Version of 5.9.03/24.7.04

Chapter 26

Change of Variable in the Integral

I suppose most courses on basic calculus still devote a substantial amount of time to practice in the techniques of integrating standard functions. Surely the most powerful single technique is that of substitution: replacing $\int g(y)dy$ by $\int g(\phi(x))\phi'(x)dx$ for an appropriate function ϕ. At this level one usually concentrates on the skills of guessing at appropriate ϕ and getting the formulae right. I will not address such questions here, except for rare special cases; in this book I am concerned rather with validating the process. For functions of one variable, it can usually be justified by an appeal to the Fundamental Theorem of Calculus, and for any particular case I would normally go first to §225 in the hope that the results there would cover it. But for functions of two or more variables some much deeper ideas are necessary.

I have already treated the general problem of integration-by-substitution in abstract measure spaces in §235. There I described conditions under which $\int g(y)dy = \int g(\phi(x))J(x)dx$ for an appropriate function J. The context there gave very little scope for suggestions as to how to compute J; at best, it could be presented as a Radon-Nikodým derivative (235M). In this chapter I give a form of the fundamental theorem for the case of Lebesgue measure, in which ϕ is a more or less differentiable function between Euclidean spaces, and J is a 'Jacobian', the modulus of the determinant of the derivative of ϕ (263D). This necessarily depends on a serious investigation of the relationship between Lebesgue measure and geometry. The first step is to establish a form of Vitali's theorem for r-dimensional space, together with r-dimensional density theorems; I do this in §261, following closely the scheme of §§221 and 223 above. We need to know quite a lot about differentiable functions between Euclidean spaces, and it turns out that the theory is intertwined with that of 'Lipschitz' functions; I treat these in §262.

In the next two sections of the chapter, I turn to a separate problem for which some of the same techniques turn out to be appropriate: the description of surface measure on (smooth) surfaces in Euclidean space, like the surface of a cone or sphere. I suppose there is no difficulty in forming a robust intuition as to what is meant by the 'area' of such a surface and of suitably simple regions within it, and there is a very strong presumption that there ought to be an expression for this intuition in terms of measure theory as presented in this book; but the details are not I think straightforward. The first point to note is that for any calculation of the area of a region G in a surface S, one would always turn at once to a parametrization of the region, that is, a bijection $\phi : D \to G$ from some subset D of Euclidean space. But obviously one needs to be sure that the result of the calculation is independent of the parametrization chosen, and while it would be possible to base the theory on results showing such independence directly, that does not seem to me to be a true reflection of the underlying intuition, which is that the area of simple surfaces, at least, is something intrinsic to their geometry. I therefore see no acceptable alternative to a theory of 'r-dimensional measure' which can be described in purely geometric terms. This is the burden of §264, in which I give the definition and most fundamental properties of Hausdorff r-dimensional measure in Euclidean spaces. With this established, we find that the techniques of §§261-263 are sufficient to relate it to calculations through parametrizations, which is what I do in §265.

The chapter ends with a brief account of the Brunn-Minkowski inequality (266C), which is an essential tool for the geometric measure theory of convex sets.

Version of 11.12.12

261 Vitali's theorem in $\mathbb{R}^r$

The main aim of this section is to give r-dimensional versions of Vitali's theorem and Lebesgue's Density Theorem, following ideas already presented in §§221 and 223. I end with a proof that Lebesgue outer measure can be defined in terms of coverings by balls instead of by intervals (261F).

261A Notation For most of this chapter, we shall be dealing with the geometry and measure of Euclidean space; it will save space to fix some notation.

Throughout this section and the two following, $r \ge 1$ will be an integer. I will use Roman letters for members of $\mathbb{R}^r$ and Greek letters for their coordinates, so that $a = (\alpha_1, \dots, \alpha_r)$, etc.; if you see any Greek letter with a subscript you should look first for a nearby vector of which it might be a coordinate. The measure under consideration will nearly always be Lebesgue measure on $\mathbb{R}^r$; so unless otherwise indicated μ should be interpreted as Lebesgue measure, and μ^* as Lebesgue outer measure. Similarly, $\int \dots dx$ will always be integration with respect to Lebesgue measure (in a dimension determined by the context).

For $x = (\xi_1, \dots, \xi_r) \in \mathbb{R}^r$, write $\|x\| = \sqrt{\xi_1^2 + \dots + \xi_r^2}$. Recall that $\|x+y\| \le \|x\| + \|y\|$ (1A2C) and that $\|\alpha x\| = |\alpha|\|x\|$ for any vectors x, y and scalar α.

I will use the same notation as in §115 for 'intervals', so that, in particular,

$$[a, b[= \{x : \alpha_i \le \xi_i < \beta_i \ \forall\, i \le r\},$$

$$]a, b[= \{x : \alpha_i < \xi_i < \beta_i \ \forall\, i \le r\},$$

$$[a, b] = \{x : \alpha_i \le \xi_i \le \beta_i \ \forall\, i \le r\}$$

whenever a, $b \in \mathbb{R}^r$.

$\mathbf{0} = (0, \dots, 0)$ will be the zero vector in $\mathbb{R}^r$, and $\mathbf{1}$ will be $(1, \dots, 1)$. If $x \in \mathbb{R}^r$ and $\delta > 0$, $B(x, \delta)$ will be the closed ball with centre x and radius δ, that is, $\{y : y \in \mathbb{R}^r, \|y - x\| \le \delta\}$. Note that $B(x, \delta) = x + B(\mathbf{0}, \delta)$; so that by the translation-invariance of Lebesgue measure we have

$$\mu B(x, \delta) = \mu B(\mathbf{0}, \delta) = \beta_r \delta^r,$$

where

$$\beta_r = \frac{1}{k!}\pi^k \text{ if } r = 2k \text{ is even,}$$
$$= \frac{2^{2k+1}k!}{(2k+1)!}\pi^k \text{ if } r = 2k+1 \text{ is odd}$$

(252Q).

261B Vitali's theorem in $\mathbb{R}^r$ Let $A \subseteq \mathbb{R}^r$ be any set, and $\mathcal{I}$ a family of closed non-trivial (that is, non-singleton, or, equivalently, non-negligible) balls in $\mathbb{R}^r$ such that every point of A is contained in arbitrarily small members of $\mathcal{I}$. Then there is a countable disjoint set $\mathcal{I}_0 \subseteq \mathcal{I}$ such that $\mu(A \setminus \bigcup \mathcal{I}_0) = 0$.

proof (a) To begin with (down to the end of (f) below), suppose that $\|x\| < M$ for every $x \in A$, and set

$$\mathcal{I}' = \{I : I \in \mathcal{I},\ I \subseteq B(\mathbf{0}, M)\}.$$

If there is a finite disjoint set $\mathcal{I}_0 \subseteq \mathcal{I}'$ such that $A \subseteq \bigcup \mathcal{I}_0$ (including the possibility that $A = \mathcal{I}_0 = \emptyset$), we can stop. So let us suppose henceforth that there is no such $\mathcal{I}_0$.

(b) In this case, if $\mathcal{I}_0$ is any finite disjoint subset of $\mathcal{I}'$, there is a $J \in \mathcal{I}'$ which is disjoint from any member of $\mathcal{I}_0$. **P** Take $x \in A \setminus \bigcup \mathcal{I}_0$. Because every member of $\mathcal{I}_0$ is closed, there is a $\delta > 0$ such that $B(x, \delta)$ does not meet any member of $\mathcal{I}_0$, and as $\|x\| < M$ we can suppose that $B(x, \delta) \subseteq B(\mathbf{0}, M)$. Let J be a member of $\mathcal{I}$, containing x, and of diameter at most δ; then $J \in \mathcal{I}'$ and $J \cap \bigcup \mathcal{I}_0 = \emptyset$. **Q**

(c) We can therefore choose a sequence $\langle \gamma_n \rangle_{n \in \mathbb{N}}$ of real numbers and a disjoint sequence $\langle I_n \rangle_{n \in \mathbb{N}}$ in $\mathcal{I}'$ inductively, as follows. Given $\langle I_j \rangle_{j<n}$ (if $n = 0$, this is the empty sequence, with no members), with $I_j \in \mathcal{I}'$ for each $j < n$, and $I_j \cap I_k = \emptyset$ for $j < k < n$, set $\mathcal{J}_n = \{I : I \in \mathcal{I}',\ I \cap I_j = \emptyset \text{ for every } j < n\}$. We know from (b) that $\mathcal{J}_n \ne \emptyset$. Set

$$\gamma_n = \sup\{\operatorname{diam} I : I \in \mathcal{J}_n\};$$

then $\gamma_n \le 2M$, because every member of $\mathcal{J}_n$ is included in $B(\mathbf{0},M)$. We can therefore find a set $I_n \in \mathcal{J}_n$ such that $\operatorname{diam} I_n \ge \frac{1}{2}\gamma_n$, and this continues the induction.

(e) Because the I_n are disjoint measurable subsets of the bounded set $B(\mathbf{0},M)$, we have

$$\sum_{n=0}^{\infty} \mu I_n \le \mu B(\mathbf{0},M) < \infty,$$

and $\lim_{n\to\infty} \mu I_n = 0$. Also $\mu I_n \ge \beta_r(\frac{1}{4}\gamma_n)^r$ for each n, so $\lim_{n\to\infty} \gamma_n = 0$.

Now define I'_n to be the closed ball with the same centre as I_n but five times the diameter, so that it contains every point within a distance γ_n of I_n. I claim that, for any n, $A \subseteq \bigcup_{j<n} I_j \cup \bigcup_{j\ge n} I'_j$. **P?** Suppose, if possible, otherwise. Take any $x \in A \setminus (\bigcup_{j<n} I_j \cup \bigcup_{j\ge n} I'_j)$. Let $\delta > 0$ be such that

$$B(x,\delta) \subseteq B(\mathbf{0},M) \setminus \bigcup_{j<n} I_j,$$

and let $J \in \mathcal{I}$ be such that $x \in J \subseteq B(x,\delta)$. Then

$$\lim_{m\to\infty} \gamma_m = 0 < \operatorname{diam} J$$

(this is where we use the hypothesis that all the balls in $\mathcal{I}$ are non-trivial); let m be the least integer greater than or equal to n such that $\gamma_m < \operatorname{diam} J$. In this case J cannot belong to $\mathcal{J}_m$, so there must be some $k < m$ such that $J \cap I_k \ne \emptyset$, because certainly $J \in \mathcal{I}'$. By the choice of δ, k cannot be less than n, so $n \le k < m$, and $\gamma_k \ge \operatorname{diam} J$. So the distance from x to the nearest point of I_k is at most $\operatorname{diam} J \le \gamma_k$. But this means that $x \in I'_k$; which contradicts the choice of x. **XQ**

(f) It follows that

$$\mu^*(A \setminus \bigcup_{j<n} I_j) \le \mu(\bigcup_{j\ge n} I'_j) \le \sum_{j=n}^{\infty} \mu I'_j \le 5^r \sum_{j=n}^{\infty} \mu I_j.$$

As

$$\sum_{j=0}^{\infty} \mu I_j \le \mu B(\mathbf{0},M) < \infty,$$

$\lim_{n\to\infty} \mu^*(A \setminus \bigcup_{j<n} I_j) = 0$ and

$$\mu(A \setminus \bigcup_{j\in\mathbb{N}} I_j) = \mu^*(A \setminus \bigcup_{j\in\mathbb{N}} I_j) = 0.$$

Thus in this case we may set $\mathcal{I}_0 = \{I_n : n \in \mathbb{N}\}$ to obtain a countable disjoint family in $\mathcal{I}$ with $\mu(A \setminus \bigcup \mathcal{I}_0) = 0$.

(g) This completes the proof if A is bounded. In general, set

$$U_n = \{x : x \in \mathbb{R}^r,\ n < \|x\| < n+1\}, \quad A_n = A \cap U_n, \quad \mathcal{J}_n = \{I : I \in \mathcal{I},\ I \subseteq U_n\},$$

for each $n \in \mathbb{N}$. Then for each n we see that every point of A_n belongs to arbitrarily small members of $\mathcal{J}_n$, so there is a countable disjoint $\mathcal{J}'_n \subseteq \mathcal{J}_n$ such that $A_n \setminus \bigcup \mathcal{J}'_n$ is negligible. Now (because the U_n are disjoint) $\mathcal{I}_0 = \bigcup_{n\in\mathbb{N}} \mathcal{J}'_n$ is disjoint, and of course $\mathcal{I}_0$ is a countable subset of $\mathcal{I}$; moreover,

$$A \setminus \bigcup \mathcal{I}_0 \subseteq (\mathbb{R}^r \setminus \bigcup_{n\in\mathbb{N}} U_n) \cup \bigcup_{n\in\mathbb{N}} (A_n \setminus \bigcup \mathcal{J}'_n)$$

is negligible. (To see that $\mathbb{R}^r \setminus \bigcup_{n\in\mathbb{N}} U_n = \{x : \|x\| \in \mathbb{N}\}$ is negligible, note that for any $n \in \mathbb{N}$ the set

$$\{x : \|x\| = n\} \subseteq B(\mathbf{0},n) \setminus B(\mathbf{0},\delta n)$$

has measure at most $\beta_r n^r - \beta_r(\delta n)^r$ for every $\delta \in [0,1[$, so must be negligible.)

261C Just as in §223, we can use the r-dimensional Vitali theorem to prove theorems on the approximation of functions by their local mean values.

Density Theorem in $\mathbb{R}^r$: integral form Let D be a subset of $\mathbb{R}^r$, and f a real-valued function which is integrable over D. Then

$$f(x) = \lim_{\delta\downarrow 0} \frac{1}{\mu B(x,\delta)} \int_{D\cap B(x,\delta)} f d\mu$$

for almost every $x \in D$.

proof (a) To begin with (down to the end of (b)), let us suppose that $D = \operatorname{dom} f = \mathbb{R}^r$.

Take $n \in \mathbb{N}$ and q, $q' \in \mathbb{Q}$ with $q < q'$, and set

$$A = A_{nqq'} = \{x : \|x\| \le n,\ f(x) \le q,\ \limsup_{\delta\downarrow 0} \frac{1}{\mu B(x,\delta)} \int_{B(x,\delta)} f d\mu > q'\}.$$

? Suppose, if possible, that $\mu^* A > 0$. Let $\epsilon > 0$ be such that $\epsilon(1 + |q|) < (q' - q)\mu^* A$, and let $\eta \in \,]0, \epsilon]$ be such that $\int_E |f| \le \epsilon$ whenever $\mu E \le \eta$ (225A). Let $G \supseteq A$ be an open set of measure at most $\mu^* A + \eta$ (134Fa). Let $\mathcal{I}$ be the set of non-trivial closed balls $B \subseteq G$ such that $\frac{1}{\mu B}\int_B f d\mu \ge q'$. Then every point of A is contained in (indeed, is the centre of) arbitrarily small members of $\mathcal{I}$. So there is a countable disjoint set $\mathcal{I}_0 \subseteq \mathcal{I}$ such that $\mu(A \setminus \bigcup \mathcal{I}_0) = 0$, by 261B; set $H = \bigcup \mathcal{I}_0$.

Because $\int_I f d\mu \ge q'\mu I$ for each $I \in \mathcal{I}_0$, we have

$$\int_H f d\mu = \sum_{I\in\mathcal{I}_0} \int_I f d\mu \ge q' \sum_{I\in\mathcal{I}_0} \mu I = q'\mu H \ge q'\mu^* A.$$

Set

$$E = \{x : x \in G,\ f(x) \le q\}.$$

Then E is measurable, and $A \subseteq E \subseteq G$; so

$$\mu^* A \le \mu E \le \mu G \le \mu^* A + \eta \le \mu^* A + \epsilon.$$

Also

$$\mu(H \setminus E) \le \mu G - \mu E \le \eta,$$

so by the choice of η, $\int_{H\setminus E} f \le \epsilon$ and

$$\int_H f \le \epsilon + \int_{H\cap E} f \le \epsilon + q\mu(H \cap E)$$
$$\le \epsilon + q\mu^* A + |q|(\mu(H \cap E) - \mu^* A) \le q\mu^* A + \epsilon(1 + |q|)$$

(because $\mu^* A = \mu^*(A \cap H) \le \mu(H \cap E) \le \mu E$)

$$< q'\mu^* A \le \int_H f,$$

which is impossible. **X**

Thus $A_{nqq'}$ is negligible. This is true for all $q < q'$ and all n, so

$$A^* = \bigcup_{q,q'\in\mathbb{Q},q<q'} \bigcup_{n\in\mathbb{N}} A_{nqq'}$$

is negligible. But

$$f(x) \ge \limsup_{\delta\downarrow 0} \frac{1}{\mu B(x,\delta)} \int_{B(x,\delta)} f$$

for every $x \in \mathbb{R}^r \setminus A^*$, that is, for almost all $x \in \mathbb{R}^r$.

(b) Similarly, or applying this result to $-f$.

$$f(x) \le \liminf_{\delta\downarrow 0} \frac{1}{\mu B(x,\delta)} \int_{B(x,\delta)} f$$

for almost every x, so

$$f(x) = \lim_{\delta\downarrow 0} \frac{1}{\mu B(x,\delta)} \int_{B(x,\delta)} f$$

for almost every x.

(c) For the (superficially) more general case enunciated in the theorem, let $\tilde f$ be a μ-integrable function extending $f{\restriction}D$, defined everywhere on $\mathbb{R}^r$, and such that $\int_F \tilde f = \int_{D\cap F} f$ for every measurable $F \subseteq \mathbb{R}^r$ (applying 214Eb to $f{\restriction}D$). Then

$$f(x) = \tilde f(x) = \lim_{\delta\downarrow 0} \frac{1}{\mu B(x,\delta)} \int_{B(x,\delta)} \tilde f = \lim_{\delta\downarrow 0} \frac{1}{\mu B(x,\delta)} \int_{D\cap B(x,\delta)} f$$

for almost every $x \in D$.

261D Corollary (a) If $D \subseteq \mathbb{R}^r$ is any set, then

$$\lim_{\delta\downarrow 0} \frac{\mu^*(D\cap B(x,\delta))}{\mu B(x,\delta)} = 1$$

for almost every $x \in D$.

(b) If $E \subseteq \mathbb{R}^r$ is a measurable set, then

$$\lim_{\delta\downarrow 0} \frac{\mu(E\cap B(x,\delta))}{\mu B(x,\delta)} = \chi E(x)$$

for almost every $x \in \mathbb{R}^r$.

(c) If $D \subseteq \mathbb{R}^r$ and $f : D \to \mathbb{R}$ is any function, then for almost every $x \in D$,

$$\lim_{\delta\downarrow 0} \frac{\mu^*(\{y:y\in D, |f(y)-f(x)|\le\epsilon\}\cap B(x,\delta))}{\mu B(x,\delta)} = 1$$

for every $\epsilon > 0$.

(d) If $D \subseteq \mathbb{R}^r$ and $f : D \to \mathbb{R}$ is measurable, then for almost every $x \in D$,

$$\lim_{\delta\downarrow 0} \frac{\mu^*(\{y:y\in D, |f(y)-f(x)|\ge\epsilon\}\cap B(x,\delta))}{\mu B(x,\delta)} = 0$$

for every $\epsilon > 0$.

proof (a) Apply 261C with $f = \chi B(\mathbf{0}, n)$ to see that, for any $n \in \mathbb{N}$,

$$\lim_{\delta\downarrow 0} \frac{\mu^*(D\cap B(x,\delta))}{\mu B(x,\delta)} = 1$$

for almost every $x \in D$ with $\|x\| < n$.

(b) Apply (a) to E to see that

$$\liminf_{\delta\downarrow 0} \frac{\mu(E\cap B(x,\delta))}{\mu B(x,\delta)} \ge \chi E(x)$$

for almost every $x \in \mathbb{R}^r$, and to $E' = \mathbb{R}^r \setminus E$ to see that

$$\limsup_{\delta\downarrow 0} \frac{\mu(E\cap B(x,\delta))}{\mu B(x,\delta)} = 1 - \liminf_{\delta\downarrow 0} \frac{\mu(E'\cap B(x,\delta))}{\mu B(x,\delta)} \le 1 - \chi E'(x) = \chi E(x)$$

for almost every x.

(c) For q, $q' \in \mathbb{Q}$, set

$$D_{qq'} = \{x : x \in D,\ q \le f(x) \le q'\},$$

$$C_{qq'} = \{x : x \in D_{qq'},\ \lim_{\delta\downarrow 0} \frac{\mu^*(D_{qq'}\cap B(x,\delta))}{\mu B(x,\delta)} = 1\};$$

now set

$$C = D \setminus \bigcup_{q,q'\in\mathbb{Q}} (D_{qq'} \setminus C_{qq'}),$$

so that $D \setminus C$ is negligible. If $x \in C$ and $\epsilon > 0$, then there are q, $q' \in \mathbb{Q}$ such that $f(x) - \epsilon \le q \le f(x) \le q' \le f(x) + \epsilon$, and now $x \in C_{qq'}$; accordingly

$$\liminf_{\delta\downarrow 0} \frac{\mu^*\{y:y\in D\cap B(x,\delta), |f(y)-f(x)|\le\epsilon\}}{\mu B(x,\delta)} \ge \liminf_{\delta\downarrow 0} \frac{\mu^*(D_{qq'}\cap B(x,\delta))}{\mu B(x,\delta)} = 1,$$

so

$$\lim_{\delta\downarrow 0} \frac{\mu^*\{y:y\in D\cap B(x,\delta), |f(y)-f(x)|\le\epsilon\}}{\mu B(x,\delta)} = 1.$$

(d) Define C as in (c). We know from (a) that $\mu(D \setminus C') = 0$, where

$$C' = \{x : x \in D, \lim_{\delta\downarrow 0} \frac{\mu^*(D\cap B(x,\delta))}{\mu B(x,\delta)} = 1\}.$$

If $x \in C \cap C'$ and $\epsilon > 0$, we know from (c) that

$$\lim_{\delta\downarrow 0}\frac{\mu^*\{y:y\in D\cap B(x,\delta),\,|f(y)-f(x)|\le\epsilon/2\}}{\mu B(x,\delta)}=1.$$

But because f is measurable, we have

$$\mu^*\{y:y\in D\cap B(x,\delta),\,|f(y)-f(x)|\ge\epsilon\}$$
$$+\mu^*\{y:y\in D\cap B(x,\delta),\,|f(y)-f(x)|\le\frac{1}{2}\epsilon\}\le\mu^*(D\cap B(x,\delta))$$

for every $\delta>0$. Accordingly

$$\limsup_{\delta\downarrow 0}\frac{\mu^*\{y:y\in D\cap B(x,\delta),\,|f(y)-f(x)|\ge\epsilon\}}{\mu B(x,\delta)}$$
$$\le\lim_{\delta\downarrow 0}\frac{\mu^*(D\cap B(x,\delta))}{\mu B(x,\delta)}-\lim_{\delta\downarrow 0}\frac{\mu^*\{y:y\in D\cap B(x,\delta),\,|f(y)-f(x)|\le\epsilon/2\}}{\mu B(x,\delta)}=0,$$

and

$$\lim_{\delta\downarrow 0}\frac{\mu^*\{y:y\in D\cap B(x,\delta),\,|f(y)-f(x)|\ge\epsilon\}}{\mu B(x,\delta)}=0$$

for every $x\in C\cap C'$, that is, for almost every $x\in D$.

261E Theorem Let f be a locally integrable function defined on a conegligible subset of $\mathbb{R}^r$. Then

$$\lim_{\delta\downarrow 0}\frac{1}{\mu B(x,\delta)}\int_{B(x,\delta)}|f(y)-f(x)|dy=0$$

for almost every $x\in\mathbb{R}^r$.

proof (Compare 223D.)

(a) Fix $n\in\mathbb{N}$ for the moment, and set $G=\{x:\|x\|<n\}$. For each $q\in\mathbb{Q}$, set $g_q(x)=|f(x)-q|$ for $x\in G\cap\operatorname{dom}f$; then g_q is integrable over G, and

$$\lim_{\delta\downarrow 0}\frac{1}{\mu B(x,\delta)}\int_{G\cap B(x,\delta)}g_q=g_q(x)$$

for almost every $x\in G$, by 261C. Setting

$$E_q=\{x:x\in G\cap\operatorname{dom}f,\ \lim_{\delta\downarrow 0}\frac{1}{\mu B(x,\delta)}\int_{G\cap B(x,\delta)}g_q=g_q(x)\},$$

we have $G\setminus E_q$ negligible for every q, so $G\setminus E$ is negligible, where $E=\bigcap_{q\in\mathbb{Q}}E_q$. Now

$$\lim_{\delta\downarrow 0}\frac{1}{\mu B(x,\delta)}\int_{G\cap B(x,\delta)}|f(y)-f(x)|dy=0$$

for every $x\in E$. **P** Take $x\in E$ and $\epsilon>0$. Then there is a $q\in\mathbb{Q}$ such that $|f(x)-q|\le\epsilon$, so that

$$|f(y)-f(x)|\le|f(y)-q|+\epsilon=g_q(y)+\epsilon$$

for every $y\in G\cap\operatorname{dom}f$, and

$$\limsup_{\delta\downarrow 0}\frac{1}{\mu B(x,\delta)}\int_{G\cap B(x,\delta)}|f(y)-f(x)|dy\le\limsup_{\delta\downarrow 0}\frac{1}{\mu B(x,\delta)}\int_{G\cap B(x,\delta)}g_q(y)+\epsilon\,dy$$
$$=\epsilon+g_q(x)\le 2\epsilon.$$

As ϵ is arbitrary,

$$\lim_{\delta\downarrow 0}\frac{1}{\mu B(x,\delta)}\int_{G\cap B(x,\delta)}|f(y)-f(x)|dy=0,$$

as required. **Q**

(b) Because G is open,

$$\lim_{\delta\downarrow 0}\frac{1}{\mu B(x,\delta)}\int_{B(x,\delta)}|f(y)-f(x)|dy=\lim_{\delta\downarrow 0}\frac{1}{\mu B(x,\delta)}\int_{G\cap B(x,\delta)}|f(y)-f(x)|dy=0$$

for almost every $x\in G$. As n is arbitrary,

$$\lim_{\delta\downarrow 0}\frac{1}{\mu B(x,\delta)}\int_{B(x,\delta)}|f(y)-f(x)|dy=0$$

for almost every $x\in\mathbb{R}^r$.

Remark The set

$$\{x:x\in\operatorname{dom}f,\,\lim_{\delta\downarrow 0}\frac{1}{\mu B(x,\delta)}\int_{B(x,\delta)}|f(y)-f(x)|dy=0\}$$

is sometimes called the **Lebesgue set** of f.

261F Another very useful consequence of 261B is the following.

Proposition Let $A\subseteq\mathbb{R}^r$ be any set, and $\epsilon>0$. Then there is a sequence $\langle B_n\rangle_{n\in\mathbb{N}}$ of closed balls in $\mathbb{R}^r$, all of radius at most ϵ, such that $A\subseteq\bigcup_{n\in\mathbb{N}}B_n$ and $\sum_{n=0}^{\infty}\mu B_n\le\mu^*A+\epsilon$. Moreover, we may suppose that the balls in the sequence whose centres do not lie in A have measures summing to at most ϵ.

proof (a) The first step is the obvious remark that if $x\in\mathbb{R}^r$, $\delta>0$ then the half-open cube $I=[x,x+\delta\mathbf{1}[$ is a subset of the ball $B(x,\delta\sqrt{r})$, which has measure $\gamma_r\delta^r=\gamma_r\mu I$, where $\gamma_r=\beta_r r^{r/2}$. It follows that if $G\subseteq\mathbb{R}^r$ is any open set, then G can be covered by a sequence of balls of total measure at most $\gamma_r\mu G$. **P** If G is empty, we can take all the balls to be singletons. Otherwise, for each $k\in\mathbb{N}$, set

$$Q_k=\{z:z\in\mathbb{Z}^r,\ \big[2^{-k}z,2^{-k}(z+\mathbf{1})\big[\subseteq G\},$$

$$E_k=\bigcup_{z\in Q_k}\big[2^{-k}z,2^{-k}(z+\mathbf{1})\big[.$$

Then $\langle E_k\rangle_{k\in\mathbb{N}}$ is a non-decreasing sequence of sets with union G, and E_0 and each of the differences $E_{k+1}\setminus E_k$ is expressible as a disjoint union of half-open cubes. Thus G also is expressible as a disjoint union of a sequence $\langle I_n\rangle_{n\in\mathbb{N}}$ of half-open cubes. Each I_n is covered by a ball B_n of measure $\gamma_r\mu I_n$; so that $G\subseteq\bigcup_{n\in\mathbb{N}}B_n$ and

$$\sum_{n=0}^{\infty}\mu B_n\le\gamma_r\sum_{n=0}^{\infty}\mu I_n=\gamma_r\mu G.\ \mathbf{Q}$$

(b) It follows at once that if $\mu A=0$ then for any $\epsilon>0$ there is a sequence $\langle B_n\rangle_{n\in\mathbb{N}}$ of balls covering A of measures summing to at most ϵ, because there is certainly an open set including A with measure at most ϵ/γ_r.

(c) Now take any set A, and $\epsilon>0$. Let $G\supseteq A$ be an open set with $\mu G\le\mu^*A+\frac{1}{2}\epsilon$. Let $\mathcal{I}$ be the family of non-trivial closed balls included in G, of radius at most ϵ and with centres in A. Then every point of A belongs to arbitrarily small members of $\mathcal{I}$, so there is a countable disjoint $\mathcal{I}_0\subseteq\mathcal{I}$ such that $\mu(A\setminus\bigcup\mathcal{I}_0)=0$. Let $\langle B'_n\rangle_{n\in\mathbb{N}}$ be a sequence of balls covering $A\setminus\bigcup\mathcal{I}_0$ with $\sum_{n=0}^{\infty}\mu B'_n\le\min(\frac{1}{2}\epsilon,\beta_r\epsilon^r)$; these surely all have radius at most ϵ. Let $\langle B_n\rangle_{n\in\mathbb{N}}$ be a sequence amalgamating $\mathcal{I}_0$ with $\langle B'_n\rangle_{n\in\mathbb{N}}$; then $A\subseteq\bigcup_{n\in\mathbb{N}}B_n$, every B_n has radius at most ϵ and

$$\sum_{n=0}^{\infty}\mu B_n=\sum_{B\in\mathcal{I}_0}\mu B+\sum_{n=0}^{\infty}\mu B'_n\le\mu G+\frac{1}{2}\epsilon\le\mu A+\epsilon,$$

while the B_n whose centres do not lie in A must come from the sequence $\langle B'_n\rangle_{n\in\mathbb{N}}$, so their measures sum to at most $\frac{1}{2}\epsilon\le\epsilon$.

Remark In fact we can (if A is not empty) arrange that the centre of *every* B_n belongs to A. This is an easy consequence of Besicovitch's Covering Lemma (see §472 in Volume 4).

261X Basic exercises (a) Show that 261C is valid for any locally integrable real-valued function f; in particular, for any $f\in\mathcal{L}^p(\mu_D)$ for any $p\ge 1$, writing μ_D for the subspace measure on D.

(b) Show that 261C, 261Dc, 261Dd and 261E are valid for complex-valued functions f.

>(c) Take three disks in the plane, each touching the other two, so that they enclose an open region R with three cusps. In R let D be a disk tangent to each of the three original disks, and R_0, R_1, R_2 the three components of $R \setminus D$. In each R_j let D_j be a disk tangent to each of the disks bounding R_j, and R_{j0}, R_{j1}, R_{j2} the three components of $R_j \setminus D_j$. Continue, obtaining 27 regions at the next step, 81 regions at the next, and so on.

Show that the total area of the residual regions converges to zero as the process continues indefinitely. (*Hint*: compare with the process in the proof of 261B.)

261Y Further exercises (a) Formulate an abstract definition of 'Vitali cover', meaning a family of sets satisfying the conclusion of 261B in some sense, and corresponding generalizations of 261C-261E, covering (at least) (b)-(d) below.

(b) For $x \in \mathbb{R}^r$, $k \in \mathbb{N}$ let $C(x,k)$ be the half-open cube of the form $\bigl[2^{-k}z, 2^{-k}(z+\mathbf{1})\bigr[$, with $z \in \mathbb{Z}^r$, containing x. Show that if f is an integrable function on $\mathbb{R}^r$ then

$$\lim_{k\to\infty} 2^{kr}\int_{C(x,k)} f = f(x)$$

for almost every $x \in \mathbb{R}^r$.

(c) Let f be a real-valued function which is integrable over $\mathbb{R}^r$. Show that

$$\lim_{\delta\downarrow 0} \frac{1}{\delta^r}\int_{[x,x+\delta\mathbf{1}[} f = f(x)$$

for almost every $x \in \mathbb{R}^r$.

(d) Give $X = \{0,1\}^{\mathbb{N}}$ its usual measure ν (254J). For $x \in X$, $k \in \mathbb{N}$ set $C(x,k) = \{y : y \in X,\ y(i) = x(i)$ for $i < k\}$. Show that if f is any real-valued function which is integrable over X then $\lim_{k\to\infty} 2^k \int_{C(x,k)} f d\nu = f(x)$, $\lim_{k\to\infty} 2^k \int_{C(x,k)} |f(y) - f(x)|\nu(dy) = 0$ for almost every $x \in X$.

(e) Let f be a real-valued function which is integrable over $\mathbb{R}^r$, and x a point in the Lebesgue set of f. Show that for every $\epsilon > 0$ there is a $\delta > 0$ such that $|f(x) - \int f(x-y)g(\|y\|)dy| \le \epsilon$ whenever $g : [0,\infty[\to [0,\infty[$ is a non-increasing function such that $\int_{\mathbb{R}^r} g(\|y\|)dy = 1$ and $\int_{B(\mathbf{0},\delta)} g(\|y\|)dy \ge 1-\delta$. (*Hint*: 223Yg.)

(f) Let $\mathfrak{T}$ be the family of those measurable sets $G \subseteq \mathbb{R}^r$ such that $\lim_{\delta\downarrow 0} \frac{\mu(G\cap B(x,\delta))}{\mu B(x,\delta)} = 1$ for every $x \in G$. Show that $\mathfrak{T}$ is a topology on $\mathbb{R}^r$, the **density topology** of $\mathbb{R}^r$. Show that a function $f : \mathbb{R}^r \to \mathbb{R}$ is measurable iff it is $\mathfrak{T}$-continuous at almost every point of $\mathbb{R}^r$.

(g) A set $A \subseteq \mathbb{R}^r$ is said to be **porous** at $x \in \mathbb{R}^r$ if $\limsup_{y\to x} \frac{\rho(y,A)}{\|y-x\|} > 0$, writing $\rho(y,A) = \inf_{z\in A} \|y-z\|$ (or ∞ if A is empty). (i) Show that if A is porous at all its points then it is negligible. (ii) Show that in the construction of 261B the residual set $A \setminus \bigcup \mathcal{I}_0$ will be porous at all its points.

(h) Let $A \subseteq \mathbb{R}^r$ be a bounded set and $\mathcal{I}$ a non-empty family of non-trivial closed balls covering A. Show that for any $\epsilon > 0$ there are disjoint $B_0, \dots, B_n \in \mathcal{I}$ such that $\mu^* A \le (3+\epsilon)^r \sum_{k=0}^n \mu B_k$.

(i) Let (X,ρ) be a metric space and $A \subseteq X$ any set, $x \mapsto \delta_x : A \to [0,\infty[$ any bounded function. Show that if $\gamma > 3$ then there is an $A' \subseteq A$ such that (i) $\rho(x,y) > \delta_x + \delta_y$ for all distinct x, $y \in A'$ (ii) $\bigcup_{x\in A} B(x,\delta_x) \subseteq \bigcup_{x\in A'} B(x,\gamma\delta_x)$, writing $B(x,\alpha)$ for the closed ball $\{y : \rho(y,x) \le \alpha\}$.

(j)(i) Let $\mathcal{C}$ be the family of those measurable sets $C \subseteq \mathbb{R}^r$ such that $\limsup_{\delta\downarrow 0} \frac{\mu(C\cap B(x,\delta))}{\mu B(x,\delta)} > 0$ for every $x \in C$. Show that $\bigcup \mathcal{C}_0 \in \mathcal{C}$ for every $\mathcal{C}_0 \subseteq \mathcal{C}$. (*Hint*: 215B(iv).) (ii) Show that any union of non-trivial closed balls in $\mathbb{R}^r$ is Lebesgue measurable.

(k) Suppose that $A \subseteq \mathbb{R}^r$ and that $\mathcal{I}$ is a family of closed subsets of $\mathbb{R}^r$ such that

for every $x \in A$ there is an $\eta > 0$ such that for every $\epsilon > 0$ there is an $I \in \mathcal{I}$ such that $x \in I$ and $0 < \eta(\operatorname{diam} I)^r \le \mu I \le \epsilon$.

Show that there is a countable disjoint set $\mathcal{I}_0 \subseteq \mathcal{I}$ such that $A \setminus \bigcup \mathcal{I}_0$ is negligible.

(l) Let $\mathfrak{T}'$ be the family of measurable sets $G \subseteq \mathbb{R}^r$ such that whenever $x \in G$ and $\epsilon > 0$ there is a $\delta > 0$ such that $\mu(G \cap I) \geq (1-\epsilon)\mu I$ whenever I is an interval containing x and included in $B(x,\delta)$. Show that $\mathfrak{T}'$ is a topology on $\mathbb{R}^r$ intermediate between the density topology (261Yf) and the Euclidean topology.

261 Notes and comments In the proofs of 261B-261E above, I have done my best to follow the lines of the one-dimensional case; this section amounts to a series of generalizations of the work of §§221 and 223.

It will be clear that the idea of 261A/261B can be used on other shapes than balls. To make it work in the form above, we need a family $\mathcal{I}$ such that there is a constant K for which

$$\mu I' \leq K\mu I$$

for every $I \in \mathcal{I}$, where we write

$$I' = \{x : \inf_{y \in I} \|x-y\| \leq \operatorname{diam}(I)\}.$$

Evidently this will be true for many classes $\mathcal{I}$ determined by the shapes of the sets involved; for instance, if $E \subseteq \mathbb{R}^r$ is any bounded set of strictly positive measure, the family $\mathcal{I} = \{x+\delta E : x \in \mathbb{R}^r,\ \delta > 0\}$ will satisfy the condition.

In 261Ya I challenge you to find an appropriate generalization of the arguments depending on the conclusion of 261B.

Another way of using 261B is to say that because sets can be essentially covered by *disjoint* sequences of balls, it ought to be possible to use balls, rather than half-open intervals, in the definition of Lebesgue measure on $\mathbb{R}^r$. This is indeed so (261F). The difficulty in using balls in the basic definition comes right at the start, in proving that if a ball is covered by finitely many balls then the sum of the volumes of the covering balls is at least the volume of the covered ball. (There is a trick, using the compactness of closed balls and the openness of open balls, to extend such a proof to infinite covers.) Of course you could regard this fact as 'elementary', on the ground that Archimedes would have noticed if it weren't true, but nevertheless it would be something of a challenge to prove it, unless you were willing to wait for a version of Fubini's theorem, as some authors do.

I have given the results in 261C-261D for arbitrary subsets D of $\mathbb{R}^r$ not because I have any applications in mind in which non-measurable subsets are significant, but because I wish to make it possible to notice when measurability matters. Of course it is necessary to interpret the integrals $\int_D f d\mu$ in the way laid down in §214. The game is given away in part (c) of the proof of 261C, where I rely on the fact that if f is integrable over D then there is an integrable $\tilde{f} : \mathbb{R}^r \to \mathbb{R}$ such that $\int_F \tilde{f} = \int_{D \cap F} f$ for every measurable $F \subseteq \mathbb{R}^r$. In effect, for all the questions dealt with here, we can replace f, D by $\tilde{f}$, $\mathbb{R}^r$.

The idea of 261C is that, for almost every x, $f(x)$ is approximated by its mean value on small balls $B(x,\delta)$, ignoring the missing values on $B(x,\delta) \setminus (D \cap \operatorname{dom} f)$; 261E is a sharper version of the same idea. The formulae of 261C-261E mostly involve the expression $\mu B(x,\delta)$. Of course this is just $\beta_r\delta^r$. But I think that leaving it unexpanded is actually more illuminating, as well as avoiding sub- and superscripts, since it makes it clearer what these density theorems are really about. In §472 of Volume 4 I will revisit this material, showing that a surprisingly large proportion of the ideas can be applied to arbitrary Radon measures on $\mathbb{R}^r$, even though Vitali's theorem (in the form stated here) is no longer valid.

Version of 11.8.15

262 Lipschitz and differentiable functions

In preparation for the main work of this chapter in §263, I devote a section to two important classes of functions between Euclidean spaces. What we really need is the essentially elementary material down to 262I, together with the technical lemma 262M and its corollaries. Theorem 262Q is not relied on in this volume, though I believe that it makes the patterns which will develop more natural and comprehensible.

As in §261, r (and here also s) will be a strictly positive integer, and 'measurable', 'negligible', 'integrable' will refer to Lebesgue measure unless otherwise stated.

262A Lipschitz functions Suppose that $\phi : D \to \mathbb{R}^s$ is a function, where $D \subseteq \mathbb{R}^r$. We say that ϕ is **γ-Lipschitz**, where $\gamma \in [0, \infty[$, if

$$\|\phi(x) - \phi(y)\| \le \gamma\|x - y\|$$

for all x, $y \in D$, writing $\|x\| = \sqrt{\xi_1^2 + \ldots + \xi_r^2}$ if $x = (\xi_1, \ldots, \xi_r) \in \mathbb{R}^r$, $\|z\| = \sqrt{\zeta_1^2 + \ldots + \zeta_s^2}$ if $z = (\zeta_1, \ldots, \zeta_s) \in \mathbb{R}^s$. In this case, γ is **a Lipschitz constant for** ϕ.

A **Lipschitz function** is a function ϕ which is γ-Lipschitz for some $\gamma \ge 0$. Note that in this case ϕ has a least Lipschitz constant (since if A is the set of Lipschitz constants for ϕ, and $\gamma_0 = \inf A$, then γ_0 is a Lipschitz constant for ϕ). Evidently a Lipschitz function is (uniformly) continuous.

262B We need the following easy facts.

Lemma Let $D \subseteq \mathbb{R}^r$ be a set and $\phi : D \to \mathbb{R}^s$ a function.

(a) ϕ is Lipschitz iff $\phi_i : D \to \mathbb{R}$ is Lipschitz for every i, writing $\phi(x) = (\phi_1(x), \ldots, \phi_s(x))$ for every $x \in D = \operatorname{dom}\phi \subseteq \mathbb{R}^r$.

(b) In this case, there is a Lipschitz function $\tilde\phi : \mathbb{R}^r \to \mathbb{R}^s$ extending ϕ.

(c) If $r = s = 1$ and $D = [a, b]$ is an interval, then ϕ is Lipschitz iff it is absolutely continuous and has a bounded derivative.

proof (a) For any x, $y \in D$ and $i \le s$,

$$|\phi_i(x) - \phi_i(y)| \le \|\phi(x) - \phi(y)\| \le \sqrt{s}\sup_{j\le s}|\phi_j(x) - \phi_j(y)|,$$

so any Lipschitz constant for ϕ will be a Lipschitz constant for every ϕ_i, and if γ_j is a Lipschitz constant for ϕ_j for each j, then $\sqrt{s}\sup_{j\le s}\gamma_j$ will be a Lipschitz constant for ϕ.

(b) By (a), it is enough to consider the case $s = 1$, for if every ϕ_i has a Lipschitz extension $\tilde\phi_i$, we can set $\tilde\phi(x) = (\tilde\phi_1(x), \ldots, \tilde\phi_s(x))$ for every x to obtain a Lipschitz extension of ϕ. Taking $s = 1$, then, note that the case $D = \emptyset$ is trivial; so suppose that $D \ne \emptyset$. Let γ be a Lipschitz constant for ϕ, and write

$$\tilde\phi(z) = \sup_{y\in D}\phi(y) - \gamma\|y - z\|$$

for every $z \in \mathbb{R}^r$. If $x \in D$, then, for any $z \in \mathbb{R}^r$ and $y \in D$,

$$\phi(y) - \gamma\|y - z\| \le \phi(x) + \gamma\|y - x\| - \gamma\|y - z\| \le \phi(x) + \gamma\|z - x\|,$$

so that $\tilde\phi(z) \le \phi(x) + \gamma\|z - x\|$; this shows, in particular, that $\tilde\phi(z) < \infty$. Also, if $z \in D$, we must have

$$\phi(z) - \gamma\|z - z\| \le \tilde\phi(z) \le \phi(z) + \gamma\|z - z\|,$$

so that $\tilde\phi$ extends ϕ. Finally, if w, $z \in \mathbb{R}^r$ and $y \in D$,

$$\phi(y) - \gamma\|y - w\| \le \phi(y) - \gamma\|y - z\| + \gamma\|w - z\| \le \tilde\phi(z) + \gamma\|w - z\|;$$

and taking the supremum over $y \in D$,

$$\tilde\phi(w) \le \tilde\phi(z) + \gamma\|w - z\|.$$

As w and z are arbitrary, $\tilde\phi$ is Lipschitz.

(c)(i) Suppose that ϕ is γ-Lipschitz. If $\epsilon > 0$ and $a \le a_1 \le b_1 \le \ldots \le a_n \le b_n \le b$ and $\sum_{i=1}^n b_i - a_i \le \epsilon/(1 + \gamma)$, then

$$\sum_{i=1}^n |\phi(b_i) - \phi(a_i)| \le \sum_{i=1}^n \gamma|b_i - a_i| \le \epsilon.$$

As ϵ is arbitrary, ϕ is absolutely continuous. If $x \in [a, b]$ and $\phi'(x)$ is defined, then

$$|\phi'(x)| = \lim_{y\to x}\tfrac{|\phi(y)-\phi(x)|}{|y-x|} \le \gamma,$$

so ϕ' is bounded.

(ii) Now suppose that ϕ is absolutely continuous and that $|\phi'(x)| \le \gamma$ for every $x \in \operatorname{dom}\phi'$, where $\gamma \ge 0$. Then whenever $a \le x \le y \le b$,

$$|\phi(y) - \phi(x)| = |\int_x^y \phi'| \le \int_x^y |\phi'| \le \gamma(y - x)$$

(using 225E for the first equality). As x and y are arbitrary, ϕ is γ-Lipschitz.

262C Remark The argument for (b) above shows that if $\phi : D \to \mathbb{R}$ is a Lipschitz function, where $D \subseteq \mathbb{R}^r$, then ϕ has an extension to $\mathbb{R}^r$ with the same Lipschitz constants. In fact it is the case that if $\phi : D \to \mathbb{R}^s$ is a Lipschitz function, then ϕ has an extension to $\tilde{\phi} : \mathbb{R}^r \to \mathbb{R}^s$ with the same Lipschitz constants; this is 'Kirszbraun's theorem' (Kirszbraun 1934, or Federer 69, 2.10.43).

262D Proposition If $\phi : D \to \mathbb{R}^r$ is a γ-Lipschitz function, where $D \subseteq \mathbb{R}^r$, then $\mu^*\phi[A] \le \gamma^r \mu^* A$ for every $A \subseteq D$, where μ is Lebesgue measure on $\mathbb{R}^r$. In particular, $\phi[D \cap A]$ is negligible for every negligible set $A \subseteq \mathbb{R}^r$.

proof Let $\epsilon > 0$. By 261F, there is a sequence $\langle B_n \rangle_{n\in\mathbb{N}} = \langle B(x_n, \delta_n)\rangle_{n\in\mathbb{N}}$ of closed balls in $\mathbb{R}^r$, covering A, such that $\sum_{n=0}^\infty \mu B_n \le \mu^* A + \epsilon$ and $\sum_{n\in\mathbb{N}\setminus K} \mu B_n \le \epsilon$, where $K = \{n : n \in \mathbb{N},\, x_n \in A\}$. Set

$$L = \{n : n \in \mathbb{N} \setminus K,\, B_n \cap D \ne \emptyset\},$$

and for $n \in L$ choose $y_n \in D \cap B_n$. Now set

$$\begin{aligned} B'_n &= B(\phi(x_n), \gamma\delta_n) \text{ if } n \in K, \\ &= B(\phi(y_n), 2\gamma\delta_n) \text{ if } n \in L, \\ &= \emptyset \text{ if } n \in \mathbb{N} \setminus (K \cup L). \end{aligned}$$

Then $\phi[B_n \cap D] \subseteq B'_n$ for every n, so $\phi[D \cap A] \subseteq \bigcup_{n\in\mathbb{N}} B'_n$, and

$$\begin{aligned} \mu^*\phi[A \cap D] &\le \sum_{n=0}^\infty \mu B'_n = \gamma^r \sum_{n\in K} \mu B_n + 2^r\gamma^r \sum_{n\in L} \mu B_n \\ &\le \gamma^r(\mu^* A + \epsilon) + 2^r\gamma^r\epsilon. \end{aligned}$$

As ϵ is arbitrary, $\mu^*\phi[A \cap D] \le \gamma^r \mu^* A$, as claimed.

262E Corollary Let $\phi : D \to \mathbb{R}^r$ be an injective Lipschitz function, where $D \subseteq \mathbb{R}^r$, and f a measurable function from a subset of $\mathbb{R}^r$ to $\mathbb{R}$.

(a) If ϕ^{-1} is defined almost everywhere in a subset H of $\mathbb{R}^r$ and f is defined almost everywhere in $\mathbb{R}^r$, then $f\phi^{-1}$ is defined almost everywhere in H.

(b) If $E \subseteq D$ is Lebesgue measurable then $\phi[E]$ is measurable.

(c) If D is measurable then $f\phi^{-1}$ is measurable.

proof Set

$$C = \operatorname{dom}(f\phi^{-1}) = \{y : y \in \phi[D],\, \phi^{-1}(y) \in \operatorname{dom} f\} = \phi[D \cap \operatorname{dom} f].$$

(a) Because f is defined almost everywhere, $\phi[D \setminus \operatorname{dom} f]$ is negligible. But now

$$C = \phi[D] \setminus \phi[D \setminus \operatorname{dom} f] = \operatorname{dom}\phi^{-1} \setminus \phi[D \setminus \operatorname{dom} f],$$

so

$$H \setminus C \subseteq (H \setminus \operatorname{dom}\phi^{-1}) \cup \phi[D \setminus \operatorname{dom} f]$$

is negligible.

(b) Now suppose that $E \subseteq D$ and that E is measurable. Let $\langle F_n\rangle_{n\in\mathbb{N}}$ be a sequence of closed bounded subsets of E such that $\mu(E \setminus \bigcup_{n\in\mathbb{N}} F_n) = 0$ (134Fb). Because ϕ is Lipschitz, it is continuous, so $\phi[F_n]$ is compact, therefore closed, therefore measurable for every n (2A2E, 115G); also $\phi[E \setminus \bigcup_{n\in\mathbb{N}} F_n]$ is negligible, by 262D, therefore measurable. So

$$\phi[E] = \phi[E \setminus \textstyle\bigcup_{n\in\mathbb{N}} F_n] \cup \bigcup_{n\in\mathbb{N}} \phi[F_n]$$

is measurable.

(c) For any $a \in \mathbb{R}$, take a measurable set $E \subseteq \mathbb{R}^r$ such that $\{x : f(x) \ge a\} = E \cap \operatorname{dom} f$. Then

$$\{y : y \in C,\, f\phi^{-1}(y) \ge a\} = C \cap \phi[D \cap E].$$

But $\phi[D \cap E]$ is measurable, by (b), so $\{y : f\phi^{-1}(y) \ge a\}$ is relatively measurable in C. As a is arbitrary, $f\phi^{-1}$ is measurable.

262F Differentiability I come now to the class of functions whose properties will take up most of the rest of the chapter.

Definitions Suppose that ϕ is a function from a subset $D = \operatorname{dom}\phi$ of $\mathbb{R}^r$ to $\mathbb{R}^s$.

(a) ϕ is **differentiable** at $x \in D$ if there is a real $s \times r$ matrix T such that

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

in this case we may write $T = \phi'(x)$.

(b) I will say that ϕ is **differentiable relative to its domain** at x, and that T is **a** derivative of ϕ at x, if $x \in D$ and for every $\epsilon > 0$ there is a $\delta > 0$ such that $\|\phi(y) - \phi(x) - T(y - x)\| \le \epsilon\|y - x\|$ for every $y \in B(x, \delta) \cap D$.

262G Remarks (a) The standard definition in 262Fa, involving an all-sided limit '$\lim_{y\to x}$', implicitly requires ϕ to be defined on some non-trivial ball centered on x, so that we can calculate $\phi(y)-\phi(x)-T(y-x)$ for all y sufficiently near x. It has the advantage that the derivative $T = \phi'(x)$ is uniquely defined (because if $\lim_{z\to\mathbf{0}} \frac{\|T_1z-T_2z\|}{\|z\|} = 0$ then

$$\frac{\|(T_1-T_2)z\|}{\|z\|} = \lim_{\alpha\to 0} \frac{\|T_1(\alpha z)-T_2(\alpha z)\|}{\|\alpha z\|} = 0$$

for every non-zero z, so $T_1 - T_2$ must be the zero matrix). For our purposes here, there is some advantage in relaxing this slightly to the form in 262Fb, so that we do not need to pay special attention to the boundary of $\operatorname{dom}\phi$. In particular we find that if T is a derivative of $\phi : D \to \mathbb{R}^s$ relative to its domain at x, and $x \in D' \subseteq D$, then T is a derivative of $\phi\restriction D'$, relative to its domain, at x.

(b) If you have not seen this concept of 'differentiability' before, but have some familiarity with partial differentiation, it is necessary to emphasize that the concept of 'differentiable' function (at least in the strict sense demanded by 262Fa) is strictly stronger than the concept of 'partially differentiable' function. For purposes of computation, the most useful method of finding true derivatives is through 262Id below. For a simple example of a function with a full set of partial derivatives, which is not everywhere differentiable, consider $\phi : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$\begin{aligned}\phi(\xi_1, \xi_2) &= \frac{\xi_1\xi_2}{\xi_1^2 + \xi_2^2} \text{ if } \xi_1^2 + \xi_2^2 \ne 0,\\ &= 0 \text{ if } \xi_1 = \xi_2 = 0.\end{aligned}$$

Then ϕ is not even continuous at $\mathbf{0}$, although both partial derivatives $\frac{\partial\phi}{\partial\xi_j}$ are defined everywhere.

(c) In the definition above, I speak of a derivative as being a matrix. Properly speaking, the derivative of a function defined on a subset of $\mathbb{R}^r$ and taking values in $\mathbb{R}^s$ should be thought of as a bounded linear operator from $\mathbb{R}^r$ to $\mathbb{R}^s$; the formulation in terms of matrices is acceptable just because there is a natural one-to-one correspondence between $s \times r$ real matrices and linear operators from $\mathbb{R}^r$ to $\mathbb{R}^s$, and all these linear operators are bounded. I use the 'matrix' description because it makes certain calculations more direct; in particular, the relationship between ϕ' and the partial derivatives of ϕ (262Ic), and the notion of the determinant $\det\phi'(x)$, used throughout §§263 and 265.

262H The norm of a matrix Some of the calculations below will rely on the notion of 'norm' of a matrix. The one I will use (in fact, for our purposes here, any norm would do) is the 'operator norm', defined by saying

$$\|T\| = \sup\{\|Tx\| : x \in \mathbb{R}^r,\ \|x\| \le 1\}$$

for any $s \times r$ matrix T. For the basic facts concerning these norms, see 2A4F-2A4G. The following will also be useful.

(a) If all the coefficients of T are small, so is $\|T\|$; in fact, if $T = \langle\tau_{ij}\rangle_{i\le s,j\le r}$, and $\|x\| \le 1$, then $|\xi_j| \le 1$ for each j, so

$$\|Tx\| = \bigl(\sum_{i=1}^{s}(\sum_{j=1}^{r}\tau_{ij}\xi_j)^2\bigr)^{1/2} \le \bigl(\sum_{i=1}^{s}(\sum_{j=1}^{r}|\tau_{ij}|)^2\bigr)^{1/2} \le r\sqrt{s}\max_{i\le s,j\le r}|\tau_{ij}|,$$

and $\|T\| \le r\sqrt{s}\max_{i\le s,j\le r}|\tau_{ij}|$. (This is a singularly crude inequality. A better one is in 262Yb. But it tells us, in particular, that $\|T\|$ is always finite.)

(b) If $\|T\|$ is small, so are all the coefficients of T; in fact, writing e_j for the jth unit vector of $\mathbb{R}^r$, then the ith coordinate of Te_j is τ_{ij}, so $|\tau_{ij}| \le \|Te_j\| \le \|T\|$.

262I Lemma Let $\phi : D \to \mathbb{R}^s$ be a function, where $D \subseteq \mathbb{R}^r$. For $i \le s$ let $\phi_i : D \to \mathbb{R}$ be its ith coordinate, so that $\phi(x) = (\phi_1(x), \dots , \phi_s(x))$ for $x \in D$.

(a) If ϕ is differentiable relative to its domain at $x \in D$, then ϕ is continuous at x.

(b) If $x \in D$, then ϕ is differentiable relative to its domain at x iff each ϕ_i is differentiable relative to its domain at x.

(c) If ϕ is differentiable at $x \in D$, then all the partial derivatives $\frac{\partial\phi_i}{\partial\xi_j}$ of ϕ are defined at x, and the derivative of ϕ at x is the matrix $\langle\frac{\partial\phi_i}{\partial\xi_j}(x)\rangle_{i\le s,j\le r}$.

(d) If all the partial derivatives $\frac{\partial\phi_i}{\partial\xi_j}$, for $i \le s$ and $j \le r$, are defined in a neighbourhood of $x \in D$ and are continuous at x, then ϕ is differentiable at x.

proof (a) Let T be a derivative of ϕ at x. Applying the definition 262Fb with $\epsilon = 1$, we see that there is a $\delta > 0$ such that

$$\|\phi(y) - \phi(x) - T(y - x)\| \le \|y - x\|$$

whenever $y \in D$ and $\|y - x\| \le \delta$. Now

$$\|\phi(y) - \phi(x)\| \le \|T(y - x)\| + \|y - x\| \le (1 + \|T\|)\|y - x\|$$

whenever $y \in D$ and $\|y - x\| \le \delta$, so ϕ is continuous at x.

(b)(i) If ϕ is differentiable relative to its domain at $x \in D$, let T be a derivative of ϕ at x. For $i \le s$ let T_i be the $1 \times r$ matrix consisting of the ith row of T. Let $\epsilon > 0$. Then we have a $\delta > 0$ such that

$$\begin{aligned}|\phi_i(y) - \phi_i(x) - T_i(y - x)| &\le \|\phi(y) - \phi(x) - T(y - x)\|\\ &\le \epsilon\|y - x\|\end{aligned}$$

whenever $y \in D$ and $\|y - x\| \le \delta$, so that T_i is a derivative of ϕ_i at x.

(ii) If each ϕ_i is differentiable relative to its domain at x, with corresponding derivatives T_i, let T be the $s \times r$ matrix with rows $T_1, \dots , T_s$. Given $\epsilon > 0$, there is for each $i \le s$ a $\delta_i > 0$ such that

$$|\phi_i(y) - \phi_i(x) - T_iy| \le \epsilon\|y - x\| \text{ whenever } y \in D,\ \|y - x\| \le \delta_i;$$

set $\delta = \min_{i\le s}\delta_i > 0$; then if $y \in D$ and $\|y - x\| \le \delta$, we shall have

$$\|\phi(y) - \phi(x) - T(y - x)\|^2 = \sum_{i=1}^{s}|\phi_i(y) - \phi_i(x) - T_i(y - x)|^2 \le s\epsilon^2\|y - x\|^2,$$

so that

$$\|\phi(y) - \phi(x) - T(y - x)\| \le \epsilon\sqrt{s}\|y - x\|.$$

As ϵ is arbitrary, T is a derivative of ϕ at x.

(c) Set $T = \phi'(x)$. We have

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

fix $j \le r$, and consider $y = x + \eta e_j$, where $e_j = (0, \dots , 0, 1, 0, \dots , 0)$ is the jth unit vector in $\mathbb{R}^r$. Then we must have

$$\lim_{\eta\to0}\frac{\|\phi(x+\eta e_j)-\phi(x)-\eta T(e_j)\|}{|\eta|} = 0.$$

Looking at the ith coordinate of $\phi(x+\eta e_j)-\phi(x)-\eta T(e_j)$, we have

$$|\phi_i(x+\eta e_j)-\phi_i(x)-\tau_{ij}\eta| \le \|\phi(x+\eta e_j)-\phi(x)-\eta T(e_j)\|,$$

where τ_{ij} is the (i,j)th coefficient of T; so that

$$\lim_{\eta\to 0}\frac{|\phi_i(x+\eta e_j)-\phi_i(x)-\tau_{ij}\eta|}{|\eta|}=0.$$

But this just says that the partial derivative $\frac{\partial\phi_i}{\partial\xi_j}(x)$ exists and is equal to τ_{ij}, as claimed.

(d) Now suppose that the partial derivatives $\frac{\partial\phi_i}{\partial\xi_j}$ are defined near x and continuous at x. Let $\epsilon>0$. Let $\delta>0$ be such that

$$|\tfrac{\partial\phi_i}{\partial\xi_j}(y)-\tau_{ij}|\le\epsilon$$

whenever $\|y-x\|\le\delta$, writing $\tau_{ij}=\frac{\partial\phi_i}{\partial\xi_j}(x)$. Now suppose that $\|y-x\|\le\delta$. Set

$$y=(\eta_1,\dots,\eta_r),\quad x=(\xi_1,\dots,\xi_r),$$

$$y_j=(\eta_1,\dots,\eta_j,\xi_{j+1},\dots,\xi_r)\text{ for }0\le j\le r,$$

so that $y_0=x$, $y_r=y$ and the line segment between y_{j-1} and y_j lies wholly within δ of x whenever $1\le j\le r$, since if $z=(\zeta_1,\dots,\zeta_r)$ lies on this segment then ζ_i lies between ξ_i and η_i for every i. By the ordinary mean value theorem for differentiable real functions, applied to the function

$$t\mapsto\phi_i(\eta_1,\dots,\eta_{j-1},t,\xi_{j+1},\dots,\xi_r),$$

there is for each $i\le s$, $j\le r$ a point z_{ij} on the line segment between y_{j-1} and y_j such that

$$\phi_i(y_j)-\phi_i(y_{j-1})=(\eta_j-\xi_j)\tfrac{\partial\phi_i}{\partial\xi_j}(z_{ij}).$$

But

$$|\tfrac{\partial\phi_i}{\partial\xi_j}(z_{ij})-\tau_{ij}|\le\epsilon,$$

so

$$|\phi_i(y_j)-\phi_i(y_{j-1})-\tau_{ij}(\eta_j-\xi_j)|\le\epsilon|\eta_j-\xi_j|\le\epsilon\|y-x\|.$$

Summing over j,

$$|\phi_i(y)-\phi_i(x)-\textstyle\sum_{j=1}^r\tau_{ij}(\eta_j-\xi_j)|\le r\epsilon\|y-x\|$$

for each i. Summing the squares and taking the square root,

$$\|\phi(y)-\phi(x)-T(y-x)\|\le\epsilon r\sqrt{s}\|y-x\|,$$

where $T=\langle\tau_{ij}\rangle_{i\le s,j\le r}$. And this is true whenever $\|y-x\|\le\delta$. As ϵ is arbitrary, $\phi'(x)=T$ is defined.

262J Remark I am not sure if I ought to apologize for the notation $\frac{\partial}{\partial\xi_j}$. In such formulae as $(\eta_j-\xi_j)\frac{\partial\phi_i}{\partial\xi_j}(z_{ij})$ above, the two appearances of ξ_j clash most violently. But I do not think that any person of good will is likely to be misled, provided that the labels ξ_j (or whatever symbols are used to represent the variables involved) are adequately described when the domain of ϕ is first introduced (and always remembering that in partial differentiation, we are not only moving one variable – a ξ_j in the present context – but holding fixed some further list of variables, not listed in the notation). I believe that the traditional notation $\frac{\partial}{\partial\xi_j}$ has survived for solid reasons, and I should like to offer a welcome to those who are more comfortable with it than with any of the many alternatives which have been proposed, but have never taken root.

262K The Cantor function revisited It is salutary to re-examine the examples of 134H-134I in the light of the present considerations. Let $f:[0,1]\to[0,1]$ be the Cantor function (134H) and set $g(x)=\frac{1}{2}(x+f(x))$ for $x\in[0,1]$. Then $g:[0,1]\to[0,1]$ is a homeomorphism (134I); set $\phi=g^{-1}:[0,1]\to[0,1]$. We see that if $0\le x\le y\le 1$ then $g(y)-g(x)\ge\frac{1}{2}(y-x)$; equivalently, $\phi(y)-\phi(x)\le 2(y-x)$ whenever $0\le x\le y\le 1$, so that ϕ is a Lipschitz function, therefore absolutely continuous (262Bc). If $D=\{x:\phi'(x)$

is defined}, then $[0,1]\setminus D$ is negligible (225Cb), so $[0,1]\setminus\phi[D]=\phi[\,[0,1]\setminus D]$ is negligible (262Da). I noted in 134I that there is a measurable function $h:[0,1]\to\mathbb{R}$ such that the composition $h\phi$ is not measurable; now $h(\phi\restriction D)=(h\phi)\restriction D$ cannot be measurable, even though $\phi\restriction D$ is differentiable.

262L It will be convenient to be able to call on the following straightforward result.

Lemma Suppose that $D\subseteq\mathbb{R}^r$ and $x\in\mathbb{R}^r$ are such that $\lim_{\delta\downarrow 0}\frac{\mu^*(D\cap B(x,\delta))}{\mu B(x,\delta)}=1$. Then $\lim_{z\to\mathbf{0}}\frac{\rho(x+z,D)}{\|z\|}=0$, where $\rho(x+z,D)=\inf_{y\in D}\|x+z-y\|$.

proof Let $\epsilon>0$. Let $\delta_0>0$ be such that

$$\mu^*(D\cap B(x,\delta))>(1-(\tfrac{\epsilon}{1+\epsilon})^r)\mu B(x,\delta)$$

whenever $0<\delta\le\delta_0$. Take any z such that $0<\|z\|\le\delta_0/(1+\epsilon)$. **?** Suppose, if possible, that $\rho(x+z,D)>\epsilon\|z\|$. Then $B(x+z,\epsilon\|z\|)\subseteq B(x,(1+\epsilon)\|z\|)\setminus D$, so

$$\begin{aligned}\mu^*(D\cap B(x,(1+\epsilon)\|z\|))&\le\mu B(x,(1+\epsilon)\|z\|)-\mu B(x+z,\epsilon\|z\|)\\&=(1-(\tfrac{\epsilon}{1+\epsilon})^r)\mu B(x,(1+\epsilon)\|z\|),\end{aligned}$$

which is impossible, as $(1+\epsilon)\|z\|\le\delta_0$. **X** Thus $\rho(x+z,D)\le\epsilon\|z\|$. As ϵ is arbitrary, this proves the result.

Remark There is a word for this; see 261Yg.

262M I come now to the first result connecting Lipschitz functions with differentiable functions. I approach it through a substantial lemma which will be the foundation of §263.

Lemma Let ϕ be a function from a subset D of $\mathbb{R}^r$ to $\mathbb{R}^s$ which is differentiable at each point of its domain. For each $x\in D$ let $T(x)$ be a derivative of ϕ. Let M_{sr} be the set of $s\times r$ matrices and $\zeta:A\to\,]0,\infty[$ a strictly positive function, where $A\subseteq M_{sr}$ is a non-empty set containing $T(x)$ for every $x\in D$. Then we can find sequences $\langle D_n\rangle_{n\in\mathbb{N}}$, $\langle T_n\rangle_{n\in\mathbb{N}}$ such that

(i) $\langle D_n\rangle_{n\in\mathbb{N}}$ is a partition of D into sets which are relatively measurable in D, that is, are intersections of D with measurable subsets of $\mathbb{R}^r$;

(ii) $T_n\in A$ for every n;

(iii) $\|\phi(x)-\phi(y)-T_n(x-y)\|\le\zeta(T_n)\|x-y\|$ for every $n\in\mathbb{N}$ and x, $y\in D_n$;

(iv) $\|T(x)-T_n\|\le\zeta(T_n)$ for every $x\in D_n$.

proof (a) The first step is to note that there is a sequence $\langle S_n\rangle_{n\in\mathbb{N}}$ in A such that

$$A\subseteq\bigcup_{n\in\mathbb{N}}\{T:T\in M_{sr},\,\|T-S_n\|<\zeta(S_n)\}.$$

P (Of course this is a standard result about separable metric spaces.) Write Q for the set of matrices in M_{sr} with rational coefficients; then there is a natural bijection between Q and $\mathbb{Q}^{sr}$, so Q and $Q\times\mathbb{N}$ are countable. Enumerate $Q\times\mathbb{N}$ as $\langle(R_n,k_n)\rangle_{n\in\mathbb{N}}$. For each $n\in\mathbb{N}$, choose $S_n\in A$ by the rule

— if there is an $S\in A$ such that $\{T:\|T-R_n\|\le 2^{-k_n}\}\subseteq\{T:\|T-S\|<\zeta(S)\}$, take such an S for S_n;

— otherwise, take S_n to be any member of A.

I claim that this works. For let $S\in A$. Then $\zeta(S)>0$; take $k\in\mathbb{N}$ such that $2^{-k}<\zeta(S)$. Take $R^*\in Q$ such that $\|R^*-S\|<\min(\zeta(S)-2^{-k},2^{-k})$; this is possible because $\|R-S\|$ will be small whenever all the coefficients of R are close enough to the corresponding coefficients of S (262Ha), and we can find rational numbers to achieve this. Let $n\in\mathbb{N}$ be such that $R^*=R_n$ and $k=k_n$. Then

$$\{T:\|T-R_n\|\le 2^{-k_n}\}\subseteq\{T:\|T-S\|<\zeta(S)\}$$

(because $\|T-S\|\le\|T-R_n\|+\|R_n-S\|$), so we must have chosen S_n by the first part of the rule above, and

$$S\in\{T:\|T-R_n\|\le 2^{-k_n}\}\subseteq\{T:\|T-S_n\|<\zeta(S_n)\}.$$

As S is arbitrary, this proves the result. **Q**

(b) Enumerate $\mathbb{Q}^r \times \mathbb{Q}^r \times \mathbb{N}$ as $\langle (q_n, q'_n, m_n)\rangle_{n\in\mathbb{N}}$. For each $n \in \mathbb{N}$, set

$$\begin{aligned} H_n &= \{x : x \in [q_n, q'_n] \cap D,\, \|\phi(y) - \phi(x) - S_{m_n}(y-x)\| \le \zeta(S_{m_n})\|y-x\| \\ &\qquad\qquad \text{for every } y \in [q_n, q'_n] \cap D\} \\ &= [q_n, q'_n] \cap D \cap \bigcap_{y\in[q_n,q'_n]\cap D} \{x : x \in D, \\ &\qquad\qquad \|\phi(y) - \phi(x) - S_{m_n}(y-x)\| \le \zeta(S_{m_n})\|y-x\|\}. \end{aligned}$$

Because ϕ is continuous, $H_n = D \cap \overline{H}_n$, writing $\overline{H}_n$ for the closure of H_n, so H_n is relatively measurable in D. Note that if x, $y \in H_n$, then $y \in D \cap [q_n, q'_n]$, so that

$$\|\phi(y) - \phi(x) - S_{m_n}(y-x)\| \le \zeta(S_{m_n})\|y-x\|.$$

Set

$$H'_n = \{x : x \in H_n,\, \|T(x) - S_{m_n}\| \le \zeta(S_{m_n})\}.$$

(c) $D = \bigcup_{n\in\mathbb{N}} H'_n$. **P** Let $x \in D$. Then $T(x) \in A$, so there is a $k \in \mathbb{N}$ such that $\|T(x) - S_k\| < \zeta(S_k)$. Let $\delta > 0$ be such that

$$\|\phi(y) - \phi(x) - T(x)(x-y)\| \le (\zeta(S_k) - \|T(x) - S_k\|)\|x-y\|$$

whenever $y \in D$ and $\|y-x\| \le \delta$. Then

$$\begin{aligned} \|\phi(y) - \phi(x) - S_k(x-y)\| &\le (\zeta(S_k) - \|T(x) - S_k\|)\|x-y\| + \|T(x) - S_k\|\|x-y\| \\ &\le \zeta(S_k)\|x-y\| \end{aligned}$$

whenever $y \in D \cap B(x,\delta)$. Let q, $q' \in \mathbb{Q}^r$ be such that $x \in [q, q'] \subseteq B(x,\delta)$. Let n be such that $q = q_n$, $q' = q'_n$ and $k = m_n$. Then $x \in H'_n$. **Q**

(d) Write

$$C_n = \{x : x \in H_n,\, \lim_{\delta\downarrow 0} \frac{\mu^*(H_n \cap B(x,\delta))}{\mu B(x,\delta)} = 1\}.$$

Then $C_n \subseteq H'_n$.

P (i) Take $x \in C_n$, and set $\tilde{T} = T(x) - S_{m_n}$. I have to show that $\|\tilde{T}\| \le \zeta(S_{m_n})$. Take $\epsilon > 0$. Let $\delta_0 > 0$ be such that

$$\|\phi(y) - \phi(x) - T(x)(y-x)\| \le \epsilon\|y-x\|$$

whenever $y \in D$ and $\|y-x\| \le \delta_0$. Since

$$\|\phi(y) - \phi(x) - S_{m_n}(y-x)\| \le \zeta(S_{m_n})\|y-x\|$$

whenever $y \in H_n$, we have

$$\|\tilde{T}(y-x)\| \le (\epsilon + \zeta(S_{m_n}))\|y-x\|$$

whenever $y \in H_n$ and $\|y-x\| \le \delta_0$.

(ii) By 262L, there is a $\delta_1 > 0$ such that $(1+2\epsilon)\delta_1 \le \delta_0$ and $\rho(x+z, H_n) \le \epsilon\|z\|$ whenever $0 < \|z\| \le \delta_1$. So if $\|z\| \le \delta_1$ there is a $y \in H_n$ such that $\|x+z-y\| \le 2\epsilon\|z\|$. (If $z = 0$ we can take $y = x$.) Now $\|x-y\| \le (1+2\epsilon)\|z\| \le \delta_0$, so

$$\begin{aligned} \|\tilde{T}z\| &\le \|\tilde{T}(y-x)\| + \|\tilde{T}(x+z-y)\| \\ &\le (\epsilon + \zeta(S_{m_n}))\|y-x\| + \|\tilde{T}\|\|x+z-y\| \\ &\le (\epsilon + \zeta(S_{m_n}))\|z\| + (\epsilon + \zeta(S_{m_n}) + \|\tilde{T}\|)\|x+z-y\| \\ &\le (\epsilon + \zeta(S_{m_n}) + 2\epsilon^2 + 2\epsilon\zeta(S_{m_n}) + 2\epsilon\|\tilde{T}\|)\|z\|. \end{aligned}$$

And this is true whenever $0 < \|z\| \le \delta_1$. But multiplying this inequality by suitable positive scalars we see that

$$\|\tilde{T}z\| \le \bigl(\epsilon + \zeta(S_{m_n}) + 2\epsilon^2 + 2\epsilon\zeta(S_{m_n}) + 2\epsilon\|\tilde{T}\|\bigr)\|z\|$$

for all $z \in \mathbb{R}^r$, and

$$\|\tilde{T}\| \le \epsilon + \zeta(S_{m_n}) + 2\epsilon^2 + 2\epsilon\zeta(S_{m_n}) + 2\epsilon\|\tilde{T}\|.$$

As ϵ is arbitrary, $\|\tilde{T}\| \le \zeta(S_{m_n})$, as claimed. **Q**

(e) By 261Da, $H_n \setminus C_n$ is negligible for every n, so $H_n \setminus H'_n$ is negligible, and

$$H'_n = D \cap (\overline{H}_n \setminus (H_n \setminus H'_n))$$

is relatively measurable in D. Set

$$D_n = H'_n \setminus \bigcup_{k<n} H'_k, \quad T_n = S_{m_n}$$

for each n; these serve.

262N Corollary Let ϕ be a function from a subset D of $\mathbb{R}^r$ to $\mathbb{R}^s$, and suppose that ϕ is differentiable relative to its domain at each point of D. Then D can be expressed as the union of a disjoint sequence $\langle D_n\rangle_{n\in\mathbb{N}}$ of relatively measurable subsets of D such that $\phi\restriction D_n$ is Lipschitz for each $n \in \mathbb{N}$.

proof In 262M, take $\zeta(T) = 1$ for every $T \in A = M_{sr}$. If $x, y \in D_n$ then

$$\begin{aligned}\|\phi(x) - \phi(y)\| &\le \|\phi(x) - \phi(y) - T_n(x-y)\| + \|T_n(x-y)\| \\ &\le \|x-y\| + \|T_n\|\|x-y\|,\end{aligned}$$

so $\phi\restriction D_n$ is $(1 + \|T_n\|)$-Lipschitz.

262O Corollary Suppose that ϕ is an injective function from a measurable subset D of $\mathbb{R}^r$ to $\mathbb{R}^r$, and that ϕ is differentiable relative to its domain at every point of D.

(a) If $A \subseteq D$ is negligible, $\phi[A]$ is negligible.

(b) If $E \subseteq D$ is measurable, then $\phi[E]$ is measurable.

(c) If D is measurable and f is a measurable function defined on a subset of $\mathbb{R}^r$, then $f\phi^{-1}$ is measurable.

(d) If $H \subseteq \mathbb{R}^r$ and ϕ^{-1} is defined almost everywhere in H, and if f is a function defined almost everywhere in $\mathbb{R}^r$, then $f\phi^{-1}$ is defined almost everywhere in H.

proof Take a sequence $\langle D_n\rangle_{n\in\mathbb{N}}$ as in 262N, and apply 262E to $\phi\restriction D_n$ for each n.

262P Corollary Let ϕ be a function from a a subset D of $\mathbb{R}^r$ to $\mathbb{R}^s$, and suppose that ϕ is differentiable relative to its domain, with a derivative $T(x)$, at each point $x \in D$. Then the function $x \mapsto T(x)$ is measurable in the sense that $\tau_{ij} : D \to \mathbb{R}$ is measurable for all $i \le s$ and $j \le r$, where $\tau_{ij}(x)$ is the (i,j)th coefficient of the matrix $T(x)$ for all i, j and x.

proof For each $k \in \mathbb{N}$, apply 262M with $\zeta(T) = 2^{-k}$ for each $T \in A = M_{sr}$, obtaining sequences $\langle D_{kn}\rangle_{n\in\mathbb{N}}$ of relatively measurable subsets of D and $\langle T_{kn}\rangle_{n\in\mathbb{N}}$ in M_{sr}. Let $\tau_{ij}^{(kn)}$ be the (i,j)th coefficient of T_{kn}. Then we have functions $f_{ijk} : D \to \mathbb{R}$ defined by setting

$$f_{ijk}(x) = \tau_{ij}^{(kn)} \text{ if } x \in D_{kn}.$$

Because the D_{kn} are relatively measurable, the f_{ijk} are measurable functions. For $x \in D_{kn}$,

$$|\tau_{ij}(x) - f_{ijk}(x)| \le \|T(x) - T_n\| \le 2^{-k},$$

so $|\tau_{ij}(x) - f_{ijk}(x)| \le 2^{-k}$ for every $x \in D$, and

$$\tau_{ij} = \lim_{k\to\infty} f_{ijk}$$

is measurable, as claimed.

***262Q** This concludes the part of the section which is essential for the rest of the chapter. However the main results of §263 will I think be better understood if you are aware of the fact that any Lipschitz function is differentiable (relative to its domain) almost everywhere in its domain. I devote the next couple of pages to a proof of this fact, which apart from its intrinsic interest is a useful exercise.

Rademacher's theorem Let ϕ be a Lipschitz function from a subset of $\mathbb{R}^r$ to $\mathbb{R}^s$. Then ϕ is differentiable relative to its domain almost everywhere in its domain.

proof (a) By 262Ba and 262Ib, it will be enough to deal with the case $s = 1$. By 262Bb, there is a Lipschitz function $\tilde{\phi} : \mathbb{R}^r \to \mathbb{R}$ extending ϕ; now ϕ is differentiable with respect to its domain at any point of $\operatorname{dom}\phi$ at which $\tilde{\phi}$ is differentiable, so it will be enough if I can show that $\tilde{\phi}$ is differentiable almost everywhere. To make the notation more agreeable to the eye, I will suppose that ϕ itself was defined everywhere in $\mathbb{R}^r$. Let γ be a Lipschitz constant for ϕ.

The proof proceeds by induction on r. If $r = 1$, we have a Lipschitz function $\phi : \mathbb{R} \to \mathbb{R}$; now ϕ is absolutely continuous in any bounded interval (262Bc), therefore differentiable almost everywhere (225Cb). Thus the induction starts. The rest of the proof is devoted to the inductive step to $r > 1$.

(b) The first step is to show that all the partial derivatives $\frac{\partial\phi}{\partial\xi_j}$ are defined almost everywhere and are Borel measurable. **P** Take $j \le r$. For $q \in \mathbb{Q} \setminus \{0\}$ set

$$\Delta_q(x) = \frac{1}{q}(\phi(x + qe_j) - \phi(x)),$$

writing e_j for the jth unit vector of $\mathbb{R}^r$. Because ϕ is continuous, so is Δ_q, so that Δ_q is a Borel measurable function for each q. Next, for any $x \in \mathbb{R}^r$,

$$D^+(x) = \limsup_{\delta\to 0} \frac{1}{\delta}(\phi(x + \delta e_j) - \phi(x)) = \lim_{n\to\infty} \sup_{q\in\mathbb{Q},0<|q|\le 2^{-n}} \Delta_q(x),$$

so that the set on which $D^+(x)$ is defined in $\mathbb{R}$ is Borel and D^+ is a Borel measurable function. Similarly,

$$D^-(x) = \liminf_{\delta\to 0} \frac{1}{\delta}(\phi(x + \delta e_j) - \phi(x))$$

is a Borel measurable function with Borel domain. So

$$E = \{x : \tfrac{\partial\phi}{\partial\xi_j}(x) \text{ exists in } \mathbb{R}\} = \{x : D^+(x) = D^-(x) \in \mathbb{R}\}$$

is a Borel set, and $\frac{\partial\phi}{\partial\xi_j}$ is a Borel measurable function.

On the other hand, if we identify $\mathbb{R}^r$ with $\mathbb{R}^J \times \mathbb{R}$, taking J to be $\{1, \dots, j-1, j+1, \dots, r\}$, then we can think of Lebesgue measure μ on $\mathbb{R}^r$ as being the product of Lebesgue measure μ_J on $\mathbb{R}^J$ with Lebesgue measure μ_1 on $\mathbb{R}$ (251N). Now for every $y \in \mathbb{R}^J$ we have a function $\phi_y : \mathbb{R} \to \mathbb{R}$ defined by writing

$$\phi_y(\sigma) = \phi(y, \sigma),$$

and E becomes

$$\{(y, \sigma) : \phi_y'(\sigma) \text{ is defined}\},$$

so that all the sections

$$\{\sigma : (y, \sigma) \in E\}$$

are conegligible subsets of $\mathbb{R}$, because every ϕ_y is Lipschitz, therefore differentiable almost everywhere, as remarked in part (a) of the proof. Since we know that E is measurable, it must be conegligible, by Fubini's theorem (apply 252D or 252F to the complement of E). Thus $\frac{\partial\phi}{\partial\xi_j}$ is defined almost everywhere, as claimed. **Q**

Write

$$H = \{x : x \in \mathbb{R}^r,\ \tfrac{\partial\phi}{\partial\xi_j}(x) \text{ exists for every } j \le r\},$$

so that H is a conegligible Borel set in $\mathbb{R}^r$.

(c) For the rest of this proof, I fix on the natural identification of $\mathbb{R}^r$ with $\mathbb{R}^{r-1}\times\mathbb{R}$, identifying $(\xi_1, \dots, \xi_r)$ with $((\xi_1, \dots, \xi_{r-1}), \xi_r)$. For $x \in H$, let $T(x)$ be the $1 \times r$ matrix $(\frac{\partial\phi}{\partial\xi_1}(x), \dots, \frac{\partial\phi}{\partial\xi_r}(x))$.

(d) Set

$$H_1=\{x:x\in H,\ \lim_{u\to\mathbf{0}\text{ in }\mathbb{R}^{r-1}}\frac{|\phi(x+(u,0))-\phi(x)-T(x)(u,0)|}{\|u\|}=0\}.$$

I claim that H_1 is conegligible in $\mathbb{R}^r$. **P** This is really the same idea as in (b). For $x\in H$, $x\in H_1$ iff

for every $\epsilon>0$ there is a $\delta>0$ such that

$$|\phi(x+(u,0))-\phi(x)-T(x)(u,0)|\le\epsilon\|u\|$$

whenever $\|u\|\le\delta$,

that is, iff

for every $m\in\mathbb{N}$ there is an $n\in\mathbb{N}$ such that

$$|\phi(x+(u,0))-\phi(x)-T(x)(u,0)|\le 2^{-m}\|u\|$$

whenever $u\in\mathbb{Q}^{r-1}$ and $\|u\|\le 2^{-n}$.

But for any particular $m\in\mathbb{N}$ and $u\in\mathbb{Q}^{r-1}$ the set

$$\{x:|\phi(x+(u,0))-\phi(x)-T(x)(u,0)|\le 2^{-m}\|u\|\}$$

is measurable, indeed Borel, because all the functions $x\mapsto\phi(x+(u,0))$, $x\mapsto\phi(x)$, $x\mapsto T(x)(u,0)$ are Borel measurable. So H_1 is of the form

$$\bigcap_{m\in\mathbb{N}}\bigcup_{n\in\mathbb{N}}\bigcap_{u\in\mathbb{Q}^{r-1},\|u\|\le 2^{-n}}E_{mnu}$$

where every E_{mnu} is a measurable set, and H_1 is therefore measurable.

Now however observe that for any $\sigma\in\mathbb{R}$, the function

$$v\mapsto\phi_\sigma(v)=\phi(v,\sigma):\mathbb{R}^{r-1}\to\mathbb{R}$$

is Lipschitz, therefore (by the inductive hypothesis) differentiable almost everywhere in $\mathbb{R}^{r-1}$; and that $(v,\sigma)\in H_1$ iff $(v,\sigma)\in H$ and $\phi'_\sigma(v)$ is defined. Consequently $\{v:(v,\sigma)\in H_1\}$ is conegligible whenever $\{v:(v,\sigma)\in H\}$ is, that is, for almost every $\sigma\in\mathbb{R}$; so that H_1, being measurable, must be conegligible. **Q**

(e) Now, for q, $q'\in\mathbb{Q}$ and $n\in\mathbb{N}$, set

$$F(q,q',n)=\{x:x\in\mathbb{R}^r,\ q\le\frac{\phi(x+(\mathbf{0},\eta))-\phi(x)}{\eta}\le q'\text{ whenever }0<|\eta|\le 2^{-n}\},$$

$$F_*(q,q',n)=\{x:x\in F(q,q',n),\ \lim_{\delta\downarrow 0}\frac{\mu^*(F(q,q',n)\cap B(x,\delta))}{\mu B(x,\delta)}=1\}.$$

By 261Da, $F(q,q',n)\setminus F_*(q,q',n)$ is negligible for all q, q', n, so that

$$H_2=H_1\setminus\bigcup_{q,q'\in\mathbb{Q},n\in\mathbb{N}}(F(q,q',n)\setminus F_*(q,q',n))$$

is conegligible.

(f) I claim that ϕ is differentiable at every point of H_2. **P** Take $x=(u,\sigma)\in H_2$. Then $\alpha=\dfrac{\partial\phi}{\partial\xi_r}(x)$ and $T=T(x)$ are defined. Let γ be a Lipschitz constant for ϕ.

Take $\epsilon>0$; take q, $q'\in\mathbb{Q}$ such that $\alpha-\epsilon\le q<\alpha<q'\le\alpha+\epsilon$. There must be an $n\in\mathbb{N}$ such that $x\in F(q,q',n)$; consequently $x\in F_*(q,q',n)$, by the definition of H_2. By 262L, there is a $\delta_0>0$ such that $\rho(x+z,F(q,q',n))\le\epsilon\|z\|$ whenever $\|z\|\le\delta_0$. Next, there is a $\delta_1>0$ such that $|\phi(x+(v,0))-\phi(x)-T(v,0)|\le\epsilon\|v\|$ whenever $v\in\mathbb{R}^{r-1}$ and $\|v\|\le\delta_1$. Set

$$\delta=\min(\delta_0,\delta_1,2^{-n})/(1+2\epsilon)>0.$$

Suppose that $z=(v,\tau)\in\mathbb{R}^r$ and that $\|z\|\le\delta$. Because $\|z\|\le\delta_0$ there is an $x'=(u',\sigma')\in F(q,q',n)$ such that $\|x+z-x'\|\le 2\epsilon\|z\|$; set $x^*=(u',\sigma)$. Now

$$\max(\|u-u'\|,|\sigma-\sigma'|)\le\|x-x'\|\le(1+2\epsilon)\|z\|\le\min(\delta_1,2^{-n})$$

and $x^*=x+(u'-u,0)$, so

$$|\phi(x^*)-\phi(x)-T(x^*-x)|\le\epsilon\|u'-u\|\le\epsilon(1+2\epsilon)\|z\|.$$

But also

$$|\phi(x')-\phi(x^*)-T(x'-x^*)|=|\phi(x')-\phi(x^*)-\alpha(\sigma'-\sigma)|\le\epsilon|\sigma'-\sigma|\le\epsilon(1+2\epsilon)\|z\|,$$

because $x' \in F(q, q', n)$ and $|\sigma - \sigma'| \le 2^{-n}$, so that (if $x' \ne x^*$)

$$\alpha - \epsilon \le q \le \frac{\phi(x^*) - \phi(x')}{\sigma - \sigma'} \le q' \le \alpha + \epsilon$$

and

$$\left|\frac{\phi(x') - \phi(x^*)}{\sigma' - \sigma} - \alpha\right| \le \epsilon.$$

Finally,

$$|\phi(x+z) - \phi(x')| \le \gamma\|x + z - x'\| \le 2\gamma\epsilon\|z\|,$$

$$|Tz - T(x' - x)| \le \|T\|\|x + z - x'\| \le 2\epsilon\|T\|\|z\|.$$

Putting all these together,

$$\begin{aligned}|\phi(x+z) - \phi x - Tz| &\le |\phi(x+z) - \phi(x')| + |T(x' - x) - Tz| \\ &\quad + |\phi(x') - \phi(x^*) - T(x' - x^*)| + |\phi(x^*) - \phi(x) - T(x^* - x)| \\ &\le 2\gamma\epsilon\|z\| + 2\epsilon\|T\|\|z\| + \epsilon(1 + 2\epsilon)\|z\| + \epsilon(1 + 2\epsilon)\|z\| \\ &= \epsilon(2\gamma + 2\|T\| + 2 + 4\epsilon)\|z\|.\end{aligned}$$

And this is true whenever $\|z\| \le \delta$. As ϵ is arbitrary, ϕ is differentiable at x. **Q**

Thus $\{x : \phi$ is differentiable at $x\}$ includes H_2 and is conegligible; and the induction continues.

262X Basic exercises (a) Let ϕ and ψ be Lipschitz functions from subsets of $\mathbb{R}^r$ to $\mathbb{R}^s$. Show that $\phi + \psi$ is a Lipschitz function from $\operatorname{dom}\phi \cap \operatorname{dom}\psi$ to $\mathbb{R}^s$.

(b) Let ϕ be a Lipschitz function from a subset of $\mathbb{R}^r$ to $\mathbb{R}^s$, and $c \in \mathbb{R}$. Show that $c\phi$ is a Lipschitz function.

(c) Suppose $\phi : D \to \mathbb{R}^s$ and $\psi : E \to \mathbb{R}^q$ are Lipschitz functions, where $D \subseteq \mathbb{R}^r$ and $E \subseteq \mathbb{R}^s$. Show that the composition $\psi\phi : D \cap \phi^{-1}[E] \to \mathbb{R}^q$ is Lipschitz.

(d) Suppose ϕ, ψ are functions from subsets of $\mathbb{R}^r$ to $\mathbb{R}^s$, and suppose that $x \in \operatorname{dom}\phi \cap \operatorname{dom}\psi$ is such that each function is differentiable relative to its domain at x, with derivatives S, T there. Show that $\phi + \psi$ is differentiable relative to its domain at x, and that $S + T$ is a derivative of $\phi + \psi$ at x.

(e) Suppose that ϕ is a function from a subset of $\mathbb{R}^r$ to $\mathbb{R}^s$, and is differentiable relative to its domain at $x \in \operatorname{dom}\phi$. Show that $c\phi$ is differentiable relative to its domain at x for every $c \in \mathbb{R}$.

>(f) Suppose $\phi : D \to \mathbb{R}^s$ and $\psi : E \to \mathbb{R}^q$ are functions, where $D \subseteq \mathbb{R}^r$ and $E \subseteq \mathbb{R}^s$; suppose that ϕ is differentiable relative to its domain at $x \in D \cap \phi^{-1}[E]$, with an $s \times r$ matrix T a derivative there, and that ψ is differentiable relative to its domain at $\phi(x)$, with a $q \times s$ matrix S a derivative there. Show that the composition $\psi\phi$ is differentiable relative to its domain at x, and that the $q \times r$ matrix ST is a derivative of $\psi\phi$ at x.

>(g) Let $\phi : \mathbb{R}^r \to \mathbb{R}^s$ be a linear operator, with associated matrix T. Show that ϕ is differentiable everywhere, with $\phi'(x) = T$ for every x.

(h) Let $G \subseteq \mathbb{R}^r$ be a convex open set, and $\phi : G \to \mathbb{R}^s$ a function such that all the partial derivatives $\frac{\partial\phi_i}{\partial\xi_j}$ are defined everywhere in G. Show that ϕ is Lipschitz iff all the partial derivatives are bounded on G.

(i) Let $\phi : \mathbb{R}^r \to \mathbb{R}^s$ be a function. Show that ϕ is differentiable at $x \in \mathbb{R}^r$ iff for every $m \in \mathbb{N}$ there are an $n \in \mathbb{N}$ and an $r \times s$ matrix T with rational coefficients such that $\|\phi(y) - \phi(x) - T(y - x)\| \le 2^{-m}\|y - x\|$ whenever $\|y - x\| \le 2^{-n}$.

(j) Suppose that f is a real-valued function which is integrable over $\mathbb{R}^r$, and that $g : \mathbb{R}^r \to \mathbb{R}$ is a bounded differentiable function such that the partial derivative $\dfrac{\partial g}{\partial\xi_j}$ is bounded, where $j \le r$. Let $f * g$ be the convolution of f and g (255E, 255L). Show that $\dfrac{\partial}{\partial\xi_j}(f*g)$ is defined everywhere and equal to $f*\dfrac{\partial g}{\partial\xi_j}$. (*Hint*: 255Xd.)

(k) Let (X,Σ,μ) be a measure space, $G\subseteq\mathbb{R}^r$ an open set, and $f:X\times G\to\mathbb{R}$ a function. Suppose that
(i) for every $x\in X$, $t\mapsto f(x,t):G\to\mathbb{R}$ is differentiable;
(ii) there is an integrable function g on X such that $|\frac{\partial f}{\partial\tau_j}(x,t)|\le g(x)$ whenever $x\in X$, $t\in G$ and $j\le r$;
(iii) $\int|f(x,t)|\mu(dx)$ exists in $\mathbb{R}$ for every $t\in G$.
Show that $t\mapsto\int f(x,t)\mu(dx):G\to\mathbb{R}$ is differentiable. (*Hint*: show first that, for a suitable M, $|f(x,t)-f(x,t')|\le M|g(x)|\|t-t'\|$ for every t, $t'\in G$ and $x\in X$.)

(l) Let $f:[a,b]\to\mathbb{R}$ be an absolutely continuous function, where $a\le b$, and $g:f[\,[a,b]\,]\to\mathbb{R}$ a Lipschitz function. Show that gf is absolutely continuous.

262Y Further exercises (a) Let L be the space of all Lipschitz functions from $\mathbb{R}^r$ to $\mathbb{R}^s$, and for $\phi\in L$ set

$$\|\phi\|=\|\phi(0)\|+\min\{\gamma:\gamma\in[0,\infty[,\ \|\phi(y)-\phi(x)\|\le\gamma\|y-x\|\text{ for every }x,\ y\in\mathbb{R}^r\}.$$

Show that $(L,\|\ \|)$ is a Banach space.

(b) Show that if $T=\langle\tau_{ij}\rangle_{i\le s,j\le r}$ is an $s\times r$ matrix then the operator norm $\|T\|$, as defined in 262H, is at most $\sqrt{\sum_{i=1}^s\sum_{j=1}^r\tau_{ij}^2}$.

(c) Let $\phi:D\to\mathbb{R}$ be any function, where $D\subseteq\mathbb{R}^r$. Show that $H=\{x:x\in D,\ \phi\text{ is differentiable relative to its domain at }x\}$ is relatively measurable in D, and that $\frac{\partial\phi}{\partial\xi_j}\restriction H$ is measurable for every $j\le r$.

(d) Let $\phi:D\to\mathbb{R}$ be a function, where $D\subseteq\mathbb{R}^r$. (i) Show that if ϕ is measurable then all its partial derivatives are measurable. (ii) Show that if ϕ is Borel measurable then all its partial derivatives are Borel measurable.

(e) A function $\phi:\mathbb{R}^r\to\mathbb{R}$ is **smooth** if all its partial derivatives $\dfrac{\partial...\partial\phi}{\partial\xi_i\partial\xi_j...\partial\xi_l}$ are defined everywhere in $\mathbb{R}^r$ and are continuous. Show that if f is integrable over $\mathbb{R}^r$ and $\phi:\mathbb{R}^r\to\mathbb{R}$ is smooth and has bounded support then the convolution $f*\phi$ is smooth. (*Hint*: 262Xj, 262Xk.)

(f) For $\delta>0$ set $\tilde\phi_\delta(x)=e^{1/(\delta^2-\|x\|^2)}$ if $\|x\|<\delta$, 0 if $\|x\|\ge\delta$; set $\alpha_\delta=\int\tilde\phi_\delta(x)dx$, $\phi_\delta(x)=\alpha_\delta^{-1}\tilde\phi_\delta(x)$ for every x. (i) Show that $\phi_\delta:\mathbb{R}^r\to\mathbb{R}$ is smooth and has bounded support. (ii) Show that if f is integrable over $\mathbb{R}^r$ then $\lim_{\delta\downarrow0}\int|f(x)-(f*\phi_\delta)(x)|dx=0$. (*Hint*: start with continuous functions f with bounded support, and use 242O.)

(g) Show that if f is integrable over $\mathbb{R}^r$ and $\epsilon>0$ there is a smooth function h with bounded support such that $\int|f-h|\le\epsilon$. (*Hint*: *either* reduce to the case in which f has bounded support and use 262Yf *or* adapt the method of 242Xi.)

(h) Suppose that f is a real function which is integrable over every bounded subset of $\mathbb{R}^r$. (i) Show that $f\times\phi$ is integrable whenever $\phi:\mathbb{R}^r\to\mathbb{R}$ is a smooth function with bounded support. (ii) Show that if $\int f\times\phi=0$ for every smooth function with bounded support then $f=0$ a.e. (*Hint*: show that $\int_{B(x,\delta)}f=0$ for every $x\in\mathbb{R}^r$ and $\delta>0$, and use 261C. Alternatively show that $\int_E f=0$ first for $E=[b,c]$, then for open sets E, then for arbitrary measurable sets E.)

(i) Let f be integrable over $\mathbb{R}^r$, and for $\delta>0$ let $\phi_\delta:\mathbb{R}^r\to\mathbb{R}$ be the function of 262Yf. Show that $\lim_{\delta\downarrow0}(f*\phi_\delta)(x)=f(x)$ for every x in the Lebesgue set of f. (*Hint*: 261Ye.)

262 Notes and comments The emphasis of this section has turned out to be on the connexions between the concepts of 'Lipschitz function' and 'differentiable function'. It is the delight of classical real analysis that such intimate relationships arise between concepts which belong to different categories. 'Lipschitz functions' clearly belong to the theory of metric spaces (I will return to this in §264), while 'differentiable functions' belong to the theory of differentiable manifolds, which is outside the scope of this volume. I have written this section out carefully just in case there are readers who have so far missed the theory of differentiable mappings between multi-dimensional Euclidean spaces; but it also gives me a chance to work through the notion of 'function differentiable relative to its domain', which will make it possible in the next section to ride smoothly past a variety of problems arising at boundaries. The difficulties I am concerned with arise in the first place with such functions as the polar-coordinate transformation

$$(\rho,\theta)\mapsto(\rho\cos\theta,\rho\sin\theta):\{(0,0)\}\cup(]0,\infty[\times]-\pi,\pi])\to\mathbb{R}^2.$$

In order to make this a bijection we have to do something rather arbitrary, and the domain of the transformation cannot be an open set. On the definitions I am using, this function is differentiable relative to its domain at every point of its domain, and we can apply such results as 262O uninhibitedly. You will observe that in this case the non-interior points of the domain form a negligible set $\{(0,0)\}\cup(]0,\infty[\times\{\pi\})$, so we can expect to be able to ignore them; and for most of the geometrically straightforward transformations that the theory is applied to, judicious excision of negligible sets will reduce problems to the case of honestly differentiable functions with open domains. But while open-domain theory will deal with a large proportion of the most important examples, there is a danger that you would be left with real misapprehensions concerning the scope of these methods.

The essence of differentiability is that a differentiable function ϕ is approximable, near any given point of its domain, by an affine function. The idea of 262M is to describe a widely effective method of dissecting $D=\operatorname{dom}\phi$ into countably many pieces on each of which ϕ is well controlled. This will be applied in §§263 and 265 to investigate the measure of $\phi[D]$; but we already have several straightforward consequences (262N-262P).

I have offered a number of results suggesting that (on the definitions I have chosen) a derivative can be expected to share at least some 'descriptive' properties with the original function; see 222Yd, 225J, 225Yg, 262Yc, 262Yd. For partial derivatives, there are complications concerning their domains (419Yd, 431Yd) which do not arise with full derivatives (225J, 262Yc).

Version of 4.4.13

263 Differentiable transformations in $\mathbb{R}^r$

This section is devoted to the proof of a single major theorem (263D) concerning differentiable transformations between subsets of $\mathbb{R}^r$. There will be a generalization of this result in §265, and those with some familiarity with the topic, or sufficient hardihood, may wish to read §264 before taking this section and §265 together. I end with a few simple corollaries and an extension of the main result which can be made in the one-dimensional case (263J).

Throughout this section, as in the rest of the chapter, μ will denote Lebesgue measure on $\mathbb{R}^r$.

263A Linear transformations I begin with the special case of linear operators, which is not only the basis of the proof of 263D, but is also one of its most important applications, and is indeed sufficient for many very striking results.

Theorem Let T be a real $r\times r$ matrix; regard T as a linear operator from $\mathbb{R}^r$ to itself. Let $J=|\det T|$ be the modulus of its determinant. Then

$$\mu T[E]=J\mu E$$

for every measurable set $E\subseteq\mathbb{R}^r$. If T is a permutation (that is, if $J\neq 0$), then

$$\mu F=J\mu(T^{-1}[F])$$

for every measurable $F\subseteq\mathbb{R}^r$, and

$$\int_F g\,d\mu = J\int_{T^{-1}[F]} gT\,d\mu$$

for every integrable function g and measurable set F.

proof (a) The first step is to show that $T[I]$ is measurable for every half-open interval $I \subseteq \mathbb{R}^r$. **P** Any non-empty half-open interval $I = [a, b[$ is a countable union of closed intervals $I_n = [a, b - 2^{-n}\mathbf{1}]$, and each I_n is compact (2A2F), so that $T[I_n]$ is compact (2A2Eb), therefore closed (2A2Ec), therefore measurable (115G), and $T[I] = \bigcup_{n\in\mathbb{N}} T[I_n]$ is measurable. **Q**

(b) Set $J^* = \mu T[\,[\mathbf{0}, \mathbf{1}[\,]$, where $\mathbf{0} = (0, \dots, 0)$ and $\mathbf{1} = (1, \dots, 1)$; because $T[\,[\mathbf{0}, \mathbf{1}[\,]$ is bounded, $J^* < \infty$. (I will eventually show that $J^* = J$.) It is convenient to deal with the case of singular T first. Recall that T, regarded as a linear transformation from $\mathbb{R}^r$ to itself, is either bijective or onto a proper linear subspace. In the latter case, take any $e \in \mathbb{R}^r \setminus T[\mathbb{R}^r]$; then the sets

$$T[\,[\mathbf{0}, \mathbf{1}[\,] + \gamma e,$$

as γ runs over $[0, 1]$, are disjoint and all of the same measure J^*, because μ is translation-invariant (134A); moreover, their union is bounded, so has finite outer measure. As there are infinitely many such γ, the common measure J^* must be zero. Now observe that

$$T[\mathbb{R}^r] = \bigcup_{z\in\mathbb{Z}^r} T[\,[\mathbf{0}, \mathbf{1}[\,] + Tz,$$

and

$$\mu(T[\,[\mathbf{0}, \mathbf{1}[\,] + Tz) = J^* = 0$$

for every $z \in \mathbb{Z}^r$, while $\mathbb{Z}^r$ is countable, so $\mu T[\mathbb{R}^r] = 0$. At the same time, because T is singular, it has zero determinant, and $J = 0$. Accordingly

$$\mu T[E] = 0 = J\mu E$$

for every measurable $E \subseteq \mathbb{R}^r$, and we're done.

(c) Henceforth, therefore, let us assume that T is non-singular. Note that it and its inverse are continuous, so that T is a homeomorphism, and $T[G]$ is open iff G is open.

If $a \in \mathbb{R}^r$ and $k \in \mathbb{N}$, then

$$\mu T[\,[a, a + 2^{-k}\mathbf{1}[\,] = 2^{-kr}J^*.$$

P Set $J_k^* = \mu T[\,[\mathbf{0}, 2^{-k}\mathbf{1}[\,]$. Now $T[\,[a, a + 2^{-k}\mathbf{1}[\,] = T[\,[\mathbf{0}, 2^{-k}\mathbf{1}[\,] + Ta$; because μ is translation-invariant, its measure also is J_k^*. Next, $[\mathbf{0}, \mathbf{1}[$ is expressible as a disjoint union of 2^{kr} sets of the form $[a, a + 2^{-k}\mathbf{1}[$; consequently, $T[\,[\mathbf{0}, \mathbf{1}[\,]$ is expressible as a disjoint union of 2^{kr} sets of the form $T[\,[a, a + 2^{-k}\mathbf{1}[\,]$, and

$$J^* = \mu T[\,[\mathbf{0}, \mathbf{1}[\,] = 2^{kr}J_k^*,$$

that is, $J_k^* = 2^{-kr}J^*$, as claimed. **Q**

(d) Consequently $\mu T[G] = J^*\mu G$ for every open set $G \subseteq \mathbb{R}^r$. **P** For each $k \in \mathbb{N}$, set

$$Q_k = \{z : z \in \mathbb{Z}^r,\ [2^{-k}z, 2^{-k}z + 2^{-k}\mathbf{1}[\ \subseteq G\},$$

$$G_k = \bigcup_{z\in Q_k} [2^{-k}z, 2^{-k}z + 2^{-k}\mathbf{1}[.$$

Then G_k is a disjoint union of $\#(Q_k)$ sets of the form $[2^{-k}z, 2^{-k}z + 2^{-k}\mathbf{1}[$, so $\mu G_k = 2^{-kr}\#(Q_k)$; also, $T[G_k]$ is a disjoint union of $\#(Q_k)$ sets of the form $T[\,[2^{-k}z, 2^{-k}z + 2^{-k}\mathbf{1}[\,]$, so has measure $2^{-kr}J^*\#(Q_k) = J^*\mu G_k$, using (c).

Observe next that $\langle G_k\rangle_{k\in\mathbb{N}}$ is a non-decreasing sequence with union G, so that

$$\mu T[G] = \lim_{k\to\infty} \mu T[G_k] = \lim_{k\to\infty} J^*\mu G_k = J^*\mu G. \text{ } \mathbf{Q}$$

(e) It follows that $\mu^*T[A] = J^*\mu^*A$ for every $A \subseteq \mathbb{R}^r$. **P** Given $A \subseteq \mathbb{R}^r$ and $\epsilon > 0$, there are open sets G, H such that $G \supseteq A$, $H \supseteq T[A]$, $\mu G \le \mu^*A + \epsilon$ and $\mu H \le \mu^*T[A] + \epsilon$ (134Fa). Set $G_1 = G \cap T^{-1}[H]$; then G_1 is open because $T^{-1}[H]$ is. Now $\mu T[G_1] = J^*\mu G_1$, so

$$\begin{aligned}\mu^*T[A] &\le \mu T[G_1] = J^*\mu G_1 \le J^*\mu^* A + J^*\epsilon\\ &\le J^*\mu G_1 + J^*\epsilon = \mu T[G_1] + J^*\epsilon \le \mu H + J^*\epsilon\\ &\le \mu^*T[A] + \epsilon + J^*\epsilon.\end{aligned}$$

As ϵ is arbitrary, $\mu^*T[A] = J^*\mu^*A$. **Q**

(f) Consequently $\mu T[E]$ exists and is equal to $J^*\mu E$ for every measurable $E \subseteq \mathbb{R}^r$. **P** Let $E \subseteq \mathbb{R}^r$ be measurable, and take any $A \subseteq \mathbb{R}^r$. Set $A' = T^{-1}[A]$. Then

$$\begin{aligned}\mu^*(A \cap T[E]) + \mu^*(A \setminus T[E]) &= \mu^*(T[A' \cap E]) + \mu^*(T[A' \setminus E])\\ &= J^*(\mu^*(A' \cap E) + \mu^*(A' \setminus E))\\ &= J^*\mu^*A' = \mu^*T[A'] = \mu^*A.\end{aligned}$$

As A is arbitrary, $T[E]$ is measurable, and now

$$\mu T[E] = \mu^*T[E] = J^*\mu^*E = J^*\mu E. \text{ **Q**}$$

(g) We are at last ready for the calculation of J^*. Recall that the matrix T must be expressible as PDQ, where P and Q are orthogonal matrices and D is diagonal, with non-negative diagonal entries (2A6C). Now we must have

$$T[\,[\mathbf{0},\mathbf{1}[\,] = P[D[Q[\,[\mathbf{0},\mathbf{1}[\,]]],$$

so, using (f),

$$J^* = J_P^* J_D^* J_Q^*,$$

where $J_P^* = \mu P[\,[\mathbf{0},\mathbf{1}[\,]$, etc. Now we find that $J_P^* = J_Q^* = 1$. **P** Let $B = B(\mathbf{0},1)$ be the unit ball of $\mathbb{R}^r$. Because B is closed, it is measurable; because it is bounded, $\mu B < \infty$; and because B includes the non-empty half-open interval $\left[\mathbf{0}, r^{-1/2}\mathbf{1}\right[$, $\mu B > 0$. Now $P[B] = Q[B] = B$, because P and Q are orthogonal matrices; so we have

$$\mu B = \mu P[B] = J_P^*\mu B,$$

and J_P^* must be 1; similarly, $J_Q^* = 1$. **Q**

(h) So we have only to calculate J_D^*. Suppose the coefficients of D are $\delta_1, \dots, \delta_r \ge 0$, so that $Dx = (\delta_1\xi_1, \dots, \delta_r\xi_r) = d \times x$. We have been assuming since the beginning of (c) that T is non-singular, so no δ_i can be 0. Accordingly

$$D[\,[\mathbf{0},\mathbf{1}[\,] = [\mathbf{0}, d[,$$

and

$$J_D^* = \mu\,[\mathbf{0}, d[= \prod_{i=1}^r \delta_i = \det D.$$

Now because P and Q are orthogonal, both have determinant ± 1, so $\det T = \pm\det D$ and $J^* = \pm\det T$; because J^* is surely non-negative, $J^* = |\det T| = J$.

(i) Thus $\mu T[E] = J\mu E$ for every Lebesgue measurable $E \subseteq \mathbb{R}^r$. As T is non-singular, we may use the above argument to show that $T^{-1}[F]$ is measurable for every measurable F, and

$$\mu F = \mu T[T^{-1}[F]] = J\mu T^{-1}[F] = \int J \times \chi(T^{-1}[F])\,d\mu,$$

identifying J with the constant function with value J. By 235A,

$$\int_F g\,d\mu = \int_{T^{-1}[F]} JgT\,d\mu = J\int_{T^{-1}[F]} gT\,d\mu$$

for every integrable function g and measurable set F.

263B Remark Perhaps I should have warned you that I should be calling on the results of §235. But if they were fresh in your mind the formulae of the statement of the theorem will have recalled them, and if not then it is perhaps better to turn back to them now rather than before reading the theorem, since they are used only in the last sentence of the proof.

I have taken the argument above at a leisurely, not to say pedestrian, pace. The point is that while the translation-invariance of Lebesgue measure, and its behaviour under simple magnification of a single coordinate, are more or less built into the definition, its behaviour under general rotations is not, since a rotation takes half-open intervals into skew cuboids. Of course the calculation of the measure of such an object is not really anything to do with the Lebesgue theory, and it will be clear that much of the argument would apply equally to any geometrically reasonable notion of r-dimensional volume.

We come now to the central result of the chapter. We have already done some of the detail work in 262M. The next basic element is the following lemma.

263C Lemma Let T be a real $r\times r$ matrix; set $J=|\det T|$. Then for any $\epsilon>0$ there is a $\zeta=\zeta(T,\epsilon)>0$ such that

(i) $|\det S-\det T|\le\epsilon$ whenever S is an $r\times r$ matrix and $\|S-T\|\le\zeta$;

(ii) whenever $D\subseteq\mathbb{R}^r$ is a bounded set and $\phi:D\to\mathbb{R}^r$ is a function such that $\|\phi(x)-\phi(y)-T(x-y)\|\le\zeta\|x-y\|$ for all x, $y\in D$, then $|\mu^*\phi[D]-J\mu^*D|\le\epsilon\mu^*D$.

proof (a) Of course (i) is the easy part. Because $\det S$ is a continuous function of the coefficients of S, and the coefficients of S must be close to those of T if $\|S-T\|$ is small (262Hb), there is surely a $\zeta_0>0$ such that $|\det S-\det T|\le\epsilon$ whenever $\|S-T\|\le\zeta_0$.

(b)(i) Write $B=B(\mathbf{0},1)$ for the unit ball of $\mathbb{R}^r$, and consider $T[B]$. We know that $\mu T[B]=J\mu B$ (263A). Let $G\supseteq T[B]$ be an open set such that $\mu G\le(J+\epsilon)\mu B$ (134Fa again). Because B is compact (2A2F again) so is $T[B]$, so there is a $\zeta_1>0$ such that $T[B]+\zeta_1 B\subseteq G$ (2A2Ed). This means that $\mu^*(T[B]+\zeta_1 B)\le(J+\epsilon)\mu B$.

(ii) Now suppose that $D\subseteq\mathbb{R}^r$ is a bounded set, and that $\phi:D\to\mathbb{R}^r$ is a function such that $\|\phi(x)-\phi(y)-T(x-y)\|\le\zeta_1\|x-y\|$ for all x, $y\in D$. Then if $x\in D$ and $\delta>0$,

$$\phi[D\cap B(x,\delta)]\subseteq\phi(x)+\delta T[B]+\delta\zeta_1 B,$$

because if $y\in D\cap B(x,\delta)$ then $T(y-x)\in\delta T[B]$ and

$$\begin{aligned}\phi(y)&=\phi(x)+T(y-x)+(\phi(y)-\phi(x)-T(y-x))\\&\in\phi(x)+\delta T[B]+\zeta_1\|y-x\|B\subseteq\phi(x)+\delta T[B]+\zeta_1\delta B.\end{aligned}$$

Accordingly

$$\begin{aligned}\mu^*\phi[D\cap B(x,\delta)]&\le\mu^*(\delta T[B]+\delta\zeta_1 B)=\delta^r\mu^*(T[B]+\zeta_1 B)\\&\le\delta^r(J+\epsilon)\mu B=(J+\epsilon)\mu B(x,\delta).\end{aligned}$$

Let $\eta>0$. Then there is a sequence $\langle B_n\rangle_{n\in\mathbb{N}}$ of balls in $\mathbb{R}^r$ such that $D\subseteq\bigcup_{n\in\mathbb{N}}B_n$, $\sum_{n=0}^\infty\mu B_n\le\mu^*D+\eta$ and the sum of the measures of those B_n whose centres do not lie in D is at most η (261F). Let K be the set of those n such that the centre of B_n lies in D. Then $\mu^*\phi[D\cap B_n]\le(J+\epsilon)\mu B_n$ for every $n\in K$. Also, of course, ϕ is $(\|T\|+\zeta_1)$-Lipschitz, so $\mu^*\phi[D\cap B_n]\le(\|T\|+\zeta_1)^r\mu B_n$ for $n\in\mathbb{N}\setminus K$ (262D). Now

$$\begin{aligned}\mu^*\phi[D]&\le\sum_{n=0}^\infty\mu^*\phi[D\cap B_n]\\&\le\sum_{n\in K}(J+\epsilon)\mu B_n+\sum_{n\in\mathbb{N}\setminus K}(\|T\|+\zeta_1)^r\mu B_n\\&\le(J+\epsilon)(\mu^*D+\eta)+\eta(\|T\|+\zeta_1)^r.\end{aligned}$$

As η is arbitrary,

$$\mu^*\phi[D]\le(J+\epsilon)\mu^*D.$$

(c) If $J = 0$, we can stop here, setting $\zeta = \min(\zeta_0, \zeta_1)$; for then we surely have $|\det S - \det T| \le \epsilon$ whenever $\|S - T\| \le \zeta$, while if $\phi : D \to \mathbb{R}^r$ is such that $\|\phi(x) - \phi(y) - T(x - y)\| \le \zeta\|x - y\|$ for all x, $y \in D$, then

$$|\mu^*\phi[D] - J\mu^*D| = \mu^*\phi[D] \le \epsilon\mu^*D.$$

If $J \ne 0$, we have more to do. Because T has non-zero determinant, it has an inverse T^{-1}, and $|\det T^{-1}| = J^{-1}$. As in (b-i) above, there is a $\zeta_2 > 0$ such that $\mu^*(T^{-1}[B] + \zeta_2 B) \le (J^{-1} + \epsilon')\mu B$, where $\epsilon' = \epsilon/J(J+\epsilon)$. Repeating (b), we see that if $C \subseteq \mathbb{R}^r$ is bounded and $\psi : C \to \mathbb{R}^r$ is such that $\|\psi(u) - \psi(v) - T^{-1}(u - v)\| \le \zeta_2\|u - v\|$ for all u, $v \in C$, then $\mu^*\psi[C] \le (J^{-1} + \epsilon')\mu^*C$.

Now suppose that $D \subseteq \mathbb{R}^r$ is bounded and $\phi : D \to \mathbb{R}^r$ is such that $\|\phi(x) - \phi(y) - T(x - y)\| \le \zeta_2'\|x - y\|$ for all x, $y \in D$, where $\zeta_2' = \min(\zeta_2, \|T^{-1}\|)/2\|T^{-1}\|^2 > 0$. Then

$$\|T^{-1}(\phi(x) - \phi(y)) - (x - y)\| \le \|T^{-1}\|\zeta_2'\|x - y\| \le \frac{1}{2}\|x - y\|$$

for all x, $y \in D$, so ϕ must be injective; set $C = \phi[D]$ and $\psi = \phi^{-1} : C \to D$. Note that C is bounded, because

$$\|\phi(x) - \phi(y)\| \le (\|T\| + \zeta_2')\|x - y\|$$

whenever x, $y \in D$. Also

$$\|T^{-1}(u - v) - (\psi(u) - \psi(v))\| \le \|T^{-1}\|\zeta_2'\|\psi(u) - \psi(v)\| \le \frac{1}{2}\|\psi(u) - \psi(v)\|$$

for all u, $v \in C$. But this means that

$$\|\psi(u) - \psi(v)\| - \|T^{-1}\|\|u - v\| \le \frac{1}{2}\|\psi(u) - \psi(v)\|$$

and $\|\psi(u) - \psi(v)\| \le 2\|T^{-1}\|\|u - v\|$ for all u, $v \in C$, so that

$$\|\psi(u) - \psi(v) - T^{-1}(u - v)\| \le 2\zeta_2'\|T^{-1}\|^2\|u - v\| \le \zeta_2\|u - v\|$$

for all u, $v \in C$.

It follows that

$$\mu^*D = \mu^*\psi[C] \le (J^{-1} + \epsilon')\mu^*C = (J^{-1} + \epsilon')\mu^*\phi[D],$$

and

$$J\mu^*D \le (1 + J\epsilon')\mu^*\phi[D].$$

(d) So if we set $\zeta = \min(\zeta_0, \zeta_1, \zeta_2') > 0$, and if $D \subseteq \mathbb{R}^r$, $\phi : D \to \mathbb{R}^r$ are such that D is bounded and $\|\phi(x) - \phi(y) - T(x - y)\| \le \zeta\|x - y\|$ for all x, $y \in D$, we shall have

$$\mu^*\phi[D] \le (J + \epsilon)\mu^*D,$$

$$\mu^*\phi[D] \ge J\mu^*D - J\epsilon'\mu^*\phi[D] \ge J\mu^*D - J\epsilon'(J + \epsilon)\mu^*D = J\mu^*D - \epsilon\mu^*D,$$

so we get the required formula

$$|\mu^*\phi[D] - J\mu^*D| \le \epsilon\mu^*D.$$

263D We are ready for the theorem.

Theorem Let $D \subseteq \mathbb{R}^r$ be any set, and $\phi : D \to \mathbb{R}^r$ a function differentiable relative to its domain at each point of D. For each $x \in D$ let $T(x)$ be a derivative of ϕ relative to D at x, and set $J(x) = |\det T(x)|$. Then

(i) $J : D \to [0, \infty[$ is a measurable function,

(ii) $\mu^*\phi[D] \le \int_D J\,d\mu$,

allowing ∞ as the value of the integral. If D is measurable, then

(iii) $\phi[D]$ is measurable.

If D is measurable and ϕ is injective, then

(iv) $\mu\phi[D] = \int_D J\,d\mu$,

(v) for every real-valued function g defined on a subset of $\phi[D]$,

$$\int_{\phi[D]} g\,d\mu = \int_D J \times g\phi\,d\mu$$

if either integral is defined in $[-\infty, \infty]$, provided we interpret $J(x)g(\phi(x))$ as zero when $J(x) = 0$ and $g(\phi(x))$ is undefined.

proof (a) To see that J is measurable, use 262P; the function $T \mapsto |\det T|$ is a continuous function of the coefficients of T, and the coefficients of $T(x)$ are measurable functions of x, by 262P, so $x \mapsto |\det T(x)|$ is measurable (121K). We also know that if D is measurable, $\phi[D]$ will be measurable, by 262Ob. Thus (i) and (iii) are done.

(b) For the moment, assume that D is bounded, and fix $\epsilon > 0$. For $r \times r$ matrices T, take $\zeta(T, \epsilon) > 0$ as in 263C. Take $\langle D_n\rangle_{n\in\mathbb{N}}$, $\langle T_n\rangle_{n\in\mathbb{N}}$ as in 262M, so that $\langle D_n\rangle_{n\in\mathbb{N}}$ is a disjoint cover of D by sets which are relatively measurable in D, and each T_n is an $r \times r$ matrix such that

$$\|T(x) - T_n\| \le \zeta(T_n, \epsilon) \text{ whenever } x \in D_n,$$

$$\|\phi(x) - \phi(y) - T_n(x - y)\| \le \zeta(T_n, \epsilon)\|x - y\| \text{ for all } x, y \in D_n.$$

Then, setting $J_n = |\det T_n|$, we have

$$|J(x) - J_n| \le \epsilon \text{ for every } x \in D_n,$$

$$|\mu^*\phi[D_n] - J_n\mu^*D_n| \le \epsilon\mu^*D_n,$$

by the choice of $\zeta(T_n, \epsilon)$. So we have

$$\int_D J\,d\mu \le \sum_{n=0}^{\infty} J_n\mu^*D_n + \epsilon\mu^*D \le \int_D J\,d\mu + 2\epsilon\mu^*D;$$

I am using here the fact that all the D_n are relatively measurable in D, so that, in particular, $\mu^*D = \sum_{n=0}^{\infty}\mu^*D_n$. Next,

$$\mu^*\phi[D] \le \sum_{n=0}^{\infty}\mu^*\phi[D_n] \le \sum_{n=0}^{\infty} J_n\mu^*D_n + \epsilon\mu^*D.$$

Putting these together,

$$\mu^*\phi[D] \le \int_D J\,d\mu + 2\epsilon\mu^*D.$$

If D is measurable and ϕ is injective, then all the D_n are measurable subsets of $\mathbb{R}^r$, so all the $\phi[D_n]$ are measurable, and they are also disjoint. Accordingly

$$\int_D J\,d\mu \le \sum_{n=0}^{\infty} J_n\mu D_n + \epsilon\mu D \le \sum_{n=0}^{\infty}(\mu\phi[D_n] + \epsilon\mu D_n) + \epsilon\mu D = \mu\phi[D] + 2\epsilon\mu D.$$

Since ϵ is arbitrary, we get

$$\mu^*\phi[D] \le \int_D J\,d\mu,$$

and if D is measurable and ϕ is injective,

$$\int_D J\,d\mu \le \mu\phi[D];$$

thus we have (ii) and (iv), on the assumption that D is bounded.

(c) For a general set D, set $B_k = B(\mathbf{0}, k)$; then

$$\mu^*\phi[D] = \lim_{k\to\infty}\mu^*\phi[D \cap B_k] \le \lim_{k\to\infty}\int_{D\cap B_k} J\,d\mu = \int_D J\,d\mu,$$

with equality if ϕ is injective and D is measurable.

(d) For part (v), I seek to show that the hypotheses of 235J are satisfied, taking $X = D$ and $Y = \phi[D]$. **P** Set $G = \{x : x \in D,\ J(x) > 0\}$.

(α) If $F \subseteq \phi[D]$ is measurable, then there are Borel sets F_1, F_2 such that $F_1 \subseteq F \subseteq F_2$ and $\mu(F_2 \setminus F_1) = 0$. Set $E_j = \phi^{-1}[F_j]$ for each j, so that $E_1 \subseteq \phi^{-1}[F] \subseteq E_2$, and both the sets E_j are measurable, because ϕ and $\operatorname{dom}\phi$ are measurable. Now, applying (iv) to $\phi\restriction E_j$,

$$\int_{E_j} J\,d\mu = \mu\phi[E_j] = \mu(F_j \cap \phi[D]) = \mu F$$

for both j, so $\int_{E_2\setminus E_1} J\,d\mu = 0$ and $J = 0$ a.e. on $E_2 \setminus E_1$. Accordingly $J \times \chi(\phi^{-1}[F]) =_{\text{a.e.}} J \times \chi E_1$, and $\int J \times \chi(\phi^{-1}[F])d\mu$ exists and is equal to $\int_{E_1} J\,d\mu = \mu F$. At the same time, $(\phi^{-1}[F] \cap G)\triangle(E_1 \cap G)$ is negligible, so $\phi^{-1}[F] \cap G$ is measurable.

(β) If $F \subseteq \phi[D]$ and $G \cap \phi^{-1}[F]$ is measurable, then we know that $\mu\phi[D \setminus G] = \int_{D\setminus G} J = 0$ (by (iv) applied to $\phi\restriction D \setminus G$), so $F \setminus \phi[G]$ must be negligible; while $F \cap \phi[G] = \phi[G \cap \phi^{-1}[F]]$ also is measurable, by (iii). Accordingly F is measurable whenever $G \cap \phi^{-1}[F]$ is measurable.

Thus all the hypotheses of 235J are satisfied. **Q** Now (v) can be read off from the conclusion of 235J.

263E Remarks (a) This is a version of the classical result on change of variable in a many-dimensional integral. What I here call $J(x)$ is the **Jacobian** of ϕ at x; it describes the change in volumes of objects near x, following the rule already established in 263A for functions with constant derivative. The idea of the proof is also the classical one: to break the set D up into small enough pieces D_m for us to be able to approximate ϕ by affine operators $y \mapsto \phi(x) + T_m(y - x)$ on each. The potential irregularity of the set D, which in this theorem may be any set, is compensated for by a corresponding freedom in choosing the sets D_m. In fact there is a further decomposition of the sets D_m hidden in part (b-ii) of the proof of 263C; each D_m is essentially covered by a disjoint family of balls, the measures of whose images we can estimate with an adequate accuracy. There is always a danger of a negligible exceptional set, and we need the crude inequalities of the proof of 262D to deal with it.

(b) Throughout the work of this chapter, from 261B to 263D, I have chosen balls $B(x, \delta)$ as the basic shapes to work with. I think it should be clear that in fact any reasonable shapes would do just as well. In particular, the 'balls'

$$B_1(x,\delta) = \{y : \textstyle\sum_{i=1}^r |\eta_i - \xi_i| \le \delta\}, \quad B_\infty(x,\delta) = \{y : |\eta_i - \xi_i| \le \delta \;\forall\, i\}$$

would serve perfectly. There are many alternatives. We could use sets of the form $C(x, k)$, for $x \in \mathbb{R}^r$ and $k \in \mathbb{N}$, defined to be the half-open cube of the form $\left[2^{-k}z, 2^{-k}(z + \mathbf{1})\right[$ with $z \in \mathbb{Z}^r$ containing x, instead; or even $C'(x, \delta) = [x, x + \delta\mathbf{1}[$. In all such cases we have versions of the density theorems (261Yb-261Yc) which support the remaining theory.

(c) I have presented 263D as a theorem about differentiable functions, because that is the normal form in which one uses it in elementary applications. However, the proof depends essentially on the fact that a differentiable function is a countable union of Lipschitz functions, and 263D would follow at once from the same theorem proved for Lipschitz functions only. Now the fact is that the theorem applies to *any* countable union of Lipschitz functions, because a Lipschitz function is differentiable almost everywhere. For more advanced work (see Federer 69 or Evans & Gariepy 92, or Chapter 47 in Volume 4) it seems clear that Lipschitz functions are the vital ones, so I spell out the result.

***263F Corollary** Let $D \subseteq \mathbb{R}^r$ be any set and $\phi : D \to \mathbb{R}^r$ a Lipschitz function. Let D_1 be the set of points at which ϕ has a derivative relative to D, and for each $x \in D_1$ let $T(x)$ be such a derivative, with $J(x) = |\det T(x)|$. Then

(i) $D \setminus D_1$ is negligible;

(ii) $J : D_1 \to [0, \infty[$ is measurable;

(iii) $\mu^*\phi[D] \le \int_D J(x)dx$.

If D is measurable, then

(iv) $\phi[D]$ is measurable.

If D is measurable and ϕ is injective, then

(v) $\mu\phi[D] = \int_D J\,d\mu$,

(vi) for every real-valued function g defined on a subset of $\phi[D]$,

$$\int_{\phi[D]} g\,d\mu = \int_D J \times g\phi\,d\mu$$

if either integral is defined in $[-\infty, \infty]$, provided we interpret $J(x)g(\phi(x))$ as zero when $J(x) = 0$ and $g(\phi(x))$ is undefined.

proof This is now just a matter of putting 262Q and 263D together, with a little help from 262D. Use 262Q to show that $D \setminus D_1$ is negligible, 262D to show that $\phi[D \setminus D_1]$ is negligible, and apply 263D to $\phi\restriction D_1$.

263G Polar coordinates in the plane I offer an elementary example with a useful consequence. Define $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ by setting $\phi(\rho, \theta) = (\rho\cos\theta, \rho\sin\theta)$ for ρ, $\theta \in \mathbb{R}^2$. Then $\phi'(\rho,\theta) = \begin{pmatrix} \cos\theta & -\rho\sin\theta \\ \sin\theta & \rho\cos\theta \end{pmatrix}$, so $J(\rho,\theta) = |\rho|$ for all ρ, θ. Of course ϕ is not injective, but if we restrict it to the domain $D = \{(0,0)\} \cup \{(\rho,\theta) : \rho > 0, -\pi < \theta \le \pi\}$ then $\phi\restriction D$ is a bijection between D and $\mathbb{R}^2$, and

$$\int g\,d\xi_1 d\xi_2 = \int_D g(\phi(\rho,\theta))\rho\,d\rho d\theta$$

for every real-valued function g which is integrable over $\mathbb{R}^2$.

Suppose, in particular, that we set

$$g(x) = e^{-\|x\|^2/2} = e^{-\xi_1^2/2} e^{-\xi_2^2/2}$$

for $x = (\xi_1, \xi_2) \in \mathbb{R}$. Then

$$\int g(x)dx = \int e^{-\xi_1^2/2} d\xi_1 \int e^{-\xi_2^2/2} d\xi_2,$$

as in 253D. Setting $I = \int e^{-t^2/2}dt$, we have $\int g = I^2$. (To see that I is well-defined in $\mathbb{R}$, note that the integrand is continuous, therefore measurable, and that

$$\int_{-1}^{1} e^{-t^2/2}dt \le 2,$$

$$\int_{-\infty}^{-1} e^{-t^2/2}dt = \int_1^\infty e^{-t^2/2}dt \le \int_1^\infty e^{-t/2}dt = \lim_{a\to\infty} \int_1^a e^{-t/2}dt = \frac{1}{2}e^{-1/2}$$

are both finite.) Now looking at the alternative expression we have

$$\begin{aligned} I^2 &= \int g(x)dx = \int_D g(\rho\cos\theta, \rho\sin\theta)\rho\,d(\rho,\theta) \\ &= \int_D e^{-\rho^2/2}\rho\,d(\rho,\theta) = \int_0^\infty \int_{-\pi}^{\pi} \rho e^{-\rho^2/2} d\theta d\rho \end{aligned}$$

(ignoring the point $(0,0)$, which has zero measure)

$$\begin{aligned} &= \int_0^\infty 2\pi\rho e^{-\rho^2/2} d\rho = 2\pi \lim_{a\to\infty} \int_0^a \rho e^{-\rho^2/2} d\rho \\ &= 2\pi \lim_{a\to\infty} (-e^{-a^2/2} + 1) = 2\pi. \end{aligned}$$

Consequently

$$\int_{-\infty}^{\infty} e^{-t^2/2}dt = I = \sqrt{2\pi},$$

which is one of the many facts every mathematician should know, and in particular is vital for Chapter 27 below.

263H Corollary If $k \in \mathbb{N}$ is odd,

$$\int_{-\infty}^{\infty} x^k e^{-x^2/2}dx = 0;$$

if $k = 2l \in \mathbb{N}$ is even, then

$$\int_{-\infty}^{\infty} x^k e^{-x^2/2}dx = \frac{(2l)!}{2^l l!}\sqrt{2\pi}.$$

proof (a) To see that all the integrals are well-defined and finite, observe that $\lim_{x\to\pm\infty} x^k e^{-x^2/4} = 0$, so that $M_k = \sup_{x\in\mathbb{R}} |x^k e^{-x^2/4}|$ is finite, and

$$\int_{-\infty}^{\infty} |x^k e^{-x^2/2}|dx \le M_k \int_{-\infty}^{\infty} e^{-x^2/4}dx < \infty.$$

(b) If k is odd, then substituting $y = -x$ we get

$$\int_{-\infty}^{\infty} x^k e^{-x^2/2}dx = -\int_{-\infty}^{\infty} y^k e^{-y^2/2}dy,$$

so that both integrals must be zero.

(c) For even k, proceed by induction. Set $I_l = \int_{-\infty}^{\infty} x^{2l} e^{-x^2/2} dx$. $I_0 = \sqrt{2\pi} = \frac{0!}{2^0 0!}\sqrt{2\pi}$ by 263G. For the inductive step to $l+1 \ge 1$, integrate by parts to see that

$$\int_{-a}^{a} x^{2l+1} \cdot x e^{-x^2/2} dx = -a^{2l+1} e^{-a^2/2} + (-a)^{2l+1} e^{-a^2/2} + \int_{-a}^{a} (2l+1) x^{2l} e^{-x^2/2} dx$$

for every $a \ge 0$. Letting $a \to \infty$,

$$I_{l+1} = (2l+1) I_l.$$

Because

$$\frac{(2(l+1))!}{2^{l+1}(l+1)!}\sqrt{2\pi} = (2l+1)\frac{(2l)!}{2^l l!}\sqrt{2\pi},$$

the induction proceeds.

263I The following is a version of 263D for non-injective transformations.

Theorem Let $D \subseteq \mathbb{R}^r$ be a measurable set, and $\phi : D \to \mathbb{R}^r$ a function differentiable relative to its domain at each point of D. For each $x \in D$ let $T(x)$ be a derivative of ϕ relative to D at x, and set $J(x) = |\det T(x)|$.

(a) Let ν be counting measure on $\mathbb{R}^r$. Then $\int_{\mathbb{R}^r} \nu(\phi^{-1}[\{y\}]) dy$ and $\int_D J\, d\mu$ are defined in $[0, \infty]$ and equal.

(b) Let g be a real-valued function defined on a subset of $\phi[D]$ such that $\int_D g(\phi(x)) \det T(x) dx$ is defined in $\mathbb{R}$, interpreting $g(\phi(x)) \det T(x)$ as zero when $\det T(x) = 0$ and $g(\phi(x))$ is undefined. Set

$$C = \{y : y \in \phi[D],\ \phi^{-1}[\{y\}] \text{ is finite}\}, \quad R(y) = \sum_{x \in \phi^{-1}[\{y\}]} \operatorname{sgn} \det T(x)$$

for $y \in C$, where $\operatorname{sgn}(0) = 0$ and $\operatorname{sgn}(\alpha) = \frac{\alpha}{|\alpha|}$ for non-zero α. If we interpret $g(y)R(y)$ as zero when $g(y) = 0$ and $R(y)$ is undefined, then $\int_{\phi[D]} g \times R\, d\mu$ is defined and equal to $\int_D g(\phi(x)) \det T(x) dx$.

proof (a) By 263D(i), J is measurable, so $\int_D J\, d\mu$ is defined in $[0, \infty]$ and $D_0 = \{x : x \in D,\ J(x) = 0\}$ is measurable. Applying 263D(ii) to $\phi \restriction D_0$, we see that $\phi[D_0]$ is negligible.

Applying 262M to $\phi \restriction D \setminus D_0$, the set A of non-singular $r \times r$-matrices and $\zeta(S) = \frac{1}{2\|S^{-1}\|}$ for $S \in A$, we have a partition $\langle E_n \rangle_{n \in \mathbb{N}}$ of $D \setminus D_0$ into measurable sets and a sequence $\langle T_n \rangle_{n \in \mathbb{N}}$ in A such that

$$\|\phi(x) - \phi(y) - T_n(x-y)\| \le \frac{1}{2\|T_n^{-1}\|}\|x - y\|, \quad \|T(x) - T_n\| \le \frac{1}{2\|T_n^{-1}\|}$$

whenever $n \in \mathbb{N}$ and $x, y \in E_n$. In this case, for $x, y \in E_n$,

$$\phi(x) = \phi(y) \implies \|x - y\| \le \|T_n^{-1}\|\|T_n(x-y)\| \le \frac{1}{2}\|x - y\| \implies x = y,$$

so $\phi \restriction E_n$ is injective, for each n. Consequently $\#(\phi^{-1}[\{y\}]) = \#(\{n : y \in \phi[E_n]\})$ for $y \in \phi[D] \setminus \phi[D_0]$, and

$$\int_{\phi[D]} \nu(\phi^{-1}[\{y\}]) dy = \int_{\phi[D] \setminus \phi[D_0]} \nu(\phi^{-1}[\{y\}]) dy = \sum_{n=0}^{\infty} \mu(\phi[E_n] \setminus \phi[D_0])$$

(applying 263D(iii) to $\phi \restriction E_n$, we know that $\phi[E_n]$ is measurable for each n)

$$= \sum_{n=0}^{\infty} \mu\phi[E_n] = \sum_{n=0}^{\infty} \int_{E_n} J$$

(applying 263D(iv) to $\phi \restriction E_n$)

$$= \int_{D \setminus D_0} J = \int_D J,$$

each sum or integral being defined in $[0, \infty]$ because the next one is.

(b)(i) Setting $D' = \phi^{-1}[\operatorname{dom} g]$, we see that $D \setminus (D_0 \cup D')$ is negligible, so (using 263D(ii) again)

$$\phi[D] \setminus \operatorname{dom} g = \phi[D] \setminus \phi[D'] \subseteq \phi[D_0] \cup \phi[D \setminus (D_0 \cup D')]$$

is negligible. Next, if we set $C' = C \cup g^{-1}[\{0\}]$, $\phi[D] \setminus C'$ is negligible. **P** For each $m \in \mathbb{N}$, set $F_m = \{y : y \in \operatorname{dom} g,\ |g(y)| \ge 2^{-m}\}$. Then

$$\phi^{-1}[F_m] \setminus D_0 = \{x : x \in D,\ J(x) \ne 0,\ g(\phi(x)) \text{ is defined and } |g(\phi(x))| \ge 2^{-m}\}$$

is measurable (because $J \times g\phi$ is measurable) and (applying (a) to $\phi\restriction\phi^{-1}[F_m]$)

$$\int_{F_m} \nu(\phi^{-1}[\{y\}])dy = \int_{\phi^{-1}[F_m]} J\,d\mu \le 2^m \int_{\phi^{-1}[F_m]} |J \times g(\phi)|\,d\mu$$

is finite. But this means that $\nu(\phi^{-1}[\{y\}])$ must be finite for almost every $y \in F_m$, that is, that $F_m \setminus C$ is negligible. As m is arbitrary, $\operatorname{dom} g \setminus C'$ and $\phi[D] \setminus C'$ are negligible. **Q**

(ii) Taking $\langle E_n\rangle_{n\in\mathbb{N}}$ and $\langle T_n\rangle_{n\in\mathbb{N}}$ as in (a), set $\epsilon_n = \operatorname{sgn}\det T_n \in \{-1,1\}$ for each n. Then $\operatorname{sgn}\det T(x) = \epsilon_n$ whenever $n \in \mathbb{N}$ and $x \in E_n$. **P** For any $\alpha \in [0,1]$,

$$\|(\alpha T(x) + (1-\alpha)T_n) - T_n\| \le \|T(x) - T_n\| \le \frac{1}{2\|T_n^{-1}\|},$$

so (using 2A4Fd) $\|(\alpha T(x)+(1-\alpha)T_n)-I_r\| \le \frac{1}{2}$, where I_r is the $r\times r$ identity matrix, and $\alpha T(x)+(1-\alpha)T_n$ is non-singular (since if $(\alpha T(x)+(1-\alpha)T_n)z = 0$, then $\|z\| \le \frac{1}{2}\|z\|$). Thus $\det(\alpha T(x)+(1-\alpha)T_n)$ is non-zero for $0 \le \alpha \le 1$. But as $\alpha \mapsto \det(\alpha T(x) + (1-\alpha)T_n)$ is continuous, $\epsilon_n = \operatorname{sgn}\det T_n = \operatorname{sgn}\det T(x)$. **Q**

(iii) Now

$$\begin{aligned}\int_D g(\phi(x))\det T(x)\,dx &= \int_{D\setminus D_0} g(\phi(x))\det T(x)\,dx\\ &= \sum_{n=0}^{\infty}\int_{E_n} g(\phi(x))\det T(x)\,dx\end{aligned}$$

(in this series of formulae, each sum and integral is well-defined because the preceding ones are)

$$\begin{aligned}&= \sum_{n=0}^{\infty}\epsilon_n\int_{E_n} g(\phi(x))J(x)\,dx = \sum_{n=0}^{\infty}\epsilon_n\int_{\phi[E_n]} g\,dx\\ &= \sum_{n=0}^{\infty}\epsilon_n\int_{\phi[E_n]\cap C'} g\,dx = \sum_{n=0}^{\infty}\epsilon_n\int_{C'} g \times \chi(\phi[E_n])\,dx\end{aligned}$$

(using 131Fa, if you like, for the last step). Since we also have

$$\infty > \int_D |g(\phi(x))\det T(x)|\,dx = \sum_{n=0}^{\infty}\int_{C'} |g| \times \chi(\phi[E_n])\,dx$$

(going through the same stages with the absolute values of integrands), we have

$$\begin{aligned}\int_D g(\phi(x))\det T(x)\,dx &= \sum_{n=0}^{\infty}\epsilon_n\int_{C'} g \times \chi(\phi[E_n])\,dx\\ &= \int_{C'} g \times \sum_{n=0}^{\infty}\epsilon_n\chi(\phi[E_n])\,dx\\ &= \int_{C'\setminus\phi[D_0]} g \times \sum_{n=0}^{\infty}\epsilon_n\chi(\phi[E_n])\,dx = \int_{C'\setminus\phi[D_0]} g \times R\,d\mu\end{aligned}$$

(because if $y \in C' \setminus \phi[D_0]$, either $g(y) = 0$ or $R(y)$ is defined and equal to $\sum_{n=0}^{\infty}\epsilon_n\chi(\phi[E_n])(y)$)

$$= \int_{\phi[D]} g \times R\,d\mu,$$

as claimed.

263J The one-dimensional case The restriction to injective functions ϕ in 263D(v) is unavoidable in the context of the result there. But in the substitutions of elementary calculus it is not always essential. In the hope of clarifying the position I give a result here which covers many of the standard tricks.

Proposition Let $I \subseteq \mathbb{R}$ be an interval with more than one point, and $\phi : I \to \mathbb{R}$ a function which is absolutely continuous on any closed bounded subinterval of I. Write $u = \inf I$, $u' = \sup I$ in $[-\infty, \infty]$, and suppose that $v = \lim_{x\downarrow u} \phi(x)$ and $v' = \lim_{x\uparrow u'} \phi(x)$ are defined in $[-\infty, \infty]$. Let g be a real function such that $\int_I g(\phi(x))\phi'(x)dx$ is defined, on the understanding that we interpret $g(\phi(x))\phi'(x)$ as 0 when $\phi'(x) = 0$ and $g(\phi(x))$ is undefined. Then $\int_v^{v'} g$ is defined and equal to $\int_I g(\phi(x))\phi'(x)dx$, where here we interpret $\int_v^{v'} g$ as $-\int_{v'}^{v} g$ if $v' < v$.

proof (a) ϕ is differentiable almost everywhere on I and $\phi[A]$ is negligible for every negligible $A \subseteq I$. **P** We can express I as the union of a sequence $\langle I_n\rangle_{n\in\mathbb{N}}$ of closed bounded intervals such that $\phi\restriction I_n$ is absolutely continuous for every n. By 225Cb and 225G, $\phi\restriction I_n$ is differentiable almost everywhere on I_n and $\phi[A]$ is negligible for every negligible $A \subseteq I_n$, for each n. So ϕ is differentiable almost everywhere on $\bigcup_{n\in\mathbb{N}} I_n = I$ and $\phi[A] = \bigcup_{n\in\mathbb{N}} \phi[A \cap I_n]$ is negligible for every negligible $A \subseteq I$. **Q**

Because $\phi\restriction J$ is continuous for every closed interval $J \subseteq I$, ϕ is continuous. By the Intermediate Value Theorem, $\phi[I]$ is an interval including $]\min(v, v'), \max(v, v')[$.

(b) Let $D \subseteq I$ be the domain of ϕ'. For $x \in D$, we can think of $\phi'(x)$ as a 1×1 matrix with determinant $\phi'(x)$. As $I \setminus D$ is negligible, $\phi[I] \setminus \phi[D] \subseteq \phi[I \setminus D]$ is negligible. Now $\int_D g(\phi(x))\phi'(x)dx = \int_I g(\phi(x))\phi'(x)dx$. Applying 263I to $\phi\restriction D$ and $g\restriction\phi[D]$, we see that $\int_{\phi[D]} g \times R$ is defined and equal to $\int_D g\phi \times \phi'$, where $R(y) = \sum_{x\in D\cap\phi^{-1}[\{y\}]} \operatorname{sgn} \phi'(x)$ whenever $y \in \phi[D]$ and $D \cap \phi^{-1}[\{y\}]$ is finite.

(c) (The key.) Set $D_0 = \{x : x \in D,\ \phi'(x) = 0\}$. By 263D(ii), applied to $\phi\restriction D_0$, $\phi[D_0]$ is negligible. Set

$$C = \{y : y \in \phi[D] \cap \operatorname{dom} g \setminus (\phi[D_0] \cup \{v, v'\}),\ \phi^{-1}[\{y\}] \text{ is finite, } g(y) \neq 0\}.$$

If $y \in C$ and $K = \phi^{-1}[\{y\}]$, then

$$\begin{aligned} R(y) = \sum_{x\in K} \operatorname{sgn} \phi'(x) &= 1 \text{ if } v < y < v', \\ &= -1 \text{ if } v' < y < v, \\ &= 0 \text{ otherwise.} \end{aligned}$$

P If $J \subseteq I \setminus K$ is an interval, $\phi(z) \neq y$ for $z \in J$; since ϕ is continuous, the Intermediate Value Theorem tells us that $\operatorname{sgn}(\phi(z) - y)$ is constant on J. Also $\phi'(x) \neq 0$ for every $x \in K$, because $y \notin \phi[D_0]$. A simple induction on $\#(K \cap]-\infty, z[)$ shows that $\operatorname{sgn}(\phi(z) - y) = \operatorname{sgn}(v - y) + 2\sum_{x\in K, x<z} \operatorname{sgn} \phi'(x)$ for every $z \in I \setminus K$; taking the limit as $z \uparrow u'$, $\sum_{x\in K} \operatorname{sgn} \phi'(x) = \frac{1}{2}(\operatorname{sgn}(v' - y) - \operatorname{sgn}(v - y))$. (Here we may have to interpret $\operatorname{sgn}(\pm\infty)$ as ±1 in the obvious way.) This turns out to be just what we need to know. **Q**

(d) So now we have

$$\begin{aligned} \int_I g\phi \times \phi' &= \int_D g\phi \times \phi' = \int_{\phi[D]} g \times R \\ &= \int_{\phi[D]} g \times R \times \chi C = \int_v^{v'} g \times \chi C = \int_v^{v'} g \end{aligned}$$

because $\phi[D] \setminus (C \cup g^{-1}[\{0\}])$ is negligible.

263X Basic exercises (a) Let (X, Σ, μ) be any measure space, $f \in \mathcal{L}^0(\mu)$ and $p \in [1, \infty[$. Show that $f \in \mathcal{L}^p(\mu)$ iff

$$\gamma = p\int_0^\infty \alpha^{p-1}\mu^*\{x : x \in \operatorname{dom} f,\ |f(x)| > \alpha\}d\alpha$$

is finite, and in this case $\|f\|_p = \gamma^{1/p}$. (*Hint*: $\int |f|^p = \int_0^\infty \mu^*\{x : |f(x)|^p > \beta\}d\beta$, by 252O; now substitute $\beta = \alpha^p$.)

(b) Let f be an integrable function defined almost everywhere in $\mathbb{R}^r$. Show that if $\alpha < r-1$ then $\sum_{n=1}^{\infty} n^{\alpha}|f(nx)|$ is finite for almost every $x \in \mathbb{R}^r$. (*Hint*: estimate $\sum_{n=0}^{\infty} n^{\alpha}\int_B |f(nx)|dx$ for balls B centered at the origin.)

(c) Let $A \subseteq \,]0,1[$ be a set such that $\mu^*A = \mu^*([0,1]\setminus A) = 1$, where μ is Lebesgue measure on $\mathbb{R}$. Set $D = A \cup \{-x : x \in \,]0,1[\,\setminus A\} \subseteq [-1,1]$, and set $\phi(x) = |x|$ for $x \in D$. Show that ϕ is injective, that ϕ is differentiable relative to its domain everywhere in D, and that $\mu^*\phi[D] < \int_D |\phi'(x)|dx$.

(d) Let $\phi : D \to \mathbb{R}^r$ be a function differentiable relative to D at each point of $D \subseteq \mathbb{R}^r$, and suppose that for each $x \in D$ there is a non-singular derivative $T(x)$ of ϕ at x. Show that D is expressible as $\bigcup_{k\in\mathbb{N}} D_k$ where $D_k = D \cap \overline{D}_k$ and $\phi{\restriction}D_k$ is injective for each k.

>(e)(i) Show that for any Lebesgue measurable $E \subseteq \mathbb{R}$ and $t \in \mathbb{R}\setminus\{0\}$, $\int_{tE}\frac{1}{|u|}du = \int_E \frac{1}{|u|}du$. (ii) For $t \in \mathbb{R}$, $u \in \mathbb{R}\setminus\{0\}$ set $\phi(t,u) = (\frac{t}{u}, u)$. Show that $\int_{\phi[E]}\frac{1}{|tu|}d(t,u) = \int_E \frac{1}{|tu|}d(t,u)$ for any Lebesgue measurable $E \subseteq \mathbb{R}^2$.

>(f) Define $\phi : \mathbb{R}^3 \to \mathbb{R}^3$ by setting

$$\phi(\rho,\theta,\alpha) = (\rho\sin\theta\sin\alpha, \rho\cos\theta\sin\alpha, \rho\cos\alpha).$$

Show that $\det\phi'(\rho,\theta,\alpha) = \rho^2\sin\alpha$.

(g) Show that if $k = 2l+1$ is odd, then $\int_0^{\infty} x^k e^{-x^2/2}dx = 2^l l!$. (Compare 252Xi.)

263Y Further exercises (a) Define a measure ν on $\mathbb{R}$ by setting $\nu E = \int_E \frac{1}{|x|}dx$ for Lebesgue measurable sets $E \subseteq \mathbb{R}$. For f, $g \in \mathcal{L}^1(\nu)$ set $(f*g)(x) = \int f(\frac{x}{t})g(t)\nu(dt)$ whenever this is defined in $\mathbb{R}$. (i) Show that $f*g = g*f \in \mathcal{L}^1(\nu)$. (ii) Show that $\int h(x)(f*g)(x)\nu(dx) = \int h(xy)f(x)f(y)\nu(dx)\nu(dy)$ for every $h \in \mathcal{L}^{\infty}(\nu)$. (iii) Show that $f*(g*h) = (f*g)*h$ for every $h \in \mathcal{L}^1(\nu)$. (*Hint*: 263Xe.)

(b) Let $E \subseteq \mathbb{R}^2$ be a measurable set such that $\limsup_{\alpha\to\infty}\frac{1}{\alpha^2}\mu_2(E\cap B(\mathbf{0},\alpha)) > 0$, writing μ_2 for Lebesgue measure on $\mathbb{R}^2$. Show that there is some $\theta \in \,]-\pi,\pi]$ such that $\mu_1 E_\theta = \infty$, where $E_\theta = \{\rho : \rho \ge 0,\ (\rho\cos\theta, \rho\sin\theta) \in E\}$. (*Hint*: show that $\mu_2(E\cap B(\mathbf{0},\alpha)) \le \alpha\int_{-\pi}^{\pi}\mu_1 E_\theta)d\theta$.) Generalize to higher dimensions and to functions other than χE.

(c) Let $E \subseteq \mathbb{R}^r$ be a measurable set, and $\phi : E \to \mathbb{R}^r$ a function differentiable relative to its domain, with a derivative $T(x)$, at each point x of E; set $J(x) = |\det T(x)|$. Show that for any integrable function g defined on $\phi[E]$,

$$\int g(y)\#(\phi^{-1}[\{y\}])dy = \int_E J(x)g(\phi(x))dx.$$

(d) Find a proof of 263J based on the ideas of §225. (*Hint*: 225Xe.)

(e) Let $f : [a,b] \to \mathbb{R}$ be a function of bounded variation, where $a < b$ in $\mathbb{R}$, with Lebesgue decomposition $f = f_p + f_{cs} + f_{ac}$ as in 226Cd; let μ be Lebesgue measure on $\mathbb{R}$. Show that the following are equiveridical: (i) f_{cs} is constant; (ii) $\mu f[\,[c,d]\,] \le \int_c^d |f'|d\mu$ whenever $a \le c \le d \le b$; (iii) $\mu^* f[A] \le \int_A |f'|d\mu$ for every $A \subseteq [a,b]$; (iv) $f[A]$ is negligible for every negligible set $A \subseteq [a,b]$. (*Hint*: for (iv)⇒(i) put 226Yd and 263D(ii) together to show that $|f(d)-f(c)| \le \int_c^d |f'|d\mu + \mathrm{Var}_{[c,d]} f_p$ whenever $a \le c \le d \le b$, and therefore that $\mathrm{Var}_{[a,b]} f \le \mathrm{Var}_{[a,b]} f_p + \mathrm{Var}_{[a,b]} f_{ac}$.)

(f) Suppose that $r = 2$ and that $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ is continuously differentiable with non-singular derivative T at $\mathbf{0}$. (i) Show that there is an $\epsilon > 0$ such that whenever Γ is a small circle with centre $\mathbf{0}$ and radius at most ϵ then $\phi{\restriction}\Gamma$ is a homeomorphism between Γ and a simple closed curve around $\mathbf{0}$. (ii) Show that if $\det T > 0$, then for such circles $\phi(x)$ runs anticlockwise around $\phi[\Gamma]$ as x runs anticlockwise around Γ. (iii) What happens if $\det T < 0$?

263 Notes and comments Yet again, approaching 263D, I find myself having to choose between giving an accessible, relatively weak result and making the extra effort to set out a theorem which is somewhere near the natural boundary of what is achievable within the concepts being developed in this volume; and, as usual, I go for the more powerful form. There are three basic sources of difficulty: (i) the fact that we are dealing with more than one dimension; (ii) the fact that we are dealing with irregular domains; (iii) the fact that we are dealing with arbitrary integrable functions. I do not think I need to apologise for (iii) in a book on measure theory. Concerning (ii), it is quite true that the principal applications of these results are to cases in which the transformation ϕ is differentiable everywhere, with continuous derivative, and the set D has negligible boundary; and in these cases there are substantial simplifications available – mostly because the sets D_m of the proof of 263D can be taken to be cubes. Nevertheless, I think any form of the result which makes such assumptions is deeply unsatisfactory at this level, being an awkward compromise between ideas natural to the Riemann integral and those natural to the Lebesgue integral. Concerning (i), it might even have been right to lay out the whole argument for the case $r = 1$ before proceeding to the general case, as I did in §§114-115, because the one-dimensional case is already important and interesting; and if you find the work above difficult – which it is – and your immediate interests are in one-dimensional integration by substitution, then I think you might find it worth your time to reproduce the $r = 1$ argument yourself, up to a proof of 263J. In fact the biggest difference is in 263A, which becomes nearly trivial; the work of 262M and 263C becomes more readable, because all the matrices turn into scalars and we can drop the word 'determinant', but I do not think we can dispense with any of the ideas, at least if we wish to obtain 263D as stated. (But see 263Yd.)

I found myself insisting, in the last paragraph, that a distinction can be made between 'ideas natural to the Riemann integral and those natural to the Lebesgue integral'. We are approaching deep questions here, like 'what are books on measure theory for?', which I do not think can be answered without some – possibly unconscious – reference to the question 'what is mathematics for?'. I do of course want to present here some of the wonderful general theorems which arise in the Lebesgue theory. But more important than any specific theorem is a general idea of what can be proved by these methods. It is the essence of modern measure theory that continuity does not matter, or, if you prefer, that measurable functions are in some sense so nearly continuous that we do not have to add hypotheses of continuity in our theorems. Now this is in a sense a great liberation, and the Lebesgue integral is now the standard one. But you must not regard the Riemann integral as outdated. The intuitions on which it is founded – for instance, that the surface of a solid body has zero volume – remain of great value in their proper context, which certainly includes the study of differentiable functions with continuous derivatives. What I am saying here is that I believe we can use these intuitions best if we maintain a division, a flexible and permeable one, of course, between the ideas of the two theories; and that when transferring a theorem from one side of the boundary to the other we should do so whole-heartedly, seeking to express the full power of the methods we are using.

I have already said that the essential difference between the one-dimensional and multi-dimensional cases lies in 263A, where the Jacobian $J = |\det T|$ enters the argument. Shorn of the technical devices necessary to deal with arbitrary Lebesgue measurable sets, this amounts to a calculation of the volume of the parallelepiped $T[I]$ where I is the interval $[\mathbf{0}, \mathbf{1}[$. I have dealt with this by a little bit of algebra, saying that the result is essentially obvious if T is diagonal, whereas if T is an isometry it follows from the fact that the unit ball is left invariant; and the algebra comes in to express an arbitrary matrix as a product of diagonal and orthogonal matrices. It is also plain from 261F that Lebesgue measure must be rotation-invariant as well as translation-invariant; that is to say, it is invariant under all isometries. Another way of looking at this will appear in the next section.

I feel myself that the centre of the argument for 263D is in the lemma 263C. This is where we turn the exact result for linear operators into an approximate result for almost-linear functions; and the whole point of differentiability is that a differentiable function is well approximated, near any point of its domain, by a linear operator. The lemma involves two rather different ideas. To show that $\mu^*\phi[D] \le (J+\epsilon)\mu^* D$, we look first at balls and then use Vitali's theorem to see that D is economically covered by balls, so that an upper bound for $\mu^*\phi[D]$ in terms of a sum $\sum_{B\in\mathcal{I}_0} \mu^*\phi[D \cap B]$ is adequate. To obtain a lower bound, we need to reverse the argument by looking at $\psi = \phi^{-1}$, which involves checking first that ϕ is invertible, and then that ψ is appropriately linked to T^{-1}. I have written out exact formulae for ϵ', ζ_2' and so on, but this is only in case you do not trust your intuition; the fact that $\|\phi^{-1}(u) - \phi^{-1}(v) - T^{-1}(u - v)\|$ is small compared with $\|u - v\|$ is pretty clearly a consequence of the hypothesis that $\|\phi(x) - \phi(y) - T(x - y)\|$ is small compared with $\|x-y\|$.

The argument of 263D itself is now a matter of breaking the set D up into appropriate pieces on each of which ϕ is sufficiently nearly linear for 263C to apply, so that

$$\mu^*\phi[D] \le \sum_{m=0}^{\infty} \mu^*\phi[D_m] \le \sum_{m=0}^{\infty} (J_m+\epsilon)\mu^* D_m.$$

With a little care (taken in 263C, with its condition (i)), we can also ensure that the Jacobian J is well approximated by J_m almost everywhere in D_m, so that $\sum_{m=0}^{\infty} J_m\mu^* D_m \simeq \int_D J(x)dx$.

These ideas, joined with the results of §262, bring us to the point

$$\int_E J\,d\mu = \mu\phi[E]$$

when ϕ is injective and $E \subseteq D$ is measurable. We need a final trick, involving Borel sets, to translate this into

$$\int_{\phi^{-1}[F]} J\,d\mu = \mu F$$

whenever $F \subseteq \phi[D]$ is measurable, which is what is needed for the application of 235J.

I hope that you long ago saw, and were delighted by, the device in 263G. Once again, this is not really Lebesgue integration; but I include it just to show that the machinery of this chapter can be turned to deal with the classical results, and that indeed we have a tiny profit from our labour, in that no apology need be made for the boundary of the set D into which the polar coordinate system maps the plane. I have already given the actual result as an exercise in 252Xi. That involved (if you chase through the references) a one-dimensional substitution (performed in 225Xh), Fubini's theorem and an application of the formulae of §235; that is to say, very much the same elements as those used above, though in a different order. I could present this with no mention of differentiation in higher dimensions because the first change of variable was in one dimension, and the second (involving the function $x \mapsto \|x\|$, in 252Xi(i)) was of a particularly simple type, so that a different method could be used to find the function J.

The function $R(y) = \sum_{x\in\phi^{-1}[\{y\}]} \operatorname{sgn}\det T(x)$ of 263Ib belongs to rather deeper notions in differential geometry than I wish to enlarge on here. In the one-dimensional context it simply counts up- and down-crossings of y (see part (c) of the proof of 263J), because we can think of each $T(x)$ as a scalar which is either positive or negative. (It is relevant that in this case we can assume that $T(x)$ is non-singular whenever $\phi(x) = y$.) In higher dimensions, I suppose the first thing to look for is a geometric interpretation of $\operatorname{sgn}\det T(x)$. (See 263Ye.) A geometric interpretation of the sum is something else again. But it is worth noting that (subject to a natural interpretation of '0×undefined') we can relate $\int_D g\phi \times \det\phi'$ to $\int_{\phi[D]} g\times R$ where R is definable from ϕ without reference to g; it counts folds in the graph of ϕ.

The abstract ideas to which this treatise is devoted do not, indeed, lead us to many particular examples on which to practise the ideas of this section. The ones which do arise tend to be very straightforward, as in 263G, 263Xa-263Xb and 263Xe. I mention the last because it provides a formula needed to discuss a new type of convolution (263Ya). In effect, this depends on the multiplicative group $\mathbb{R}\setminus\{0\}$ in place of the additive group $\mathbb{R}$ treated in §255. The formula $\frac{1}{x}$ in the definition of ν is of course the derivative of $\ln x$, and ln is an isomorphism between $(]0,\infty[\,,\cdot,\nu)$ and $(\mathbb{R},+,\text{Lebesgue measure})$.

Version of 12.5.03

264 Hausdorff measures

The next topic I wish to approach is the question of 'surface measure'; a useful example to bear in mind throughout this section and the next is the notion of area for regions on the sphere, but any other smoothly curved two-dimensional surface in three-dimensional space will serve equally well. It is I think more than plausible that our intuitive concepts of 'area' for such surfaces should correspond to appropriate measures. But formalizing this intuition is non-trivial, especially if we seek the generality that simple geometric ideas lead us to; I mean, not contenting ourselves with arguments that depend on the special nature of the sphere, for instance, to describe spherical surface area. I divide the problem into two parts. In this section I will describe a construction which enables us to define the r-dimensional measure of an r-dimensional surface –

among other things – in s-dimensional space. In the next section I will set out the basic theorems making it possible to calculate these measures effectively in the leading cases.

264A Definitions Let $s \ge 1$ be an integer, and $r > 0$. (I am primarily concerned with integral r, but will not insist on this until it becomes necessary, since there are some very interesting ideas which involve non-integral 'dimension' r.) For any $A \subseteq \mathbb{R}^s$, $\delta > 0$ set

$$\theta_{r\delta}A = \inf\{\sum_{n=0}^{\infty}(\operatorname{diam} A_n)^r : \langle A_n\rangle_{n\in\mathbb{N}} \text{ is a sequence of subsets of } \mathbb{R}^s \text{ covering } A,$$

$$\operatorname{diam} A_n \le \delta \text{ for every } n \in \mathbb{N}\}.$$

It is convenient in this context to say that $\operatorname{diam}\emptyset = 0$. Now set

$$\theta_r A = \sup_{\delta>0} \theta_{r\delta}A;$$

θ_r is r**-dimensional Hausdorff outer measure** on $\mathbb{R}^s$.

264B Of course we must immediately check the following:

Lemma θ_r, as defined in 264A, is always an outer measure.

proof You should be used to these arguments by now, but there is an extra step in this one, so I spell out the details.

(a) Interpreting the diameter of the empty set as 0, we have $\theta_{r\delta}\emptyset = 0$ for every $\delta > 0$, so $\theta_r\emptyset = 0$.

(b) If $A \subseteq B \subseteq \mathbb{R}^s$, then every sequence covering B also covers A, so $\theta_{r\delta}A \le \theta_{r\delta}B$ for every δ and $\theta_r A \le \theta_r B$.

(c) Let $\langle A_n\rangle_{n\in\mathbb{N}}$ be a sequence of subsets of $\mathbb{R}^s$ with union A, and take any $a < \theta_r A$. Then there is a $\delta > 0$ such that $a \le \theta_{r\delta}A$. Now $\theta_{r\delta}A \le \sum_{n=0}^{\infty}\theta_{r\delta}(A_n)$. **P** Let $\epsilon > 0$, and for each $n \in \mathbb{N}$ choose a sequence $\langle A_{nm}\rangle_{m\in\mathbb{N}}$ of sets, covering A_n, with $\operatorname{diam} A_{nm} \le \delta$ for every m and $\sum_{m=0}^{\infty}(\operatorname{diam} A_{nm})^r \le \theta_{r\delta}A_n + 2^{-n}\epsilon$. Then $\langle A_{nm}\rangle_{m,n\in\mathbb{N}}$ is a cover of A by countably many sets of diameter at most δ, so

$$\theta_{r\delta}A \le \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}(\operatorname{diam} A_{nm})^r \le \sum_{n=0}^{\infty}\theta_{r\delta}A_n + 2^{-n}\epsilon = 2\epsilon + \sum_{n=0}^{\infty}\theta_{r\delta}A_n.$$

As ϵ is arbitrary, we have the result. **Q**

Accordingly

$$a \le \theta_{r\delta}A \le \sum_{n=0}^{\infty}\theta_{r\delta}A_n \le \sum_{n=0}^{\infty}\theta_r A_n.$$

As a is arbitrary,

$$\theta_r A \le \sum_{n=0}^{\infty}\theta_r A_n;$$

as $\langle A_n\rangle_{n\in\mathbb{N}}$ is arbitrary, θ_r is an outer measure.

264C Definition If $s \ge 1$ is an integer, and $r > 0$, then **Hausdorff** r**-dimensional measure** on $\mathbb{R}^s$ is the measure μ_{Hr} on $\mathbb{R}^s$ defined by Carathéodory's method from the outer measure θ_r of 264A-264B.

264D Remarks (a) It is important to note that the sets used in the definition of the $\theta_{r\delta}$ need not be balls; even in $\mathbb{R}^2$ not every set A can be covered by a ball of the same diameter as A.

(b) In the definitions above I require $r > 0$. It is sometimes appropriate to take μ_{H0} to be counting measure. This is nearly the result of applying the formulae above with $r = 0$, but there can be difficulties if we interpret them over-literally.

(c) All Hausdorff measures must be complete, because they are defined by Carathéodory's method (212A). For $r > 0$, they are atomless (264Yg). In terms of the other criteria of §211, however, they are very ill-behaved; for instance, if r, s are integers and $1 \le r < s$, then μ_{Hr} on $\mathbb{R}^s$ is not semi-finite. (I will give a proof of this in 439H in Volume 4.) Nevertheless, they do have some striking properties which make them reasonably tractable.

(d) In 264A, note that $\theta_{r\delta}A \le \theta_{r\delta'}A$ when $0 < \delta' \le \delta$; consequently, for instance, $\theta_r A = \lim_{n\to\infty} \theta_{r,2^{-n}}A$. I have allowed arbitrary sets A_n in the covers, but it makes no difference if we restrict our attention to covers consisting of open sets or of closed sets (264Xc).

264E Theorem Let $s \ge 1$ be an integer, and $r \ge 0$; let μ_{Hr} be Hausdorff r-dimensional measure on $\mathbb{R}^s$, and Σ_{Hr} its domain. Then every Borel subset of $\mathbb{R}^s$ belongs to Σ_{Hr}.

proof This is trivial if $r = 0$; so suppose henceforth that $r > 0$.

(a) The first step is to note that if A, B are subsets of $\mathbb{R}^s$ and $\eta > 0$ is such that $\|x - y\| \ge \eta$ for all $x \in A$, $y \in B$, then $\theta_r(A\cup B) = \theta_r A + \theta_r B$, where θ_r is r-dimensional Hausdorff outer measure on $\mathbb{R}^s$. **P** Of course $\theta_r(A \cup B) \le \theta_r A + \theta_r B$, because θ_r is an outer measure. For the reverse inequality, we may suppose that $\theta_r(A \cup B) < \infty$, so that $\theta_r A$ and $\theta_r B$ are both finite. Let $\epsilon > 0$ and let δ_1, $\delta_2 > 0$ be such that

$$\theta_r A + \theta_r B \le \theta_{r\delta_1} A + \theta_{r\delta_2} B + \epsilon.$$

Set $\delta = \min(\delta_1, \delta_2, \frac{1}{2}\eta) > 0$ and let $\langle A_n\rangle_{n\in\mathbb{N}}$ be a sequence of sets of diameter at most δ, covering $A \cup B$, and such that $\sum_{n=0}^{\infty}(\operatorname{diam} A_n)^r \le \theta_{r\delta}(A \cup B) + \epsilon$. Set

$$K = \{n : A_n \cap A \ne \emptyset\}, \quad L = \{n : A_n \cap B \ne \emptyset\}.$$

Because

$$\|x - y\| \ge \eta > \operatorname{diam} A_n$$

whenever $x \in A$, $y \in B$ and $n \in \mathbb{N}$, $K \cap L = \emptyset$; and of course $A \subseteq \bigcup_{n\in K} A_k$, $B \subseteq \bigcup_{n\in L} A_n$. Consequently

$$\begin{aligned}\theta_r A + \theta_r B &\le \epsilon + \theta_{r\delta_1} A + \theta_{r\delta_2} B \\ &\le \epsilon + \sum_{n\in K}(\operatorname{diam} A_n)^r + \sum_{n\in L}(\operatorname{diam} A_n)^r \\ &\le \epsilon + \sum_{n=0}^{\infty}(\operatorname{diam} A_n)^r \le 2\epsilon + \theta_{r\delta}(A \cup B) \le 2\epsilon + \theta_r(A \cup B).\end{aligned}$$

As ϵ is arbitrary, $\theta_r(A \cup B) \ge \theta_r A + \theta_r B$, as required. **Q**

(b) It follows that $\theta_r A = \theta_r(A \cap G) + \theta_r(A \setminus G)$ whenever $A \subseteq \mathbb{R}^s$ and G is open. **P** As usual, it is enough to consider the case $\theta_r A < \infty$ and to show that in this case $\theta_r(A \cap G) + \theta_r(A \setminus G) \le \theta_r A$. Set

$$A_n = \{x : x \in A,\ \|x - y\| \ge 2^{-n} \text{ for every } y \in A \setminus G\},$$

$$B_0 = A_0, \quad B_n = A_n \setminus A_{n-1} \text{ for } n > 1.$$

Observe that $A_n \subseteq A_{n+1}$ for every n and $\bigcup_{n\in\mathbb{N}} A_n = \bigcup_{n\in\mathbb{N}} B_n = A \cap G$. The point is that if m, $n \in \mathbb{N}$ and $n \ge m + 2$, and if $x \in B_m$ and $y \in B_n$, then there is a $z \in A \setminus G$ such that $\|y - z\| < 2^{-n+1} \le 2^{-m-1}$, while $\|x - z\|$ must be at least 2^{-m}, so $\|x - y\| \ge \|x - z\| - \|y - z\| \ge 2^{-m-1}$. It follows that for any $k \ge 0$

$$\textstyle\sum_{m=0}^{k} \theta_r B_{2m} = \theta_r(\bigcup_{m\le k} B_{2m}) \le \theta_r(A \cap G) < \infty,$$

$$\textstyle\sum_{m=0}^{k} \theta_r B_{2m+1} = \theta_r(\bigcup_{m\le k} B_{2m+1}) \le \theta_r(A \cap G) < \infty,$$

(inducing on k, using (a) above for the inductive step). Consequently $\sum_{n=0}^{\infty} \theta_r B_n < \infty$.

But now, given $\epsilon > 0$, there is an m such that $\sum_{n=m}^{\infty} \theta_r B_m \le \epsilon$, so that

$$\begin{aligned}\theta_r(A \cap G) + \theta_r(A \setminus G) &\le \theta_r A_m + \sum_{n=m}^{\infty} \theta_r B_n + \theta_r(A \setminus G) \\ &\le \epsilon + \theta_r A_m + \theta_r(A \setminus G) = \epsilon + \theta_r(A_m \cup (A \setminus G))\end{aligned}$$

(by (a) again, since $\|x - y\| \ge 2^{-m}$ for $x \in A_m$, $y \in A \setminus G$)

$$\le \epsilon + \theta_r A.$$

As ϵ is arbitrary, $\theta_r(A \cap G) + \theta_r(A \setminus G) \le \theta_r A$, as required. **Q**

(c) Part (b) shows exactly that open sets belong to Σ_{Hr}. It follows at once that the Borel σ-algebra of $\mathbb{R}^s$ is included in Σ_{Hr}, as claimed.

264F Proposition Let $s \ge 1$ be an integer, and $r > 0$; let θ_r be r-dimensional Hausdorff outer measure on $\mathbb{R}^s$, and write μ_{Hr} for r-dimensional Hausdorff measure on $\mathbb{R}^s$, Σ_{Hr} for its domain. Then

(a) for every $A \subseteq \mathbb{R}^s$ there is a Borel set $E \supseteq A$ such that $\mu_{Hr}E = \theta_r A$;

(b) $\theta_r = \mu^*_{Hr}$, the outer measure defined from μ_{Hr};

(c) if $E \in \Sigma_{Hr}$ is expressible as a countable union of sets of finite measure, there are Borel sets E', E'' such that $E' \subseteq E \subseteq E''$ and $\mu_{Hr}(E'' \setminus E') = 0$.

proof (a) If $\theta_r A = \infty$ this is trivial – take $E = \mathbb{R}^s$. Otherwise, for each $n \in \mathbb{N}$ choose a sequence $\langle A_{nm}\rangle_{m\in\mathbb{N}}$ of sets of diameter at most 2^{-n}, covering A, and such that $\sum_{m=0}^{\infty}(\operatorname{diam} A_{nm})^r \le \theta_{r,2^{-n}}A + 2^{-n}$. Set $F_{nm} = \overline{A}_{nm}$, $E = \bigcap_{n\in\mathbb{N}}\bigcup_{m\in\mathbb{N}} F_{nm}$; then E is a Borel set in $\mathbb{R}^s$. Of course

$$A \subseteq \bigcap_{n\in\mathbb{N}}\bigcup_{m\in\mathbb{N}} A_{mn} \subseteq \bigcap_{n\in\mathbb{N}}\bigcup_{m\in\mathbb{N}} F_{nm} = E.$$

For any $n \in \mathbb{N}$,

$$\operatorname{diam} F_{nm} = \operatorname{diam} A_{nm} \le 2^{-n} \text{ for every } m \in \mathbb{N},$$

$$\sum_{m=0}^{\infty}(\operatorname{diam} F_{nm})^r = \sum_{m=0}^{\infty}(\operatorname{diam} A_{nm})^r \le \theta_{r,2^{-n}}A + 2^{-n},$$

so

$$\theta_{r,2^{-n}}E \le \theta_{r,2^{-n}}A + 2^{-n}.$$

Letting $n \to \infty$,

$$\theta_r E = \lim_{n\to\infty}\theta_{r,2^{-n}}E \le \lim_{n\to\infty}\theta_{r,2^{-n}}A + 2^{-n} = \theta_r A;$$

of course it follows that $\theta_r A = \theta_r E$, because $A \subseteq E$. Now by 264E we know that $E \in \Sigma_{Hr}$, so we can write $\mu_{Hr}E$ in place of $\theta_r E$.

(b) This follows at once, because we have

$$\mu^*_{Hr}A = \inf\{\mu_{Hr}E : E \in \Sigma_{Hr},\ A \subseteq E\} = \inf\{\theta_r E : E \in \Sigma_{Hr},\ A \subseteq E\} \ge \theta_r A$$

for every $A \subseteq \mathbb{R}^s$. On the other hand, if $A \subseteq \mathbb{R}^s$, we have a Borel set $E \supseteq A$ such that $\theta_r A = \mu_{Hr}E$, so that $\mu^*_{Hr}A \le \mu_{Hr}E = \theta_r A$.

(c)(i) Suppose first that $\mu_{Hr}E < \infty$. By (a), there are Borel sets $E'' \supseteq E$, $H \supseteq E'' \setminus E$ such that $\mu_{Hr}E'' = \theta_r E$,

$$\mu_{Hr}H = \theta_r(E'' \setminus E) = \mu_{Hr}(E'' \setminus E) = \mu_{Hr}E'' - \mu_{Hr}E = \mu_{Hr}E'' - \theta_r E = 0.$$

So setting $E' = E'' \setminus H$, we obtain a Borel set included in E, and

$$\mu_{Hr}(E'' \setminus E') \le \mu_{Hr}H = 0.$$

(ii) For the general case, express E as $\bigcup_{n\in\mathbb{N}} E_n$ where $\mu_{Hr}E_n < \infty$ for each n; take Borel sets E'_n, E''_n such that $E'_n \subseteq E_n \subseteq E''_n$ and $\mu_{Hr}(E''_n \setminus E'_n) = 0$ for each n; and set $E' = \bigcup_{n\in\mathbb{N}} E'_n$, $E'' = \bigcup_{n\in\mathbb{N}} E''_n$.

264G Lipschitz functions The definition of Hausdorff measure is exactly adapted to the following result, corresponding to 262D.

Proposition Let m, $s \ge 1$ be integers, and $\phi : D \to \mathbb{R}^s$ a γ-Lipschitz function, where D is a subset of $\mathbb{R}^m$. Then for any $A \subseteq D$ and $r \ge 0$,

$$\mu^*_{Hr}(\phi[A]) \le \gamma^r \mu^*_{Hr}A$$

for every $A \subseteq D$, writing μ_{Hr} for r-dimensional Hausdorff outer measure on either $\mathbb{R}^m$ or $\mathbb{R}^s$.

proof (a) The case $r = 0$ is trivial, since then $\gamma^r = 1$ and $\mu^*_{Hr}A = \mu_{H0}A = \#(A)$ if A is finite, ∞ otherwise, while $\#(\phi[A]) \le \#(A)$.

(b) If $r > 0$, then take any $\delta > 0$. Set $\eta = \delta/(1+\gamma)$ and consider $\theta_{r\eta} : \mathcal{P}\mathbb{R}^m \to [0,\infty]$, defined as in 264A. We know from 264Fb that

$$\mu^*_{Hr}A = \theta_r A \ge \theta_{r\eta}A,$$

so there is a sequence $\langle A_n\rangle_{n\in\mathbb{N}}$ of sets, all of diameter at most η, covering A, with $\sum_{n=0}^{\infty}(\operatorname{diam} A_n)^r \le \mu^*_{Hr}A + \delta$. Now $\phi[A] \subseteq \bigcup_{n\in\mathbb{N}} \phi[A_n \cap D]$ and

$$\operatorname{diam} \phi[A_n \cap D] \le \gamma \operatorname{diam} A_n \le \gamma\eta \le \delta$$

for every n. Consequently

$$\theta_{r\delta}(\phi[A]) \le \sum_{n=0}^{\infty}(\operatorname{diam} \phi[A_n])^r \le \sum_{n=0}^{\infty} \gamma^r(\operatorname{diam} A_n)^r \le \gamma^r(\mu^*_{Hr}A + \delta),$$

and

$$\mu^*_{Hr}(\phi[A]) = \lim_{\delta\downarrow 0} \theta_{r\delta}(\phi[A]) \le \gamma^r \mu^*_{Hr}A,$$

as claimed.

264H The next step is to relate r-dimensional Hausdorff measure on $\mathbb{R}^r$ to Lebesgue measure on $\mathbb{R}^r$. The basic fact we need is the following, which is even more important for the idea in its proof than for the result.

Theorem Let $r \ge 1$ be an integer, and A a bounded subset of $\mathbb{R}^r$; write μ_r for Lebesgue measure on $\mathbb{R}^r$ and $d = \operatorname{diam} A$. Then

$$\mu^*_r(A) \le \mu_r B(\mathbf{0}, \frac{d}{2}) = 2^{-r}\beta_r d^r,$$

where $B(\mathbf{0}, \frac{d}{2})$ is the ball with centre $\mathbf{0}$ and diameter d, so that $B(\mathbf{0}, 1)$ is the unit ball in $\mathbb{R}^r$, and has measure

$$\begin{aligned}\beta_r &= \frac{1}{k!}\pi^k \text{ if } r = 2k \text{ is even,}\\ &= \frac{2^{2k+1}k!}{(2k+1)!}\pi^k \text{ if } r = 2k+1 \text{ is odd.}\end{aligned}$$

proof (a) For the calculation of β_r, see 252Q or 252Xi.

(b) The case $r = 1$ is elementary, for in this case A is included in an interval of length $\operatorname{diam} A$, so that $\mu^*_1 A \le \operatorname{diam} A$. So henceforth let us suppose that $r \ge 2$.

(c) For $1 \le i \le r$ let $S_i : \mathbb{R}^r \to \mathbb{R}^r$ be reflection in the ith coordinate, so that $S_i x = (\xi_1, \dots, \xi_{i-1}, -\xi_i, \xi_{i+1}, \dots, \xi_r)$ for every $x = (\xi_1, \dots, \xi_r) \in \mathbb{R}^r$. Let us say that a set $C \subseteq \mathbb{R}^r$ is **symmetric in coordinates in** J, where $J \subseteq \{1, \dots, r\}$, if $S_i[C] = C$ for $i \in J$. Now the centre of the argument is the following fact: if $C \subseteq \mathbb{R}$ is a bounded set which is symmetric in coordinates in J, where J is a proper subset of $\{1, \dots, r\}$, and $j \in \{1, \dots, r\} \setminus J$, then there is a set D, symmetric in coordinates in $J \cup \{j\}$, such that $\operatorname{diam} D \le \operatorname{diam} C$ and $\mu^*_r C \le \mu^*_r D$.

P (i) Because Lebesgue measure is invariant under permutation of coordinates, it is enough to deal with the case $j = r$. Start by writing $F = \overline{C}$, so that $\operatorname{diam} F = \operatorname{diam} C$ and $\mu_r F \ge \mu^*_r C$. Note that because S_i is a homeomorphism for every i,

$$S_i[F] = S_i[\overline{C}] = \overline{S_i[C]} = \overline{C} = F$$

for $i \in J$, and F is symmetric in coordinates in J.

For $y = (\eta_1, \dots, \eta_{r-1}) \in \mathbb{R}^{r-1}$, set

$$F_y = \{\xi : (\eta_1, \dots, \eta_{r-1}, \xi) \in F\}, \quad f(y) = \mu_1 F_y,$$

where μ_1 is Lebesgue measure on $\mathbb{R}$. Set

$$D = \{(y, \xi) : y \in \mathbb{R}^{r-1},\ |\xi| < \frac{1}{2}f(y)\} \subseteq \mathbb{R}^r.$$

(ii) If $H \subseteq \mathbb{R}^r$ is measurable and $H \supseteq D$, then, writing μ_{r-1} for Lebesgue measure on $\mathbb{R}^{r-1}$, we have

$$\mu_r H = \int \mu_1\{\xi : (y,\xi) \in H\}\mu_{r-1}(dy)$$

(using 251N and 252D)

$$\geq \int \mu_1\{\xi : (y,\xi) \in D\}\mu_{r-1}(dy) = \int f(y)\mu_{r-1}(dy)$$
$$= \int \mu_1\{\xi : (y,\xi) \in F\}\mu_{r-1}(dy) = \mu_r F \geq \mu_r^* C.$$

As H is arbitrary, $\mu_r^* D \geq \mu_r^* C$.

(iii) The next step is to check that $\operatorname{diam} D \leq \operatorname{diam} C$. If x, $x' \in D$, express them as (y, ξ_r) and (y', ξ'_r). Because F is a bounded closed set in $\mathbb{R}^r$, F_y and $F_{y'}$ are bounded closed subsets of $\mathbb{R}$. Also both $f(y)$ and $f(y')$ must be greater than 0, so that F_y, $F_{y'}$ are both non-empty. Consequently

$$\alpha = \inf F_y, \quad \beta = \sup F_y, \quad \alpha' = \inf F_{y'}, \quad \beta' = \sup F_{y'}$$

are all defined in $\mathbb{R}$, and α, $\beta \in F_y$, while α' and β' belong to $F_{y'}$. We have

$$|\xi_r - \xi'_r| \leq |\xi_r| + |\xi'_r| < \frac{1}{2}f(y) + \frac{1}{2}f(y')$$
$$= \frac{1}{2}(\mu_1 F_y + \mu_1 F_{y'}) \leq \frac{1}{2}(\beta - \alpha + \beta' - \alpha') \leq \max(\beta' - \alpha, \beta - \alpha').$$

So taking (ξ, ξ') to be one of (α, β') or (β, α'), we can find $\xi \in F_y$, $\xi' \in F_{y'}$ such that $|\xi - \xi'| \geq |\xi_r - \xi'_r|$. Now $z = (y, \xi)$, $z' = (y', \xi')$ both belong to F, so

$$\|x - x'\|^2 = \|y - y'\|^2 + |\xi_r - \xi'_r|^2 \leq \|y - y'\|^2 + |\xi - \xi'|^2 = \|z - z'\|^2 \leq (\operatorname{diam} F)^2,$$

and $\|x - x'\| \leq \operatorname{diam} F$. As x and x' are arbitrary, $\operatorname{diam} D \leq \operatorname{diam} F = \operatorname{diam} C$, as claimed.

(iv) Evidently $S_r[D] = D$. Moreover, if $i \in J$, then (interpreting S_i as an operator on $\mathbb{R}^{r-1}$)

$$F_{S_i(y)} = F_y \text{ for every } y \in \mathbb{R}^{r-1},$$

so $f(S_i(y)) = f(y)$ and, for $\xi \in \mathbb{R}$, $y \in \mathbb{R}^{r-1}$,

$$(y, \xi) \in D \iff |\xi| < \tfrac{1}{2}f(y) \iff |\xi| < \tfrac{1}{2}f(S_i(y)) \iff (S_i(y), \xi) \in D,$$

so that $S_i[D] = D$. Thus D is symmetric in coordinates in $J \cup \{r\}$. **Q**

(d) The rest is easy. Starting from any bounded $A \subseteq \mathbb{R}^r$, set $A_0 = A$ and construct inductively $A_1, \dots, A_r$ such that

$$d = \operatorname{diam} A = \operatorname{diam} A_0 \geq \operatorname{diam} A_1 \geq \ldots \geq \operatorname{diam} A_r,$$

$$\mu_r^* A = \mu_r^* A_0 \leq \ldots \leq \mu_r^* A_r,$$

$$A_j \text{ is symmetric in coordinates in } \{1, \dots, j\} \text{ for every } j \leq r.$$

At the end, we have A_r symmetric in coordinates in $\{1, \dots, r\}$. But this means that if $x \in A_r$ then

$$-x = S_1 S_2 \ldots S_r x \in A_r,$$

so that

$$\|x\| = \frac{1}{2}\|x - (-x)\| \leq \frac{1}{2}\operatorname{diam} A_r \leq \frac{d}{2}.$$

Thus $A_r \subseteq B(\mathbf{0}, \frac{d}{2})$, and

$$\mu_r^* A \leq \mu_r^* A_r \leq \mu_r B(\mathbf{0}, \frac{d}{2}),$$

as claimed.

264I Theorem Let $r \ge 1$ be an integer; let μ be Lebesgue measure on $\mathbb{R}^r$, and let μ_{Hr} be r-dimensional Hausdorff measure on $\mathbb{R}^r$. Then μ and μ_{Hr} have the same measurable sets and

$$\mu E = 2^{-r}\beta_r\mu_{Hr}E$$

for every measurable set $E \subseteq \mathbb{R}^r$, where $\beta_r = \mu B(\mathbf{0},1)$, so that the normalizing factor is

$$2^{-r}\beta_r = \frac{1}{2^{2k}k!}\pi^k \text{ if } r = 2k \text{ is even,}$$

$$= \frac{k!}{(2k+1)!}\pi^k \text{ if } r = 2k+1 \text{ is odd.}$$

proof (a) Of course if $B = B(x,\alpha)$ is any ball of radius α,

$$2^{-r}\beta_r(\operatorname{diam} B)^r = \beta_r\alpha^r = \mu B.$$

(b) The point is that $\mu^* = 2^{-r}\beta_r\mu^*_{Hr}$. **P** Let $A \subseteq \mathbb{R}^r$.

(i) Let δ, $\epsilon > 0$. By 261F, there is a sequence $\langle B_n\rangle_{n\in\mathbb{N}}$ of balls, all of diameter at most δ, such that $A \subseteq \bigcup_{n\in\mathbb{N}} B_n$ and $\sum_{n=0}^\infty \mu B_n \le \mu^*A + \epsilon$. Now, defining $\theta_{r\delta}$ as in 264A,

$$2^{-r}\beta_r\theta_{r\delta}(A) \le 2^{-r}\beta_r\sum_{n=0}^\infty(\operatorname{diam} B_n)^r = \sum_{n=0}^\infty \mu B_n \le \mu^*A + \epsilon.$$

Letting $\delta \downarrow 0$,

$$2^{-r}\beta_r\mu^*_{Hr}A \le \mu^*A + \epsilon.$$

As ϵ is arbitrary, $2^{-r}\beta_r\mu^*_{Hr}A \le \mu^*A$.

(ii) Let $\epsilon > 0$. Then there is a sequence $\langle A_n\rangle_{n\in\mathbb{N}}$ of sets of diameter at most 1 such that $A \subseteq \bigcup_{n\in\mathbb{N}} A_n$ and $\sum_{n=0}^\infty(\operatorname{diam} A_n)^r \le \theta_{r1}A + \epsilon$, so that

$$\mu^*A \le \sum_{n=0}^\infty \mu^*A_n \le \sum_{n=0}^\infty 2^{-r}\beta_r(\operatorname{diam} A_n)^r \le 2^{-r}\beta_r(\theta_{r1}A + \epsilon) \le 2^{-r}\beta_r(\mu^*_{Hr}A + \epsilon)$$

by 264H. As ϵ is arbitrary, $\mu^*A \le 2^{-r}\beta_r\mu^*_{Hr}A$. **Q**

(c) Because μ, μ_{Hr} are the measures defined from their respective outer measures by Carathéodory's method, it follows at once that $\mu = 2^{-r}\beta_r\mu_{Hr}$ in the strict sense required.

***264J The Cantor set** I remarked in 264A that fractional 'dimensions' r were of interest. I have no space for these here, and they are off the main lines of this volume, but I will give one result for its intrinsic interest.

Proposition Let C be the Cantor set in $[0,1]$. Set $r = \ln 2/\ln 3$. Then the r-dimensional Hausdorff measure of C is 1.

proof (a) Recall that $C = \bigcap_{n\in\mathbb{N}} C_n$, where each C_n consists of 2^n closed intervals of length 3^{-n}, and C_{n+1} is obtained from C_n by deleting the middle (open) third of each interval of C_n. (See 134G.) Because C is closed, $\mu_{Hr}C$ is defined (264E). Note that $3^r = 2$.

(b) If $\delta > 0$, take n such that $3^{-n} \le \delta$; then C can be covered by 2^n intervals of diameter 3^{-n}, so

$$\theta_{r\delta}C \le 2^n(3^{-n})^r = 1.$$

Consequently

$$\mu_{Hr}C = \mu^*_{Hr}C = \lim_{\delta\downarrow 0}\theta_{r\delta}C \le 1.$$

(c) We need the following elementary fact: if α, β, $\gamma \ge 0$ and $\max(\alpha,\gamma) \le \beta$, then $\alpha^r + \gamma^r \le (\alpha+\beta+\gamma)^r$. **P** Because $0 < r \le 1$,

$$\xi \mapsto (\xi+\eta)^r - \xi^r = r\int_0^\eta(\xi+\zeta)^{r-1}d\zeta$$

is non-increasing for every $\eta \ge 0$. Consequently

$$(\alpha+\beta+\gamma)^r-\alpha^r-\gamma^r \ge (\beta+\beta+\gamma)^r-\beta^r-\gamma^r$$
$$\ge (\beta+\beta+\beta)^r-\beta^r-\beta^r=\beta^r(3^r-2)=0,$$

as required. **Q**

(d) Now suppose that $I\subseteq\mathbb{R}$ is any interval, and $m\in\mathbb{N}$; write $j_m(I)$ for the number of the intervals composing C_m which are included in I. Then $2^{-m}j_m(I)\le(\operatorname{diam}I)^r$. **P** If I does not meet C_m, this is trivial. Otherwise, induce on

$$l=\min\{i: I \text{ meets only one of the intervals composing } C_{m-i}\}.$$

If $l=0$, so that I meets only one of the intervals composing C_m, then $j_m(I)\le 1$, and if $j_m(I)=1$ then $\operatorname{diam}I\ge 3^{-m}$ so $(\operatorname{diam}I)^r\ge 2^{-m}$; thus the induction starts. For the inductive step to $l>1$, let J be the interval of C_{m-l} which meets I, and J', J'' the two intervals of C_{m-l+1} included in J, so that I meets both J' and J'', and

$$j_m(I)=j_m(I\cap J)=j_m(I\cap J')+j_m(I\cap J'').$$

By the inductive hypothesis,

$$(\operatorname{diam}(I\cap J'))^r+(\operatorname{diam}(I\cap J''))^r\ge 2^{-m}j_m(I\cap J')+2^{-m}j_m(I\cap J'')=2^{-m}j_m(I).$$

On the other hand, by (c),

$$(\operatorname{diam}(I\cap J'))^r+(\operatorname{diam}(I\cap J''))^r\le(\operatorname{diam}(I\cap J')+3^{-m+l-1}+\operatorname{diam}(I\cap J''))^r$$
$$=(\operatorname{diam}(I\cap J))^r\le(\operatorname{diam}I)^r$$

because J', J'' both have diameter at most $3^{-(m-l+1)}$, the length of the interval between them. Thus the induction continues. **Q**

(e) Now suppose that $\epsilon>0$. Then there is a sequence $\langle A_n\rangle_{n\in\mathbb{N}}$ of sets, covering C, such that

$$\textstyle\sum_{n=0}^{\infty}(\operatorname{diam}A_n)^r<\mu_{Hr}C+\epsilon.$$

Take $\eta_n>0$ such that $\sum_{n=0}^{\infty}(\operatorname{diam}A_n+\eta_n)^r\le\mu_{Hr}C+\epsilon$, and for each n take an open interval $I_n\supseteq A_n$ of length at most $\operatorname{diam}A_n+\eta_n$ and with neither endpoint belonging to C; this is possible because C does not include any non-trivial interval. Now $C\subseteq\bigcup_{n\in\mathbb{N}}I_n$; because C is compact, there is a $k\in\mathbb{N}$ such that $C\subseteq\bigcup_{n\le k}I_n$. Next, there is an $m\in\mathbb{N}$ such that no endpoint of any I_n, for $n\le k$, belongs to C_m. Consequently each of the intervals composing C_m must be included in some I_n, and (in the terminology of (d) above) $\sum_{n=0}^{k}j_m(I_n)\ge 2^m$. Accordingly

$$1\le\textstyle\sum_{n=0}^{k}2^{-m}j_m(I_n)\le\sum_{n=0}^{k}(\operatorname{diam}I_n)^r\le\sum_{n=0}^{\infty}(\operatorname{diam}A_n+\eta_n)^r\le\mu_{Hr}C+\epsilon.$$

As ϵ is arbitrary, $\mu_{Hr}C\ge 1$, as required.

***264K General metric spaces** While this chapter deals exclusively with Euclidean spaces, readers familiar with the general theory of metric spaces may find the nature of the theory clearer if they use the language of metric spaces in the basic definitions and results. I therefore repeat the definition here, and spell out the corresponding results in the exercises 264Yb-264Yl.

Let (X,ρ) be a metric space, and $r>0$. For any $A\subseteq X$, $\delta>0$ set

$$\theta_{r\delta}A=\inf\{\sum_{n=0}^{\infty}(\operatorname{diam}A_n)^r:\langle A_n\rangle_{n\in\mathbb{N}}\text{ is a sequence of subsets of }X\text{ covering }A,$$
$$\operatorname{diam}A_n\le\delta\text{ for every }n\in\mathbb{N}\},$$

interpreting the diameter of the empty set as 0, and $\inf\emptyset$ as ∞, so that $\theta_{r\delta}A=\infty$ if A cannot be covered by a sequence of sets of diameter at most δ. Say that $\theta_rA=\sup_{\delta>0}\theta_{r\delta}A$ is the ***r*-dimensional Hausdorff outer measure** of A, and take the measure μ_{Hr} defined by Carathéodory's method from this outer measure to be ***r*-dimensional Hausdorff measure** on X.

264X Basic exercises >(a) Show that all the functions $\theta_{r\delta}$ of 264A are outer measures. Show that in that context, $\theta_{r\delta}(A) = 0$ iff $\theta_r(A) = 0$, for any $\delta > 0$ and any $A \subseteq \mathbb{R}^s$.

(b) Let $s \ge 1$ be an integer, and θ an outer measure on $\mathbb{R}^s$ such that $\theta(A \cup B) = \theta A + \theta B$ whenever A, B are non-empty subsets of $\mathbb{R}^s$ and $\inf_{x\in A, y\in B} \|x - y\| > 0$. Show that every Borel subset of $\mathbb{R}^s$ is measured by the measure defined from θ by Carathéodory's method.

>(c) Let $s \ge 1$ be an integer and $r > 0$; define $\theta_{r\delta}$ as in 264A. Show that for any $A \subseteq \mathbb{R}^s$, $\delta > 0$,

$$\begin{aligned}\theta_{r\delta}A = \inf\{\sum_{n=0}^{\infty}(\operatorname{diam} F_n)^r : \langle F_n\rangle_{n\in\mathbb{N}} &\text{ is a sequence of closed subsets of } X \\ &\text{covering } A, \operatorname{diam} F_n \le \delta \text{ for every } n \in \mathbb{N}\} \\ = \inf\{\sum_{n=0}^{\infty}(\operatorname{diam} G_n)^r : \langle G_n\rangle_{n\in\mathbb{N}} &\text{ is a sequence of open subsets of } X \\ &\text{covering } A, \operatorname{diam} G_n \le \delta \text{ for every } n \in \mathbb{N}\}.\end{aligned}$$

>(d) Let $s \ge 1$ be an integer and $r \ge 0$; let μ_{Hr} be r-dimensional Hausdorff measure on $\mathbb{R}^s$. Show that for every $A \subseteq \mathbb{R}^s$ there is a G_δ set (that is, a set expressible as the intersection of a sequence of open sets) $H \supseteq A$ such that $\mu_{Hr}H = \mu^*_{Hr}A$. (*Hint*: use 264Xc.)

>(e) Let $s \ge 1$ be an integer, and $0 \le r < r'$. Show that if $A \subseteq \mathbb{R}^s$ and the r-dimensional Hausdorff outer measure $\mu^*_{Hr}A$ of A is finite, then $\mu^*_{Hr'}A$ must be zero.

(f)(i) Suppose that $f : [a, b] \to \mathbb{R}$ has graph $\Gamma_f \subseteq \mathbb{R}^2$, where $a \le b$ in $\mathbb{R}$. Show that the outer measure $\mu^*_{H1}(\Gamma_f)$ of Γ for one-dimensional Hausdorff measure on $\mathbb{R}^2$ is at most $b - a + \operatorname{Var}_{[a,b]}(f)$. (*Hint*: if f has finite variation, show that $\operatorname{diam}(\Gamma_{f\restriction]t,u[}) \le u - t + \operatorname{Var}_{]t,u[}(f)$; then use 224E.) (ii) Let $f : [0, 1] \to [0, 1]$ be the Cantor function (134H). Show that $\mu_{H1}(\Gamma_f) = 2$. (*Hint*: 264G.)

(g) In 264A, show that

$$\begin{aligned}\theta_{r\delta}A = \inf\{\sum_{n=0}^{\infty}(\operatorname{diam} A_n)^r : \langle A_n\rangle_{n\in\mathbb{N}} &\text{ is a sequence of convex sets covering } A, \\ &\operatorname{diam} A_n \le \delta \text{ for every } n \in \mathbb{N}\}\end{aligned}$$

for any $A \subseteq \mathbb{R}^s$.

264Y Further exercises (a) Let θ_{11} be the outer measure on $\mathbb{R}^2$ defined in 264A, with $r = \delta = 1$, and μ_{11} the measure derived from θ_{11} by Carathéodory's method, Σ_{11} its domain. Show that any set in Σ_{11} is either negligible or conegligible.

(b) Let (X, ρ) be a metric space and $r \ge 0$. Show that if $A \subseteq X$ and $\mu^*_{Hr}A < \infty$, then A is separable.

(c) Let (X, ρ) be a metric space, and θ an outer measure on X such that $\theta(A \cup B) = \theta A + \theta B$ whenever A, B are non-empty subsets of X and $\inf_{x\in A, y\in B} \rho(x, y) > 0$. (Such an outer measure is called a **metric outer measure**.) Show that every open subset of X is measured by the measure defined from θ by Carathéodory's method.

(d) Let (X, ρ) be a metric space and $r > 0$; define $\theta_{r\delta}$ as in 264K. Show that for any $A \subseteq X$,

$$\begin{aligned}\mu^*_{Hr}A = \sup_{\delta>0}\inf\{\sum_{n=0}^{\infty}(\operatorname{diam} F_n)^r : \langle F_n\rangle_{n\in\mathbb{N}} &\text{ is a sequence of closed subsets of } X \\ &\text{covering } A, \operatorname{diam} F_n \le \delta \text{ for every } n \in \mathbb{N}\} \\ = \sup_{\delta>0}\inf\{\sum_{n=0}^{\infty}(\operatorname{diam} G_n)^r : \langle G_n\rangle_{n\in\mathbb{N}} &\text{ is a sequence of open subsets of } X \\ &\text{covering } A, \operatorname{diam} G_n \le \delta \text{ for every } n \in \mathbb{N}\}.\end{aligned}$$

(e) Let (X,ρ) be a metric space and $r\ge 0$; let μ_{Hr} be r-dimensional Hausdorff measure on X. Show that for every $A\subseteq X$ there is a G_δ set $H\supseteq A$ such that $\mu_{Hr}H=\mu^*_{Hr}A$ is the r-dimensional Hausdorff outer measure of A.

(f) Let (X,ρ) be a metric space and $r\ge 0$; let Y be any subset of X, and give Y its induced metric ρ_Y. (i) Show that the r-dimensional Hausdorff outer measure $\mu^{(Y)*}_{Hr}$ on Y is just the restriction to $\mathcal{P}Y$ of the outer measure μ^*_{Hr} on X. (ii) Show that if *either* $\mu^*_{Hr}Y<\infty$ *or* μ_{Hr} measures Y then r-dimensional Hausdorff measure $\mu^{(Y)}_{Hr}$ on Y is just the subspace measure on Y induced by the measure μ_{Hr} on X.

(g) Let (X,ρ) be a metric space and $r>0$. Show that r-dimensional Hausdorff measure on X is atomless. (*Hint*: Let $E\in\operatorname{dom}\mu_{Hr}$. (i) If E is not separable, there is an open set G such that $E\cap G$ and $E\setminus G$ are both non-separable, therefore both non-negligible. (ii) If there is an $x\in E$ such that $\mu_{Hr}(E\cap B(x,\delta))>0$ for every $\delta>0$, then one of these sets has non-negligible complement in E. (iii) Otherwise, $\mu_{Hr}E=0$.)

(h) Let (X,ρ) be a metric space and $r\ge 0$; let μ_{Hr} be r-dimensional Hausdorff measure on X. Show that if $\mu_{Hr}E<\infty$ then $\mu_{Hr}E=\sup\{\mu_{Hr}F: F\subseteq E\text{ is closed and totally bounded}\}$. (*Hint*: given $\epsilon>0$, use 264Yd to find a closed totally bounded set F such that $\mu_{Hr}(F\setminus E)=0$ and $\mu_{Hr}(E\setminus F)\le\epsilon$, and now apply 264Ye to $F\setminus E$.)

(i) Let (X,ρ) be a complete metric space and $r\ge 0$; let μ_{Hr} be r-dimensional Hausdorff measure on X. Show that if $\mu_{Hr}E<\infty$ then $\mu_{Hr}E=\sup\{\mu_{Hr}F: F\subseteq E\text{ is compact}\}$.

(j) Let (X,ρ) and (Y,σ) be metric spaces. If $D\subseteq X$ and $\phi:D\to Y$ is a function, then ϕ is γ-**Lipschitz** if $\sigma(\phi(x),\phi(x'))\le\gamma\rho(x,x')$ for every x, $x'\in D$. (i) Show that in this case, if $r\ge 0$, $\mu^*_{Hr}(\phi[A])\le\gamma^r\mu^*_{Hr}A$ for every $A\subseteq D$, writing μ^*_{Hr} for r-dimensional Hausdorff outer measure on either X or Y. (ii) Show that if X is complete and $\mu_{Hr}E$ is defined and finite, then $\mu_{Hr}(\phi[E])$ is defined. (*Hint*: 264Yi.)

(k) Let (X,ρ) be a metric space, and for $r\ge 0$ let μ_{Hr} be Hausdorff r-dimensional measure on X. Show that there is a unique $\Delta=\Delta(X)\in[0,\infty]$ such that $\mu_{Hr}X=\infty$ if $r\in[0,\Delta[$, 0 if $r\in\,]\Delta,\infty[$.

(l) Let (X,ρ) be a metric space and $\phi:I\to X$ a continuous function, where $I\subseteq\mathbb{R}$ is an interval. Write μ_{H1} for one-dimensional Hausdorff measure on X. Show that

$$\mu_{H1}(\phi[I])\le\sup\{\textstyle\sum_{i=1}^n\rho(\phi(t_i),\phi(t_{i-1})): t_0,\dots,t_n\in I,\ t_0\le\dots\le t_n\},$$

the length of the curve ϕ, with equality if ϕ is injective.

(m) Set $r=\ln 2/\ln 3$, as in 264J, and write μ_{Hr} for r-dimensional Hausdorff measure on the Cantor set C. Let λ be the usual measure on $\{0,1\}^{\mathbb{N}}$ (254J). Define $\phi:\{0,1\}^{\mathbb{N}}\to C$ by setting $\phi(x)=\frac{2}{3}\sum_{n=0}^\infty 3^{-n}x(n)$ for $x\in\{0,1\}^{\mathbb{N}}$. Show that ϕ is an isomorphism between $(\{0,1\}^{\mathbb{N}},\lambda)$ and (C,μ_{Hr}), so that μ_{Hr} is the subspace measure on C induced by 'Cantor measure' as defined in 256Hc.

(n) Set $r=\ln 2/\ln 3$ and write μ_{Hr} for r-dimensional Hausdorff measure on the Cantor set C. Let $f:[0,1]\to[0,1]$ be the Cantor function and let μ be Lebesgue measure on $\mathbb{R}$. Show that $\mu f[E]=\mu_{Hr}E$ for every $E\in\operatorname{dom}\mu_{Hr}$ and $\mu_{Hr}(C\cap f^{-1}[F])=\mu F$ for every Lebesgue measurable set $F\subseteq[0,1]$.

(o) Let (X,ρ) be a metric space and $h:[0,\infty[\,\to[0,\infty[$ a non-decreasing function. For $A\subseteq X$ set

$$\theta_h A=\sup_{\delta>0}\inf\{\sum_{n=0}^\infty h(\operatorname{diam}A_n):\langle A_n\rangle_{n\in\mathbb{N}}\text{ is a sequence of subsets of }X$$
$$\text{covering }A,\ \operatorname{diam}A_n\le\delta\text{ for every }n\in\mathbb{N}\},$$

interpreting $\operatorname{diam}\emptyset$ as 0, $\inf\emptyset$ as ∞ as usual. Show that θ_h is an outer measure on X. State and prove theorems corresponding to 264E and 264F. Look through 264X and 264Y for further results which might be generalizable, perhaps on the assumption that h is continuous on the right.

(p) Let (X,ρ) be a metric space. Let us say that if $a<b$ in $\mathbb{R}$ and $f:[a,b]\to X$ is a function, then f is **absolutely continuous** if for every $\epsilon>0$ there is a $\delta>0$ such that $\sum_{i=1}^{n}\rho(f(a_i),f(b_i))\le\epsilon$ whenever $a\le a_0\le b_0\le\ldots\le a_n\le b_n\le b$ and $\sum_{i=0}^{n}b_i-a_i\le\delta$. Show that $f:[a,b]\to X$ is absolutely continuous iff it is continuous and of bounded variation (in the sense of 224Ye) and $\mu_{H1}f[A]=0$ whenever $A\subseteq[a,b]$ is Lebesgue negligible, where μ_{H1} is 1-dimensional Hausdorff measure on X. (Compare 225M.) Show that in this case $\mu_{H1}f[\,[a,b]\,]<\infty$.

(q) Let $s\ge 1$ be an integer, and $r\in[1,\infty[$. For x, $y\in\mathbb{R}^s$ set $\rho(x,y)=\|x-y\|^{s/r}$. (i) Show that ρ is a metric on $\mathbb{R}^s$ inducing the Euclidean topology. (ii) Let μ_{Hr} be the associated r-dimensional Hausdorff measure. Show that $\mu_{Hr}B(\mathbf{0},1)=2^s$.

264 Notes and comments In this section we have come to the next step in 'geometric measure theory'. I am taking this very slowly, because there are real difficulties in the subject, and for the purposes of this volume we do not need to master very much of it. The idea here is to find a definition of r-dimensional Lebesgue measure which will be 'geometric' in the strict sense, that is, dependent only on the metric structure of $\mathbb{R}^r$, and therefore applicable to sets which have a metric structure but no linear structure. As has happened before, the definition of Hausdorff measure from an outer measure gives no problems – the only new idea in 264A-264C is that of using a supremum $\theta_r=\sup_{\delta>0}\theta_{r\delta}$ of outer measures – and the difficult part is proving that our new measure has any useful properties. Concerning the properties of Hausdorff measure, there are two essential objectives; first, to check that these measures, in general, share a reasonable proportion of the properties of Lebesgue measure; and second, to justify the term 'r-dimensional measure' by relating Hausdorff r-dimensional measure on $\mathbb{R}^r$ to Lebesgue measure on $\mathbb{R}^r$.

As for the properties of general Hausdorff measures, we have to go rather carefully. I do not give counter-examples here because they involve concepts which belong to Volumes 4 and 5 rather than this volume, but I must warn you to expect the worst. However, we do at least have open sets measurable, so that all Borel sets are measurable (264E). The outer measure of a set A can be defined in terms of the Borel sets including A (264Fa), though not in general in terms of the open sets including A; but the measure of a measurable set E is not necessarily the supremum of the measures of the Borel sets included in E, unless E has finite measure (264Fc). We do find that the outer measure θ_r defined in 264A is the outer measure defined from μ_{Hr} (264Fb), so that the phrase 'r-dimensional Hausdorff outer measure' is unambiguous. A crucial property of Lebesgue measure is the fact that the measure of a measurable set E is the supremum of the measures of the compact subsets of E; this is not generally shared by Hausdorff measures, but is valid for sets E of finite measure in complete spaces (264Yi). Concerning subspaces, there are no problems with the outer measures, and for sets of finite measure the subspace measures are also consistent (264Yf). Because Hausdorff measure is defined in metric terms, it behaves regularly for Lipschitz maps (264G); one of the most natural classes of functions to consider when studying metric spaces is that of 1-Lipschitz functions, so that (in the language of 264G) $\mu_{Hr}^*\phi[A]\le\mu_{Hr}^*A$ for every A.

The second essential feature of Hausdorff measure, its relation with Lebesgue measure in the appropriate dimension, is Theorem 264I. Because both Hausdorff measure and Lebesgue measure are translation-invariant, this can be proved by relatively elementary means, except for the evaluation of the normalizing constant; all we need to know is that $\mu\,[0,1[^r=1$ and $\mu_{Hr}\,[0,1[^r$ are both finite and non-zero, and this is straightforward. (The arguments of part (a) of the proof of 261F are relevant.) For the purposes of this chapter, we do not I think have to know the value of the constant; but I cannot leave it unsettled, and therefore give Theorem 264H, the **isodiametric inequality**, to show that it is just the Lebesgue measure of an r-dimensional ball of diameter 1, as one would hope and expect. The critical step in the argument of 264H is in part (c) of the proof. This is called 'Steiner symmetrization'; the idea is that given a set A, we transform A through a series of steps, at each stage lowering, or at least not increasing, its diameter, and raising, or at least not decreasing, its outer measure, progressively making A more symmetric, until at the end we have a set which is sufficiently constrained to be amenable. The particular symmetrization operation used in this proof is important enough; but the idea of progressive regularization of an object is one of the most powerful methods in measure theory, and you should give all your attention to mastering any example you encounter. In my experience, the idea is principally useful when seeking an inequality involving disparate quantities – in the present example, the diameter and volume of a set.

Of course it is awkward having two measures on $\mathbb{R}^r$, differing by a constant multiple, and for the purposes of the next section it would actually have been a little more convenient to follow Federer 69 in using 'normalized Hausdorff measure' $2^{-r}\beta_r\mu_{Hr}$. (For non-integral r, we could take $\beta_r = \pi^{r/2}/\Gamma(1+\frac{r}{2})$, as suggested in 252Xi.) However, I believe this to be a minority position, and the striking example of Hausdorff measure on the Cantor set (264J, 264Ym-264Yn) looks much better in the non-normalized version.

Hausdorff $(\ln 2/\ln 3)$-dimensional measure on the Cantor set is of course but one, perhaps the easiest, of a large class of examples. Because the Hausdorff r-dimensional outer measure of a set A, regarded as a function of r, behaves dramatically (falling from ∞ to 0) at a certain critical value $\Delta(A)$ (see 264Xe, 264Yk), it gives us a metric space invariant of A; $\Delta(A)$ is the **Hausdorff dimension** of A. Evidently the Hausdorff dimension of C is $\ln 2/\ln 3$, while that of r-dimensional Euclidean space is r.

Version of 3.9.13

265 Surface measures

In this section I offer a new version of the arguments of §263, this time not with the intention of justifying integration-by-substitution, but instead to give a practically effective method of computing the Hausdorff r-dimensional measure of a smooth r-dimensional surface in an s-dimensional space. The basic case to bear in mind is $r=2$, $s=3$, though any other combination which you can easily visualize will also be a valuable aid to intuition. I give a fundamental theorem (265E) providing a formula from which we can hope to calculate the r-dimensional measure of a surface in s-dimensional space which is parametrized by a differentiable function, and work through some of the calculations in the case of the r-sphere (265F-265H).

265A Normalized Hausdorff measure As I remarked at the end of the last section, Hausdorff measure, as defined in 264A-264C, is not quite the most appropriate measure for our work here; so in this section I will use **normalized Hausdorff measure**, meaning $\nu_r = 2^{-r}\beta_r\mu_{Hr}$, where μ_{Hr} is r-dimensional Hausdorff measure (interpreted in whichever space is under consideration) and $\beta_r = \mu_r B(\mathbf{0},1)$ is the Lebesgue measure of any ball of radius 1 in $\mathbb{R}^r$. It will be convenient to take $\beta_0 = 1$. As shown in 264H-264I, this normalization makes ν_r on $\mathbb{R}^r$ agree with Lebesgue measure μ_r. Observe that of course $\nu_r^* = 2^{-r}\beta_r\mu_{Hr}^*$ (264Fb).

265B Linear subspaces Just as in §263, the first step is to deal with linear operators.

Theorem Suppose that r, s are integers with $1 \le r \le s$, and that T is a real $s\times r$ matrix; regard T as a linear operator from $\mathbb{R}^r$ to $\mathbb{R}^s$. Set $J = \sqrt{\det T^\top T}$, where $T^\top$ is the transpose of T. Write ν_r for normalized r-dimensional Hausdorff measure on $\mathbb{R}^s$, T_r for its domain, and μ_r for Lebesgue measure on $\mathbb{R}^r$. Then

$$\nu_r T[E] = J\mu_r E$$

for every measurable set $E \subseteq \mathbb{R}^r$. If T is injective (that is, if $J \ne 0$), then

$$\nu_r F = J\mu_r T^{-1}[F]$$

whenever $F \in \mathrm{T}_r$ and $F \subseteq T[\mathbb{R}^r]$.

proof The formula for J assumes that $\det T^\top T$ is non-negative, which is a fact not in evidence; but the argument below will establish it adequately soon.

(a) Let V be the linear subspace of $\mathbb{R}^s$ consisting of vectors $y = (\eta_1,\dots,\eta_s)$ such that $\eta_i = 0$ whenever $r < i \le s$. Let R be the $r\times s$ matrix $\langle\rho_{ij}\rangle_{i\le r,j\le s}$, where $\rho_{ij} = 1$ if $i = j \le r$, 0 otherwise; then the $s\times r$ matrix $R^\top$ may be regarded as a bijection from $\mathbb{R}^r$ to V. Let W be an r-dimensional linear subspace of $\mathbb{R}^s$ including $T[\mathbb{R}^r]$, and let P be an orthogonal $s\times s$ matrix such that $P[W] = V$. Then $S = RPT$ is an $r\times r$ matrix. We have $R^\top Ry = y$ for $y \in V$, so $R^\top RPT = PT$ and

$$S^\top S = T^\top P^\top R^\top RPT = T^\top P^\top PT = T^\top T;$$

accordingly

$$\det T^\top T = \det S^\top S = (\det S)^2 \ge 0$$

and $J = |\det S|$. At the same time,

$$P^\top R^\top S = P^\top R^\top RPT = P^\top PT = T.$$

Observe that $J = 0$ iff S is not injective, that is, T is not injective.

(b) If we consider the $s \times r$ matrix $P^\top R^\top$ as a map from $\mathbb{R}^r$ to $\mathbb{R}^s$, we see that $\phi = P^\top R^\top$ is an isometry between $\mathbb{R}^r$ and W, with inverse $\phi^{-1} = RP{\restriction}W$. It follows that ϕ is an isomorphism between the measure spaces $(\mathbb{R}^r, \mu_{Hr}^{(r)})$ and $(W, \mu_{HrW}^{(s)})$, where $\mu_{Hr}^{(r)}$ is r-dimensional Hausdorff measure on $\mathbb{R}^r$ and $\mu_{HrW}^{(s)}$ is the subspace measure on W induced by r-dimensional Hausdorff measure $\mu_{Hr}^{(s)}$ on $\mathbb{R}^s$.

P **(i)** If $A \subseteq \mathbb{R}^r$ and $A' \subseteq W$,

$$\mu_{Hr}^{(s)*}(\phi[A]) \le \mu_{Hr}^{(r)*}(A), \quad \mu_{Hr}^{(r)*}(\phi^{-1}[A']) \le \mu_{Hr}^{(s)*}(A'),$$

using 264G twice. Thus $\mu_{Hr}^{(s)*}(\phi[A]) = \mu_{Hr}^{(r)*}(A)$ for every $A \subseteq \mathbb{R}^r$.

(ii) Now because W is closed, therefore in the domain of $\mu_{Hr}^{(s)}$ (264E), the subspace measure $\mu_{HrW}^{(s)}$ is just the measure induced by $\mu_{Hr}^{(s)*}{\restriction}W$ by Carathéodory's method (214H(b-ii)). Because ϕ is an isomorphism between $(\mathbb{R}^r, \mu_{Hr}^{(r)*})$ and $(W, \mu_{Hr}^{(s)*}{\restriction}W)$, it is an isomorphism between $(\mathbb{R}^r, \mu_{Hr}^{(r)})$ and $(W, \mu_{HrW}^{(s)})$. **Q**

(c) It follows that ϕ is also an isomorphism between the normalized versions $(\mathbb{R}^r, \mu_r)$ and (W, ν_{rW}), writing ν_{rW} for the subspace measure on W induced by ν_r.

Now if $E \subseteq \mathbb{R}^r$ is Lebesgue measurable, we have $\mu_r S[E] = J\mu_r E$, by 263A; so that

$$\nu_r T[E] = \nu_r(P^\top R^\top[S[E]]) = \nu_r(\phi[S[E]]) = \mu_r S[E] = J\mu_r E.$$

If T is injective, then $S = \phi^{-1}T$ must also be injective, so that $J \ne 0$ and

$$\nu_r F = \mu_r(\phi^{-1}[F]) = J\mu_r(S^{-1}[\phi^{-1}[F]]) = J\mu_r T^{-1}[F]$$

whenever $F \in \mathrm{T}_r$ and $F \subseteq W = T[\mathbb{R}^r]$.

265C Corollary Under the conditions of 265B,

$$\nu_r^* T[A] = J\mu_r^* A$$

for every $A \subseteq \mathbb{R}^r$.

proof (a) If E is Lebesgue measurable and $A \subseteq E$, then $T[A] \subseteq T[E]$, so

$$\nu_r^* T[A] \le \nu_r T[E] = J\mu_r E;$$

as E is arbitrary, $\nu_r^* T[A] \le J\mu_r^* A$.

(b) If $J = 0$ we can stop. If $J \ne 0$ then T is injective, so if $F \in \mathrm{T}_r$ and $T[A] \subseteq F$ we shall have

$$J\mu_r^* A \le J\mu_r T^{-1}[F \cap T[\mathbb{R}^r]] = \nu_r(F \cap T[\mathbb{R}^r]) \le \nu_r F;$$

as F is arbitrary, $J\mu_r^* A \le \nu_r^* T[A]$.

265D I now proceed to the lemma corresponding to 263C.

Lemma Suppose that $1 \le r \le s$ and that T is an $s \times r$ matrix; set $J = \sqrt{\det T^\top T}$, and suppose that $J \ne 0$. Then for any $\epsilon > 0$ there is a $\zeta = \zeta(T, \epsilon) > 0$ such that

(i) $|\sqrt{\det S^\top S} - J| \le \epsilon$ whenever S is an $s \times r$ matrix and $\|S - T\| \le \zeta$, defining the norm of a matrix as in 262H;

(ii) whenever $D \subseteq \mathbb{R}^r$ is a bounded set and $\phi : D \to \mathbb{R}^s$ is a function such that $\|\phi(x) - \phi(y) - T(x - y)\| \le \zeta\|x - y\|$ for all x, $y \in D$, then $|\nu_r^*\phi[D] - J\mu_r^* D| \le \epsilon\mu_r^* D$.

proof (a) Because $\det S^\top S$ is a continuous function of the coefficients of S, 262Hb tells us that there must be a $\zeta_0 > 0$ such that $|J - \sqrt{\det S^\top S}| \le \epsilon$ whenever $\|S - T\| \le \zeta_0$.

(b) Because $J \ne 0$, T is injective, and there is an $r \times s$ matrix T^* such that T^*T is the identity $r \times r$ matrix. Take $\zeta > 0$ such that $\zeta \le \zeta_0$, $\zeta\|T^*\| < 1$, $J(1 + \zeta\|T^*\|)^r \le J + \epsilon$ and $1 - J^{-1}\epsilon \le (1 - \zeta\|T^*\|)^r$.

Let $\phi : D \to \mathbb{R}^s$ be such that $\|\phi(x) - \phi(y) - T(x-y)\| \le \zeta\|x-y\|$ whenever $x, y \in D$. Set $\psi = \phi T^*$, so that $\phi = \psi T$. Then

$$\|\psi(u) - \psi(v)\| \le (1 + \zeta\|T^*\|)\|u - v\|, \quad \|u - v\| \le (1 - \zeta\|T^*\|)^{-1}\|\psi(u) - \psi(v)\|$$

whenever $u, v \in T[D]$. **P** Take $x, y \in D$ such that $u = Tx$, $v = Ty$; of course $x = T^*u$, $y = T^*v$. Then

$$\begin{aligned}\|\psi(u) - \psi(v)\| &= \|\phi(T^*u) - \phi(T^*v)\| = \|\phi(x) - \phi(y)\| \\ &\le \|T(x-y)\| + \zeta\|x-y\| \\ &= \|u - v\| + \zeta\|T^*u - T^*v\| \le \|u - v\|(1 + \zeta\|T^*\|).\end{aligned}$$

Next,

$$\begin{aligned}\|u - v\| = \|Tx - Ty\| &\le \|\phi(x) - \phi(y)\| + \zeta\|x - y\| \\ &= \|\psi(u) - \psi(v)\| + \zeta\|T^*u - T^*v\| \\ &\le \|\psi(u) - \psi(v)\| + \zeta\|T^*\|\|u - v\|,\end{aligned}$$

so that $(1 - \zeta\|T^*\|)\|u - v\| \le \|\psi(u) - \psi(v)\|$ and $\|u - v\| \le (1 - \zeta\|T^*\|)^{-1}\|\psi(u) - \psi(v)\|$. **Q**

(c) Now from 264G and 265C we see that

$$\nu_r^*\phi[D] = \nu_r^*\psi[T[D]] \le (1 + \zeta\|T^*\|)^r\nu_r^*T[D] = (1 + \zeta\|T^*\|)^r J\mu_r^*D \le (J + \epsilon)\mu_r^*D,$$

and (provided $\epsilon \le J$)

$$(J - \epsilon)\mu_r^*D = (1 - J^{-1}\epsilon)\nu_r^*T[D] \le (1 - J^{-1}\epsilon)(1 - \zeta\|T^*\|)^{-r}\nu_r^*\psi[T[D]]$$

(applying 264G to $\psi^{-1} : \psi[T[D]] \to T[D]$)

$$\le \nu_r^*\psi[T[D]] = \nu_r^*\phi[D].$$

(Of course, if $\epsilon \ge J$, then surely $(J - \epsilon)\mu_r^*D \le \nu_r^*\phi[D]$.) Thus

$$(J - \epsilon)\mu_r^*D \le \nu_r^*\phi[D] \le (J + \epsilon)\mu_r^*D$$

as required, and we have an appropriate ζ.

265E Theorem Suppose that $1 \le r \le s$; write μ_r for Lebesgue measure on $\mathbb{R}^r$, ν_r for normalized Hausdorff measure on $\mathbb{R}^s$, and T_r for the domain of ν_r. Let $D \subseteq \mathbb{R}^r$ be any set, and $\phi : D \to \mathbb{R}^s$ a function differentiable relative to its domain at each point of D. For each $x \in D$ let $T(x)$ be a derivative of ϕ at x relative to D, and set $J(x) = \sqrt{\det T(x)^\top T(x)}$. Set $D' = \{x : x \in D, J(x) > 0\}$. Then

(i) $J : D \to [0, \infty[$ is a measurable function;

(ii) $\nu_r^*\phi[D] \le \int_D J(x)\mu_r(dx)$,

allowing ∞ as the value of the integral;

(iii) $\nu_r^*\phi[D \setminus D'] = 0$.

If D is Lebesgue measurable, then

(iv) $\phi[D] \in \mathrm{T}_r$.

If D is measurable and ϕ is injective, then

(v) $\nu_r\phi[D] = \int_D J\,d\mu_r$;

(vi) for any set $E \subseteq \phi[D]$, $E \in \mathrm{T}_r$ iff $\phi^{-1}[E] \cap D'$ is Lebesgue measurable, and in this case

$$\nu_r E = \int_{\phi^{-1}[E]} J(x)\mu_r(dx) = \int_D J \times \chi(\phi^{-1}[E])d\mu_r;$$

(vii) for every real-valued function g defined on a subset of $\phi[D]$,

$$\int_{\phi[D]} g\,d\nu_r = \int_D J \times g\phi\,d\mu_r$$

if either integral is defined in $[-\infty, \infty]$, provided we interpret $J(x)g(\phi(x))$ as zero when $J(x) = 0$ and $g(\phi(x))$ is undefined.

proof I seek to follow the line laid out in the proof of 263D.

(a) Just as in 263D, we know that $J : D \to \mathbb{R}$ is measurable, since $J(x)$ is a continuous function of the coefficients of $T(x)$, all of which are measurable, by 262P. If D is Lebesgue measurable, then there is a sequence $\langle F_n \rangle_{n\in\mathbb{N}}$ of compact subsets of D such that $D \setminus \bigcup_{n\in\mathbb{N}} F_n$ is μ_r-negligible. Now $\phi[F_n]$ is compact, therefore belongs to T_r, for each $n \in \mathbb{N}$. As for $\phi[D \setminus \bigcup_{n\in\mathbb{N}} F_n]$, this must be ν_r-negligible by 264G, because ϕ is a countable union of Lipschitz functions (262N). So

$$\phi[D] = \bigcup_{n\in\mathbb{N}} \phi[F_n] \cup \phi[D \setminus \bigcup_{n\in\mathbb{N}} F_n] \in \mathrm{T}_r.$$

This deals with (i) and (iv).

(b) For the moment, assume that D is bounded and that $J(x) > 0$ for every $x \in D$, and fix $\epsilon > 0$. Let M^*_{sr} be the set of $s \times r$ matrices T such that $\det T^\top T \neq 0$, that is, the corresponding map $T : \mathbb{R}^r \to \mathbb{R}^s$ is injective. For $T \in M^*_{sr}$ take $\zeta(T, \epsilon) > 0$ as in 265D.

Take $\langle D_n \rangle_{n\in\mathbb{N}}$, $\langle T_n \rangle_{n\in\mathbb{N}}$ as in 262M, with $A = M^*_{sr}$, so that $\langle D_n \rangle_{n\in\mathbb{N}}$ is a partition of D into sets which are relatively measurable in D, and each T_n is an $s \times r$ matrix such that

$$\|T(x) - T_n\| \le \zeta(T_n, \epsilon) \text{ whenever } x \in D_n,$$

$$\|\phi(x) - \phi(y) - T_n(x - y)\| \le \zeta(T_n, \epsilon)\|x - y\| \text{ for all } x,\, y \in D_n.$$

Then, setting $J_n = \sqrt{\det T_n^\top T_n}$, we have

$$|J(x) - J_n| \le \epsilon \text{ for every } x \in D_n,$$

$$|\nu^*_r \phi[D_n] - J_n \mu^*_r D_n| \le \epsilon \mu^*_r D_n,$$

by the choice of $\zeta(T_n, \epsilon)$. So

$$\nu^*_r \phi[D] \le \sum_{n=0}^{\infty} \nu^*_r \phi[D_n]$$

(because $\phi[D] = \bigcup_{n\in\mathbb{N}} \phi[D_n]$)

$$\le \sum_{n=0}^{\infty} J_n \mu^*_r D_n + \epsilon \mu^*_r D_n \le \epsilon \mu^*_r D + \sum_{n=0}^{\infty} J_n \mu^*_r D_n$$

(because the D_n are disjoint and relatively measurable in D)

$$= \epsilon \mu^*_r D + \int_D \sum_{n=0}^{\infty} J_n \chi D_n d\mu$$
$$\le \epsilon \mu^*_r D + \int_D J(x) + \epsilon \mu_r(dx) = 2\epsilon \mu^*_r D + \int_D J \, d\mu_r.$$

If D is measurable and ϕ is injective, then all the D_n are Lebesgue measurable subsets of $\mathbb{R}^r$, so all the $\phi[D_n]$ are measured by ν_r, and they are also disjoint. Accordingly

$$\int_D J \, d\mu \le \sum_{n=0}^{\infty} J_n \mu_r D_n + \epsilon \mu_r D$$
$$\le \sum_{n=0}^{\infty} (\nu_r \phi[D_n] + \epsilon \mu_r D_n) + \epsilon \mu_r D = \nu_r \phi[D] + 2\epsilon \mu_r D.$$

Since ϵ is arbitrary, we get

$$\nu^*_r \phi[D] \le \int_D J \, d\mu_r,$$

and if D is measurable and ϕ is injective,

$$\int_D J \, d\mu_r \le \nu_r \phi[D];$$

thus we have (ii) and (v), on the assumption that D is bounded and $J > 0$ everywhere on D.

(c) Just as in the proof of 263D, we can now relax the assumption that D is bounded by considering $B_k = B(\mathbf{0}, k) \subseteq \mathbb{R}^r$; provided $J > 0$ everywhere on D, we get

$$\nu_r^*\phi[D] = \lim_{k\to\infty} \nu_r^*\phi[D \cap B_k] \le \lim_{k\to\infty} \int_{D\cap B_k} J\,d\mu_r = \int_D J\,d\mu_r,$$

with equality if D is measurable and ϕ is injective.

(d) Now we find that $\nu_r^*\phi[D \setminus D'] = 0$.

P$(\boldsymbol{\alpha})$ Let $\eta \in \,]0,1]$. Define $\psi_\eta : D \to \mathbb{R}^{s+r}$ by setting $\psi_\eta(x) = (\phi(x), \eta x)$, identifying $\mathbb{R}^{s+r}$ with $\mathbb{R}^s \times \mathbb{R}^r$. ψ_η is differentiable relative to its domain at each point of D, with derivative $\tilde{T}_\eta(x)$, being the $(s+r) \times r$ matrix in which the top s rows consist of the $s \times r$ matrix $T(x)$, and the bottom r rows are ηI_r, writing I_r for the $r \times r$ identity matrix. (Use 262Ib.) Now of course $\tilde{T}_\eta(x)$, regarded as a map from $\mathbb{R}^r$ to $\mathbb{R}^{s+r}$, is injective, so

$$\tilde{J}_\eta(x) = \sqrt{\det \tilde{T}_\eta(x)^\top \tilde{T}_\eta(x)} = \sqrt{\det(T(x)^\top T(x) + \eta^2 I)} > 0.$$

We have $\lim_{\eta\downarrow 0} \tilde{J}_\eta(x) = J(x) = 0$ for $x \in D \setminus D'$.

$(\boldsymbol{\beta})$ Express $T(x)$ as $\langle \tau_{ij}(x)\rangle_{i\le s, j\le r}$ for each $x \in D$. Set

$$C_m = \{x : x \in D,\ \|x\| \le m,\ |\tau_{ij}(x)| \le m \text{ for all } i \le s,\ j \le r\}$$

for each $m \ge 1$. For $x \in C_m$, all the coefficients of $\tilde{T}_\eta(x)$ have moduli at most m; consequently (giving the crudest and most immediately available inequalities) all the coefficients of $\tilde{T}_\eta(x)^\top \tilde{T}_\eta(x)$ have moduli at most $(r+s)m^2$ and $\tilde{J}_\eta(x) \le \sqrt{r!(s+r)^r}m^r$. Consequently we can use Lebesgue's Dominated Convergence Theorem to see that

$$\lim_{\eta\downarrow 0} \int_{C_m\setminus D'} \tilde{J}_\eta d\mu_r = 0.$$

$(\boldsymbol{\gamma})$ Let $\tilde{\nu}_r$ be normalized Hausdorff r-dimensional measure on $\mathbb{R}^{s+r}$. Applying (b) of this proof to $\psi_\eta \restriction C_m \setminus D'$, we see that

$$\tilde{\nu}_r^*\psi_\eta[C_m \setminus D'] \le \int_{C_m\setminus D'} \tilde{J}_\eta d\mu_r.$$

Now we have a natural map $P : \mathbb{R}^{s+r} \to \mathbb{R}^s$ given by setting $P(\xi_1, \dots, \xi_{s+r}) = (\xi_1, \dots, \xi_s)$, and P is 1-Lipschitz, so by 264G once more we have (allowing for the normalizing constants $2^{-r}\beta_r$)

$$\nu_r^* P[A] \le \tilde{\nu}_r^* A$$

for every $A \subseteq \mathbb{R}^{s+r}$. In particular,

$$\nu_r^*\phi[C_m \setminus D'] = \nu_r^* P[\psi_\eta[C_m \setminus D']] \le \tilde{\nu}_r^*\psi_\eta[C_m \setminus D'] \le \int_{C_m\setminus D'} \tilde{J}_\eta d\mu_r \to 0$$

as $\eta \downarrow 0$. But this means that $\nu_r^*\phi[C_m \setminus D'] = 0$. As $D = \bigcup_{m\ge 1} C_m$, $\nu_r^*\phi[D \setminus D'] = 0$, as claimed. **Q**

(e) This proves (iii) of the theorem. But of course this is enough to give (ii) and (v), because (applying (b)-(c) to $\phi\restriction D'$) we must have

$$\nu_r^*\phi[D] = \nu_r^*\phi[D'] \le \int_{D'} J\,d\mu_r = \int_D J\,d\mu_r,$$

with equality if D (and therefore also D') is measurable and ϕ is injective.

(f) So let us turn to part (vi). Assume that D is measurable and that ϕ is injective.

$(\boldsymbol{\alpha})$ Suppose that $E \subseteq \phi[D]$ belongs to T_r. Let

$$H_k = \{x : x \in D,\ \|x\| \le k,\ J(x) \le k\}$$

for each k; then each H_k is Lebesgue measurable, so (applying (iii) to $\phi\restriction H_k$) $\phi[H_k] \in \mathrm{T}_r$, and

$$\nu_r\phi[H_k] \le k\mu_r H_k < \infty.$$

Thus $\phi[D]$ can be covered by a sequence of sets of finite measure for ν_r, which of course are of finite measure for r-dimensional Hausdorff measure on $\mathbb{R}^s$. By 264Fc, there are Borel sets E_1, $E_2 \subseteq \mathbb{R}^s$ such that

$E_1 \subseteq E \subseteq E_2$ and $\nu_r(E_2 \setminus E_1) = 0$. Now $F_1 = \phi^{-1}[E_1]$, $F_2 = \phi^{-1}[E_2]$ are Lebesgue measurable subsets of D, and

$$\int_{F_2 \setminus F_1} J \, d\mu_r = \nu_r \phi[F_2 \setminus F_1] = \nu_r(\phi[D] \cap E_2 \setminus E_1) = 0.$$

Accordingly $\mu_r(D' \cap (F_2 \setminus F_1)) = 0$. But as

$$D' \cap F_1 \subseteq D' \cap \phi^{-1}[E] \subseteq D' \cap F_2,$$

it follows that $D' \cap \phi^{-1}[E]$ is measurable, and that

$$\begin{aligned}\int_{\phi^{-1}[E]} J \, d\mu_r &= \int_{D' \cap \phi^{-1}[E]} J \, d\mu_r = \int_{D' \cap F_1} J \, d\mu_r \\ &= \int_{D \cap F_1} J \, d\mu_r = \nu_r \phi[D \cap F_1] = \nu_r E_1 = \nu_r E.\end{aligned}$$

Moreover, $J \times \chi(\phi^{-1}[E]) = J \times \chi(D' \cap \phi^{-1}[E])$ is measurable, so we can write $\int J \times \chi(\phi^{-1}[E])$ in place of $\int_{\phi^{-1}[E]} J$.

(β) If $E \subseteq \phi[D]$ and $D' \cap \phi^{-1}[E]$ is measurable, then of course

$$E = \phi[D' \cap \phi^{-1}[E]] \cup \phi[(D \setminus D') \cap \phi^{-1}[E]] \in \mathrm{T}_r,$$

because $\phi[G] \in \mathrm{T}_r$ for every measurable $G \subseteq D$ and $\phi[D \setminus D']$ is ν_r-negligible.

(g) Finally, (vii) follows at once from (vi), applying 235J to μ_r and the subspace measure induced by ν_r on $\phi[D]$.

265F The surface of a sphere To show how these ideas can be applied to one of the basic cases, I give the details of a method of describing spherical surface measure in s-dimensional space. Take $r \geq 1$ and $s = r+1$. Write S_r for $\{z : z \in \mathbb{R}^{r+1}, \|z\| = 1\}$, the r-sphere. Then we have a parametrization ϕ_r of S_r given by setting

$$\phi_r \begin{pmatrix} \xi_1 \\ \xi_2 \\ \dots \\ \dots \\ \dots \\ \xi_r \end{pmatrix} = \begin{pmatrix} \sin\xi_1 \sin\xi_2 \sin\xi_3 \dots \sin\xi_r \\ \cos\xi_1 \sin\xi_2 \sin\xi_3 \dots \sin\xi_r \\ \cos\xi_2 \sin\xi_3 \dots \sin\xi_r \\ \dots \\ \cos\xi_{r-2} \sin\xi_{r-1} \sin\xi_r \\ \cos\xi_{r-1} \sin\xi_r \\ \cos\xi_r \end{pmatrix}.$$

I choose this formulation because I wish to use an inductive argument based on the fact that

$$\phi_{r+1} \begin{pmatrix} x \\ \xi \end{pmatrix} = \begin{pmatrix} \sin\xi \, \phi_r(x) \\ \cos\xi \end{pmatrix}$$

for $x \in \mathbb{R}^r$, $\xi \in \mathbb{R}$. Every ϕ_r is differentiable, by 262Id. If we set

$$\begin{aligned} D_r = \{x : \xi_1 \in \left]-\pi, \pi\right], \xi_2, \dots, \xi_r \in [0, \pi], \\ \text{if } \xi_j \in \{0, \pi\} \text{ then } \xi_i = 0 \text{ for } i < j\},\end{aligned}$$

then it is easy to check that D_r is a Borel subset of $\mathbb{R}^r$ and that $\phi_r \restriction D_r$ is a bijection between D_r and S_r. Now let $T_r(x)$ be the $(r+1) \times r$ matrix $\phi_r'(x)$. Then

$$T_{r+1} \begin{pmatrix} x \\ \xi \end{pmatrix} = \begin{pmatrix} \sin\xi \, T_r(x) & \cos\xi \, \phi_r(x) \\ \mathbf{0} & -\sin\xi \end{pmatrix}.$$

So

$$(T_{r+1} \begin{pmatrix} x \\ \xi \end{pmatrix})^\top T_{r+1} \begin{pmatrix} x \\ \xi \end{pmatrix} = \begin{pmatrix} \sin^2\xi \, T_r(x)^\top T_r(x) & \sin\xi \cos\xi \, T_r(x)^\top \phi_r(x) \\ \cos\xi \sin\xi \, \phi_r(x)^\top T_r(x) & \cos^2\xi \phi_r(x)^\top \phi_r(x) + \sin^2\xi \end{pmatrix}.$$

But of course $\phi_r(x)^\top\phi_r(x) = \|\phi_r(x)\|^2 = 1$ for every x, and (differentiating with respect to each coordinate of x, if you wish) $T_r(x)^\top\phi_r(x) = \mathbf{0}$, $\phi_r(x)^\top T_r(x) = \mathbf{0}$. So we get

$$(T_{r+1}\begin{pmatrix} x \\ \xi \end{pmatrix})^\top T_{r+1}\begin{pmatrix} x \\ \xi \end{pmatrix} = \begin{pmatrix} \sin^2\xi\, T_r(x)^\top T_r(x) & \mathbf{0} \\ \mathbf{0} & 1 \end{pmatrix},$$

and writing $J_r(x) = \sqrt{\det T_r(x)^\top T_r(x)}$,

$$J_{r+1}\begin{pmatrix} x \\ \xi \end{pmatrix} = |\sin^r\xi| J_r(x).$$

At this point we induce on r to see that

$$J_r(x) = |\sin^{r-1}\xi_r \sin^{r-2}\xi_{r-1}\dots\sin\xi_2|$$

(since of course the induction starts with the case $r = 1$,

$$\phi_1(x) = \begin{pmatrix} \sin x \\ \cos x \end{pmatrix}, \quad T_1(x) = \begin{pmatrix} \cos x \\ -\sin x \end{pmatrix}, \quad T_1(x)^\top T_1(x) = 1, \quad J_1(x) = 1).$$

To find the surface measure of S_r, we need to calculate

$$\begin{aligned}\int_{D_r} J_r d\mu_r &= \int_0^\pi \dots \int_0^\pi \int_{-\pi}^\pi \sin^{r-1}\xi_r \dots \sin\xi_2 d\xi_1 d\xi_2 \dots d\xi_r \\ &= 2\pi \prod_{k=2}^r \int_0^\pi \sin^{k-1} t\,dt = 2\pi \prod_{k=1}^{r-1} \int_{-\pi/2}^{\pi/2} \cos^k t\,dt\end{aligned}$$

(substituting $\frac{\pi}{2} - t$ for t). But in the language of 252Q, this is just

$$2\pi \textstyle\prod_{k=1}^{r-1} I_k = 2\pi\beta_{r-1},$$

where β_{r-1} is the volume of the unit ball of $\mathbb{R}^{r-1}$ (interpreting β_0 as 1, if you like).

265G The surface area of a sphere can also be calculated through the following result.

Theorem Let μ_{r+1} be Lebesgue measure on $\mathbb{R}^{r+1}$, and ν_r normalized r-dimensional Hausdorff measure on $\mathbb{R}^{r+1}$. If f is a locally μ_{r+1}-integrable real-valued function, $y \in \mathbb{R}^{r+1}$ and $\delta > 0$,

$$\textstyle\int_{B(y,\delta)} f d\mu_{r+1} = \int_0^\delta \int_{\partial B(y,t)} f d\nu_r dt,$$

where I write $\partial B(y,t)$ for the sphere $\{x : \|x - y\| = t\}$ and the integral $\int \dots dt$ is to be taken with respect to Lebesgue measure on $\mathbb{R}$.

proof Take any differentiable function $\phi : \mathbb{R}^r \to S_r$ with a Borel set $F \subseteq \mathbb{R}^r$ such that $\phi{\restriction}F$ is a bijection between F and S_r; such a pair (ϕ, F) is described in 265F. Define $\psi : \mathbb{R}^r \times \mathbb{R} \to \mathbb{R}^{r+1}$ by setting $\psi(z,t) = y + t\phi(z)$; then ψ is differentiable and $\psi{\restriction}F \times]0,\delta]$ is a bijection between $F \times]0,\delta]$ and $B(y,\delta) \setminus \{y\}$. For $t \in]0,\delta]$, $z \in \mathbb{R}^r$ set $\psi_t(z) = \psi(z,t)$; then $\psi_t{\restriction}F$ is a bijection between F and the sphere $\{x : \|x - y\| = t\} = \partial B(y,t)$.

The derivative of ϕ at z is an $(r+1) \times r$ matrix $T_1(z)$ say, and the derivative $T_t(z)$ of ψ_t at z is just $tT_1(z)$; also the derivative of ψ at (z,t) is the the $(r+1) \times (r+1)$ matrix $T(z,t) = (\,tT_1(z) \quad \phi(z)\,)$, where $\phi(z)$ is interpreted as a column vector. If we set

$$J_t(z) = \sqrt{\det T_t(z)^\top T_t(z)}, \quad J(z,t) = |\det T(z,t)|,$$

then

$$\begin{aligned}J(z,t)^2 = \det T(z,t)^\top T(z,t) &= \det \begin{pmatrix} tT_1(z)^\top \\ \phi(z)^\top \end{pmatrix} (\,tT_1(z) \quad \phi(z)\,) \\ &= \det \begin{pmatrix} t^2 T_1(z)^\top T_1(z) & \mathbf{0} \\ \mathbf{0} & 1 \end{pmatrix} = J_t(z)^2,\end{aligned}$$

because when we come to calculate the $(i, r+1)$-coefficient of $T(z,t)^\top T(z,t)$, for $1 \le i \le r$, it is

$$\sum_{j=1}^{r+1} t\frac{\partial\phi_j}{\partial\zeta_i}(z)\phi_j(z) = \frac{t}{2}\frac{\partial}{\partial\zeta_i}(\sum_{j=1}^{r+1}\phi_j(z)^2) = 0,$$

where ϕ_j is the jth coordinate of ϕ; while the $(r+1,r+1)$-coefficient of $T(z,t)^\top T(z,t)$ is just $\sum_{j=1}^{r+1}\phi_j(z)^2 = 1$. So in fact $J(z,t) = J_t(z)$ for all $z \in \mathbb{R}^r$, $t > 0$.

Now, given $f \in \mathcal{L}^1(\mu_{r+1})$, we can calculate

$$\int_{B(y,\delta)} f d\mu_{r+1} = \int_{B(y,\delta)\setminus\{y\}} f d\mu_{r+1}$$
$$= \int_{F\times]0,\delta]} f(\psi(z,t))J(z,t)\mu_{r+1}(d(z,t))$$

(by 263D)

$$= \int_0^\delta \int_F f(\psi_t(z))J_t(z)\mu_r(dz)dt$$

(where μ_r is Lebesgue measure on $\mathbb{R}^r$, by Fubini's theorem, 252B)

$$= \int_0^\delta \int_{\partial B(y,t)} f d\nu_r dt$$

by 265E(vii).

265H Corollary If ν_r is normalized r-dimensional Hausdorff measure on $\mathbb{R}^{r+1}$, then $\nu_r S_r = (r+1)\beta_{r+1}$.

proof In 265G, take $y = \mathbf{0}$, $\delta = 1$, and $f = \chi B(\mathbf{0}, 1)$; then

$$\beta_{r+1} = \int f d\mu_{r+1} = \int_0^1 \nu_r(\partial B(\mathbf{0},t))dt = \int_0^1 t^r \nu_r S_r dt = \frac{1}{r+1}\nu_r S_r,$$

this time applying 264G to the maps $x \mapsto tx$, $x \mapsto \frac{1}{t}x$ from $\mathbb{R}^{r+1}$ to itself to see that $\nu_r(\partial B(\mathbf{0},t)) = t^r \nu_r S_r$ for $t > 0$.

265X Basic exercises (a) Let $r \ge 1$, and let $S_r(\alpha) = \{z : z \in \mathbb{R}^{r+1}, \|z\| = \alpha\}$ be the r-sphere of radius α. Show that $\nu_r S_r(\alpha) = 2\pi\beta_{r-1}\alpha^r = (r+1)\beta_{r+1}\alpha^r$ for every $\alpha \ge 0$.

>(b) Let $r \ge 1$, and for $a \in [-1, 1]$ set $C_a = \{z : z \in \mathbb{R}^{r+1}, \|z\| = 1, \zeta_{r+1} \ge a\}$, writing $z = (\zeta_1, \dots, \zeta_{r+1})$ as usual. (i) Show that

$$\nu_r C_a = r\beta_r \int_0^{\arccos a} \sin^{r-1} t\, dt.$$

(ii) Compute the integral in the cases $r = 2$, $r = 4$.

>(c) Again write $C_a = \{z : z \in S_r, \zeta_{r+1} \ge a\}$, where $S_r \subseteq \mathbb{R}^{r+1}$ is the unit sphere. Show that, for any $a \in]0, 1]$, $\nu_r C_a \le \frac{\nu_r S_r}{2(r+1)a^2}$. (*Hint*: calculate $\sum_{i=1}^{r+1}\int_{S_r}\|\xi_i\|^2\nu_r(dx)$.)

>(d) Let $\phi :]0, 1[\to \mathbb{R}^r$ be an injective differentiable function. Show that the 'length' or one-dimensional Hausdorff measure of $\phi[]0, 1[]$ is just $\int_0^1 \|\phi'(t)\|dt$.

(e)(i) Show that if I is the identity $r \times r$ matrix and $z \in \mathbb{R}^r$, then $\det(I + zz^\top) = 1 + \|z\|^2$. (*Hint*: induce on r.) (ii) Write U_{r-1} for the open unit ball in $\mathbb{R}^{r-1}$, where $r \ge 2$. Define $\phi : U_{r-1} \times \mathbb{R} \to S_r$ by setting

$$\phi\begin{pmatrix} x \\ \xi \end{pmatrix} = \begin{pmatrix} x \\ \theta(x)\cos\xi \\ \theta(x)\sin\xi \end{pmatrix},$$

where $\theta(x) = \sqrt{1 - \|x\|^2}$. Show that

$$\phi'\begin{pmatrix} x \\ \xi \end{pmatrix}^\top \phi'\begin{pmatrix} x \\ \xi \end{pmatrix} = \begin{pmatrix} I + \frac{1}{\theta(x)^2}xx^\top & \mathbf{0} \\ \mathbf{0} & \theta(x)^2 \end{pmatrix},$$

so that $J\begin{pmatrix} x \\ \xi \end{pmatrix} = 1$ for all $x \in U_{r-1}$, $\xi \in \mathbb{R}$. (iii) Hence show that the normalized r-dimensional Hausdorff measure of $\{y : y \in S_r,\ \sum_{i=1}^{r-1} \eta_i^2 < 1\}$ is just $2\pi\beta_{r-1}$, where β_{r-1} is the Lebesgue measure of U_{r-1}. (iv) By considering $\psi z = \begin{pmatrix} z \\ 0 \\ 0 \end{pmatrix}$ for $z \in S_{r-2}$, or otherwise, show that the normalized r-dimensional Hausdorff measure of S_r is $2\pi\beta_{r-1}$. (v) This time setting $C_a = \{z : z \in \mathbb{R}^{r+1},\ \|z\| = 1,\ \zeta_1 \ge a\}$, show that $\nu_r C_a = 2\pi\mu_{r-1}\{x : x \in \mathbb{R}^{r-1},\ \|x\| \le 1,\ \xi_1 \ge a\}$ for every $a \in [-1, 1]$.

(f) Suppose that $r \ge 2$. Identifying $\mathbb{R}^r$ with $\mathbb{R}^{r-1} \times \mathbb{R}$, let C_r be the cylinder $B_{r-1} \times [-1, 1] \supseteq B_r$, and $\partial C_r = (B_{r-1} \times \{-1, 1\}) \cup (S_{r-2} \times [-1, 1])$ its boundary. Show that

$$\frac{\mu_r B_r}{\mu_r C_r} = \frac{\nu_{r-1} S_{r-1}}{\nu_{r-1}(\partial C_r)}.$$

(The case $r = 3$ is due to Archimedes.)

265Y Further exercises (a) Take $a < b$ in $\mathbb{R}$. (i) Show that $\phi : [a, b] \to \mathbb{R}^r$ is absolutely continuous in the sense of 264Yp iff all its coordinates $\phi_i : [a, b] \to \mathbb{R}$, for $i \le r$, are absolutely continuous in the sense of §225. (ii) Let $\phi : [a, b] \to \mathbb{R}^r$ be a continuous function, and set $F = \{x : x \in \left]a, b\right[,\ \phi \text{ is differentiable at } x\}$. Show that ϕ is absolutely continuous iff $\int_F \|\phi'(x)\| dx$ is finite and $\nu_1(\phi[[a, b] \setminus F]) = 0$, where ν_1 is (normalized) Hausdorff one-dimensional measure on $\mathbb{R}^r$. (*Hint*: 225K.) (iii) Show that if $\phi : [a, b] \to \mathbb{R}^r$ is absolutely continuous then $\nu_1^*(\phi[D]) \le \int_D \|\phi'(x)\| dx$ for every $D \subseteq [a, b]$, with equality if D is measurable and $\phi\restriction D$ is injective.

(b) Suppose that $a \le b$ in $\mathbb{R}$, and that $f : [a, b] \to \mathbb{R}$ is a continuous function of bounded variation with graph Γ_f. Show that the one-dimensional Hausdorff measure of Γ_f is $\mathrm{Var}_{[a,b]}(f) + \int_a^b (\sqrt{1 + (f')^2} - |f'|)$.

265 Notes and comments The proof of 265B seems to call on most of the second half of the alphabet. The idea is supposed to be straightforward enough. Because $T[\mathbb{R}^r]$ has dimension at most r, it can be rotated by an orthogonal transformation P into a subspace of the canonical r-dimensional subspace V, which is a natural copy of $\mathbb{R}^r$; the matrix R represents the copying process from V to $\mathbb{R}^r$, and ϕ or $P^\top R^\top$ is a copy of $\mathbb{R}^r$ onto a subspace including $T[\mathbb{R}^r]$. All this copying back and forth is designed to turn T into a linear operator $S : \mathbb{R}^r \to \mathbb{R}^r$ to which we can apply 263A, and part (b) of the proof is the check that we are copying the measures as well as the linear structures.

In 265D-265E I have tried to follow 263C-263D as closely as possible. In fact only one new idea is needed. When $s = r$, we have a special argument available to show that $\mu_r^*\phi[D] \le J\mu_r^* D + \epsilon\mu_r^* D$ (in the language of 263C) which applies whether or not $J = 0$. When $s > r$, this approach fails, because we can no longer approximate $\nu_r T[B]$ by $\nu_r G$ where $G \supseteq T[B]$ is open. (See part (b-i) of the proof of 263C.) I therefore turn to a different argument, valid only when $J > 0$, and accordingly have to find a separate method to show that $\{\phi(x) : x \in D,\ J(x) = 0\}$ is ν_r-negligible. Since we are working without restrictions on the dimensions r, s except that $r \le s$, we can use the trick of approximating $\phi : D \to \mathbb{R}^s$ by $\psi_\eta : D \to \mathbb{R}^{s+r}$, as in part (d) of the proof of 265E.

I give three methods by which the area of the r-sphere can be calculated; a bare-hands approach (265F), the surrounding-cylinder method (265Xe) and an important repeated-integral theorem (265G). The first two provide formulae for the area of a cap (265Xb, 265Xe(v)). The surrounding-cylinder method is attractive because the Jacobian comes out to be 1, that is, we have an inverse-measure-preserving function. I note that despite having developed a technique which allows irregular domains, I am still forced by the singularity in the function θ of 265Xe to take the sphere in two bites. Theorem 265G is a special case of the Coarea Theorem (Evans & Gariepy 92, §3.4; Federer 69, 3.2.12).

For the next steps in the geometric theory of measures on Euclidean space, see Chapter 47 in Volume 4.

Version of 28.1.09

*266 The Brunn-Minkowski inequality

We now have most of the essential ingredients for a proof of the Brunn-Minkowski inequality (266C) in a strong form. I do not at present expect to use it in this treatise, but it is one of the basic results of geometric measure theory and from where we now stand is not difficult, so I include it here. The preliminary results on arithmetic and geometric means (266A) and essential closures (266B) are of great importance for other reasons.

266A Arithmetic and geometric means We shall need the following standard result.

Proposition If $u_0,\dots,u_n,p_0,\dots,p_n\in[0,\infty[$ and $\sum_{i=0}^n p_i=1$, then $\prod_{i=0}^n u_i^{p_i}\le\sum_{i=0}^n p_iu_i$.

proof Induce on n. For $n=0$, $p_0=1$ the result is trivial. If $n=1$, then if $u_1=0$ the result is trivial (even if, as is standard in this book, we interpret 0^0 as 1). Otherwise, set $t=\frac{u_0}{u_1}$; then

$$t^{p_0}\le p_0t+1-p_0=p_0t+p_1$$

(as in part (a) of the proof of 244E), so

$$u_0^{p_0}u_1^{p_1}=t^{p_0}u_1\le p_0tu_1+p_1u_1=p_0u_0+p_1u_1.$$

For the inductive step to $n\ge 2$, if $p_0=\ldots=p_{n-1}=0$ the result is trivial. Otherwise, set $q=p_0+\ldots+p_{n-1}=1-p_n$; then

$$\prod_{i=0}^n u_i^{p_i}=(\prod_{i=0}^{n-1}u_i^{p_i/q})^q u_n^{p_n}\le(\sum_{i=0}^{n-1}\frac{p_i}{q}u_i)^q u_n^{p_n}$$

(by the inductive hypothesis)

$$\le q(\sum_{i=0}^{n-1}\frac{p_i}{q}u_i)+p_nu_n$$

(by the two-term case just examined)

$$=\sum_{i=0}^n p_iu_i,$$

and the induction continues.

266B Proposition For any set $D\subseteq\mathbb{R}^r$ set

$$\mathrm{cl}^*D=\{x:\limsup_{\delta\downarrow 0}\frac{\mu^*(D\cap B(x,\delta))}{\mu B(x,\delta)}>0\},$$

where μ is Lebesgue measure on $\mathbb{R}^r$.

(a) $D\setminus\mathrm{cl}^*D$ is negligible.

(b) $\mathrm{cl}^*D\subseteq\overline{D}$.

(c) cl^*D is a Borel set.

(d) $\mu(\mathrm{cl}^*D)=\mu^*D$.

(e) If $C\subseteq\mathbb{R}$ then $\overline{C}+\mathrm{cl}^*D\subseteq\mathrm{cl}^*(C+D)$, writing $C+D$ for $\{x+y:x\in C,\,y\in D\}$.

proof (a) 261Da.

(b) If $x\in\mathbb{R}^r\setminus\overline{D}$ then $D\cap B(x,\delta)=\emptyset$ for all small δ.

(c) The point is just that $(x,\delta)\mapsto\mu^*(D\cap B(x,\delta))$ is continuous. **P** For any x, $y\in\mathbb{R}^r$ and δ, $\eta\ge 0$ we have

$$|\mu^*(D \cap B(y,\eta)) - \mu^*(D \cap B(x,\delta))| \le \mu(B(y,\eta) \triangle B(x,\delta))$$
$$= 2\mu(B(x,\delta) \cup B(y,\eta)) - \mu B(x,\delta) - \mu B(y,\eta)$$
$$\le \beta_r\bigl(2(\max(\delta,\eta) + \|x-y\|)^r - \delta^r - \eta^r\bigr)$$

(where $\beta_r = \mu B(\mathbf{0},1)$)

$$\to 0$$

as $(y,\eta) \to (x,\delta)$. **Q** So

$$x \mapsto \limsup_{\delta \downarrow 0} \frac{\mu^*(D \cap B(x,\delta))}{\mu B(x,\delta)} = \inf_{\alpha \in \mathbb{Q}, \alpha > 0} \sup_{\beta \in \mathbb{Q}, 0 < \beta \le \alpha} \frac{1}{\beta_r \beta^r} \mu^*(D \cap B(x,\beta))$$

is Borel measurable, and

$$\mathrm{cl}^* D = \{x : \limsup_{\delta \downarrow 0} \frac{\mu^*(D \cap B(x,\delta))}{\mu B(x,\delta)} > 0\}$$

is a Borel set.

(d) By (c), $\mu(\mathrm{cl}^* D)$ is defined; by (a), $\mu(\mathrm{cl}^* D) \ge \mu^* D$. On the other hand, let E be a measurable envelope of D (132Ee); then 261Db tells us that

$$\limsup_{\delta \downarrow 0} \frac{\mu^*(D \cap B(x,\delta))}{\mu B(x,\delta)} \le \limsup_{\delta \downarrow 0} \frac{\mu(E \cap B(x,\delta))}{\mu B(x,\delta)} = 0$$

for almost every $x \in \mathbb{R}^r \setminus E$, so $\mathrm{cl}^* D \setminus E$ is negligible and

$$\mu(\mathrm{cl}^* D) \le \mu E = \mu^* D.$$

(e) If $x \in \overline{C}$ and $y \in \mathrm{cl}^* D$, set

$$\gamma = \frac{1}{3} \limsup_{\delta \downarrow 0} \frac{\mu^*(D \cap B(y,\delta))}{\mu B(y,\delta)} > 0.$$

For any $\eta > 0$, there is a $\delta \in \,]0,\eta]$ such that $\mu^*(D \cap B(y,\delta)) \ge 2\gamma \mu B(x,\delta)$. Let $\delta_1 \in [0,\delta[$ be such that $\delta^r - \delta_1^r \le \gamma \delta^r$. Then there is an $x' \in C$ such that $\|x - x'\| \le \delta - \delta_1$. In this case,

$$\mu^*((C+D) \cap B(x+y,\delta)) \ge \mu^*((x'+D) \cap B(x'+y,\delta_1)) = \mu^*(D \cap B(y,\delta_1))$$
$$\ge \mu^*(D \cap B(y,\delta)) - \mu B(y,\delta) + \mu B(y,\delta_1)$$
$$\ge 2\beta_r \gamma \delta^r - \beta_r \delta^r + \beta_r \delta_1^r \ge \beta_r \gamma \delta^r.$$

As η is arbitrary,

$$\limsup_{\delta \downarrow 0} \frac{\mu^*((C+D) \cap B(x+y,\delta))}{\mu B(y,\delta)} \ge \gamma$$

and $x + y \in \mathrm{cl}^*(C+D)$; as x and y are arbitrary, $\overline{C} + \mathrm{cl}^* D \subseteq \mathrm{cl}^*(C+D)$.

Remark In this context, $\mathrm{cl}^* D$ is called the **essential closure** of D.

266C Theorem Let A, $B \subseteq \mathbb{R}^r$ be non-empty sets, where $r \ge 1$ is an integer. If μ is Lebesgue measure on $\mathbb{R}^r$, and $A + B = \{x + y : x \in A, y \in B\}$, then $\mu^*(A+B)^{1/r} \ge (\mu^* A)^{1/r} + (\mu^* B)^{1/r}$.

proof (a) Consider first the case in which $A = [a, a'[$ and $B = [b, b'[$ are half-open intervals. In this case $A + B = [a+b, a'+b'[$; writing $a = (\alpha_1, \dots, \alpha_r)$, etc., as in §115, set

$$u_i = \frac{\alpha_i' - \alpha_i}{\alpha_i' + \beta_i' - \alpha_i - \beta_i}, \quad v_i = \frac{\beta_i' - \beta_i}{\alpha_i' + \beta_i' - \alpha_i - \beta_i}$$

for each i. Then we have

$$
\begin{aligned}
(\mu A)^{1/r}+(\mu B)^{1/r} &= \prod_{i=0}^{r}(\alpha_i'-\alpha_i)^{1/r}+\prod_{i=0}^{r}(\beta_i'-\beta_i)^{1/r}\\
&= \mu(A+B)^{1/r}(\prod_{i=1}^{r}u_i^{1/r}+\prod_{i=1}^{r}v_i^{1/r})\\
&\le \mu(A+B)^{1/r}(\frac{1}{r}\sum_{i=1}^{r}u_i+\frac{1}{r}\sum_{i=1}^{r}v_i)
\end{aligned}
$$

(266A)

$$
= \mu(A+B)^{1/r}.
$$

(b) Now I show by induction on $m+n$ that if $A=\bigcup_{j=0}^{m}A_j$ and $B=\bigcup_{j=0}^{n}B_j$, where $\langle A_j\rangle_{j\le m}$ and $\langle B_j\rangle_{j\le n}$ are both disjoint families of non-empty half-open intervals, then $\mu(A+B)^{1/r}\ge(\mu A)^{1/r}+(\mu B)^{1/r}$. **P** The induction starts with the case $m=n=0$, dealt with in (a). For the inductive step to $m+n=l\ge 1$, one of m, n is non-zero; the argument is the same in both cases; suppose the former. Since $A_0\cap A_1=\emptyset$, there must be some $j\le r$ and $\alpha\in\mathbb{R}$ such that A_0 and A_1 are separated by the hyperplane $\{x:\xi_j=\alpha\}$. Set $A'=\{x:x\in A,\ \xi_j<\alpha\}$ and $A''=\{x:x\in A,\ \xi_j\ge\alpha\}$; then both A' and A'' are non-empty and can be expressed as the union of at most $m-1$ disjoint half-open intervals. Set $\gamma=\frac{\mu A'}{\mu A}\in\,]0,1[$. The function $\beta\mapsto\mu\{x:x\in B,\ \xi_j<\beta\}$ is continuous, so there is a $\beta\in\mathbb{R}$ such that $\mu B'=\gamma\mu B$, where $B'=\{x:x\in B,\ \xi_j<\beta\}$; set $B''=B\setminus B$. Then B' and B'' can be expressed as unions of at most n half-open intervals. By the inductive hypothesis,

$$
\mu(A'+B')^{1/r}\ge(\mu A')^{1/r}+(\mu B')^{1/r},\quad \mu(A''+B'')^{1/r}\ge(\mu A'')^{1/r}+(\mu B'')^{1/r}.
$$

Now $A'+B'\subseteq\{x:\xi_j<\alpha+\beta\}$, while $A''+B''\subseteq\{x:\xi_j\ge\alpha+\beta\}$. So

$$
\begin{aligned}
\mu(A+B) &\ge \mu(A'+B')+\mu(A''+B'')\\
&\ge \bigl((\mu A')^{1/r}+(\mu B')^{1/r}\bigr)^r+\bigl((\mu A'')^{1/r}+(\mu B'')^{1/r}\bigr)^r\\
&= \bigl((\gamma\mu A)^{1/r}+(\gamma\mu B)^{1/r}\bigr)^r+\bigl(((1-\gamma)\mu A)^{1/r}+((1-\gamma)\mu B)^{1/r}\bigr)^r\\
&= \bigl((\mu A)^{1/r}+(\mu B)^{1/r}\bigr)^r.
\end{aligned}
$$

Taking rth roots, $\mu(A+B)^{1/r}\ge(\mu A)^{1/r}+(\mu B)^{1/r}$ and the induction proceeds. **Q**

(c) Now suppose that A and B are compact non-empty subsets of $\mathbb{R}^r$. Then $\mu(A+B)^{1/r}\ge(\mu A)^{1/r}+(\mu B)^{1/r}$. **P** $A+B$ is compact (because $A\times B\subseteq\mathbb{R}^r\times\mathbb{R}^r$ is compact, being closed and bounded, and addition is continuous, so we can use 2A2Eb). Let $\epsilon>0$. Let $G\supseteq A+B$ be an open set such that $\mu G\le\mu(A+B)+\epsilon$ (134Fa); then there is a $\delta>0$ such that $B(x,2\delta)\subseteq G$ for every $x\in A+B$ (2A2Ed). Let $n\in\mathbb{N}$ be such that $2^{-n}\sqrt{r}\le\delta$, and let A_1 be the union of all the half-open intervals of the form $[2^{-n}z,2^{-n}z+2^{-n}e[$ which meet A, where $z\in\mathbb{Z}^r$ and $e=(1,1,\dots,1)$. Then A_1 is a finite disjoint union of half-open intervals, $A\subseteq A_1$ and every point of A_1 is within a distance δ of some point of A. Similarly, we can find a set B_1, a finite disjoint union of half-open intervals, including B and such that every point of B_1 is within δ of some point of B. But this means that every point of A_1+B_1 is within a distance 2δ of some point of $A+B$, and belongs to G. Accordingly

$$
(\mu(A+B)+\epsilon)^{1/r}\ge(\mu G)^{1/r}\ge\mu(A_1+B_1)^{1/r}\ge(\mu A_1)^{1/r}+(\mu B_1)^{1/r}
$$

(by (b))

$$
\ge(\mu A)^{1/r}+(\mu B)^{1/r}.
$$

As ϵ is arbitrary, $\mu(A+B)^{1/r}\ge(\mu A)^{1/r}+(\mu B)^{1/r}$. **Q**

(d) Next suppose that A, $B\subseteq\mathbb{R}^r$ are Lebesgue measurable. Then

$$(\mu A)^{1/r} + (\mu B)^{1/r} = \sup\{(\mu K)^{1/r} + (\mu L)^{1/r} : K \subseteq A \text{ and } L \subseteq B \text{ are compact}\}$$

(134Fb)

$$\leq \sup\{\mu(K+L)^{1/r} : K \subseteq A \text{ and } L \subseteq B \text{ are compact}\}$$

(by (c))

$$\leq \mu^*(A+B)^{1/r}.$$

(e) For the penultimate step, suppose that A, $B \subseteq \mathbb{R}^r$ have non-zero outer Lebesgue measure. Consider cl^*A, cl^*B and $\mathrm{cl}^*(A+B)$ as defined in 266B. Then cl^*A and cl^*B are non-empty and their sum is included in $\mathrm{cl}^*(A+B)$, by 266Bb and 266Be. So we have

$$(\mu^* A)^{1/r} + (\mu^* B)^{1/r} = \mu(\mathrm{cl}^*A)^{1/r} + \mu(\mathrm{cl}^*B)^{1/r}$$

(266Bd)

$$\leq \mu^*(\mathrm{cl}^*A + \mathrm{cl}^*B)^{1/r}$$

(by (d) here)

$$\leq \mu(\mathrm{cl}^*(A+B))^{1/r} = \mu^*(A+B)^{1/r}.$$

(f) Finally, for arbitrary non-empty sets A, $B \subseteq \mathbb{R}^r$, note that if (for instance) A is negligible then we can take any $x \in A$ and see that

$$\mu^*(A+B)^{1/r} \geq \mu^*(x+B)^{1/r} = (\mu B)^{1/r} = (\mu^* A)^{1/r} + (\mu^* B)^{1/r},$$

and the result is similarly trivial if B is negligible. So all cases are covered.

266X Basic exercises (a) Let D, D' be subsets of $\mathbb{R}^r$. Show that (i) $\mathrm{cl}^*(D \cup D') = \mathrm{cl}^*D \cup \mathrm{cl}^*D'$ (ii) $\mathrm{cl}^*D = \mathrm{cl}^*D'$ iff D and D' have a common measurable envelope (iii) $\mathrm{cl}^*D \setminus \mathrm{cl}^*(\mathbb{R}^r \setminus D') \subseteq \mathrm{cl}^*(D \cap D')$ (iv) D is Lebesgue measurable iff $\mathrm{cl}^*D \cap \mathrm{cl}^*(\mathbb{R}^r \setminus D)$ is Lebesgue negligible (v) $D \cup \mathrm{cl}^*D$ is a measurable envelope of D (vi) $\mathrm{cl}^*(\mathrm{cl}^*D) = \mathrm{cl}^*D$.

(b) Show that, for a measurable set $E \subseteq \mathbb{R}$, cl^*E is just the set of real numbers which are not density points of $\mathbb{R} \setminus E$.

(c) In 266C, show that if A and B are similar convex sets in the same orientation then $A+B$ is a convex set similar to both and $\mu(A+B)^{1/r} = (\mu A)^{1/r} + (\mu B)^{1/r}$.

(d) Show that if $r \geq 1$, μ is Lebesgue measure on $\mathbb{R}^r$ and $A_0, \dots, A_n$ are non-empty subsets of $\mathbb{R}^r$, then $\mu^*(A_0 + \ldots + A_n)^{1/r} \geq \sum_{i=0}^{n} (\mu^* A_i)^{1/r}$.

(e) In 266C, show that if $p \in [0,1]$ then (subject to an appropriate interpretation of ∞^0) $\mu^*(pA + (1-p)B) \geq (\mu^* A)^p (\mu^* B)^{1-p}$.

266 Notes and comments The proof of 266C is taken from Federer 69. There is a slightly specious generality in the form given here. If the sets A and B are at all irregular, then $\mu^*(A+B)^{1/r}$ is likely to be much greater than $(\mu^* A)^{1/r} + (\mu^* B)^{1/r}$. The critical case, in which A and B are similar convex sets, is much easier (266Xc). The theorem is therefore most useful when A and B are non-similar convex sets and we get a non-trivial estimate which may be hard to establish by other means. For this case we do not need 266B. Theorem 266C is an instructive example of the way in which the dimension r enters formulae when we seek results applying to general Euclidean spaces. There will be many more when I return to geometric measure theory in Chapter 47 of Volume 4.

Version of 17.12.12

References for Volume 2

Alexits G. [78] (ed.) *Fourier Analysis and Approximation Theory.* North-Holland, 1978 (Colloq. Math. Soc. Janos Bolyai 19).

Antonov N.Yu. [96] 'Convergence of Fourier series', East J. Approx. 7 (1996) 187-196. [§286 *notes*.]

Arias de Reyna J. [02] *Pointwise Convergence of Fourier Series.* Springer, 2002 (Lecture Notes in Mathematics 1785). [§286 *notes*.]

Baker R. [04] '"Lebesgue measure" on $\mathbb{R}^\infty$, II', Proc. Amer. Math. Soc. 132 (2004) 2577-2591. [254Yb.]

Bergelson V., March P. & Rosenblatt J. [96] (eds.) *Convergence in Ergodic Theory and Probability.* de Gruyter, 1996.

Bogachev V.I. [07] *Measure theory.* Springer, 2007.

du Bois-Reymond P. [1876] 'Untersuchungen über die Convergenz und Divergenz der Fouriersche Darstellungformeln', Abh. Akad. München 12 (1876) 1-103. [§282 *notes*.]

Bourbaki N. [66] *General Topology.* Hermann/Addison-Wesley, 1968. [2A5F.]

Bourbaki N. [87] *Topological Vector Spaces.* Springer, 1987. [2A5E.]

Carleson L. [66] 'On convergence and growth of partial sums of Fourier series', Acta Math. 116 (1966) 135-157. [§282 *notes*, §286 *intro.*, §286 *notes*.]

Clarkson J.A. [1936] 'Uniformly convex spaces', Trans. Amer. Math. Soc. 40 (1936) 396-414. [244O.]

Defant A. & Floret K. [93] *Tensor Norms and Operator Ideals*, North-Holland, 1993. [§253 *notes*.]

Doob J.L. [53] *Stochastic Processes.* Wiley, 1953.

Dudley R.M. [89] *Real Analysis and Probability.* Wadsworth & Brooks/Cole, 1989. [§282 *notes*.]

Dunford N. & Schwartz J.T. [57] *Linear Operators I.* Wiley, 1957 (reprinted 1988). [§244 *notes*, 2A5J.]

Enderton H.B. [77] *Elements of Set Theory.* Academic, 1977. [§2A1.]

Engelking R. [89] *General Topology.* Heldermann, 1989 (Sigma Series in Pure Mathematics 6). [2A5F.]

Etemadi N. [96] 'On convergence of partial sums of independent random variables', pp. 137-144 in Bergelson March & Rosenblatt 96. [272V.]

Evans L.C. & Gariepy R.F. [92] *Measure Theory and Fine Properties of Functions.* CRC Press, 1992. [263Ec, §265 *notes*.]

Federer H. [69] *Geometric Measure Theory.* Springer, 1969 (reprinted 1996). [262C, 263Ec, §264 *notes*, §265 *notes*, §266 *notes*.]

Feller W. [66] *An Introduction to Probability Theory and its Applications,* vol. II. Wiley, 1966. [Chap. 27 *intro.*, 274H, 275Xc, 285N.]

Fremlin D.H. [74] *Topological Riesz Spaces and Measure Theory.* Cambridge U.P., 1974. [§232 *notes*, 241F, §244 *notes*, §245 *notes*, §247 *notes*.]

Fremlin D.H. [93] 'Real-valued-measurable cardinals', pp. 151-304 in Judah 93. [232H.]

Haimo D.T. [67] (ed.) *Orthogonal Expansions and their Continuous Analogues.* Southern Illinois University Press, 1967.

Hall P. [82] *Rates of Convergence in the Central Limit Theorem.* Pitman, 1982. [274H.]

Halmos P.R. [50] *Measure Theory.* Van Nostrand, 1950. [§251 *notes*, §252 *notes*, 255Yn.]

Halmos P.R. [60] *Naive Set Theory.* Van Nostrand, 1960. [§2A1.]

Hanner O. [56] 'On the uniform convexity of L^p and l^p', Arkiv för Matematik 3 (1956) 239-244. [244O.]

Henle J.M. [86] *An Outline of Set Theory.* Springer, 1986. [§2A1.]

Hoeffding W. [63] 'Probability inequalities for sums of bounded random variables', J. Amer. Statistical Association 58 (1963) 13-30. [272W.]

Hunt R.A. [67] 'On the convergence of Fourier series', pp. 235-255 in Haimo 67. [§286 *notes*.]

Jorsbøe O.G. & Mejlbro L. [82] *The Carleson-Hunt Theorem on Fourier Series.* Springer, 1982 (Lecture Notes in Mathematics 911). [§286 *notes*.]

Judah H. [93] (ed.) Proceedings of the Bar-Ilan Conference on Set Theory and the Reals, 1991. Amer. Math. Soc. (Israel Mathematical Conference Proceedings 6), 1993.

Kelley J.L. [55] *General Topology.* Van Nostrand, 1955. [2A5F.]

Kelley J.L. & Namioka I. [76] *Linear Topological Spaces.* Springer, 1976. [2A5C.]

Kirszbraun M.D. [1934] 'Über die zusammenziehenden und Lipschitzian Transformationen', Fund. Math. 22 (1934) 77-108. [262C.]

Kolmogorov A.N. [1926] 'Une série de Fourier-Lebesgue divergente partout', C. R. Acad. Sci. Paris 183 (1926) 1327-1328. [§282 *notes*.]

Komlós J. [67] 'A generalization of a problem of Steinhaus', Acta Math. Acad. Sci. Hung. 18 (1967) 217-229. [276H.]

Körner T.W. [88] *Fourier Analysis.* Cambridge U.P., 1988. [§282 *notes*.]

Köthe G. [69] *Topological Vector Spaces I.* Springer, 1969. [2A5C, 2A5E, 2A5J.]

Krivine J.-L. [71] *Introduction to Axiomatic Set Theory.* D. Reidel, 1971. [§2A1.]

Lacey M.T. [05] *Carleson's Theorem: Proof, Complements, Variations.* `http://arxiv.org/pdf/math/0307008v4.pdf`. [§286 *notes*.]

Lacey M.T. & Thiele C.M. [00] 'A proof of boundedness of the Carleson operator', Math. Research Letters 7 (2000) 1-10. [§286 *intro*., 286H.]

Lebesgue H. [72] *Oeuvres Scientifiques.* L'Enseignement Mathématique, Institut de Mathématiques, Univ. de Genève, 1972. [Chap. 27 *intro*.]

Liapounoff A. [1901] 'Nouvelle forme du théorème sur la limite de probabilité', Mém. Acad. Imp. Sci. St-Pétersbourg 12(5) (1901) 1-24. [274Xh.]

Lighthill M.J. [59] *Introduction to Fourier Analysis and Generalised Functions.* Cambridge U.P., 1959. [§284 *notes*.]

Lindeberg J.W. [1922] 'Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung', Math. Zeitschrift 15 (1922) 211-225. [274H, §274 *notes*.]

Lipschutz S. [64] *Set Theory and Related Topics.* McGraw-Hill, 1964 (Schaum's Outline Series). [§2A1.]

Loève M. [77] *Probability Theory I.* Springer, 1977. [Chap. 27 *intro*., 274H.]

Luxemburg W.A.J. & Zaanen A.C. [71] *Riesz Spaces I.* North-Holland, 1971. [241F.]

Mozzochi C.J. [71] *On the Pointwise Convergence of Fourier Series.* Springer, 1971 (Lecture Notes in Mathematics 199). [§286 *notes*.]

Naor A. [04] 'Proof of the uniform convexity lemma', `http://www.cims.nyu.edu/~naor/homepage files/inequality.pdf`, 26.2.04. [244O.]

Rényi A. [70] *Probability Theory.* North-Holland, 1970. [274H.]

Roitman J. [90] *An Introduction to Set Theory.* Wiley, 1990. [§2A1.]

Roselli P. & Willem M. [02] 'A convexity inequality', Amer. Math. Monthly 109 (2002) 64-70. [244Ym.]

Saks S. [1924] 'Sur les nombres dérivés des fonctions', Fund

Schipp F. [78] 'On Carleson's method', pp. 679-695 in Alexits 78. [§286 *notes*.]

Semadeni Z. [71] *Banach spaces of continuous functions I.* Polish Scientific Publishers, 1971. [§253 *notes*.]

Shiryayev A.N. [84] *Probability.* Springer, 1984. [285N.]

Steele J.M. [86] 'An Efron-Stein inequality of nonsymmetric statistics', Annals of Statistics 14 (1986) 753-758. [274Ya.]

Zaanen A.C. [83] *Riesz Spaces II.* North-Holland, 1983. [241F.]

Zygmund A. [59] *Trigonometric Series.* Cambridge U.P, 1959. [§244 *notes*, §282 *notes*, 284Xk.]