THE FORMULA FOR CHANGE OF VARIABLES IN MULTIPLE INTEGRALS

D. A. EDWARDS

1. Introduction

The object of this note is to offer a reasonably self-contained proof of the following well known theorem, which, despite its usefulness, is often omitted from elementary accounts of Lebesgue integration. Lebesgue measure in $\mathbb{R}^k$ is denoted by λ.

Theorem 1. *Let U and V be open subsets of Euclidean k-dimensional space $\mathbb{R}^k$ and let ϕ be a bijection of U onto V such that ϕ and its inverse ϕ^{-1} are continuous and have continuous derivatives. Then, for every measurable function $f : V \to \overline{\mathbb{R}}_+$, we have*

$$\int_V f(x)\lambda(dx) = \int_U (f \circ \phi)(x)|\det \phi'(x)|\lambda(dx).$$

(*The case in which both integrals are infinite is not excluded.*)

If, on the other hand, a function $f : V \to \overline{\mathbb{R}}$ is given, then $f \in L^1(V)$ if and only if $(f \circ \phi)|\det \phi'| \in L^1(U)$, and in that case the above formula remains valid.

The prerequisites for reading this paper are a first course in Lebesgue integration theory, treated via the Carathéodory extension theorem, and just a little linear algebra and multivariable analysis. No originality is claimed for the demonstration given here, which has simply been pieced together from a number of sources.

We first recall some matters that are not always treated in elementary accounts of Lebesgue integration.

2. Regularity of Lebesgue measure

Given a set Ω and a family $\mathcal{E}$ of subsets of Ω, we denote by $\sigma(\mathcal{E})$ the smallest σ-algebra of subsets of Ω that contains $\mathcal{E}$, and we say that $\sigma(\mathcal{E})$ is the σ-algebra of subsets of Ω **generated** by $\mathcal{E}$. When Ω is a topological space, we denote by $\mathfrak{B}(\Omega)$ the σ-algebra of subsets of Ω generated by the family of all open subsets of Ω, or, equivalently, by the family of all closed subsets of Ω. The members of $\mathfrak{B}(\Omega)$ are termed **Borel subsets** of Ω, or simply **Borel sets** if there is no risk of confusion.

By a **cell** in $\mathbb{R}^k$ we shall mean a set of the form

$$P := \prod_{i=1}^{k} [a_i, b_i)$$

where $a := (a_1, \ldots, a_k)$ and $b := (b_1, \ldots, b_k)$ are points of $\mathbb{R}^k$ and $a_i \leq b_i$ for all i. Note that if $a_i = b_i$ for some i then $P = \emptyset$. The **volume**, $\mathrm{vol}\,(P)$, of P is, by definition, $\prod_{i=1}^{k}(b_i - a_i)$. The Lebesgue outer measure $\lambda^*(E)$ of a set $E \subseteq \mathbb{R}^k$ can be defined as $\inf \sum_{n=1}^{\infty} \mathrm{vol}\,(P_n)$, where the infimum is taken over all sequences (P_n) of cells such that $E \subseteq \bigcup_{n=1}^{\infty} P_n$. Lebesgue measure for $\mathbb{R}^k$ can then be obtained from λ^* by the usual Carathéodory procedure, in the course of which it is shown that every cell is a measurable set. We prove below that that all Borel subsets of $\mathbb{R}^k$ are measurable (with respect to λ). Suppose our cell P is non-empty. Then it has interior $\mathring{P} = \prod_{i=1}^{k}(a_i, b_i)$, and closure $\overline{P} = \prod_{i=1}^{k}[a_i, b_i]$; and if also E is a set satisfying $\mathring{P} \subseteq E \subseteq \overline{P}$ then E is measurable and $\lambda(E) = \mathrm{vol}\,(P)$.

Theorem 2 (Regularity theorem)**.** *Let E be a measurable subset of $\mathbb{R}^k$ and suppose that $\epsilon > 0$.*

(a) *Then there exist an open set G and a closed set F such that $F \subseteq E \subseteq G$, and $\lambda(G \setminus F) < \epsilon$.*

(b) *If $\lambda(E) < \infty$ we can find an open set G and a compact set K such that $K \subseteq E \subseteq G$ and $\lambda(G \setminus K) < \epsilon$.*

Proof. (a) Consider first the case of the cell $P = \prod_{i=1}^{k}[a_i, b_i)$. The product

$$G_n := \prod_{i=1}^{k}(a_i - n^{-1}, b_i)$$

is an open set and $(G_n \setminus P) \downarrow \emptyset$ as $n \to \infty$. Since $\lambda(G_1 \setminus P) < \infty$, it follows that $\lambda(G_n \setminus P) \downarrow 0$ as $n \to \infty$.

Now let E be a measurable set with $\lambda(E) < \infty$, and suppose that $\epsilon > 0$. We can find a sequence of cells (P_n) such that $E \subseteq \bigcup_{n=1}^{\infty} P_n$ and $\sum_{n=1}^{\infty} \lambda(P_n) < \lambda(E) + \epsilon/2^2$. Now choose open sets G_n such that $P_n \subseteq G_n$ and $\lambda(G_n \setminus P_n) < \epsilon/2^{n+2}$. Then $G := \bigcup_{n=1}^{\infty} G_n$ is open, $E \subseteq G$, and

$$\lambda(G) < \sum_{n=1}^{\infty} \lambda(G_n) < \sum_{n=1}^{\infty}(\lambda(P_n) + \lambda(G_n \setminus P_n)) < \lambda(E) + \epsilon/2,$$

and hence $\lambda(G \setminus E) < \epsilon/2$.

Next, let E be a measurable set with $\lambda(E) = \infty$ and for each integer $n \geq 1$ let $E_n = E \cap B_n$, where B_n is the Borel set $\{ x \in \mathbb{R}^k : (n-1) \leq \|x\| < n \}$. For each n choose an open set G_n with $E_n \subseteq G_n$ and $\lambda(G_n \setminus E_n) < \epsilon/2^{n+1}$. Then $G := \bigcup_{n=1}^{\infty} G_n$ is open, $E \subseteq G$ and $\lambda(G \setminus E) < \epsilon/2$.

To approximate E from the inside by closed sets, first approximate $\complement E$ from outside by open sets, then pass to complements to obtain a closed set F such that $E \supseteq F$ and $\lambda(E \setminus F) < \epsilon/2$. Then $\lambda(G \setminus F) < \epsilon$.

(b) Now suppose that $\lambda(E) < \infty$ and let $E_n = \{ x \in E : \|x\| < n \}$. Then $\lambda(E \setminus E_n) \downarrow 0$ as $n \to \infty$. Hence $\lambda(E \setminus E_N) < \epsilon/4$ for some N. And $\lambda(E_N \setminus K) < \epsilon/4$ for some closed and bounded (and hence compact) subset K of E_N. Hence $\lambda(E \setminus K) < \epsilon/2$. If now G is constructed as in part (a) of this proof then we see that $K \subseteq E \subseteq G$ and $\lambda(G \setminus K) < \epsilon$. □

.

Corollary 3. *If E is a measurable subset of $\mathbb{R}^k$ then*

$$\begin{aligned}\lambda(E) &= \inf\{ \lambda(G) : G \text{ open},\ E \subseteq G \} \\ &= \sup\{ \lambda(K) : K \text{ compact},\ K \subseteq E \}.\end{aligned}$$

Proof. (i) If $\lambda(E) = \infty$ then it is obvious that $\lambda(E) = \inf\{ \lambda(G) : G \text{ is open},\ E \subseteq G \}$. Next, suppose that $\lambda(E) < \infty$ and that $\epsilon > 0$. Then we can find an open set G such that $E \subseteq G$ and $\lambda(G \setminus E) < \epsilon$. But then

$$\lambda(E) \leq \lambda(G \setminus E) + \lambda(E) < \lambda(E) + \epsilon.$$

Hence $\lambda(E) = \inf\{ \lambda(G) : G \text{ open},\ E \subseteq G \}$.

(ii) Suppose that $\mathbb{R} \ni t < \lambda(E)$, and for each integer $n \geq 1$ let $E_n = \{ x \in E : \|x\| \leq n \}$. Then $\lambda(E_N) > t$ for large enough N. Now choose a closed set K such that $K \subseteq E_N$ and $\lambda(E_N \setminus K) < \lambda(E_N) - t$. Then K, being both closed and bounded, is a compact subset of E and $\lambda(K) > t$. □

Corollary 4. *Let E be a subset of $\mathbb{R}^k$. Then E is measurable if and only if there exist Borel sets A and B such that $A \subseteq E \subseteq B$ and $\lambda(B \setminus A) = 0$.*

Proof. Suppose that E is measurable. Then, for each integer $n \geq 1$, we can find an open set G_n such that $E \subseteq G_n$ and $\lambda(G_n \setminus E) < n^{-1}$. We can arrange that the sequence (G_n) is decreasing. Let B be the Borel set $\bigcap_{n=1}^{\infty} G_n$. Then $E \subseteq B$, and $(G_n \setminus E) \downarrow (B \setminus E)$ as $n \to \infty$. It follows that $\lambda(B \setminus E) = \lim_{n\to\infty} \lambda(G_n \setminus E) = 0$. Similarly, approximating E

from the inside by closed sets, we obtain a Borel set A such that $A \subseteq E$ and $\lambda(E \setminus A) = 0$. Consequently $\lambda(B \setminus A) = 0$.

Suppose, conversely, that E is a subset of $\mathbb{R}^k$ for which there exist Borel sets A and B such that $A \subseteq E \subseteq B$ and $\lambda(B \setminus A) = 0$. Then, by the completeness of Lebesgue measure, the set $E \setminus A$ is measurable because it is a subset of the null set $B \setminus A$. But A is a measurable set, because it is Borel. Hence $E = A \cup (E \setminus A)$ is measurable. □

3. Borel sets

Let U be an open subset of $\mathbb{R}^k$. The relation between $\mathfrak{B}(U)$ and $\mathfrak{B}(\mathbb{R}^k)$ is given by the following lemma.

Lemma 5. *If U is an open subset of $\mathbb{R}^k$ then*

$$\mathfrak{B}(U) = \{\, B \cap U : B \in \mathfrak{B}(\mathbb{R}^k)\,\} = \{A \in \mathfrak{B}(\mathbb{R}^k) : A \subseteq U\,\}.$$

Proof. Denote by $\mathcal{U}$ the set of all open subsets of U and by $\mathcal{G}$ the set of all open subsets of $\mathbb{R}^k$. Observe that

$$\mathcal{U} \subseteq \{\, B \cap U : B \in \mathfrak{B}(\mathbb{R}^k)\,\}.$$

It is easy to see that $\{\, B \cap U : B \in \mathfrak{B}(\mathbb{R}^k)\,\}$ is a σ-algebra of subsets of U, so it follows that

$$\mathfrak{B}(U) \subseteq \{\, B \cap U : B \in \mathfrak{B}(\mathbb{R}^k)\,\}. \tag{1}$$

To obtain the reverse inclusion, consider $\mathcal{E} := \{\, E : E \cap U \in \mathfrak{B}(U)\,\}$. This is a σ-algebra of subsets of $\mathbb{R}^k$ and, clearly, $\mathcal{G} \subseteq \mathcal{E}$. Hence $\mathfrak{B}(\mathbb{R}^k) \subseteq \mathcal{E}$. But that means that

$$\{\, B \cap U : B \in \mathfrak{B}(\mathbb{R}^k)\,\} \subseteq \mathfrak{B}(U),$$

so that, by (1), we in fact have equality. Finally, because $U \in \mathfrak{B}(\mathbb{R}^k)$, the truth of the equation $\{\, B \cap U : B \in \mathfrak{B}(\mathbb{R}^k)\,\} = \{A \in \mathfrak{B}(\mathbb{R}^k) : A \subseteq U\,\}$ is obvious. □

Lemma 6. *If X, Y are topological spaces and $h : X \to Y$ is a continuous map then $h^{-1}(\mathfrak{B}(Y)) \subseteq \mathfrak{B}(X)$. If h is a homeomorphism of X onto Y then $h(\mathfrak{B}(X)) = \mathfrak{B}(Y)$ and $h^{-1}(\mathfrak{B}(Y)) = \mathfrak{B}(X)$.*

Proof. Let $h : X \to Y$ be continuous and let $\mathcal{E} = \{\, E : E \subseteq Y, h^{-1}(E) \in \mathfrak{B}(X)\,\}$. Then $\mathcal{E}$ is a σ-algebra of subsets of Y that contains the open sets. Hence $\mathcal{E} \supseteq \mathfrak{B}(Y)$ and therefore $h^{-1}(\mathfrak{B}(Y)) \subseteq \mathfrak{B}(X)$.

Now assume that h is a homeomrphism. Then, by what we have proved, $h(\mathfrak{B}(X)) \subseteq \mathfrak{B}(Y)$. Hence

$$\mathfrak{B}(X) = h^{-1}h(\mathfrak{B}(X)) \subseteq h^{-1}(\mathfrak{B}(Y)) \subseteq \mathfrak{B}(X),$$

so $h^{-1}(\mathfrak{B}(Y)) \subseteq \mathfrak{B}(X)$. Similarly, $h(\mathfrak{B}(X)) = \mathfrak{B}(Y)$. □

4. DYADIC CUBES

We shall denote by W the cell $[0,1)^k$. By a **dyadic cube** in $\mathbb{R}^k$ we shall mean a cell of the form $Q = 2^{-n}a + 2^{-n}W$, where n is a integer ≥ 0 and $a \in \mathbb{Z}^k$. The number n is called the **order** of the cube, the number 2^{-n} is termed the **edge-length** of Q, and the point $2^{-n}a$ will be termed its **vertex**. Thus W is a dyadic cube of order zero, of edge-length 1, and vertex the origin. Note that a cube of edge-length h has diameter $h^k\sqrt{k}$. For each n the set Δ_n of all dyadic cubes of order n is a disjoint cover of $\mathbb{R}^k$. The set Δ_n is countably infinite, and hence the set $\Delta := \bigcup_{n=0}^{\infty} \Delta_n$ of all dyadic cubes is countably infinite.

Lemma 7. *Let Q, Q' be dyadic cubes of orders n, n' respectively, and suppose that $n \geq n'$. Then the following statements are equivalent*

(i) $Q \subseteq Q'$;
(ii) $Q \cap Q' \neq \emptyset$;
(iii) $v(Q) \in Q'$, *where $v(Q)$ is the vertex of Q.*

Proof. By translation and scaling we can suppose for the proof that $Q' = W$.
(i)⇒(ii): This implication is trivial.
(ii)⇒(iii): Suppose that $Q \cap W \neq \emptyset$, where $Q = 2^{-n}a + 2^{-n}W$. Then there exist $u, v \in W$ such that $2^{-n}a + 2^{-n}u = v$, or $a = 2^n v - u$. Hence $-1 < a_i < 2^n$ with $a_i \in \mathbb{Z}$ for each i. Therefore $a_i \in \{0, 1, \dots, 2^n - 1\}$, and hence $v(Q) = 2^{-n}a \in W$.
(iii)⇒(i): Suppose that $v(Q) = 2^{-n}a = 2^{-n}(a_1, \dots a_k) \in W$. Then, for each i, we have $a_i \in 2^n W \cap \mathbb{Z}$ and so $a_i \in \{0, 1, \dots, 2^n - 1\}$. Hence $Q = 2^{-n}a + 2^{-n}W \subseteq W$. □

Corollary 8. *Let Q, Q' be dyadic cubes. Then at least one of the following assertions is true:* (i) $Q \subseteq Q'$; (ii) $Q' \subseteq Q$; (iii) $Q \cap Q' = \emptyset$. *If two cubes are of the same order then they are either equal or disjoint.*

Proof. Obvious, by the preceding Lemma. □

Theorem 9. *Let G be a non-empty open subset of $\mathbb{R}^k$. Then there exists a disjoint (infinite) sequence (Q_n) of dyadic cubes such that* (a) $G = \bigcup_{n=1}^{\infty} Q_n$ *and* (b) $\overline{Q}_n \subseteq G$ *for all n.*

Proof. For each $n \geq 0$ let $\mathcal{C}_n$ denote the set of all dyadic cubes of order n whose closures are subsets of G. Let $\mathcal{A}_0 = \mathcal{C}_0$. and let $A_0 = \bigcup\{Q : Q \in \mathcal{A}_0\}$. Next, let $\mathcal{A}_1$ be the set of all the cubes in $\mathcal{C}_1$ that have empty intersection with A_0, and let $A_1 = \bigcup\{Q : Q \in \mathcal{A}_1\}$. And for $n > 1$ let $\mathcal{A}_n$ be the set of all cubes in $\mathcal{C}_n$ that have empty intersection with $A_1 \cup \cdots \cup A_{n-1}$, and let $A_n := \bigcup\{Q : Q \in \mathcal{A}_n\}$. For each n let $B_n = A_1 \cup \cdots \cup A_n$.

I claim that $\bigcup_{n=0}^{\infty} \mathcal{A}_n$ is a disjoint covering of G by dyadic cubes. Disjointness is clear, so it will suffice to show that $\bigcup_{n=0}^{\infty} B_n = G$. To see this, let $x \in G$. The distance of x from $\complement G$ is > 0, so we can find a dyadic cube Q such that $x \in Q \subseteq \overline{Q} \subseteq G$, and we can suppose that Q has order $n \geq 1$. If $Q \cap B_{n-1} = \emptyset$ then, by definition, $Q \in \mathcal{A}_n$, so $Q \subseteq A_n \subseteq B_n$. If $Q \cap B_{n-1} \neq \emptyset$ then Q has non-empty intersection with some larger cube Q' belonging to the family $\mathcal{A}_0 \cup \cdots \cup \mathcal{A}_{n-1}$. But then we have $Q \subseteq Q' \subseteq B_{n-1} \subseteq B_n$. Thus in any case $x \in Q \subseteq B_n$. Since $B_n \subseteq G$ for all n, we have thus shown that $\bigcup_{n=0}^{\infty} B_n = G$.

The family of cubes $\bigcup_{n=0}^{\infty} \mathcal{A}_n$ is countably infinite, since otherwise G could be expressed as a finite union $\bigcup_{n=0}^{N} C_n$ in which each term is a dyadic cube whose closure is a subset of G. But that would imply that $\bigcup_{n=0}^{N} \overline{C_n} = G$, and hence that G is both open and compact. But that is impossible, because $\mathbb{R}^k$ is connected and $G \neq \emptyset$.

Taking (Q_n) now to be any sequence that enumerates the elements of $\bigcup_{n=0}^{\infty} \mathcal{A}_n$, we obtain a disjoint infinite sequence of dyadic cubes with the desired properties (a) and (b). □

Given an open subset G of $\mathbb{R}^k$, we denote by $\Delta(G)$ the set of all dyadic cubes Q such that $\overline{Q} \subseteq G$.

Theorem 10. (i) *If* $\mathfrak{C}$ *denotes the set of all cells in* $\mathbb{R}^k$*, then* $\mathfrak{B}(\mathbb{R}^k) = \sigma(\mathfrak{C}) = \sigma(\Delta)$*. Hence all Borel subsets of* $\mathbb{R}^k$ *are measurable.* (ii) *Let* U *be an non-empty open subset of* $\mathbb{R}^k$*. Then the* σ*-algebra of subsets of* U *generated by* $\Delta(U)$ *is* $\mathfrak{B}(U)$*.*

Proof. (i) Let P be the cell $\prod_{i=1}^{k}[a_i, b_i) \neq \emptyset$. Then P is σ-compact, and hence Borel, because $P = \bigcup_{n=N}^{\infty}[a_i, b_i - n^{-1}]$ for large N. Thus $\Delta \subseteq \mathfrak{C} \subseteq \mathfrak{B}(\mathbb{R}^k)$ and hence $\sigma(\Delta) \subseteq \sigma(\mathfrak{C}) \subseteq \mathfrak{B}(\mathbb{R}^k)$.

On the other hand, if $\mathcal{G}$ denotes the set of all open subsets of $\mathbb{R}^k$, then by the preceding theorem $\sigma(\Delta) \supseteq \mathcal{G}$. Hence $\sigma(\Delta) \supseteq \sigma(\mathcal{G}) = \mathfrak{B}(\mathbb{R}^k)$. Putting together these inclusions, we have $\mathfrak{B}(\mathbb{R}^k) = \sigma(\mathfrak{C}) = \sigma(\Delta)$.

It follows that all Borel subsets of $\mathbb{R}^k$ are measurable since, as we noted in §2, all sets belonging to $\sigma(\mathfrak{C})$ are measurable.

(ii) Let $\mathcal{A}$ be the σ-algebra of subsets of U generated by $\Delta(U)$ and let $\mathcal{U}$ be the set of open subsets of U. By Theorem 9 we have $\mathcal{A} \supseteq \mathcal{U}$, and hence $\mathcal{A} \supseteq \mathfrak{B}(U)$. On the other hand $\Delta(U) \subseteq \mathfrak{B}(U)$ so $\mathcal{A} \subseteq \mathcal{U}$. Therefore $\mathcal{A} = \mathcal{U}$.

□

5. Linear transformations

By an **elementary transformation** in $\mathbb{R}^k$ we shall mean an invertible linear map $T : \mathbb{R}^k \to \mathbb{R}^k$ of one of the following three types:

(i) T is a permutation of coordinates. That is to say

$$T(x_1, \dots, x_k) = (x_{\pi 1}, \dots, x_{\pi k}),$$

where π is a permutation of the set $\{1, 2, \dots, k\}$.
(ii) T is of the form

$$T(x_1, \dots, x_k) = (\alpha x_1, x_2, \dots, x_k),$$

where α is a non-zero scalar.
(iii) T adds the second coordinate to the first, leaving all others unchanged, thus

$$T(x_1, \dots, x_k) = (x_1 + x_2, x_2, \dots, x_k).$$

Lemma 11. *If $T : \mathbb{R}^k \to \mathbb{R}^k$ is an elementary transformation and Q is a dyadic cube, then Q and TQ are Borel sets and*

$$\lambda(TQ) = |\det T|\, \lambda(Q).$$

Proof. Since $T : \mathbb{R}^k \to \mathbb{R}^k$ is continuous and Q, being a cell, is σ-compact, the image TQ is σ-compact. Hence both Q and TQ are Borel sets.

We first prove the theorem for the case $Q = W$, taking the three types of elementary transformation in turn. Note that $\lambda(W) = 1$.

(i) In this case $\det T = \pm 1$, $TW = W$, and so $\lambda(TW) = \lambda(W) = |\det T|$.

(ii) Here

$$TW = \{x : x_1 \in J,\ 0 \le x_i < 1 \text{ for } i = 2, \dots, k\},$$

where $J = [0, \alpha)$ if $\alpha > 0$, and $J = (\alpha, 0]$ if $\alpha < 0$. In both cases $\lambda(TW) = |\alpha| = |\det T|$.

(iii) Here

$$TW = \{x : 0 \le x_2 \le x_1 < x_2 + 1,\ 0 \le x_i < 1 \text{ for } i = 2, \dots, k\}.$$

Let $A_1 = \{x \in TW : x_1 < 1\}$, $A_2 = TW \setminus A_1$. Denote by e_1 the first vector in the standard basis for $\mathbb{R}^k$:

$$e_1 = (1, 0, \dots, 0).$$

Then W is the disjoint union $A_1 \cup (A_2 - e_1)$ and

$$\begin{aligned}\lambda(TW) &= \lambda(A_1 \cup A_2) = \lambda(A_1) + \lambda(A_2) \\ &= \lambda(A_1) + \lambda(A_2 - e_1) \\ &= \lambda(A_1 \cup (A_2 - e_1)) = \lambda(W) = 1.\end{aligned}$$

Here we have assumed that the sets in play are measurable. But that is easily proved. For A_1 is the intersection of the two Borel sets TW and

$\{ x : x_1 < 1 \}$ and hence it is Borel. Consequently, A_2 is also a Borel set; and finally $(A_2 - e_1)$ is a Borel set because it is equal to $W \setminus A_1$.

Since $\det T = 1$ in this case, we again have $\lambda(TW) = |\det T|$.

Now let $Q = 2^{-n}W$ and let T be an elementary transformation of any one of the three types defined above.. Then W is the disjoint union of 2^{nk} translates of Q, so, by the translation-invariance of λ, $2^{nk}\lambda(Q) = \lambda(W) = 1$, hence $\lambda(Q) = 2^{-nk}$. For all $v \in \mathbb{R}^k$, $\lambda(T(v + Q)) = \lambda(TQ)$. Now TW is the disjoint union of 2^{nk} sets of the form $T(v + Q)$. Therefore $2^{nk}\lambda(TQ) = \lambda(TW) = |\det T|$, so

$$\lambda(TQ) = |\det T|\, 2^{-nk} = |\det T|\, \lambda(Q).$$

By the translation-invariance of λ, this formula remains valid if we have $Q = v + 2^{-n}W$ instead of $Q = 2^{-n}W$. □

We pass now to consideration of an arbitrary invertible linear transformation $T : \mathbb{R}^k \to \mathbb{R}^k$. Note that such a T is a homeomorphism of $\mathbb{R}^k$, and hence it preserves open sets, and, by Lemma 6, also Borel sets.

Theorem 12. *Let B be a Borel set in $\mathbb{R}^k$ and $T : \mathbb{R}^k \to \mathbb{R}^k$ an invertible linear transformation. Then TB is a Borel set and*

$$\lambda(TB) = |\det T|\, \lambda(B).$$

Proof. We prove first that if T is an elementary linear transformation and G an open set in $\mathbb{R}^k$ then $\lambda(TG) = |\det T|\, \lambda(G)$.

Dismissing the trivial case where $G = \emptyset$, we suppose that $G \neq \emptyset$. Then there exists a disjoint sequence of dyadic cubes (Q_n) whose union is G. Then, by the preceding lemma,

$$\lambda(TG) = \sum_{n=1}^{\infty} \lambda(TQ_n) = \sum_{n=1}^{\infty} |\det T|\, \lambda(Q_n) = |\det T|\, \lambda(G).$$

Suppose next that T_1, T_2 are invertible linear transformations such that $\lambda(T_r G) = |\det T_r|\lambda(G)$ for open G and $r = 1, 2$. Noting that $T_2 G$ is an open set, we see that

$$\begin{aligned}\lambda(T_1 T_2 G) &= |\det T_1|\, \lambda(T_2 G) \\ &= |\det T_1||\det T_2|\, \lambda(G) = |\det(T_1 T_2)|\, \lambda(G).\end{aligned}$$

This reasoning can be extended to finite products. But a theorem of elementary algebra states that an arbitrary invertible linear transformation can be represented as the product of a finite sequence of elementary transformations. Thus, if $T = T_1 T_2 \dots T_n$ is such a product, we shall have

$$\lambda(TG) = |\det(T_1| \times \cdots \times |\det T_n)|\, \lambda(G) = |\det T|\, \lambda(G)$$

for all open G. We have already noted that if $T : \mathbb{R}^k \to \mathbb{R}^k$ an invertible linear transformation then $TB \in \mathfrak{B}(\mathbb{R}^k)$ for all $B \in \mathfrak{B}(\mathbb{R}^k)$. For such B we have, by the regularity of λ,

$$\lambda(B) = \inf\{\, \lambda(G) : G \text{ open}, G \supseteq B \,\}$$

and

$$\lambda(TB) = \inf\{\, \lambda(O) : O \text{ open}, O \supseteq TB \,\}.$$

But T is a homeomorphism of $\mathbb{R}^k$, so

$$\{\, O : O \text{ open}, O \supseteq TB \,\} = \{\, TG : G \text{ open}, G \supseteq B \,\}$$

and hence

$$\begin{aligned} \lambda(TB) &= \inf\{\, \lambda(TG) : G \text{ open}, G \supseteq B \,\} \\ &= \inf\{\, |\det T|\, \lambda(G) : G \text{ open}, G \supseteq B \,\} \\ &= |\det T| \times \inf\{\, \lambda(G) : G \text{ open}, G \supseteq B \,\} \\ &= |\det T|\, \lambda(B) \end{aligned}$$

□

6. Some important estimates

Theorem 13. *Suppose that $x_0 \in U$ and $\epsilon > 0$ Then there exists $\delta > 0$ such that, for every cube Q for which $x_0 \in \overline{Q} \subseteq U$ and $e(Q) < \delta$, we have*

$$\frac{\lambda(\phi(Q))}{\lambda(Q)} < |\det \phi'(x_0)| + \epsilon.$$

Proof. In this proof the norm in $\mathbb{R}^k$, denoted simply by $\|\cdot\|$, will be $\|\cdot\|_\infty$.

Case (i). Suppose that $x_0 = 0$, $\phi(x_0) = 0$, and $\phi'(x_0) = I$. So, for small x,

$$\phi(x) = x + \|x\|\rho(x),$$

where $\rho(x) \to 0$ as $x \to 0$. For $r > 0$ and $y \in U$ let $C(y, r)$ be the closed cube $\{\, x : \|x - y\| \le r \,\}$. Choose $\eta > 0$ to satisfy

$$(1 + 2\eta)^k < 1 + \epsilon,$$

and let $\delta > 0$ be such that $x \in U$ and $\|\rho(x)\| < \eta$, whenever $\|x\| < \delta$.

Let Q be a cube for which $0 \in \overline{Q} \subseteq U$ and $s := e(Q) < \delta$, and let a denote the centre of Q. Then $\overline{Q} = C(a; \frac{s}{2})$. Since $0 \in C(a; \frac{s}{2})$ we have $\|a\| = \|0 - a\| \le \frac{s}{2}$. Hence, for $x \in Q$,

$$\|x\| = \|x - a\| + \|a\| \le \frac{s}{2} + \frac{s}{2} = s < \delta,$$

and therefore $\|\rho(x)\| < \eta$.

Thus for $x \in Q$ we have

$$\begin{aligned}\|\phi(x) - a\| &\leq \|x - a\| + \|x\|\|\rho(x)\| \\ &< \tfrac{s}{2} + s\eta = \tfrac{s}{2}(1 + 2\eta).\end{aligned}$$

Consequently, $\phi(x) \in C(a, \frac{s}{2}(1 + 2\eta))$.

So we have proved that

$$\phi(Q) \subseteq C(a, \tfrac{s}{2}(1 + 2\eta))$$

Comparing the Lebesgue measures of these terms we see that

$$\lambda(\phi(Q)) \leq s^k (1 + 2\eta)^k.$$

But $\lambda(Q) = s^k$, so

$$\frac{\lambda(\phi(Q))}{\lambda(Q)} \leq (1 + 2\eta)^k < 1 + \epsilon,$$

as desired.

Case (ii). Assume now that $x_0 = 0$, $\phi(x_0) = 0$, and let $T = \phi'(x_0)$ and let $\chi = T^{-1}\phi$. Then χ is a C^1 diffeomorphism of U onto $T^{-1}V$ and $\chi'(x_0) = I$. By our proof for Case (i) there exists $\delta > 0$ such that, for every cube Q such that $x_0 \in \overline{Q}$ with $e(Q) < \delta$, we have $\overline{Q} \subseteq U$ and

$$\frac{\lambda(\chi(Q))}{\lambda(Q)} < 1 + \frac{\epsilon}{|\det T|}.$$

But

$$\begin{aligned}\lambda(\chi(Q)) &= \lambda(T^{-1}\phi(Q)) \\ &= |\det T^{-1}|\lambda(\phi(Q)) = \frac{\lambda(\phi(Q))}{|\det T|}.\end{aligned}$$

Therefore

$$\frac{\lambda(\phi(Q))}{\lambda(Q)} < |\det T| + \epsilon.$$

Case (iii). Now we consider the general case. So assume that $x_0 \in \overline{Q} \subseteq U$ and, as above, denote by T the linear operator $\phi'(x_0)$. For $x \in U - x_0$ let $\sigma(x) := \phi(x + x_0) - \phi(x_0)$. Then σ is a C^1 diffeomorphism of $U_0 := U - x_0$ onto $V_0 := V - \phi(x_0)$, $0 \in U_0$, $\sigma(0) = 0$, and $\sigma'(0) = T$.

Now write $Q_0 := Q - x_0$. Then $0 \in \overline{Q}_0 \subseteq U_0$, and by Case (ii) there exists $\delta > 0$ such that when $e(Q_0) < \delta$ we have

$$\frac{\lambda(\sigma(Q_0))}{\lambda(Q_0)} < |\det T| + \epsilon.$$

Now $e(Q) = e(Q_0)$, and $\lambda(Q) = \lambda(Q_0)$. Moreover $\sigma(Q_0) = \phi(Q) - \phi(x_0)$, so $\lambda(\sigma(Q_0)) = \lambda(\phi(Q))$. Thus, for $e(Q) < \delta$, the preceding inequality yields

$$\frac{\lambda(\phi(Q))}{\lambda(Q)} < |\det \phi'(x_0)| + \epsilon,$$

as desired. □

Corollary 14. *Let (Q_n) be a decreasing sequence of cubes in $\mathbb{R}^k$ such that $\overline{Q}_n \subseteq U$ for all n, and and suppose that $e(Q_n) \to 0$ as $n \to \infty$. Let x_0 be the unique point that belongs to every $\overline{Q}_n$. Then*

$$\limsup_{n\to\infty} \frac{\lambda(\phi(Q_n))}{\lambda(Q_n)} \le |\det \phi'(x_0)|.$$

Proof. By Theorem 13 we have, for all $\epsilon > 0$,

$$\limsup_{n\to\infty} \frac{\lambda(\phi(Q_n))}{\lambda(Q_n)} \le |\det \phi'(x_0)| + \epsilon.$$

□

7. An upper bound for $\lambda(\phi(E))$

Lemma 15. *If Q is a dyadic cube such that $\overline{Q} \subseteq U$ then*

$$\lambda(\phi(Q)) \le \int_Q |\det \phi'(x)| \lambda(dx). \tag{2}$$

Proof. Suppose that there is a dyadic cube Q such that $\overline{Q} \subseteq U$ and for which the equation (2) is false. Then for some $\epsilon > 0$ we shall have

$$\lambda(\phi(Q)) > \int_Q |\det \phi'(x)| \lambda(dx) + \epsilon\lambda(Q). \tag{3}$$

Divide Q into a disjoint family $\{Q_i\}$ of 2^k congruent little dyadic cubes with $e(Q_i) = 2^{-1}e(Q)$. I claim that for at least one of these, Q_1 say, we shall then have

$$\lambda(\phi(Q_1)) > \int_{Q_1} |\det \phi'(x)| \lambda(dx) + \epsilon\lambda(Q_1)$$

For otherwise each little cube Q_i would satisfy

$$\lambda(\phi(Q_i)) \le \int_{Q_i} |\det \phi'(x)| \lambda(dx) + \epsilon\lambda(Q_i).$$

By summation that would lead to

$$\lambda(\phi(Q)) \le \int_Q |\det \phi'(x)| \lambda(dx) + \epsilon\lambda(Q),$$

which contradicts the inequality (3).

Now write $C_1 := Q$, $C_2 = Q_1$. Continue thus to obtain, by successively subdividing, a decreasing sequence $C_1 \supseteq C_2 \supseteq C_3 \supseteq \dots$ of dyadic cubes such that $e(C_{n+1}) = 2^{-1}e(C_n)$ for all $n \geq 1$ and for which

$$\frac{\lambda(\phi(C_n))}{\lambda(C_n)} > \frac{1}{\lambda(C_n)} \int_{C_n} |\det \phi'(x)| \lambda(dx) + \epsilon \tag{4}$$

for all n. By the continuity of $|\det \phi'|$ we have

$$\lim_{n\to\infty} \frac{1}{\lambda(C_n)} \int_{C_n} |\det \phi'(x)|\lambda(dx) = |\det \phi'(x_0)|.$$

where $\{x_0\} = \bigcap_{n=1}^{\infty} \overline{C}_n$. And, by Corollary 14,

$$\limsup_{n\to\infty} \frac{\lambda(\phi(C_n))}{\lambda(C_n)} \leq |\det \phi'(x_0)|.$$

These two limits lead, via the inequality (4), to the impossible conclusion that

$$|\det \phi'(x_0)| \geq |\det \phi'(x_0)| + \epsilon.$$

We conclude that every dyadic cube Q such that $\overline{Q} \subseteq U$ satisfies the equation (2). □

Lemma 16. *If G is a open subset of U then $\phi(G)$ is a open subset of V and*

$$\lambda(\phi(G)) \leq \int_G |\det \phi'(x)|\lambda(dx).$$

(*The case in which both terms are infinite is not excluded.*)

Proof. Let G be a non-empty open subset of U. Then we can find a disjoint sequence (Q_n) of cubes whose union is G and which satisfy $\overline{Q_n} \subseteq G$ for all n. Then

$$\lambda(\phi(Q_n)) \leq \int_{Q_n} |\det \phi'(x)|\lambda(dx).$$

Summing, we see that

$$\lambda(\phi(G)) \leq \int_G |\det \phi'(x)|\lambda(dx).$$

□

Lemma 17. *There exists a sequence (K_n) of compact subsets of U such that $\bigcup_{n=1}^{\infty} K_n = U$ and $K_n \subseteq \mathring{K}_{n+1}$ for all n.*

Proof. It suffices to take

$$K_n := \{ x \in U : d(x, \complement U) \geq 1/n, \|x\| \leq n \},$$

where $d(x, \complement U)$ denotes the distance of x from $\complement U$. □

Theorem 18. *If E is a measurable subset of U then $\phi(E)$ is a measurable subset of V and*

$$\lambda(\phi(E)) \leq \int_E |\det \phi'(x)| \, \lambda(dx).$$

Proof. Consider first the case in which E is a Borel set such that $E \subset\subset U$. Then $\overline{E}$ is a compact subset of $U = \bigcup_{n=1}^{\infty} \mathring{K}_n$, so there exists an integer N such that $\overline{E} \subseteq \mathring{K}_N$. By the outer regularity of λ there exists a decreasing sequence (G_n) of open sets such that $E \subseteq G_n \subseteq \mathring{K}_N$ for all n, with $\lambda(G_n) \downarrow \lambda(E)$ as $n \to \infty$. Then $\lambda(G_1) \leq \lambda(K_N) < \infty$, so $\lambda(G_n) \downarrow \lambda(F)$ as $n \to \infty$, where $F = \bigcap_{n=1}^{\infty} G_n$. Thus $E \subseteq F$ with $\lambda(E) = \lambda(F) < \infty$, so $\lambda(F \setminus E) = 0$, and therefore $\mu(E) = \mu(F)$, where

$$\mu(S) = \int_S |\det \phi'(x)| \, \lambda(dx)$$

for each measurable subset S of U.

Observe now that $\mu(G_1) \leq \mu(K_N) < \infty$. Hence, by the countable additivity of the indefinite integral, $\mu(G_n) \downarrow \mu(F) = \mu(E)$ as $n \to \infty$. But, by Lemma 6, $\phi(E) \in \mathfrak{B}(V)$ and by Lemma 16 we have $\lambda(\phi(E)) \leq \lambda(\phi(G_n)) \leq \mu(G_n)$. and therefore $\lambda(\phi(E)) \leq \lim_n \mu(G_n) = \mu(E)$.

Now let E an arbitrary Borel subset of U. Writing $E_n = E \cap K_n$, we see that E_n is a Borel set and that $E_n \subset\subset U$, and thus $\lambda(\phi(E_n)) \leq \mu(E_n)$ for all n. Passing to the limit as $n \to \infty$, we see that $\lambda(\phi(E)) \leq \mu(E)$, as claimed. Note the consequence that if $E \in \mathfrak{B}(U)$ and $\lambda(E) = 0$ then $\lambda(\phi(E)) = 0$.

Now suppose that E is a measurable subset of $\mathbb{R}^k$. Then, by Lemma 4, there exist $A, B \in \mathfrak{B}(\mathbb{R}^k)$ such that $A \subseteq E \subseteq B$ and $\lambda(B \setminus A) = 0$. If also $E \subseteq U$ then, replacing B by $B \cap U$ if necessary, we can suppose that $B \subseteq U$. But then, by Lemma 5, $A, B \in \{ S \in \mathfrak{B}(\mathbb{R}^k) : S \subseteq U \} = \mathfrak{B}(U)$. Consequently, by Lemma 6, $\phi(A), \phi(B) \in \mathfrak{B}(V) \subseteq \mathfrak{B}(\mathbb{R}^k)$. Moreover, $\phi(A) \subseteq \phi(E) \subseteq \phi(B)$ and $\lambda(\phi(B) \setminus \phi(A)) = \lambda(\phi(B \setminus A)) = 0$. Hence, by Lemma 4, the set $\phi(E)$ is measurable and we have

$$\lambda(\phi(E)) = \lambda(\phi(A)) \leq \mu(A) = \mu(E). \qquad \square$$

8. Proof of Theorem 1

Now let F be a measurable subset of V. Reversing the roles of U and V we see, by Theorem 16, that $E := \psi(F)$ is a measurable subset of U. Since $1_E = 1_F \circ \phi$, we have

$$\int_V 1_F(x)\lambda(dx) = \lambda(\phi(E)) \leq \int_U 1_E(x)|\det\phi'(x)|dx$$
$$= \int_U (1_F \circ \phi(x))|\det\phi'(x)|dx.$$

We deduce immediately that

$$\int_V f(x)\lambda(dx) \leq \int_U (f\circ\phi)(x)|\det\phi'(x)|\lambda(dx) \tag{5}$$

for every simple measurable function $f \geq 0$ on V. Then, by passing to the limit via an increasing sequence of simple measurable functions, we conclude that the equation (5) is true for every measurable function $f : V \to \overline{\mathbb{R}}_+$. Now write $g(y) = (f \circ \phi(y))|\det\phi'(y)|$. Then, reversing again the roles of U and V we have

$$\begin{aligned}\int_V f(x)\lambda(dx) &\leq \int_U (f\circ\phi)(x)|\det\phi'(x)|\lambda(dx)\\ &= \int_U g(x)\lambda(dx)\\ &\leq \int_V (g\circ\psi)(x)|\det\psi'(x)|\lambda(dx)\\ &= \int_V (f\circ\phi\circ\psi)(x)|\det\phi'\circ(\psi)||\det\psi'(x)|\lambda(dx)\\ &= \int_V f(x)\lambda(dx)\end{aligned}$$

If f is integrable then so, obviously, is the function $(f\circ\phi)\,|det\phi'|$ over U. The case of a function $f : U \to \overline{\mathbb{R}}$ now follows by consideration of its positive and negative parts. We have thus proved Theorem 1.

9. Concluding remarks

The present subject has a large literature, and the references that follow are only a representative sample. Various proofs of Therem 1 or variants thereof can be found in [1, 2, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19]. Some of these authors obtain variants of Theorem 1 that are valid under weaker conditions than those studied here. For the argument of the present note I have drawn particularly on passages in [4, 6, 11, 12].

It should perhaps be mentioned that some of the terminology in this note is not standard. It should also be noted that, whereas we have defined a **cell** as a product of the form $\prod_{i=1}^k [a_i, b_i)$, many authors, especially in probability theory, prefer to work with products of the form $\prod_{i=1}^{k}(a_i, b_i]$.

References

[1] H. Amann and J. Escher, *Analysis* III, Zweite Auflage, Birkhäuser, Basel, 2008.

[2] E. Asplund and L. Bungart, *A First Course in Integration*, Holt, Rinehart and Winston, New York, 1966.

[3] H. Bauer, *Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie*, W. de Gruyter, Berlin, 1968.

[4] P. Billingsley, *Probability and Measure*, Wiley, New York, 1979.

[5] V. I. Bogachev, *Measure Theory, vol.* I, Springer, Berlin, 2007.

[6] S. D. Chatterji, *Cours d'Analyse:* 1*, Analyse vectorielle*, Presses polytechniques et universitaires romandes, Lausanne, 1997.

[7] J. A. Dieudonné, *Éléments d'analyse, Tome* III, Gauthiers-Villars, Paris, 1970.

[8] W. H. Fleming, *Functions of Several Variables*, Addison-Wesley, Reading, Massachusetts, 1965.

[9] D. H. Fremlin, *Measure Theory*, vol. 2, Torres Fremlin, Colchester, 2001.

[10] F. Jones, *Lebesgue Integration on Euclidean Space*, Jones and Bartlett, Boston MA, 1993.

[11] B. Makarov and A. Podkorytov, *Real Analysis: Measures, Integrals and Applications*, Springer, Heidelberg, 2011.

[12] W. Rudin, *Real and Complex Analysis*, McGraw-Hill, New York, 1966.

[13] R. L. Schilling, *Measures, Integrals and Martingales*, Cambridge University Press, London, 2005.

[14] J. Schwartz, The Formula for Change in Variables in a Multiple Integral, *American Math. Monthly* **61**, No.2, (1954) 81–85.

[15] K. T. Smith, *Primer of Modern Analysis*, Springer, New York, 1983.

[16] K. R. Stromberg, *Introduction to Classical Real Analysis*, Wadsworth, Belmont CA, 1981.

[17] D. W. Strook, *A Concise Introduction to the Theory of Integration*, 2nd edition, Birkhäuser, Boston MA, 1994.

[18] A. J. Weir, *Lebesgue Integration and Measure*, Cambridge University Press, London, 1973.

[19] J. H. Williamson, *Lebesgue Integration*, Holt, Rinehart and Winston, New York, 1962.

Mathematical Institute,
Radcliffe Observatory Quarter,
Woodstock Road
Oxford OX2 6GG

November 2016

David.Edwards@maths.ox.ac.uk