LEMMAS FOR CARATHÉODORY MEASURE THEORY

Since I have decided to take this approach to measure theory rather than the more historical geometric approach of Wheeden & Zygmund's Chapter 3, it is my responsibility to make sure that the development remains consistent and free from circular reasoning. These notes will pretty much be a laundry list of propositions; their purpose is to make sure that we all know what we are taking for granted and what we are trying to prove.

1. *Volume in* $\mathbb{R}^n$*.*

Wheeden & Zygmund *define* the **volume** of a parallelepiped, whether rectangular or not, by the determinantal formula: their volume of a parallelepiped is "*by definition* the absolute value of the $n \times n$ determinant having [the edges of the parallelepiped] as rows" (their italics). My feeling is that while one may define the **volume of an interval** $I = [a_1, b_1] \times \cdots \times [a_n, b_n] \subset \mathbb{R}^n$ by $v(I) = \prod_{k=1}^{n}(b_k - a_k)$, the determinantal formula for a parallelepiped whose faces are not parallel to the coördinate hyperplanes (solutions of $x_j =$ const. for some $j = 1, \ldots, n$) ought to be a measure-theoretic theorem rather than a definition.[1] For measure-theoretic purposes we can take this volume definition to apply to all intervals, whether open, closed or neither: once we have n-dimensional Lebesgue measure we shall see that the faces of the interval have measure zero, so they play no rôle in measure-theoretic constructions, and in the definition of (Lebesgue n-dimensional) outer measure, we shall see that it doesn't really matter whether one uses open, closed, or semi-open intervals.

2. *Lebesgue Outer Measure as a Carathéodory Outer Measure.*

The plan is to follow Wheeden & Zygmund's Chapter 11 rather than their Chapter 3. We will also use somewhat different notation: they really do overwork the vertical bars $|\cdot|$ and I would rather use the more standard m^* as the name of **Lebesgue outer measure** (the dimension n will either be free or be clear from the context). The definition is the same, of course: for any $E \subseteq \mathbb{R}^n$, one may cover E with a countable family of intervals $\mathcal{S} = \{I_k\}_{k=1}^{\infty}$ which we shall always write as if it were infinite but which may be finite, form the sum $\sigma(\mathcal{S}) = \sum_{k=1}^{\infty} v(I_k)$ (order of summation is unimportant in a series of nonnegative terms) and then put

$$m^*(E) = \inf\{\sigma(\mathcal{S}) : E \subseteq \bigcup \mathcal{S}\}\ .$$

This is what everybody does. It does not matter whether one uses open, closed, or semi-open intervals in this definition, because the infimum will be the same. Given $\mathcal{S}$, replacing its elements by their closures will not change the volume; on the other hand, if $\mathcal{S}$ is composed of non-open intervals, then for given $\epsilon > 0$ we may replace each $I_k \in \mathcal{S}$ by a slightly larger open $I_k^* \supset \overline{I_k}$ with $v(I_k^*) \le v(I_k) + \epsilon/2^k$. If the collection of the I_k^*'s is called $\mathcal{S}^*$, then evidently $\sigma(\mathcal{S}) \le \sigma(\mathcal{S}^*) \le \sigma(\mathcal{S}) + \epsilon$; the ϵ makes no difference if the infimum is $+\infty$, and otherwise (since $\epsilon > 0$ is in our hands) we see that the two possibilities for computing the infimum yield the same result.

We need Wheeden & Zygmund's **Theorem (3.2)** that $m^*(I) = v(I)$ when I is an interval. The inequality $m^*(I) \le v(I)$ holds because I covers itself. For the reverse inequality: given $\{I_k\}_{k=1}^{\infty}$ with $I \subseteq \bigcup_{k=1}^{\infty} I_k$, take $0 < \epsilon < v(I)$, replace I by a closed interval $I_* \subseteq I$ with $v(I_*) \ge v(I) - \epsilon/2$ and each I_k by an open interval $I_k^* \supseteq I_k$ with $v(I_k^*) \le v(I_k) + \epsilon/2^{k+1}$. The Heine-Borel theorem lets one choose a finite cover of I_* from the I_k^*'s, say $\{I_k^*\}_{k=1}^{N}$. "Clearly" $v(I) \le \sum_{k=1}^{N} v(I_k^*)$, except that to me (and to most authors) it is not so clear. A combinatorial argument can be given; it would be an interesting exercise for the reader to find one in dimensions 1 and 2. For a general proof in a situation in which pictures are hard to draw, we could use characteristic functions.[2] For an open interval $J = \prod_{j=1}^{n}(a_j, b_j)$ or a closed interval $\overline{J} = \prod_{j=1}^{n}[a_j, b_j]$, the characteristic function of the n-dimensional interval is the product of the characteristic functions of the 1-dimensional intervals of which it is the product:

$$\chi_J(x_1, \ldots, x_n) = \prod_{j=1}^{n} \chi_{(a_j,b_j)}(x_j) \quad \text{and} \quad \chi_{\overline{J}}(x_1, \ldots, x_n) = \prod_{j=1}^{n} \chi_{[a_j,b_j]}(x_j)\ .$$

---

[1] It is so easy to prove that this point of view will not unduly burden the reader.

[2] Professional probabilists and statisticians may call these functions "indicator functions," reserving the name "characteristic function" for the Fourier transform of a probability distribution function (or measure).

The iterated one-dimensional Riemann integral (*not* the n-dimensional Riemann integral, which [in principle] we do not have) of either of these gives the n-dimensional volume of the interval by direct computation (integrability is clear from what we know about the 1-dimensional Riemann integral): for any $a < b$ for which $a < a_j$ and $b_j < b$ hold for all $j = 1, \ldots, n$, we have

$$\int_a^b \left[\cdots \left[\int_a^b \chi_J(x_1, \ldots, x_n)\, dx_1 \right] \cdots \right] dx_n = \prod_{j=1}^{n} (b_j - a_j) = v(J)$$

and similarly for $\overline{J}$. Returning to our Lebesgue-measure problem, we see that $I_* \subseteq \bigcup_{k=1}^{N} I_k^*$ implies that $\chi_{I_*}(\mathbf{x}) \leq \sum_{k=1}^{N} \chi_{I_k^*}(\mathbf{x})$ must hold for all $\mathbf{x} \in \mathbb{R}^n$, because if $\mathbf{x} \in I_*$ then at least one of the terms on the r. h. s. of that relation must equal 1. We may now take $[a, b] \subset \mathbb{R}$ sufficiently large to contain all the intervals of which any I_k^* or I_* might be a product, and then iterated Riemann integration and the linearity and positivity of the 1-dimensional Riemann integral give

$$\begin{aligned} v(I) - \frac{\epsilon}{2} \leq v(I_*) &= \int_a^b \left[\cdots \left[\int_a^b \chi_{I_*}(x_1, \ldots, x_n)\, dx_1 \right] \cdots \right] dx_n \\ &\leq \int_a^b \left[\cdots \left[\int_a^b \sum_{k=1}^{N} \chi_{I_k^*}(x_1, \ldots, x_n)\, dx_1 \right] \cdots \right] dx_n \\ &= \sum_{k=1}^{N} \int_a^b \left[\cdots \left[\int_a^b \chi_{I_k^*}(x_1, \ldots, x_n)\, dx_1 \right] \cdots \right] dx_n = \sum_{k=1}^{N} v(I_k^*) \leq \frac{\epsilon}{2} + \sum_{k=1}^{N} v(I_k) \\ v(I) &\leq \epsilon + \sum_{k=1}^{N} v(I_k) \leq \epsilon + \sum_{k=1}^{\infty} v(I_k) \end{aligned}$$

and since $\epsilon > 0$ is in our hands, we have the reverse inequality $v(I) \leq m^*(I)$.

There are a couple of observations we can make about the Lebesgue outer measure of certain sets: *e.g.*, because of the translation-invariance of volumes of intervals, the Lebesgue outer measure of a translate of a set is the same as that of the set. A **homothety** $\mathbf{x} \mapsto \lambda \mathbf{x}$ of $\mathbb{R}^n$ onto itself (for $\lambda \neq 0$) multiplies the volume of every interval by $|\lambda|^n$ and hence does the same for Lebesgue outer measure. Any subset of a hyperplane parallel to a coördinate hyperplane has Lebesgue outer measure zero. This follows from the countable subadditivity of Lebesgue outer measure (**Theorem (3.3)**) and the fact that an $(n-1)$-dimensional interval $\prod_{j=1}^{n-1} [a_j, b_j]$ in the hyperplane $x_n = 0$ can be enclosed in n-dimensional intervals $I = \prod_{j=1}^{n-1} [a_j, b_j] \times [-\epsilon, \epsilon]$ of arbitrarily small n-dimensional volume (the corresponding statement for other coördinates is equally true).

As a result of Wheeden & Zygmund's **Theorems (3.3)** and **(3.4)**, we now know that Lebesgue n-dimensional outer measure $m^*(\cdot)$ satisfies the Carathéodory axioms for an outer measure with which their Chapter 11 begins (p. 193). Any differences between what follows and the treatments in Chapter 11 will be largely notational—I find measures named gamma hard to live with.

3. *Carathéodory Outer Measures.*

We start with a $[0, \infty]$-valued set function $m^*(\cdot)$ defined on all subsets of a nonempty set X and satisfying the following axioms:

(i) $m^*(\emptyset) = 0$, and for all $A \in \mathbf{2}^X$ one has $0 \leq m^*(A) \leq \infty$;

(ii) $A_1 \subseteq A_2 \Rightarrow m^*(A_1) \leq m^*(A_2)$ (monotonicity);

(iii) For countable $\{A_k\} \subseteq \mathbf{2}^X$, one has $m^*\left(\bigcup A_k\right) \leq \sum m^*(A_k)$ (countable subadditivity).

Once we have the σ-algebra $\mathfrak{M}$ of measurable sets selected out of $\mathbf{2}^X$ we shall use $m(\cdot)$ to denote the restriction of m^* to $\mathfrak{M}$.

A set $E \in \mathbf{2}^X$ is defined to be (m^*-)**measurable** if the relation

$$m^*(A) = m^*(A \cap E) + m^*(A \cap \complement E)$$

holds for every $A \in \mathbf{2}^X$ (without restriction on A); the family of all measurable sets will be denoted by $\mathfrak{M}$. Since

$$m^*(A) \leq m^*(A \cap E) + m^*(A \cap \complement E)$$

holds automatically by subadditivity of m^*, one only has to verify

$$m^*(A) \geq m^*(A \cap E) + m^*(A \cap \complement E) \ ;$$

moreover, since that relation holds automatically when $m^*(A) = \infty$, it only has to be checked under the assumption $m^*(A) < \infty$.

It is obvious that $\emptyset \in \mathfrak{M}$, $X \in \mathfrak{M}$, and $X \in \mathfrak{M} \iff \complement X \in \mathfrak{M}$. A set is called ($m^*$-)**null** if $m^*(E) = 0$. Any subset of a m^*-null set is null, and therefore any null set E is measurable because for any $A \subseteq X$ one has $0 \leq m^*(A \cap E) \leq m^*(E) = 0$ and $m^*(A) \geq m^*(A \cap \complement E)$ (both by monotonicity) and therefore $m^*(A) \geq m^*(A \cap E) + m^*(A \cap \complement E)$. These facts are not very thrilling; to get more, one has to do some rather unintuitive work and prove some lemmas. The first we attacked in class is

Lemma [Wheeden & Zygmund **(11.2)**]: If $\{E_k\}$ is a countable disjoint family of measurable sets, then its union is measurable and the relations

$$m^*(A) = \sum m^*(A \cap E_k) + m^*\left(A \setminus \bigcup E_k\right) \quad \text{and} \quad m^*\left(A \cap \bigcup E_k\right) = \sum m^*(A \cap E_k)$$

hold.

Proof sketch. First for two sets E_1, E_2: let $A \in \mathbf{2}^X$ be given, split A with (say) E_1:

$$m^*(A) = m^*(A \cap E_1) + m^*(A \cap \complement E_1)$$

and then split $A \cap \complement E_1$ with E_2:

$$\begin{aligned} m^*(A \cap \complement E_1) &= m^*(A \cap \complement E_1 \cap E_2) + m^*(A \cap \complement E_1 \cap \complement E_2) \\ &= m^*(A \cap E_2) + m^*(A \cap \complement E_1 \cap \complement E_2) \ . \end{aligned}$$

Putting these together gives

$$\begin{aligned} m^*(A) &= m^*(A \cap E_1) + m^*(A \cap E_2) + m^*(A \cap \complement E_1 \cap \complement E_2) \\ m^*(A) &= m^*(A \cap E_1) + m^*(A \cap E_2) + m^*(A \cap \complement(E_1 \cup E_2)) \qquad (*) \end{aligned}$$

which is the case in which there are two E_k's of the first relation we are trying to establish. It implies the measurability of $E_1 \cup E_2$ because subadditivity gives

$$m^*(A) = m^*(A \cap E_1) + m^*(A \cap E_2) + m^*(A \cap \complement(E_1 \cup E_2)) \geq m^*(A \cap (E_1 \cup E_2)) + m^*(A \cap \complement(E_1 \cup E_2)).$$

The case of any finite number of E_k's follows by (finite) mathematical induction. In the case in which there are N sets $\{E_k\}_{k=1}^N$, the relation $(*)$ becomes

$$m^*(A) = \sum_{k=1}^{N} m^*(A \cap E_k) + m^*\left[A \cap \complement\left(\bigcup_{k=1}^{N} E_k\right)\right], \quad \text{which by monotonicity implies}$$

$$m^*(A) \geq \sum_{k=1}^{N} m^*(A \cap E_k) + m^*\left[A \cap \complement\left(\bigcup_{k=1}^{\infty} E_k\right)\right].$$

Letting $N \to \infty$ in the second of these relations gives

$$m^*(A) \geq \sum_{k=1}^{\infty} m^*(A \cap E_k) + m^*\left[A \cap \complement\left(\bigcup_{k=1}^{\infty} E_k\right)\right]$$

from which we may deduce by subadditivity

$$m^*(A) \geq \sum_{k=1}^{\infty} m^*(A \cap E_k) + m^*\left[A \cap \complement\left(\bigcup_{k=1}^{\infty} E_k\right)\right] \geq m^*\left[A \cap \bigcup_{k=1}^{\infty} E_k\right] + m^*\left[A \cap \complement\left(\bigcup_{k=1}^{\infty} E_k\right)\right] .$$

The left and right ends of this inequality tell us that $\bigcup_{k=1}^{\infty} E_k$ and its complement are measurable; but then equality holds among all three of these expressions, which gives the first assertion of the lemma. Since $A \in \mathbf{2}^X$ was arbitrary we may replace it by $A \cap \bigcup_{k=1}^{\infty} E_k$ throughout, and then the last set-off line above becomes

$$m^*\left(A \cap \bigcup_{k=1}^{\infty} E_k\right) \geq \sum_{k=1}^{\infty} m^*(A \cap E_k) + m^*(\emptyset) \geq m^*\left(A \cap \bigcup_{k=1}^{\infty} E_k\right) + m^*(\emptyset)$$

which is a somewhat ornamented version of the second assertion of the lemma.

The traditional next thing to prove is that $\mathfrak{M}$ is closed under relative complementation. Proving that it is closed under (not necessarily disjoint) union is just as easy and perhaps slightly more efficient.

Lemma: If both E_1 and E_2 are m^*-measurable, then so is $E_1 \cup E_2$.

Proof. Let $A \in \mathbf{2}^X$ with $m^*(A) < \infty$ be given. We can split $A \cap \complement E_1$ using E_2:

$$m^*(A \cap \complement E_1) = m^*(A \cap \complement E_1 \cap E_2) + m^*(A \cap \complement E_1 \cap \complement E_2)$$

and then, disjointizing $A \cap (E_1 \cup E_2)$ into $(A \cap E_1) \cup (A \cap \complement E_1 \cap E_2)$, write

$$m^*(A \cap (E_1 \cup E_2)) \leq m^*(A \cap E_1) + m^*(A \cap \complement E_1 \cap E_2)$$

by subadditivity. We now have

$$\begin{aligned} m^*(A \cap (E_1 \cup E_2)) + m^*(A \cap \complement(E_1 \cup E_2)) &= m^*(A \cap (E_1 \cup E_2)) + m^*(A \cap \complement E_1 \cap \complement E_2) \\ &\leq \{m^*(A \cap E_1) + m^*(A \cap \complement E_1 \cap E_2)\} + \{m^*(A \cap \complement E_1) - m^*(A \cap \complement E_1 \cap E_2)\} \\ &= m^*(A \cap E_1) + m^*(A \cap \complement E_1) = m^*(A) \end{aligned}$$

which shows that $E_1 \cup E_2$ is measurable.

Corollary: Intersections of pairs of m^*-measurable sets are measurable. The relative complement of one measurable set in another is measurable. Countable (not necessarily disjoint) unions of measurable sets are measurable, as are countable intersections of measurable sets.

Should we dignify it by the name proof? Since measurability is preserved under union and complementation,

$$E_1 \cap E_2 = \complement(\complement E_1 \cup \complement E_2)$$

shows that it is preserved under intersection. Then $E_1 \setminus E_2 = E_1 \cap \complement E_2$ shows that measurability is preserved under relative complementation. If $\{E_k\}_{k=1}^{\infty}$ is a sequence of measurable sets, then successive disjointization into $F_1 = E_1, \ldots, F_k = E_k \setminus \bigcup_{j=1}^{k-1} F_j$ gives measurable sets whose union equals the union of the $\{E_k\}_{k=1}^{\infty}$, which is therefore measurable. For countable intersections we also have the DeMorgan-law $\bigcap_{k=1}^{\infty} E_k = \complement\left(\bigcup_{k=1}^{\infty} \complement E_k\right)$.

We have thus shown that for a given outer measure on a set X that satisfies the three Carathéodory axioms, the family $\mathfrak{M}$ of Carathéodory-measurable sets is a **σ-algebra of sets**, *i.e.*, a subfamily of $\mathbf{2}^X$ containing $\emptyset$ and closed under complementation, countable union and countable intersection. The set function $m = m^*|\mathfrak{M}$ is a **measure**; the triple $(X, \mathfrak{M}, m)$ is an (abstract) **measure space** consisting of a set, a σ-algebra of subsets of the set, and a countably additive $[0, \infty]$-valued set function on the σ-algebra. While we shall only be using this fact for Lebesgue (outer) measure for the time being, we have given the arguments that establish it in general.

4. *Back to Wheeden & Zygmund Chapter 3.*

We now have to check that the remaining theorems—and, in some cases, definitions—of this chapter are theorems in the present context. Throughout this § we shall work in an $\mathbb{R}^n$ whose dimension n will be clear from context; $\mathfrak{M}$ will denote the σ-algebra of sets that are measurable for Lebesgue (outer) measure, and m will denote Lebesgue measure (on $\mathfrak{M}$).

First notice that the calculations of the measures of the various Cantor sets on the line are now rigorously established: the lengths of the complementary intervals were computed previously, we now know that the sum of their lengths is the measure of each $[0,1] \setminus C$ because Lebesgue measure is countably additive, and for the same reason we have $m(C) = \prod_{k=1}^{\infty}(1 - r_k)$.

Theorem (3.6), the formula **(3.7)**, and **(3.8)** of p. 36 hold by the definition of Lebesgue outer measure (provided that we use open intervals to compute it, as we may do), and so the Wheeden & Zygmund proofs apply without modification. We shall defer proof of **Theorem (3.10)** to a more general theorem below (the image of any Lebesgue-measurable set under a nonsingular linear transformation A gets its measure multiplied by $|\det A|$).

Some of the theorems of Wheeden & Zygmund's §2 are now known, whereas some of their more "obvious" propositions require proof. For example, we have to prove their **Example 1**, p. 37.

Proposition: Every open subset of $\mathbb{R}^n$ is measurable.

Proof. One can deduce this fact from observing that m is a **metric outer measure** in the sense of Chapter 11, §2, but we can also proceed directly. It suffices to prove that every closed coördinate half-space $H = \{\mathbf{x} = (x_1, \ldots, x_n) : c \le x_j\}$ (where j and $c \in \mathbb{R}$ are fixed, $1 \le j \le n$) is measurable, because then every open half-space $\{\mathbf{x} = (x_1, \ldots, x_n) : x_j < c\}$ will be measurable, the fact that $\mathfrak{M}$ is a σ-algebra will imply that half-spaces $\{\mathbf{x} = (x_1, \ldots, x_n) : c < x_j\} = \displaystyle\bigcup_{k=1}^{\infty} \{\mathbf{x} = (x_1, \ldots, x_n) : c + \frac{1}{k} \le x_j\}$ and $\{\mathbf{x} = (x_1, \ldots, x_n) : x_j \le c\}$ are measurable, then that every interval is measurable (as the intersection of 2^n measurable sets) and so that every open set is measurable (as a countable union of intervals). So let an arbitrary subset $A \subseteq \mathbb{R}^n$ with $m^*(A) < \infty$ be given; we must show that $m^*(A) \ge m^*(A \cap \complement H) + m^*(A \cap H)$. For any $\epsilon > 0$ we may cover A with a countable family of intervals $\{I_k\}$ for which $m^*(A) \le \sum_k v(I_k) + \epsilon$. The intervals I_k divide into three classes: those lying entirely in the half-space $H = \{\mathbf{x} = (x_1, \ldots, x_n) : c \le x_j\}$—let R denote the set of indices k for which this happens; those lying entirely in its complement $\complement H = \{\mathbf{x} = (x_1, \ldots, x_n) : x_j < c\}$—let L denote the set of their indices; and those that have nonempty intersection with the hyperplane $\{\mathbf{x} = (x_1, \ldots, x_n) : c = x_j\}$—let their set of indices be denoted by M. In this case $j = n$, for example, an I_k with $k \in M$ would have the form $I_k = \displaystyle\prod_{i=1}^{n-1} [a_i, b_i] \times [a_n, b_n]$ where $a_n \le c \le b_n$. Such an I_k can be rewritten as the disjoint union

$$I_k = \prod_{i=1}^{n-1} [a_i, b_i] \times [a_n, c) \cup \prod_{i=1}^{n-1} [a_i, b_i] \times [c, b_n] = I_k' \cup I_k''$$

and similarly for other coördinates. Clearly $A \cap \complement H \subseteq \bigcup_{k\in L} I_k \cup \bigcup_{k\in M} I'_k$, while $A \cap H \subseteq \bigcup_{k\in R} I_k \cup \bigcup_{k\in M} I''_k$, and therefore

$$\begin{aligned} m^*(A \cap \complement H) + m^*(A \cap H) &\le \left[\sum_{k\in L} v(I_k) + \sum_{k\in M} v(I'_k)\right] + \left[\sum_{k\in R} v(I_k) + \sum_{k\in M} v(I''_k)\right] \\ &= \sum_{k\in L} v(I_k) + \sum_{k\in R} v(I_k) + \left[\sum_{k\in M} v(I'_k) + \sum_{k\in M} v(I''_k)\right] \\ &= \sum_{k\in L} v(I_k) + \sum_{k\in R} v(I_k) + \sum_{k\in M} v(I_k) = \sum_{k} v(I_k) \le m^*(A) + \epsilon \end{aligned}$$

and that suffices for the proof, since $\epsilon > 0$ was arbitrary.

Corollary [Wheeden & Zygmund **(3.13)**]: Intervals are measurable, with measure equal to their volumes.

Corollary [Wheeden & Zygmund **(3.14)**]: Closed sets are measurable.

Of course we already knew that (**Example 2**) every set of outer measure zero was measurable. On the other hand, **(3.11)** cannot be a definition, but has to be established:

Proposition: A set $E \subseteq \mathbb{R}^n$ is (Lebesgue-)measurable if and only if for every $\epsilon > 0$ there exists open $G \subseteq \mathbb{R}^n$ for which $E \subseteq G$ and $m^*(G \setminus E) < \epsilon$.

Proof. This condition is sufficient for measurability, because if we take a decreasing sequence $\{G_k\}_{k=1}^{\infty}$ of open supersets of E with $m^*(G_k \setminus E) \to 0$, then $m^*\left[\left(\bigcap_{k=1}^{\infty} G_k\right) \setminus E\right] = 0$, whence E is the intersection of a G_δ and the complement of a null set. Since each of those is measurable, so is E. For necessity, suppose first that $m(E) < \infty$. Then for any $\epsilon > 0$ one has a countable family $\{I_k\}_{k=1}^{\infty}$ of open intervals with $E \subseteq \bigcup_{k=1}^{\infty} I_k$ and $\sum_k m(I_k) < m(E) + \epsilon$. If we set $G = \bigcup_k I_k$ then we have $E \subseteq G$ and $m(G \setminus E) = m(G) - m(E) \le \sum_k m(I_k) - m(E) < \epsilon$. In any other case we may write $\mathbb{R}^n$ as the union of an increasing sequence $\{F_j\}_{j=1}^{\infty}$ of bounded measurable sets and, given $\epsilon > 0$, apply the reasoning just given to the sets $E_j = E \cap F_j$, obtaining open sets G_j with $E_j \subseteq G_j$ and $m(G_j \setminus E_j) < \epsilon/2^j$. Letting $G = \bigcup_{j=1}^{\infty} G_j$ we will then have $G \setminus E \subseteq \bigcup_{j=1}^{\infty} (G_j \setminus E_j)$ and therefore $m(G \setminus E) \le \sum_{j=1}^{\infty} m(G_j \setminus E_j) < \epsilon$, as we would desire.

Wheeden & Zygmund's **(3.15)** follows from the measurability of intervals and the fact that their boundaries are null sets: the measure = volume of a finite (or even countable: see **(3.24)**) union of nonoverlapping intervals is the same as that of their (disjoint) interiors, and therefore equals the sum of their respective volumes. The proof of their **(3.16)** uses only the definition of Lebesgue outer measure and thus is valid in our present context. **(3.17)** through **(3.20)** are already known to hold as a result of the Carathéodory development, and **(3.21)** holds for the same reasons that Wheeden & Zygmund give—*i.e.*, almost by definition.

In Wheeden & Zygmund's §3, their proof of **(3.22)** is valid in the present context because we just showed that their definition of measurable set characterized sets that were Carathéodory-measurable with respect to outer Lebesgue measure. We already have **(3.23)** through **(3.27)**, either from the Carathéodory development or because the proofs of Wheeden & Zygmund are valid in this context. The arguments of their §4 are also valid in this context, and of course we've been using **(3.30)** as a definition all along. So from now on—except for a question about linear transformations that we shall settle below—there is no need to distinguish between the two constructions of Lebesgue measure.

5. *Conclusion.*

The considerations about Lipschitz transformations in Wheeden & Zygmund's §5 require no emendation. We still have open, however, the question of the measure of non-rectangular parallelepipeds and the way in which Lebesgue measure transforms under linear transformations of $\mathbb{R}^n$.

It is a theorem of elementary linear algebra (proved by row-reducing a given matrix to diagonal form and observing that row operations can be performed through left-multiplication by elementary matrices) that any nonsingular $n \times n$ matrix over a field $\mathbb{K}$ can be written as a finite product of **elementary matrices**, which are matrices of three types: diagonal matrices $E_{(1)}$ (with nonzero elements on their diagonals, all but perhaps one of which equal 1), permutation matrices $E_{(2)}$ (that interchange two coördinate directions, leaving the others fixed), and "shear" or "transvection" matrices $E_{(3)}$ of the respective forms

$$E_{(1)} = \begin{pmatrix} 1 & \cdots & \cdots & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & \lambda & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \vdots & \vdots & \vdots & 1 \end{pmatrix}, \quad E_{(2)} = \begin{array}{cccccccc} & & & i & & j & & \\ & 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ i & 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ & \vdots & \cdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ j & 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ & \vdots & \cdots & \vdots & \cdots & \vdots & \ddots & \vdots \\ & 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{array},$$

$$E_{(3)} = \begin{array}{cccccccc} & & & & & j & & \\ & 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ i & 0 & \cdots & 1 & \cdots & \lambda & \cdots & 0 \\ & \vdots & \cdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ & 0 & \cdots & \cdots & \cdots & 1 & \cdots & 0 \\ & \vdots & \cdots & \vdots & \cdots & \vdots & \ddots & \vdots \\ & 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{array}$$

where in form $E_{(1)}$ one must have $\lambda \neq 0$, but in form $E_{(3)}$ any $\lambda \in \mathbb{K}$ is possible ($\det(E_{(3)}) = 1$ in all cases). It is easily verified that a diagonal matrix $E_{(1)}$ with all its diagonal elements but the i-th equal to 1 and the i-th equal to $\lambda > 0$ sends any interval $\prod_{j=1}^{n}[a_j, b_j]$ to an interval $\prod_{j<i}[a_j, b_j] \times [\lambda a_i, \lambda b_i] \times \prod_{j>i}[a_j, b_j]$ (empty products interpreted in the obvious way) whose volume is λ-times that of the original interval; if $\lambda < 0$ then the i-th interval becomes $[-\lambda b_i, -\lambda a_i]$ and the volume is multiplied by $|\lambda|$; either way it is multiplied by $|\det(E_{(1)})|$. A permutation matrix $E_{(2)}$ that interchanges exactly two coördinates has determinant -1, sends intervals to intervals, and does not alter the product that defines the volume of an interval. In a shear matrix of the form $E_{(3)}$, the interesting activity takes place in the ij-plane: the image of an interval $J = \prod_{k=1}^{n}[a_k, b_k]$ acted upon by such a matrix retains its product structure in the coördinates different from the i-th and j-th. It follows that it suffices to analyze the $n = 2$ case, because one can compute the measure of the image under the shear of an interval in $\mathbb{R}^n$ by covering its projection in the ij-plane with (finitely or countably many) intervals approximating its 2-dimensional measure, then taking their (Cartesian) products with the intervals $[a_k, b_k]$ of J in coördinate directions that are not affected by the shear.

The $i = 1$, $j = 2$ case is typical, so we may consider matrices of the form $E_{(3)} = \begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix}$, and by translation to the origin we may consider intervals of the form $[0, a) \times [0, b)$. Such an interval is the convex hull of the four points $(0, 0)$, $(a, 0)$, $(0, b)$ and (a, b) with the top and right-hand sides omitted, and so its image[3] is the convex hull of the four points $(0, 0)$, $(a, 0)$, $(\lambda b, b)$ and $(a + \lambda b, b)$ with the corresponding line segments omitted; the case in which $\lambda > 0$ is shown in Fig. (i), and it suffices to consider this case because one can change the sign of λ by conjugating with a reflection: $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 1 & -\lambda \\ 0 & 1 \end{pmatrix}$.

---

[3] Matrices are interpreted as acting on column vectors by left multiplication.

Fig. (i) Fig. (ii)

This parallelogram $E_{(3)}$-image of the original interval can be completed to the interval with extreme points $(0,0)$, $(a+\lambda b, 0)$, $(a+\lambda b, b)$, and $(0,b)$, as in Fig. (ii). Translating the triangle with vertices $(a,0)$, $(a+\lambda b, 0)$ and $(a+\lambda b, b)$ by $(-a, 0)$ will combine it and the triangle with vertices $(0,0)$, $(\lambda b, b)$ and $(0,b)$ into an interval $[0, \lambda b] \times [0, b]$ with volume = Lebesgue measure λb^2. If the measure of the parallelogram is P, we then have $\lambda b^2 + P = (a + \lambda b)b = ab + \lambda b^2$, so $P = ab$ and the Lebesgue measure of the $E_{(3)}$-image of the original interval is ab, the same as the measure of the original interval.

It follows that when the matrix A of a nonsingular linear transformation of $\mathbb{R}^n$ onto itself is written as a finite product of elementary matrices, the measure of the image $A[I]$ of an interval I is the product of the measure of the original interval with the product of absolute values of all the diagonal elements of the matrices of type $E_{(1)}$ that occur in the factorization. It then follows from the well-known multiplicative property of determinants that that factor is $|\det A|$ (and thus independent of the particular factorization of A). Since a countable cover by intervals $\{I_k\}$ of any set $E \subseteq \mathbb{R}^n$ will be transformed into a countable cover $\{A[I_k]\}$ of $A[E]$ by measurable sets (every interval is a K_σ, a property preserved by continuous mappings), we must have $m^*(A[E]) \leq |\det A| \cdot \sum_k m(I_k)$ and therefore (by taking infima) $m^*(A[E]) \leq |\det A| \cdot m^*(E)$. On the other hand, the same reasoning is applicable to A^{-1}, so $m^*(E) \leq |\det A^{-1}| \cdot m^*(A[E])$ or $|\det A| \cdot m^*(E) \leq m^*(A[E])$, and that establishes the relation $m^*(A[E]) = |\det A| \cdot m^*(E)$ in general.

A mundane but useful consequence of this fact is that, since any hyperplane in $\mathbb{R}^n$ can be mapped to a coördinate hyperplane by a suitably chosen nonsingular linear transformation, every proper affine subspace[4] of $\mathbb{R}^n$ is a null set for n-dimensional Lebesgue measure.

6. *Non-Lebesgue-measurable Subsets of* $\mathbb{R}$.

These objects can be constructed by explicit use of the choice axiom. Consider the cosets of the subgroup $(\mathbb{Q}, +)$ of $(\mathbb{R}, +)$. Each intersects $[0,1]$ in a nonempty (dense!) set, so the choice axiom will pick a complete set E of coset representatives for $\mathbb{Q}$ from $[0,1]$. E cannot be Lebesgue measurable. Every element $r \in \mathbb{R}$ has the form $r = e + q$ for uniquely determined $e \in E$ and $q \in \mathbb{Q}$, so $\mathbb{R} = \bigcup_{q \in \mathbb{Q}} (E + q)$ is a countable union of disjoint sets of outer measure measure equal to that of E. One cannot therefore have $m^*(E) = 0$. On the other hand, if $q \in [0,1] \cap \mathbb{Q}$ then $e + q \in [0,2]$, so if S is a finite set of rationals and $S \subseteq [0,1] \cap \mathbb{Q}$, then $\bigcup_{q \in S} (E + q) \subseteq [0,2]$. If E (and therefore every $E + q$) were measurable, then the measure of the (disjoint) union would be $(\#S) \cdot m(E) \leq 2$, which is clearly violated for $\#S$ sufficiently large. It follows that E cannot be measurable.

If the axiom of choice is negated, then it is not inconsistent to assume that every subset of $\mathbb{R}$ is Lebesgue-measurable. The first results in this direction were obtained by R. M. Solovay. See, *e.g.*, *A model of set theory in which every set of reals is Lebesgue measurable*, Ann. Math. **92** (1970), 1–56.

---

[4] An *affine subspace* is a translate of a vector subspace, or (equivalently) the solution set of a finite set of linear (perhaps inhomogeneous) equations.