel set, is included in U, has edges parallel to the coordinate axes, and is such that*

$$\|G'(x) - I\| \le \varepsilon$$

holds at each x in C. Then the image $G(C)$ of C under G satisfies

$$\lambda^*(G(C)) \le (1+\varepsilon)^d \lambda(C).$$

Proof. Let x_0 be the center of C and let b be the length of the edges of C. Then each x in C satisfies $\|x - x_0\|_\infty \le b/2$, and so Lemma 6.1.6, applied to the function $x \mapsto G(x) - x$, implies that each x in C satisfies

$$\|(G(x) - x) - (G(x_0) - x_0)\|_\infty \le \varepsilon \|x - x_0\|_\infty$$

and hence satisfies

$$\|G(x) - G(x_0)\|_\infty \le (1+\varepsilon)\|x - x_0\|_\infty \le \frac{1}{2}(1+\varepsilon)b.$$

Thus $G(C)$ is a subset of the closed cube (with edges parallel to the coordinate axes) whose center is at $G(x_0)$ and whose edges are of length $(1+\varepsilon)b$. Since this cube has measure $(1+\varepsilon)^d b^d$, while C has measure b^d, the lemma follows. □

Lemma 6.1.9. *Let U, V, and T be as in the statement of Theorem 6.1.7. Suppose that a is a positive number and that B is a Borel subset of U.*

(a) *If $|J_T(x)| \le a$ holds at each x in B, then $\lambda(T(B)) \le a\lambda(B)$.*
(b) *If $|J_T(x)| \ge a$ holds at each x in B, then $\lambda(T(B)) \ge a\lambda(B)$.*

Proof. First suppose that b is a positive number and that W is an open subset of U such that

(a) $\overline{W}$ is compact and included in U, and
(b) $|J_T(x)| < b$ holds at each x in W.

Let ε be a positive number. Since $\overline{W}$ is compact and T is of class C^1, the functions that take x to the components[3] of $T'(x)$—that is, to the partial derivatives of the

[3]Here we are dealing with the components of the matrices of these operators with respect to the usual ordered basis of $\mathbb{R}^d$.

components of T—are uniformly continuous on W (part (a) of Theorem C.12). A similar argument shows that the components of $(T'(x))^{-1}$ are bounded on W. Thus (see Eq. (3)) we can choose first a positive number M such that

$$\|(T'(x))^{-1}\| \le M \tag{6}$$

holds at each x in W and then a positive number δ such that

$$\|T'(x) - T'(x_0)\| \le \frac{\varepsilon}{M} \tag{7}$$

holds whenever x and x_0 belong to W and satisfy $\|x - x_0\| \le \delta$.

According to Lemma 1.4.2 the set W is the union of a countable family $\{C_i\}$ of disjoint half-open cubes with edges parallel to the coordinate axes. By subdividing these cubes, if necessary, we can assume that each has edges of length at most 2δ. Let C be one of these cubes, let x_0 be its center, and define $G\colon U \to \mathbb{R}^d$ by

$$G = (T'(x_0))^{-1} \circ T.$$

The chain rule implies that for each x in U we have

$$\begin{aligned} G'(x) - I &= (T'(x_0))^{-1} \circ T'(x) - I \\ &= (T'(x_0))^{-1} \circ (T'(x) - T'(x_0)), \end{aligned}$$

and so (6), (7), and Exercise 3.5.1 imply that

$$\begin{aligned} \|G'(x) - I\| &\le \|(T'(x_0))^{-1}\| \cdot \|T'(x) - T'(x_0)\| \\ &\le M \cdot \frac{\varepsilon}{M} = \varepsilon \end{aligned}$$

holds at each x in C. It now follows from Lemma 6.1.8 that $\lambda(G(C)) \le (1+\varepsilon)^d \lambda(C)$. If we use Proposition 6.1.3 and the fact that $T = T'(x_0) \circ G$, we find

$$\begin{aligned} \lambda(T(C)) &= |\det(T'(x_0))| \lambda(G(C)) \\ &\le b(1+\varepsilon)^d \lambda(C). \end{aligned}$$

Since C was an arbitrary one of the cubes C_i, it follows that

$$\begin{aligned} \lambda(T(W)) &= \sum_i \lambda(T(C_i)) \\ &\le \sum_i b(1+\varepsilon)^d \lambda(C_i) = b(1+\varepsilon)^d \lambda(W) \end{aligned}$$

holds for each ε, and hence that $\lambda(T(W)) \le b\lambda(W)$.

Now suppose that W is an arbitrary open subset of U such that $|J_T(x)| < b$ holds at each x in W. We can choose an increasing sequence $\{W_n\}$ of open sets such that $W = \cup_n W_n$ and such that the closure of each W_n is compact and included in U (the details are left to the reader). Then each W_n satisfies $\lambda(T(W_n)) \le b\lambda(W_n)$, and so we have

$$\lambda(T(W)) = \lim_n \lambda(T(W_n)) \le \lim_n b\lambda(W_n) = b\lambda(W). \tag{8}$$

More generally, let B be a Borel subset of U such that $|J_T(x)| \le a$ holds at each x in B. Let b be a number such that $a < b$. If W is an open subset of U that includes B and if W_b is defined by $W_b = \{x \in W : |J_T(x)| < b\}$, then $B \subseteq W_b$ and inequality (8) implies that

$$\lambda(T(B)) \le \lambda(T(W_b)) \le b\lambda(W_b) \le b\lambda(W).$$

Since b can be made arbitrarily close to a and since λ is regular (Proposition 1.4.1), part (a) of the lemma follows.

We will prove part (b) by applying part (a) to the function $T^{-1}: V \to U$. If $|J_T(x)| \ge a$ holds at each x in B, then $|J_{T^{-1}}(y)| \le 1/a$ holds at each y in $T(B)$, and so part (a) of the lemma implies that $\lambda(T^{-1}(T(B))) \le (1/a)\lambda(T(B))$ or, equivalently, that $a\lambda(B) \le \lambda(T(B))$. □

Proof of Theorem 6.1.7. First suppose that B is a Borel subset of U for which $\lambda(B)$ is finite. For each positive integer n define sets $B_{n,k}$, $k = 1, 2, \dots$, by

$$B_{n,k} = \left\{x \in B : \frac{k-1}{n} \le |J_T(x)| < \frac{k}{n}\right\}.$$

It follows from Lemma 6.1.9 that

$$\frac{k-1}{n}\lambda(B_{n,k}) \le \lambda(T(B_{n,k})) \le \frac{k}{n}\lambda(B_{n,k}) \tag{9}$$

and from the definition of $B_{n,k}$ that

$$\frac{k-1}{n}\lambda(B_{n,k}) \le \int_{B_{n,k}} |J_T(x)|\,\lambda(dx) \le \frac{k}{n}\lambda(B_{n,k}). \tag{10}$$

We conclude from (9) and (10) that

$$\left|\lambda(T(B_{n,k})) - \int_{B_{n,k}} |J_T(x)|\,\lambda(dx)\right| \le \frac{k}{n}\lambda(B_{n,k}) - \frac{k-1}{n}\lambda(B_{n,k}) = \frac{1}{n}\lambda(B_{n,k})$$

and, from this, since $B = \cup_k B_{n,k}$, that

$$\left|\lambda(T(B)) - \int_B |J_T(x)|\,\lambda(dx)\right| \le \frac{1}{n}\lambda(B).$$

However, n is arbitrary and $\lambda(B)$ is finite, and so

$$\lambda(T(B)) = \int_B |J_T(x)|\,\lambda(dx).$$

Thus (4) is proved in the case where $\lambda(B)$ is finite.

If B is an arbitrary Borel subset of U, then it is the union of an increasing sequence $\{B_k\}$ of Borel sets of finite measure, and taking limits over k in the relation

$$\lambda(T(B_k)) = \int_{B_k} |J_T(x)|\,\lambda(dx)$$

yields

$$\lambda(T(B)) = \int_{B} |J_T(x)|\,\lambda(dx).$$

This completes the proof of (4).

We turn to the proof of (5). If f is the characteristic function of a Borel subset C of V and if $B = T^{-1}(C)$, then (5) reduces to (4). The linearity of the integral and the monotone convergence theorem now imply that (5) holds for all nonnegative Borel functions. The case of an arbitrary Borel function f reduces to this through the decomposition $f = f^+ - f^-$. □

With more work it is possible to prove somewhat strengthened versions of Theorem 6.1.7 (see, for example, Theorem 8.26 in Rudin [105]). The version given here, however, seems adequate for most purposes.

Example 6.1.10. Let us apply Theorem 6.1.7 to polar coordinates in $\mathbb{R}^2$. Let R be a positive number, let

$$U = \{(r,\theta) : 0 < r < R \text{ and } 0 < \theta < 2\pi\},$$

let

$$V = \{(x,y) : x^2 + y^2 < R^2\},$$

and let V_0 be the set consisting of those points in V that do not lie on the nonnegative x-axis. Define $T\colon U \to \mathbb{R}^2$ by $T(r,\theta) = (r\cos\theta, r\sin\theta)$. Then T, U, and V_0 satisfy the hypotheses of Theorem 6.1.7. Furthermore, $J_T(r,\theta) = r$. Since V and V_0 differ only by a Lebesgue null set, each integrable function $f\colon V \to \mathbb{R}$ satisfies

$$\int_V f\,d\lambda = \int_{V_0} f\,d\lambda = \int_0^{2\pi}\int_0^R f(r\cos\theta, r\sin\theta)\,r\,dr\,d\theta.$$

This is, of course, the standard formula for the evaluation of integrals by means of polar coordinates. □

Exercises

1. Show that if $T\colon \mathbb{R}^d \to \mathbb{R}^d$ is an invertible linear map, then T can be decomposed as specified in the second paragraph of the proof of Proposition 6.1.3. (Hint: Let

A be the matrix of T with respect to the usual basis of $\mathbb{R}^d$, and recall hows one can use Gaussian elimination to find the inverse of A by performing row operations on the d by $2d$ matrix $(A|I)$, consisting of A followed by the d by d identity matrix. How are linear maps satisfying conditions (a), (b), and (c) of the proof of Proposition 6.1.3 related to row operations?)

2. Show that if F is differentiable at x_0, then $F'(x_0)$ is unique. (Hint: Check that if S and T are both derivatives of F at x_0, then

$$\lim_{x \to x_0} \frac{\|(S-T)(x-x_0)\|}{\|x-x_0\|} = 0;$$

conclude that $S = T$.)

3. Show that if F is differentiable at x_0, then F is continuous at x_0. (Hint: Use Eq. (2), together with the continuity of $x \mapsto F'(x_0)(x)$, to verify that $\lim_{x \to x_0} \|F(x) - F(x_0)\| = 0$.)
4. Show that if $T\colon X \to Y$ is a continuous linear map from one Banach space to another, then T is differentiable, with derivative T, at each point in X. (Hint: Simplify the expression $T(x) - T(x_0) - T(x-x_0)$.)
5. Prove the chain rule, Proposition 6.1.4. (Hint: Let $y_0 = G(x_0)$ and define remainders R_{F,y_0} and R_{G,x_0} by

$$F(y) = F(y_0) + F'(y_0)(y-y_0) + R_{F,y_0}(y-y_0)$$

and

$$G(x) = G(x_0) + G'(x_0)(x-x_0) + R_{G,x_0}(x-x_0);$$

then compute $F(G(x))$ in terms of $F(G(x_0))$, $G'(x_0)$, $F'(G(x_0))$, R_{G,x_0}, R_{F,y_0}, and $x - x_0$. Consider the behavior of the remainders as x approaches x_0.)

6. Let $\|\cdot\|_2$ and $\|\cdot\|_\infty$ be the norms on $\mathbb{R}^d$ defined by $\|x\|_2 = (\sum_i x_i^2)^{1/2}$ and $\|x\|_\infty = \max_i |x_i|$.
   (a) Show that each x in $\mathbb{R}^d$ satisfies $\|x\|_\infty \le \|x\|_2 \le \sqrt{d}\,\|x\|_\infty$.
   (b) Use part (a) to show that the open sets determined by $\|\cdot\|_2$ are the same as those determined by $\|\cdot\|_\infty$.
7. Verify Eq. (3). (Hint: Suppose that $x \in \mathbb{R}^d$ and $y = T(x)$, and calculate an upper bound for $|y_i|$ in terms of $\|x\|_\infty$ and the elements of the matrix (a_{ij}). Also note how to construct a vector x that satisfies $\|x\|_\infty = 1$ and $\|T(x)\|_\infty = \max_i \sum_j |a_{ij}|$ by letting x be an appropriate sequence of 1's and -1's.)
8. Prove Lemma 6.1.5. (Hint: First consider the derivatives (as linear operators from $\mathbb{R}^d$ to $\mathbb{R}$) of the components f_i of F. Let x and x_0 belong to U, and define points u_j, for $j = 0, \ldots, d$, by letting the first j components of u_j agree with the corresponding components of x and letting the remaining components of u_j agree with the corresponding components of x_0. If x_0 is fixed and x is sufficiently close to x_0, then each u_i belongs to U. Use the formula $f_i(x) - f_i(x_0) =$

$\sum_{j=1}^{d}(f_i(u_j) - f_i(u_{j-1}))$, together with the mean value theorem (Theorem C.14), to show that there are points $v_1, \ldots, v_d$ such that[4]

$$f_i(x) - f_i(x_0) = \sum_{j=1}^{d} (\partial f_i(v_j)/\partial x_j)(x_j - x_{0,j}) \tag{11}$$

and such that for each j the point v_j lies on the line segment connecting u_{j-1} and u_j. Deduce the differentiability of f_i at x_0 and compute the matrix of $f_i'(x_0)$. Finally, turn to F.)

9. Prove Lemma 6.1.6. (Hint: Let $f_1, \ldots, f_d$ be the components of F. It is enough to show that $|f_i(x_1) - f_i(x_0)| \le C\|x_1 - x_0\|_\infty$ holds for each i. Use the chain rule to compute the derivative of the function $t \mapsto f_i(x_0 + t(x_1 - x_0))$, and then use the mean value theorem (Theorem C.14) and Exercise 3.5.1 to obtain the required bound for $|f_i(x_1) - f_i(x_0)|$.)

6.2 Differentiation of Measures

Let $\mathscr{C}$ be the family consisting of those nondegenerate closed cubes in $\mathbb{R}^d$ whose edges are parallel to the coordinate axes. In other words, let $\mathscr{C}$ be the collection of all sets of the form

$$[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_d, b_d],$$

where $[a_1, b_1], \ldots, [a_d, b_d]$ are closed subintervals of $\mathbb{R}$ that have a common nonzero length. For each cube C in $\mathscr{C}$ let $e(C)$ be the length of the edges of C.

Suppose that A is a subset of $\mathbb{R}^d$. A *Vitali covering* of A is a subfamily $\mathscr{V}$ of $\mathscr{C}$ such that for each x in A and each positive number δ there is a cube C that belongs to $\mathscr{V}$, contains x, and satisfies $e(C) < \delta$.

The following fact about Vitali coverings forms the basis for our treatment of differentiation theory. The reader should note, however, that differentiation theory can also be based on the "rising sun lemma" of F. Riesz; see, for example, Chapter I of Riesz and Nagy [99].

Theorem 6.2.1 (Vitali Covering Theorem). *Let A be an arbitrary nonempty subset of $\mathbb{R}^d$, and let $\mathscr{V}$ be a Vitali covering of A. Then there is a finite or infinite sequence $\{C_n\}$ of disjoint sets that belong to $\mathscr{V}$ and are such that $\cup_n C_n$ contains λ-almost every point in A.*

Proof. First consider the case where the set A is bounded. Choose a bounded open subset U of $\mathbb{R}^d$ that includes A, and let $\mathscr{V}_0$ consist of those cubes in $\mathscr{V}$ that are included in U. It is clear that $\mathscr{V}_0$ is a Vitali covering of A. Let

$$\delta_1 = \sup\{e(C) : C \in \mathscr{V}_0\}.$$

[4] The symbols x_j and $x_{0,j}$ in (11) refer to the jth components of the vectors x and x_0.

Then δ_1 satisfies $0 < \delta_1 < +\infty$ (recall that A is nonempty and U is bounded), and we can choose a cube C_1 that belongs to $\mathscr{V}_0$ and satisfies $e(C_1) > \delta_1/2$. We continue this construction inductively, producing sequences $\{\delta_n\}$ and $\{C_n\}$ as follows. If $A \subseteq \cup_{k=1}^n C_k$, then the construction is complete. Otherwise there are points in A that lie outside $\cup_{k=1}^n C_k$, and so, since $\cup_{k=1}^n C_k$ is closed and $\mathscr{V}_0$ is a Vitali covering of A, there are cubes in $\mathscr{V}_0$ that are disjoint from $\cup_{k=1}^n C_k$. Thus the quantity δ_{n+1} defined by

$$\delta_{n+1} = \sup\{e(C) : C \in \mathscr{V}_0 \text{ and } C \cap (\cup_{k=1}^n C_k) = \varnothing\}$$

satisfies $0 < \delta_{n+1} < +\infty$, and we can choose a cube C_{n+1} in $\mathscr{V}_0$ that satisfies $e(C_{n+1}) > \delta_{n+1}/2$ and is disjoint from $\cup_{k=1}^n C_k$. This completes the induction step in the construction of the sequences $\{\delta_n\}$ and $\{C_n\}$.

If this construction terminates in a finite number, say N, of steps, then $A \subseteq \cup_{n=1}^N C_n$ and $\{C_n\}_{n=1}^N$ is the required sequence. We turn to the case in which the construction does not terminate.

Since the sets C_n are disjoint and included in the bounded set U, the series $\sum_n \lambda(C_n)$ must be convergent; thus $\lim_n \lambda(C_n) = 0$ and hence $\lim_n \delta_n = 0$. For each n let D_n be the cube in $\mathscr{C}$ with the same center as C_n but with edges 5 times as long as those of C_n. Then, since $\lambda(D_n) = 5^d \lambda(C_n)$, the series $\sum_n \lambda(D_n)$ is also convergent. We will show that

$$A - \cup_{n=1}^N C_n \subseteq \cup_{n=N+1}^\infty D_n \tag{1}$$

holds for each positive integer N. This inclusion implies that

$$\lambda^*(A - \cup_{n=1}^\infty C_n) \le \lambda^*(A - \cup_{n=1}^N C_n) \le \sum_{n=N+1}^\infty \lambda(D_n);$$

since the convergence of $\sum_{n=1}^\infty \lambda(D_n)$ implies that $\lim_N \sum_{n=N+1}^\infty \lambda(D_n) = 0$, it follows that $\lambda^*(A - \cup_{n=1}^\infty C_n) = 0$ and hence that $\{C_n\}$ is the required sequence.

We turn to the proof of (1). Suppose that x belongs to $A - \cup_{n=1}^N C_n$. Since $\cup_{n=1}^N C_n$ is closed and $\mathscr{V}_0$ is a Vitali covering of A, there are cubes in $\mathscr{V}_0$ that contain x and are disjoint from $\cup_{n=1}^N C_n$. Let C be such a cube. Then C meets $\cup_{n=1}^k C_n$ for some k, since otherwise we would have $e(C) \le \delta_k$ for all k, contradicting $\lim_n \delta_n = 0$. Let k_0 be the smallest of those positive integers k for which C meets $\cup_{n=1}^k C_n$. Then $e(C) \le \delta_{k_0}$ and $\delta_{k_0}/2 \le e(C_{k_0})$, and it follows that $e(C) \le 2e(C_{k_0})$. The definition of the sets D_n, the inequality $e(C) \le 2e(C_{k_0})$, and the fact that $C \cap C_{k_0} \neq \varnothing$ together imply that $C \subseteq D_{k_0}$. Since C was chosen to be disjoint from $\cup_{n=1}^N C_n$, it follows that $k_0 \ge N+1$, and so

$$x \in C \subseteq D_{k_0} \subseteq \bigcup_{n=N+1}^\infty D_n.$$

Relation (1) follows, since x was an arbitrary element of $A - \cup_{n=1}^N C_n$. This completes the proof of the theorem in the case where A is bounded.

Now suppose that the set A is unbounded. Let $U_1, U_2, \ldots$ be disjoint bounded open subsets of $\mathbb{R}^d$ such that $\lambda(\mathbb{R}^d - (\cup_{k=1}^\infty U_k)) = 0$; for example, the open cubes whose edges have length 1 and whose corners have integer coordinates will do. For each k such that $A \cap U_k \neq \varnothing$ we can use the preceding argument to choose a sequence $\{C_{k,j}\}_j$ of disjoint cubes that belong to $\mathscr{V}$ and are such that $\cup_j C_{k,j}$ is included in U_k and contains almost every point in $A \cap U_k$. Merging the resulting sequences into one sequence completes the proof. □

Let μ be a finite Borel measure on $\mathbb{R}^d$. Then $(\overline{D}\mu)(x)$, the *upper derivate of* μ *at* x, is defined by

$$(\overline{D}\mu)(x) = \lim_{\varepsilon \downarrow 0} \sup \left\{ \frac{\mu(C)}{\lambda(C)} : C \in \mathscr{C}, x \in C, \text{ and } e(C) < \varepsilon \right\}, \tag{2}$$

and $(\underline{D}\mu)(x)$, the *lower derivate of* μ *at* x, is defined by

$$(\underline{D}\mu)(x) = \lim_{\varepsilon \downarrow 0} \inf \left\{ \frac{\mu(C)}{\lambda(C)} : C \in \mathscr{C}, x \in C, \text{ and } e(C) < \varepsilon \right\}. \tag{3}$$

The *upper derivate* and the *lower derivate* of μ are the $[0, +\infty]$-valued functions $\overline{D}\mu$ and $\underline{D}\mu$ whose values at x are given by (2) and (3). The measure μ is *differentiable* at x if $(\overline{D}\mu)(x)$ and $(\underline{D}\mu)(x)$ are finite and equal, and at each such x the *derivative* $(D\mu)(x)$ *of* μ *at* x is defined by

$$(D\mu)(x) = (\overline{D}\mu)(x) = (\underline{D}\mu)(x). \tag{4}$$

The *derivative* of μ is the function $D\mu$ that is defined by (4) at each x at which μ is differentiable and is undefined elsewhere.

Lemma 6.2.2. *Let μ be a finite Borel measure on $\mathbb{R}^d$. Then $\overline{D}\mu$, $\underline{D}\mu$, and $D\mu$ are Borel measurable.*

Proof. Let $\mathscr{U}$ be the collection of all open cubes in $\mathbb{R}^d$ whose edges are parallel to the coordinate axes, and for each U in $\mathscr{U}$ let $e(U)$ be the length of the edges of U. Note that for each cube C in $\mathscr{C}$ there is a decreasing sequence $\{U_n\}$ of cubes in $\mathscr{U}$ for which $C = \cap_n U_n$ and hence (Proposition 1.2.5) for which $\mu(C)/\lambda(C) = \lim_n \mu(U_n)/\lambda(U_n)$. Likewise for each cube U in $\mathscr{U}$ there is an increasing sequence $\{C_n\}$ of cubes in $\mathscr{C}$ for which $U = \cup_n C_n$ and hence for which $\mu(U)/\lambda(U) = \lim_n \mu(C_n)/\lambda(C_n)$. It follows that $(\overline{D}\mu)(x)$ is given by

$$(\overline{D}\mu)(x) = \lim_{\varepsilon \downarrow 0} \sup \left\{ \frac{\mu(U)}{\lambda(U)} : U \in \mathscr{U}, x \in U, \text{ and } e(U) < \varepsilon \right\}.$$

For each positive ε let us define a function $s_\varepsilon : \mathbb{R}^d \to [0, \infty]$ whose value at x is the supremum considered above:

$$s_\varepsilon(x) = \sup \left\{ \frac{\mu(U)}{\lambda(U)} : U \in \mathscr{U}, x \in U, \text{ and } e(U) < \varepsilon \right\}.$$

Then for each a in $\mathbb{R}$ we have

$$\{x \in \mathbb{R}^d : s_\varepsilon(x) > a\} = \bigcup \left\{ U \in \mathscr{U} : e(U) < \varepsilon \text{ and } \frac{\mu(U)}{\lambda(U)} > a \right\},$$

and so s_ε is Borel measurable. If $\{\varepsilon_n\}$ is a sequence of numbers that decreases to 0, then $\overline{D}\mu$ is the pointwise limit of the sequence of functions $\{s_{\varepsilon_n}\}$ and so is Borel measurable. The measurability of $\underline{D}\mu$ can be proved in a similar way.

The measurability of $D\mu$ is a consequence of Proposition 2.1.3 and the measurability of $\overline{D}\mu$ and $\underline{D}\mu$. □

The following theorem is the main result of this section.

Theorem 6.2.3. *Let μ be a finite Borel measure on $\mathbb{R}^d$. Then μ is differentiable at λ-almost every point in $\mathbb{R}^d$, and the function defined by*

$$x \mapsto \begin{cases} (D\mu)(x) & \text{if } \mu \text{ is differentiable at } x, \\ 0 & \text{otherwise} \end{cases}$$

is a Radon–Nikodym derivative of the absolutely continuous part of μ.

We will need the following two lemmas for the proof of Theorem 6.2.3.

Lemma 6.2.4. *Let μ be a finite Borel measure on $\mathbb{R}^d$, let a be a positive real number, and let A be a Borel subset of $\mathbb{R}^d$ such that $(\overline{D}\mu)(x) \ge a$ holds at each x in A. Then $\mu(A) \ge a\lambda(A)$.*

Proof. We can certainly assume that A is nonempty. Let U be an open set that includes A, let ε satisfy $0 < \varepsilon < a$, and let $\mathscr{V}$ be the family consisting of those cubes C in $\mathscr{C}$ that are included in U and satisfy $\mu(C) \ge (a-\varepsilon)\lambda(C)$. Since $(\overline{D}\mu)(x) \ge a$ holds at each x in A, the family $\mathscr{V}$ is a Vitali covering of A. Thus the Vitali covering theorem (Theorem 6.2.1) provides a sequence $\{C_n\}$ of disjoint cubes that belong to $\mathscr{V}$ and satisfy $\lambda(A - \cup_n C_n) = 0$. If we use the fact that the sets C_n are disjoint and included in U, the fact that each C_n satisfies $\mu(C_n) \ge (a-\varepsilon)\lambda(C_n)$, and finally the fact that $\lambda(A - \cup_n C_n) = 0$, we find

$$\mu(U) \ge \sum_n \mu(C_n) \ge \sum_n (a-\varepsilon)\lambda(C_n)$$
$$= (a-\varepsilon)\lambda\left(\bigcup_n C_n\right) \ge (a-\varepsilon)\lambda(A).$$

Since μ is regular (Proposition 1.5.6) and ε can be made arbitrarily close to 0, the inequality $\mu(A) \ge a\lambda(A)$ follows. □

Lemma 6.2.5. *Let μ be a finite Borel measure on $\mathbb{R}^d$ that is absolutely continuous with respect to Lebesgue measure, let a be a positive real number, and let A be a Borel subset of $\mathbb{R}^d$ such that $(\underline{D}\mu)(x) \le a$ holds at each x in A. Then $\mu(A) \le a\lambda(A)$.*

Proof. We can again assume that A is not empty. Let U be an open set that includes A, and let ε be a positive number. Arguments similar to those used in the proof of the preceding lemma show that there is a sequence $\{C_n\}$ of disjoint closed cubes that are included in U, satisfy $\mu(C_n) \le (a+\varepsilon)\lambda(C_n)$, and are such that $\cup_n C_n$ contains λ-almost every point in A. Since μ is absolutely continuous with respect to λ, the union of the sets C_n also contains μ-almost every point in A. It follows that

$$(a+\varepsilon)\lambda(U) \ge (a+\varepsilon)\sum_n \lambda(C_n) \ge \sum_n \mu(C_n) = \mu(\cup_n C_n) \ge \mu(A).$$

Since λ is regular (Proposition 1.4.1) and ε is arbitrary, it follows that $\mu(A) \le a\lambda(A)$. □

Proof of Theorem 6.2.3. We begin with the case where μ is singular with respect to Lebesgue measure. Let N be a Borel set such that $\lambda(N) = 0$ and $\mu(N^c) = 0$. For each n define a subset B_n of N^c by

$$B_n = \{x \in N^c : (\overline{D}\mu)(x) \ge 1/n\}.$$

Then Lemma 6.2.4 (with a equal to $1/n$) implies that

$$\lambda(B_n) \le n\mu(B_n) \le n\mu(N^c) = 0$$

holds for each n. Thus $\{x \in \mathbb{R}^d : (\overline{D}\mu)(x) > 0\}$, since it is a subset of $N \cup (\cup_n B_n)$, has Lebesgue measure 0; since also $0 \le \underline{D}\mu \le \overline{D}\mu$, it follows that μ is differentiable, with derivative 0, λ-almost everywhere.

Next let us consider the case where μ is absolutely continuous with respect to Lebesgue measure. We start by proving that in this case $\underline{D}\mu$ and $\overline{D}\mu$ are equal almost everywhere. For positive rational numbers p and q such that $p < q$, define $A(p,q)$ by

$$A(p,q) = \{x \in \mathbb{R}^d : (\underline{D}\mu)(x) \le p < q \le (\overline{D}\mu)(x)\}.$$

Lemmas 6.2.4 and 6.2.5 imply that

$$q\lambda(A(p,q)) \le \mu(A(p,q)) \le p\lambda(A(p,q));$$

it follows from this first that $\lambda(A(p,q))$ is finite and then, since $p < q$, that $\lambda(A(p,q)) = 0$. Since $(\underline{D}\mu)(x) \le (\overline{D}\mu)(x)$ holds for every x, while

$$\{x \in \mathbb{R}^d : (\underline{D}\mu)(x) < (\overline{D}\mu)(x)\} = \bigcup_{p,q} A(p,q),$$

it follows that $\underline{D}\mu$ and $\overline{D}\mu$ are equal λ-almost everywhere. (Note that we have not yet shown that they are finite almost everywhere.)

We continue to assume that $\mu \ll \lambda$. Let f be a Radon–Nikodym derivative of μ with respect to λ. An easy modification of the argument in the preceding paragraph shows that $f \le \underline{D}\mu$ holds λ-almost everywhere (use the fact that whenever a is a positive number and A is a Borel set such that $f(x) \ge a$ holds at each x in A, then $\mu(A) = \int_A f\,d\lambda \ge a\lambda(A)$). A similar argument shows that $f \ge \overline{D}\mu$ holds λ-a .e.

Since in addition f is finite almost everywhere, it follows that f, $\underline{D}\mu$, and $\overline{D}\mu$ are finite and equal almost everywhere and hence that μ is differentiable, with derivative f, almost everywhere.

Finally, let μ be an arbitrary finite Borel measure on $\mathbb{R}^d$, let $\mu = \mu_a + \mu_s$ be its Lebesgue decomposition, and let f be a Radon–Nikodym derivative of μ_a with respect to λ. Then

$$(D\mu)(x) = (D\mu_a)(x) + (D\mu_s)(x) = f(x) + 0 = f(x)$$

holds at almost every x, and the proof is complete. □

Let E be a Lebesgue measurable subset of $\mathbb{R}^d$. A point x in $\mathbb{R}^d$ is a *point of density* of E if for each positive ε there is a positive δ such that

$$\left| \frac{\lambda(E \cap C)}{\lambda(C)} - 1 \right| < \varepsilon$$

holds whenever C is a cube that belongs to $\mathscr{C}$, contains x, and satisfies $e(C) < \delta$. Less formally, x is a point of density of E if $\lim \lambda(E \cap C)/\lambda(C) = 1$, where the limit is taken as C approaches x (through the collection of cubes in $\mathscr{C}$ that contain x). A point x is a *point of dispersion* of E if it is a point of density of E^c. Equivalently, x is a point of dispersion of E if $\lim \lambda(E \cap C)/\lambda(C) = 0$ holds as the cube C approaches x.

Corollary 6.2.6 (Lebesgue Density Theorem). *Let E be a Lebesgue measurable subset of $\mathbb{R}^d$. Then λ-almost every point in E is a point of density of E, and λ-almost every point in E^c is a point of dispersion of E.*

Proof. First suppose that $\lambda(E) < +\infty$, and define a finite Borel measure μ on $\mathbb{R}^d$ by $\mu(A) = \lambda(A \cap E)$. Choose a Borel subset E_0 of E such that $\lambda(E - E_0) = 0$ (see Lemma 1.5.3). Since $\mu \ll \lambda$ and since χ_{E_0} is a Radon–Nikodym derivative of μ with respect to λ, Theorem 6.2.3 implies that almost every x in E satisfies $(D\mu)(x) = 1$ and so is a point of density of E.

If $\lambda(E)$ is infinite and if $\{E_n\}$ is a sequence of Lebesgue measurable sets of finite measure such that $E = \cup_n E_n$, then almost every point of E is a point of density of some E_n and so is a point of density of E. Finally, almost every point of E^c is a point of density of E^c and so is a point of dispersion of E. □

Exercises

1. Show that the union of an arbitrary family of closed cubes with edges parallel to the coordinate axes is Lebesgue measurable. (Hint: Use the Vitali covering theorem.)
2. Let I be the line segment in $\mathbb{R}^2$ that connects the points $(0,0)$ and $(1,1)$. Define a finite Borel measure μ on $\mathbb{R}^2$ by letting $\mu(A)$ be the one-dimensional Lebesgue

measure of $A \cap I$. (More precisely, let T be the map of the interval $[0, \sqrt{2}]$ onto I given by $T(t) = (t/\sqrt{2})(1,1)$, and define μ by $\mu(A) = \lambda(T^{-1}(A))$.) Find $\overline{D}\mu$ and $\underline{D}\mu$.

3. Let f be a nonnegative function in $\mathscr{L}^1(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d), \lambda, \mathbb{R})$, and let μ be the finite Borel measure on $\mathbb{R}^d$ given by $\mu(A) = \int_A f\, d\lambda$.
   (a) Show that $(D\mu)(x) = f(x)$ holds at each x at which f is continuous.
   (b) Show by example that the equation $(D\mu)(x) = f(x)$ need not hold at every x in $\mathbb{R}^d$.
4. Show by example that the assumption that $\mu \ll \lambda$ cannot be omitted in Lemma 6.2.5.

6.3 Differentiation of Functions

Let us apply the results of Sect. 6.2 to the differentiation of functions of a real variable. We begin with two lemmas.

Lemma 6.3.1. *Let μ be a finite Borel measure on $\mathbb{R}$, and let $F\colon \mathbb{R} \to \mathbb{R}$ be defined by $F(x) = \mu((-\infty, x])$. If μ is differentiable at x_0, then F is differentiable at x_0, and $F'(x_0) = (D\mu)(x_0)$.*

Proof. The differentiability of μ at x_0 implies that $\mu(\{x_0\}) = 0$ and hence that F is continuous at x_0. Thus $\frac{F(x)-F(x_0)}{x-x_0}$ is equal to $\frac{\mu([x_0,x])}{\lambda([x_0,x])}$ if $x_0 < x$ and to $\frac{\mu((x,x_0])}{\lambda((x,x_0])}$ if $x < x_0$. Now apply the definitions of $(D\mu)(x_0)$ and $F'(x_0)$ (note that the half-open interval $(x, x_0]$ causes no difficulty, since its measure is the limit of the measure of $[x + \frac{1}{n}, x_0]$ as n approaches infinity). □

Lemma 6.3.2. *Let $F\colon \mathbb{R} \to \mathbb{R}$ be nondecreasing. Then*

(a) *the one-sided limits $F(x-)$ and $F(x+)$ exist at each x in $\mathbb{R}$,*
(b) *the set of points at which F fails to be continuous is at most countably infinite, and*
(c) *the function $G\colon \mathbb{R} \to \mathbb{R}$ defined by $G(x) = F(x+)$ is nondecreasing and right-continuous, and agrees with F at each point at which F is continuous.*

Proof. Since F is nondecreasing, the limits $F(x-)$ and $F(x+)$ exist and are given by $F(x-) = \sup\{F(t) : t < x\}$ and $F(x+) = \inf\{F(t) : t > x\}$. For each x we have $F(x-) \le F(x) \le F(x+)$, and so F is continuous at x if and only if $F(x-) = F(x+)$. Let D be the set of points at which F is not continuous, and for each x in D choose a rational number r_x that satisfies $F(x-) < r_x < F(x+)$. Then r_x and r_y are distinct whenever x and y are distinct elements of D, and the countability of D follows from the countability of $\mathbb{Q}$.

Now suppose that G is defined by $G(x) = F(x+)$. Then G satisfies the relation $G(x) = \inf\{F(t) : t > x\}$, which implies that G is nondecreasing and right-continuous. Since $F(x) = F(x+)$ holds if F is continuous at x, the proof is complete. □

The following is one of the basic theorems of differentiation theory.

Theorem 6.3.3 (Lebesgue). *Let $F: \mathbb{R} \to \mathbb{R}$ be nondecreasing. Then F is differentiable λ-almost everywhere.*

Proof. First suppose that F is bounded, nondecreasing, and right-continuous, and that it vanishes at $-\infty$. Then there is a finite Borel measure μ such that $F(x) = \mu((-\infty,x])$ holds at each x in $\mathbb{R}$ (Proposition 1.3.10), and so Theorem 6.2.3 and Lemma 6.3.1 imply that F is differentiable almost everywhere.

Now remove the requirement that F be right-continuous, and define $G: \mathbb{R} \to \mathbb{R}$ by $G(x) = F(x+)$. Then G is right-continuous (Lemma 6.3.2) and so, by what we have just proved, differentiable almost everywhere. Note that F and G are continuous at the same points and they agree at each point at which they are continuous; furthermore, if $F(x_0) = G(x_0)$, then $\frac{F(x)-F(x_0)}{x-x_0}$ lies between $\frac{G(x)-G(x_0)}{x-x_0}$ and $\frac{G(x-)-G(x_0)}{x-x_0}$. Hence if G is differentiable at x_0, then F is differentiable at x_0, and $F'(x_0) = G'(x_0)$. The almost everywhere differentiability of F follows.

Finally, suppose that F is an arbitrary nondecreasing function. It is enough to prove that F is differentiable almost everywhere on an arbitrary bounded open interval (a,b). Since we can reduce this to the preceding case by considering the function

$$x \mapsto \begin{cases} 0 & \text{if } x \le a, \\ F(x) - F(a) & \text{if } a < x < b, \\ F(b) - F(a) & \text{if } b \le x, \end{cases}$$

the proof is complete. □

Corollary 6.3.4. *Let $F: \mathbb{R} \to \mathbb{R}$ be of finite variation. Then F is differentiable λ-almost everywhere.*

Proof. Since each function of finite variation is the difference of two nondecreasing functions (Proposition 4.4.2), this is an immediate consequence of Theorem 6.3.3. □

Proposition 6.3.5 (Fubini). *Let $F_n: \mathbb{R} \to \mathbb{R}$, $n = 1, 2, \ldots$, be nondecreasing functions such that the series $\sum_n F_n(x)$ converges at each x in $\mathbb{R}$. Define $F: \mathbb{R} \to \mathbb{R}$ by $F(x) = \sum_n F_n(x)$. Then $F' = \sum_n F_n'$ holds λ-almost everywhere.*

Proof. First suppose that the functions F_n, for $n = 1, 2, \ldots$, are bounded, nondecreasing, and right-continuous, that they vanish at $-\infty$, and that the function F (defined by $F(x) = \sum_n F_n(x)$) is bounded. Let μ_n, $n = 1, 2, \ldots$, be the finite Borel measures corresponding to the functions F_n, and define a Borel measure μ on $\mathbb{R}$ by $\mu(A) = \sum_n \mu_n(A)$ (check that μ is a measure). Since we are temporarily assuming that F is bounded and since $\mu((-\infty,x]) = F(x)$ holds at each x, the measure μ is finite.

For each n form the Lebesgue decomposition $\mu_n = \mu_{n,a} + \mu_{n,s}$ of μ_n with respect to Lebesgue measure,[5] let f_n be a Radon–Nikodym derivative of $\mu_{n,a}$ with respect to λ, and let N_n be a Borel set of Lebesgue measure zero on which $\mu_{n,s}$ is concentrated. It is easy to check that $\sum_n \mu_{n,s}$ is concentrated on $\cup_n N_n$ and that $\sum_n \mu_{n,a}(A) = \int_A \sum_n f_n \, d\lambda$ holds for each A in $\mathscr{B}(\mathbb{R})$. Thus the Lebesgue decomposition of μ is given by $\mu = (\sum_n \mu_{n,a}) + (\sum_n \mu_{n,s})$, and $\sum_n f_n$ is a Radon–Nikodym derivative of $\sum_n \mu_{n,a}$ with respect to λ. It now follows from Theorem 6.2.3 and Lemma 6.3.1 that

$$\sum_n F_n'(x) = \sum_n (D\mu_n)(x) = \sum_n f_n(x) = (D\mu)(x) = F'(x)$$

holds at almost every x in $\mathbb{R}$.

Arguments similar to those given in the second and third paragraphs of the proof of Theorem 6.3.3 allow one to reduce the proposition to the case just considered; the details are left to the reader. □

Theorem 6.3.6 (Lebesgue). *Suppose that f belongs to $\mathscr{L}^1(\mathbb{R}, \mathscr{M}_{\lambda^*}, \lambda, \mathbb{R})$ and that $F\colon \mathbb{R} \to \mathbb{R}$ is defined by $F(x) = \int_{-\infty}^{x} f(t)\, dt$. Then F is differentiable, and its derivative is given by $F'(x) = f(x)$, at λ-almost every x in $\mathbb{R}$.*

Proof. First suppose that f is nonnegative, and define a finite Borel measure μ on $\mathbb{R}$ by $\mu(A) = \int_A f \, d\lambda$. Let f_0 be a Borel measurable function that agrees with f almost everywhere (see Proposition 2.2.5). Then Theorem 6.2.3 and Lemma 6.3.1 imply that

$$F'(x) = (D\mu)(x) = f_0(x) = f(x)$$

holds at almost every x, and so the proof is complete in the case where f is nonnegative.

An arbitrary f in $\mathscr{L}^1(\mathbb{R}, \mathscr{M}_{\lambda^*}, \lambda, \mathbb{R})$ can be dealt with through the decomposition $f = f^+ - f^-$. □

We will often need to know that almost everywhere derivatives of reasonable functions are measurable. This is given by the following lemma.

Lemma 6.3.7. *Let $F\colon [a,b] \to \mathbb{R}$ be a Lebesgue measurable function that is differentiable almost everywhere. Suppose that $g\colon [a,b] \to \mathbb{R}$ satisfies $g(x) = F'(x)$ almost everywhere. Then g is Lebesgue measurable, as is F' (whose domain is the set where F is differentiable).*

Proof. Extend F to the interval $[a, +\infty)$ by letting $F(x)$ be equal to $F(b)$ if $x > b$. Since

$$g(x) = \overline{\lim_n}\, n\Big(F\big(x + \frac{1}{n}\big) - F(x)\Big)$$

[5] Thus $\mu_{n,a} \ll \lambda$ and $\mu_{n,s} \perp \lambda$.

holds at almost every x in $[a,b]$, it follows from Propositions 2.1.5 and 2.2.2 that g is Lebesgue measurable. Since the set of points where F is differentiable is the complement in $[a,b]$ of a Lebesgue null set, it follows that F' is also Lebesgue measurable. □

We can now derive the following characterization of the absolutely continuous[6] functions on a closed bounded interval.

Corollary 6.3.8. *A function $F: [a,b] \to \mathbb{R}$ is absolutely continuous if and only if it is differentiable λ-almost everywhere, F' is integrable, and F can be reconstructed from its derivative through the formula*

$$F(x) = F(a) + \int_a^x F'(t)\,dt. \tag{1}$$

Proof. First suppose that F is absolutely continuous. Then F is also of finite variation (Exercise 4.4.5), and so Proposition 4.4.6, applied to the function

$$x \mapsto \begin{cases} 0 & \text{if } x \le a, \\ F(x) - F(a) & \text{if } a < x < b, \\ F(b) - F(a) & \text{if } b \le x, \end{cases}$$

provides a function f in $\mathscr{L}^1(\mathbb{R}, \mathscr{B}(\mathbb{R}), \lambda, \mathbb{R})$ such that

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

holds at each x in $[a,b]$. Theorem 6.3.6 then implies that F is differentiable, with derivative given by $F'(x) = f(x)$, at almost every such x; hence (1) follows.

The other half of the proof is easy; Proposition 4.4.6 (see also Proposition 2.2.5 or Exercise 2) implies that each F that has an integrable derivative and satisfies (1) is absolutely continuous. □

We are now in a position to prove the following version of integration by parts.

Corollary 6.3.9. *Let F and G be absolutely continuous functions on the interval $[a,b]$. Then*

$$F(b)G(b) - F(a)G(a) = \int_a^b F(t)G'(t)\,dt + \int_a^b F'(t)G(t)\,dt.$$

Proof. We begin by showing that the function FG is absolutely continuous. Since the functions F and G are continuous and the interval $[a,b]$ is compact, there are positive numbers M and N such that $|F(x)| \le M$ and $|G(x)| \le N$ hold at each x

[6] It is easy to modify the definition of absolute continuity for functions on $\mathbb{R}$ to make it apply to functions on $[a,b]$.

in $[a,b]$. Suppose that $\{(s_i,t_i)\}$ is a finite sequence of disjoint open subintervals of $[a,b]$. Then for each i we have

$$\begin{aligned}|F(t_i)G(t_i)-F(s_i)G(s_i)| &\le |F(t_i)-F(s_i)|\,|G(t_i)|+|F(s_i)|\,|G(t_i)-G(s_i)|\\ &\le N|F(t_i)-F(s_i)|+M|G(t_i)-G(s_i)|,\end{aligned}$$

and so

$$\sum_i |F(t_i)G(t_i)-F(s_i)G(s_i)| \le N\sum_i |F(t_i)-F(s_i)|+M\sum_i |G(t_i)-G(s_i)|.$$

Since F and G are absolutely continuous, we can make the sums on the right side of this inequality small by making $\sum_i(t_i-s_i)$ small. The absolute continuity of FG follows.

Thus Corollary 6.3.8 can be applied to the function FG. Since FG' and $F'G$ are integrable (check this) and since

$$(FG)'(x)=F(x)G'(x)+F'(x)G(x)$$

holds at almost every x in $[a,b]$, the proof is complete. □

Theorem 6.2.3 also implies the following strengthened version of Theorem 6.3.6 (see also Exercises 3 and 4).

Proposition 6.3.10. *Suppose that f belongs to $\mathscr{L}^1(\mathbb{R},\mathscr{M}_{\lambda^*},\lambda,\mathbb{R})$. Then*

$$\lim_I \frac{1}{\lambda(I)}\int_I |f(t)-f(x)|\,dt=0 \tag{2}$$

holds at λ-almost every x in $\mathbb{R}$; here I is a closed interval that contains x, and the limit is taken as the length of I approaches zero.

Points x at which (2) holds are called *Lebesgue points*[7] of f, and the set of all Lebesgue points of f is called the *Lebesgue set* of f.

Proof. It is enough to choose an arbitrary bounded open interval (a,b) and to show that (2) holds at almost every x in (a,b).

Let us first suppose that the integrable function f is in fact Borel measurable. For each rational number r let μ_r be the finite Borel measure on $\mathbb{R}$ defined by

$$\mu_r(A)=\int_{A\cap(a,b)}|f(t)-r|\,dt.$$

---

[7] Some authors use the condition $\lim_{h\to 0^+}\frac{1}{h}\int_0^h |f(x+t)+f(x-t)-2f(x)|\,dt=0$ as the defining condition for being a Lebesgue point; of course each point that satisfies (2) is also a Lebesgue point in this sense.

Theorem 6.2.3 implies that there is a Lebesgue null set N_r such that $(D\mu_r)(x) = |f(x) - r|$ holds at each x in $(a,b) - N_r$. Let N be the Lebesgue null set $\cup_{r\in\mathbb{Q}} N_r$. Suppose that x belongs to $(a,b) - N$, that I is a closed subinterval of (a,b) that contains x, and that r is a rational number. Then

$$\int_I |f(t) - f(x)|\, dt \le \int_I |f(t) - r|\, dt + \int_I |r - f(x)|\, dt,$$

and so if we divide the terms of this inequality by $\lambda(I)$ and let the length of I approach 0, we find

$$\overline{\lim_I} \frac{1}{\lambda(I)} \int_I |f(t) - f(x)|\, dt \le (D\mu_r)(x) + |r - f(x)| = 2|f(x) - r|.$$

Since $|f(x) - r|$ can be made arbitrarily small by an appropriate choice of the rational number r, Eq. (2) follows.

In case f is Lebesgue measurable, rather than Borel measurable, we can complete the proof by applying the preceding argument to a Borel measurable function that agrees with f almost everywhere (see Proposition 2.2.5). □

It is of course of interest to have easily verified conditions that imply the absolute continuity of a function. One might conjecture that a continuous function on a closed bounded interval is absolutely continuous if it is differentiable almost everywhere, if it is differentiable almost everywhere and its derivative is integrable, or if it is differentiable everywhere. These conjectures all fail (see Exercises 5 and 6), but the following related result holds.

Theorem 6.3.11. *Let $F\colon [a,b] \to \mathbb{R}$ be a continuous function such that*

(a) *F is differentiable at all except countably many of the points in $[a,b]$, and*
(b) *F' is integrable.*

Then F is absolutely continuous, and so $F(x) = F(a) + \int_a^x F'(t)\, dt$ holds at each x in $[a,b]$.

Theorem 6.3.11 would fail if condition (b) were removed (see Exercise 6), and so condition (a) does not imply condition (b). There is, however, an analogue to Theorem 6.3.11 for the Henstock–Kurzweil integral in which condition (b) is not needed; see Exercise 23 in Appendix H.

For the proof we need the following definitions and lemmas.

A function $f\colon \mathbb{R} \to (-\infty, +\infty]$ is *lower semicontinuous* if for each x in $\mathbb{R}$ and each real number A such that $A < f(x)$ there is a positive number δ such that $A < f(t)$ holds whenever t satisfies $|t - x| < \delta$. A function $f\colon \mathbb{R} \to [-\infty, +\infty)$ is *upper semicontinuous* if $-f$ is lower semicontinuous. In other words, f is upper semicontinuous if for each x in $\mathbb{R}$ and each real number A such that $f(x) < A$ there is a positive number δ such that $f(t) < A$ holds whenever t satisfies $|t - x| < \delta$.

Of course, a function $f\colon \mathbb{R} \to \mathbb{R}$ is continuous if and only if it is both lower semicontinuous and upper semicontinuous. Furthermore, it is easy to check that

(a) a function $f\colon \mathbb{R} \to (-\infty, +\infty]$ is lower semicontinuous if and only if for each real number A, the set $\{x \in \mathbb{R} : A < f(x)\}$ is open,
(b) a function $f\colon \mathbb{R} \to [-\infty, +\infty)$ is upper semicontinuous if and only if for each real number A, the set $\{x \in \mathbb{R} : f(x) < A\}$ is open,
(c) if U is an open subset of $\mathbb{R}$, then the characteristic function χ_U is lower semicontinuous,
(d) if C is a closed subset of $\mathbb{R}$, then the characteristic function χ_C is upper semicontinuous,
(e) if f and g are lower semicontinuous, then $f+g$ is lower semicontinuous, and
(f) if $\{f_n\}$ is an increasing sequence of lower semicontinuous functions, then $\lim_n f_n$ is lower semicontinuous.

It follows (from (a) and (b)) that the upper semicontinuous functions and the lower semicontinuous functions are Borel measurable.

Lemma 6.3.12. *Let $f\colon [a,b] \to [-\infty, +\infty]$ be Lebesgue integrable. Then for each positive ε there is a lower semicontinuous function $g\colon \mathbb{R} \to (-\infty, +\infty]$ that is integrable on $[a,b]$ and satisfies*

(a) *$f(t) \le g(t)$ holds at each t in $[a,b]$, and*
(b) *$\int_a^b g(t)\,dt < \int_a^b f(t)\,dt + \varepsilon$.*

Proof. Let ε be a positive number. First suppose that f is nonnegative. There is a nondecreasing sequence $\{f_n\}$ of nonnegative simple measurable functions such that $f = \lim_n f_n$ (Proposition 2.1.8), and so we can find Lebesgue measurable sets A_k, $k = 1, 2, \ldots$, and positive real numbers a_k such that $f = \sum_k a_k \chi_{A_k}$ (write each $f_n - f_{n-1}$ as a sum of positive multiples of characteristic functions). Use the regularity of Lebesgue measure (Proposition 1.4.1) to choose open sets U_k, $k = 1, 2, \ldots$, that include the corresponding A_k's and satisfy $\sum_k a_k \lambda(U_k) < \sum_k a_k \lambda(A_k) + \varepsilon/2$. Then the formula $f^\infty = \sum_k a_k \chi_{U_k}$ defines a lower semicontinuous function f^∞ that satisfies

$$\int_a^b f^\infty(t)\,dt < \int \sum_k a_k \chi_{A_k}\,dt + \varepsilon/2 = \int_a^b f(t)\,dt + \varepsilon/2$$

and is such that $f(t) \le f^\infty(t)$ holds for each t in $[a,b]$.

Now suppose that f is an arbitrary integrable function on $[a,b]$. For each n define a function h_n by $h_n(x) = \max(f(x), -n)$. The dominated convergence theorem implies that $\int_a^b f(t)\,dt = \lim_n \int_a^b h_n(t)\,dt$ and hence that we can choose a positive integer N such that $\int_a^b h_N(t)\,dt < \int_a^b f(t)\,dt + \varepsilon/2$. If we apply the argument in the preceding paragraph to the nonnegative function $h_N + N$, producing the lower semicontinuous function f^∞, then the required function g is given by $g = f^\infty - N\chi_{[a,b]}$. □

Lemma 6.3.13. *Let $H\colon [a,b] \to \mathbb{R}$ be continuous, and let C be a countable subset of $[a,b]$. Suppose that for each x in $[a,b) - C$ there is a positive number δ_x such that $H(t) > H(x)$ holds at each t in the interval $(x, x+\delta_x)$. Then H is strictly increasing.*

Proof. It suffices to prove that H is nondecreasing (why?), and for this it is enough to show that numbers x_1 and x_2 in $[a,b]$ that satisfy $x_1 < x_2$ and $H(x_1) > H(x_2)$ do not exist. Assume that such numbers do exist, and for each y between $H(x_1)$ and $H(x_2)$ define a number x_y by

$$x_y = \sup\{x \in [x_1, x_2] : H(x) \geq y\}.$$

It is easy to check that each x_y satisfies $H(x_y) = y$ and belongs to the countable exceptional set C. Since there are uncountably many such points x_y, we have reached a contradiction, and the proof is complete. □

Proof of Theorem 6.3.11. Suppose that the function F satisfies the hypotheses of Theorem 6.3.11 and that C is a countable subset of $[a,b)$ such that F is differentiable at each point of $[a,b) - C$. Let ε be a positive number. Lemma 6.3.12 (applied to the function that agrees with F' where F is differentiable and that vanishes elsewhere) provides a lower semicontinuous function g such that $F'(t) \leq g(t)$ holds at each t in $[a,b) - C$ and such that $\int_a^b g(t)\,dt < \int_a^b F'(t)\,dt + \varepsilon$. By adding a small positive continuous function to g, if necessary, we can assume that $F'(t) < g(t)$ holds at each t in $[a,b) - C$. Define $G\colon [a,b] \to \mathbb{R}$ by $G(x) = F(a) + \int_a^x g(t)\,dt$. The lower semicontinuity of g implies that

$$\varliminf_{y \downarrow x} \frac{G(y) - G(x)}{y - x} \geq g(x)$$

holds at each x in $[a,b)$. Thus

$$\varliminf_{y \downarrow x} \frac{(G(y) - F(y)) - (G(x) - F(x))}{y - x} \geq g(x) - F'(x) > 0$$

holds at each x in $[a,b) - C$, and so Lemma 6.3.13 implies that $G - F$ is nondecreasing. Since furthermore $G(a) = F(a)$, it follows that $F \leq G$. This and the inequality $\int_a^b g(t)\,dt < \int_a^b F'(t)\,dt + \varepsilon$ imply that

$$\begin{aligned}
F(x) \leq G(x) &= F(a) + \int_a^x g(t)\,dt \\
&= F(a) + \int_a^x F'(t)\,dt + \int_a^x (g(t) - F'(t))\,dt \\
&\leq F(a) + \int_a^x F'(t)\,dt + \varepsilon
\end{aligned}$$

holds at each x in $[a,b]$. Since ε was arbitrary,

$$F(x) \leq F(a) + \int_a^x F'(t)\,dt$$

must hold at each such x. The reverse inequality can be proved by applying the same argument to $-F$, and so the proof is complete. □

Exercises

1. Let $F\colon \mathbb{R} \to \mathbb{R}$ be nondecreasing. Show that if ε is a positive number, then each bounded interval contains only a finite number of values x such that $F(x+) - F(x-) \geq \varepsilon$. Use this observation to give a second proof of part (b) of Lemma 6.3.2.
2. Prove the following modified version of Lemma 6.3.7: If $F\colon [a,b] \to \mathbb{R}$ is continuous and if D is the set consisting of those points in $[a,b]$ at which F is differentiable, then $D \in \mathscr{B}(\mathbb{R})$ and F' (as a function from D to $\mathbb{R}$) is Borel measurable.
3. Derive Theorem 6.3.6 from Proposition 6.3.10.
4. Let f and F be as in Theorem 6.3.6. Show by example that there can be points x that are not Lebesgue points of f, but are such that $F'(x)$ exists and is equal to $f(x)$.
5. Show that the Cantor function provides a counterexample to two of the three conjectures suggested just before the statement of Theorem 6.3.11.
6. Define $F\colon [0,1] \to \mathbb{R}$ by

$$F(x) = \begin{cases} 0 & \text{if } x = 0, \\ x^2 \sin \dfrac{1}{x^2} & \text{if } 0 < x \leq 1. \end{cases}$$

   Show that F is differentiable everywhere on $[0,1]$ but is not absolutely continuous.
7. Show that there is a strictly increasing continuous function $F\colon [0,1] \to \mathbb{R}$ such that $F'(x) = 0$ holds at λ-almost every x in $[0,1]$. (Hint: Let F be the sum of a suitable series of functions, and use Proposition 6.3.5.)

Notes

The proof of Theorem 6.1.7 presented here was inspired by one given by A.M. Gleason in some unpublished notes on advanced calculus.

Munroe [92], Rudin [105], and Wheeden and Zygmund [127] carry the study of the differentiation of measures and functions a bit farther than it is taken here. See Bruckner [21], Bruckner [22], de Guzmán [33], Hayes and Pauc [56], Kölzow [74], and Saks [106] for more advanced treatments of differentiation theory.

The proof of Theorem 6.3.11 given here is taken from Walker [123].

Chapter 7
Measures on Locally Compact Spaces

Let $\mathscr{K}(\mathbb{R})$ be the vector space consisting of those continuous functions $f\colon \mathbb{R} \to \mathbb{R}$ that have compact support—that is, for which the set $\{x \in \mathbb{R} : f(x) \neq 0\}$ has a compact closure. Then $f \mapsto \int_{\mathbb{R}} f\,d\lambda$ defines a positive[1] linear functional on $\mathscr{K}(\mathbb{R})$. More generally, if μ is a measure on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ that has finite values on the compact subsets of $\mathbb{R}$, then $f \mapsto \int_{\mathbb{R}} f\,d\mu$ defines a positive linear functional on $\mathscr{K}(\mathbb{R})$. According to a special case of the Riesz representation theorem (see Theorem 7.2.8), the converse also holds: for every positive linear functional $I\colon \mathscr{K}(\mathbb{R}) \to \mathbb{R}$ there is a Borel measure μ on $\mathbb{R}$ that is finite on compact sets and represents I, in the sense that $I(f) = \int_{\mathbb{R}} f\,d\mu$ holds for each f in $\mathscr{K}(\mathbb{R})$.

This chapter is devoted to the Riesz representation theorem and related results. The first section (Sect. 7.1) contains some basic facts about locally compact Hausdorff spaces, the spaces that provide the natural setting for the Riesz representation theorem, while the second section (Sect. 7.2) gives a proof of the Riesz representation theorem. The next two sections (Sects. 7.3 and 7.4) contain some useful and relatively basic related material. The results of Sects. 7.5 and 7.6 are needed for dealing with large locally compact Hausdorff spaces; for relatively small locally compact Hausdorff spaces (those that have a countable base), Proposition 7.6.2 is the only result from these sections that one really needs (see also Proposition 7.2.5 and Theorems 4.5.1 and 5.2.2).

The Daniell–Stone integral gives another way to deal with integration on locally compact Hausdorff spaces. A measure-theoretic version of the basic result of the Daniell–Stone theory is given by Theorem 7.7.4; the general Daniell–Stone setup is outlined in the exercises at the end of Sect. 7.7.

---

[1]Recall that a linear functional I on a vector space of functions is *positive* if $I(f) \geq 0$ holds for each nonnegative function f in the domain of I.

D.L. Cohn, *Measure Theory: Second Edition*, Birkhäuser Advanced Texts Basler Lehrbücher, DOI 10.1007/978-1-4614-6956-8_7,