1 Measure Theory

> The sets whose measure we can define by virtue of the preceding ideas we will call measurable sets; we do this without intending to imply that it is not possible to assign a measure to other sets.
>
> *E. Borel,* 1898

This chapter is devoted to the construction of Lebesgue measure in $\mathbb{R}^d$ and the study of the resulting class of measurable functions. After some preliminaries we pass to the first important definition, that of exterior measure for any subset E of $\mathbb{R}^d$. This is given in terms of approximations by unions of cubes that cover E. With this notion in hand we can define measurability and thus restrict consideration to those sets that are measurable. We then turn to the fundamental result: the collection of measurable sets is closed under complements and countable unions, and the measure is additive if the subsets in the union are disjoint.

The concept of measurable functions is a natural outgrowth of the idea of measurable sets. It stands in the same relation as the concept of continuous functions does to open (or closed) sets. But it has the important advantage that the class of measurable functions is closed under pointwise limits.

1 Preliminaries

We begin by discussing some elementary concepts which are basic to the theory developed below.

The main idea in calculating the "volume" or "measure" of a subset of $\mathbb{R}^d$ consists of approximating this set by unions of other sets whose geometry is simple and whose volumes are known. It is convenient to speak of "volume" when referring to sets in $\mathbb{R}^d$; but in reality it means "area" in the case $d = 2$ and "length" in the case $d = 1$. In the approach given here we shall use rectangles and cubes as the main building blocks of the theory: in $\mathbb{R}$ we use intervals, while in $\mathbb{R}^d$ we take products of intervals. In all dimensions rectangles are easy to manipulate and have a standard notion of volume that is given by taking the product of the length of all sides.

Next, we prove two simple theorems that highlight the importance of these rectangles in the geometry of open sets: in $\mathbb{R}$ every open set is a countable union of disjoint open intervals, while in $\mathbb{R}^d$, $d \geq 2$, every open set is "almost" the disjoint union of closed cubes, in the sense that only the boundaries of the cubes can overlap. These two theorems motivate the definition of exterior measure given later.

We shall use the following standard notation. A **point** $x \in \mathbb{R}^d$ consists of a d-tuple of real numbers

$$x = (x_1, x_2, \ldots, x_d), \qquad x_i \in \mathbb{R}, \text{ for } i = 1, \ldots, d.$$

Addition of points is componentwise, and so is multiplication by a real scalar. The **norm** of x is denoted by $|x|$ and is defined to be the standard Euclidean norm given by

$$|x| = \left(x_1^2 + \cdots + x_d^2\right)^{1/2}.$$

The **distance** between two points x and y is then simply $|x - y|$.

The **complement** of a set E in $\mathbb{R}^d$ is denoted by E^c and defined by

$$E^c = \{x \in \mathbb{R}^d : x \notin E\}.$$

If E and F are two subsets of $\mathbb{R}^d$, we denote the complement of F in E by

$$E - F = \{x \in \mathbb{R}^d : x \in E \text{ and } x \notin F\}.$$

The **distance** between two sets E and F is defined by

$$d(E, F) = \inf |x - y|,$$

where the infimum is taken over all $x \in E$ and $y \in F$.

Open, closed, and compact sets

The **open ball** in $\mathbb{R}^d$ centered at x and of radius r is defined by

$$B_r(x) = \{y \in \mathbb{R}^d : |y - x| < r\}.$$

A subset $E \subset \mathbb{R}^d$ is **open** if for every $x \in E$ there exists $r > 0$ with $B_r(x) \subset E$. By definition, a set is **closed** if its complement is open.

We note that any (not necessarily countable) union of open sets is open, while in general the intersection of only finitely many open sets is open. A similar statement holds for the class of closed sets, if one interchanges the roles of unions and intersections.

A set E is **bounded** if it is contained in some ball of finite radius. A bounded set is **compact** if it is also closed. Compact sets enjoy the Heine-Borel covering property:

- Assume E is compact, $E \subset \bigcup_\alpha \mathcal{O}_\alpha$, and each $\mathcal{O}_\alpha$ is open. Then there are finitely many of the open sets, $\mathcal{O}_{\alpha_1}, \mathcal{O}_{\alpha_2}, \ldots, \mathcal{O}_{\alpha_N}$, such that $E \subset \bigcup_{j=1}^N \mathcal{O}_{\alpha_j}$.

In words, *any* covering of a compact set by a collection of open sets contains a *finite* subcovering.

A point $x \in \mathbb{R}^d$ is a **limit point** of the set E if for every $r > 0$, the ball $B_r(x)$ contains points of E. This means that there are points in E which are arbitrarily close to x. An **isolated point** of E is a point $x \in E$ such that there exists an $r > 0$ where $B_r(x) \cap E$ is equal to $\{x\}$.

A point $x \in E$ is an **interior point** of E if there exists $r > 0$ such that $B_r(x) \subset E$. The set of all interior points of E is called the **interior** of E. Also, the **closure** $\overline{E}$ of the E consists of the union of E and all its limit points. The **boundary** of a set E, denoted by ∂E, is defined as $\overline{E} - E$.

Note that the closure of a set is a closed set; every point in E is a limit point of E; and a set is closed if and only if it contains all its limit points. Finally, a closed set E is **perfect** if E does not have any isolated points.

Rectangles and cubes

A (closed) **rectangle** R in $\mathbb{R}^d$ is given by the product of d one-dimensional closed and bounded intervals

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

where $a_j \leq b_j$ are real numbers, $j = 1, 2, \ldots, d$. In other words, we have

$$R = \{(x_1, \ldots, x_d) \in \mathbb{R}^d : \ a_j \leq x_j \leq b_j \quad \text{for all } j = 1, 2, \ldots, d\}.$$

We remark that in our definition, a rectangle is *closed* and has sides *parallel* to the coordinate axis. In $\mathbb{R}$, the rectangles are precisely the closed and bounded intervals, while in $\mathbb{R}^2$ they are the usual four-sided rectangles. In $\mathbb{R}^3$ they are the closed parallelepipeds.

We say that the lengths of the sides of the rectangle R are $b_1 - a_1, \ldots, b_d - a_d$. The **volume** of the rectangle R is denoted by $|R|$, and

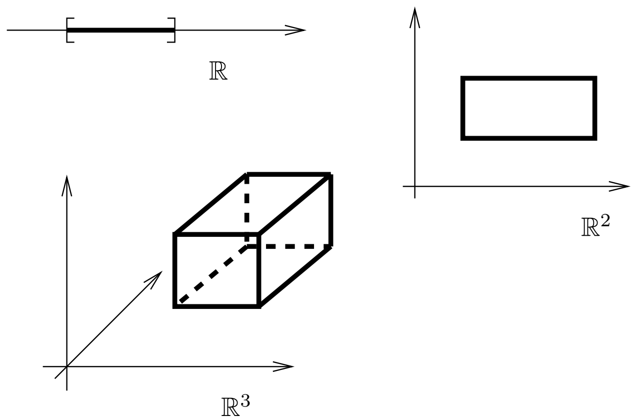

Figure 1. Rectangles in $\mathbb{R}^d$, $d = 1, 2, 3$

is defined to be

$$|R| = (b_1 - a_1) \cdots (b_d - a_d).$$

Of course, when $d = 1$ the "volume" equals length, and when $d = 2$ it equals area.

An open rectangle is the product of open intervals, and the interior of the rectangle R is then

$$(a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_d, b_d).$$

Also, a **cube** is a rectangle for which $b_1 - a_1 = b_2 - a_2 = \cdots = b_d - a_d$. So if $Q \subset \mathbb{R}^d$ is a cube of common side length ℓ, then $|Q| = \ell^d$.

A union of rectangles is said to be **almost disjoint** if the interiors of the rectangles are disjoint.

In this chapter, coverings by rectangles and cubes play a major role, so we isolate here two important lemmas.

Lemma 1.1 *If a rectangle is the almost disjoint union of finitely many other rectangles, say* $R = \bigcup_{k=1}^N R_k$, *then*

$$|R| = \sum_{k=1}^{N} |R_k|.$$

Proof. We consider the grid formed by extending indefinitely the sides of all rectangles $R_1, \ldots, R_N$. This construction yields finitely many rectangles $\tilde{R}_1, \ldots, \tilde{R}_M$, and a partition $J_1, \ldots, J_N$ of the integers between 1 and M, such that the unions

$$R = \bigcup_{j=1}^{M} \tilde{R}_j \quad \text{and} \quad R_k = \bigcup_{j \in J_k} \tilde{R}_j, \quad \text{for } k = 1, \ldots, N$$

are almost disjoint (see the illustration in Figure 2).

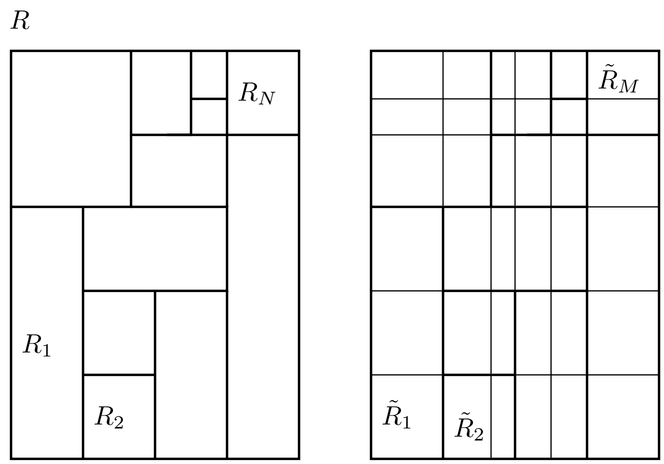

Figure 2. The grid formed by the rectangles R_k

For the rectangle R, for example, we see that $|R| = \sum_{j=1}^{M} |\tilde{R}_j|$, since the grid actually partitions the sides of R and each $\tilde{R}_j$ consists of taking products of the intervals in these partitions. Thus when adding the volumes of the $\tilde{R}_j$ we are summing the corresponding products of lengths of the intervals that arise. Since this also holds for the other rectangles $R_1, \ldots, R_N$, we conclude that

$$|R| = \sum_{j=1}^{M} |\tilde{R}_j| = \sum_{k=1}^{N} \sum_{j \in J_k} |\tilde{R}_j| = \sum_{k=1}^{N} |R_k|.$$

A slight modification of this argument then yields the following:

Lemma 1.2 *If $R, R_1, \ldots, R_N$ are rectangles, and $R \subset \bigcup_{k=1}^N R_k$, then*

$$|R| \leq \sum_{k=1}^{N} |R_k|.$$

The main idea consists of taking the grid formed by extending all sides of the rectangles $R, R_1, \ldots, R_N$, and noting that the sets corresponding to the J_k (in the above proof) need not be disjoint any more.

We now proceed to give a description of the structure of open sets in terms of cubes. We begin with the case of $\mathbb{R}$.

Theorem 1.3 *Every open subset $\mathcal{O}$ of $\mathbb{R}$ can be writen uniquely as a countable union of disjoint open intervals.*

Proof. For each $x \in \mathcal{O}$, let I_x denote the largest open interval containing x and contained in $\mathcal{O}$. More precisely, since $\mathcal{O}$ is open, x is contained in some small (non-trivial) interval, and therefore if

$$a_x = \inf\{a < x : (a, x) \subset \mathcal{O}\} \quad \text{and} \quad b_x = \sup\{b > x : (x, b) \subset \mathcal{O}\}$$

we must have $a_x < x < b_x$ (with possibly infinite values for a_x and b_x). If we now let $I_x = (a_x, b_x)$, then by construction we have $x \in I_x$ as well as $I_x \subset \mathcal{O}$. Hence

$$\mathcal{O} = \bigcup_{x \in \mathcal{O}} I_x.$$

Now suppose that two intervals I_x and I_y intersect. Then their union (which is also an open interval) is contained in $\mathcal{O}$ and contains x. Since I_x is maximal, we must have $(I_x \cup I_y) \subset I_x$, and similarly $(I_x \cup I_y) \subset I_y$. This can happen only if $I_x = I_y$; therefore, any two distinct intervals in the collection $\mathcal{I} = \{I_x\}_{x \in \mathcal{O}}$ must be disjoint. The proof will be complete once we have shown that there are only countably many distinct intervals in the collection $\mathcal{I}$. This, however, is easy to see, since every open interval I_x contains a rational number. Since different intervals are disjoint, they must contain distinct rationals, and therefore $\mathcal{I}$ is countable, as desired.

Naturally, if $\mathcal{O}$ is open and $\mathcal{O} = \bigcup_{j=1}^\infty I_j$, where the I_j's are disjoint open intervals, the measure of $\mathcal{O}$ ought to be $\sum_{j=1}^\infty |I_j|$. Since this representation is unique, we could take this as a definition of measure; we would then note that whenever $\mathcal{O}_1$ and $\mathcal{O}_2$ are open and disjoint, the measure of their union is the sum of their measures. Although this provides a natural notion of measure for an open set, it is not immediately clear how to generalize it to other sets in $\mathbb{R}$. Moreover, a similar approach in higher dimensions already encounters complications even when defining measures of open sets, since in this context the direct analogue of Theorem 1.3 is not valid (see Exercise 12). There is, however, a substitute result.

Theorem 1.4 *Every open subset $\mathcal{O}$ of $\mathbb{R}^d$, $d \geq 1$, can be written as a countable union of almost disjoint closed cubes.*

Proof. We must construct a countable collection $\mathcal{Q}$ of closed cubes whose interiors are disjoint, and so that $\mathcal{O} = \bigcup_{Q \in \mathcal{Q}} Q$.

As a first step, consider the grid in $\mathbb{R}^d$ formed by taking all closed cubes of side length 1 whose vertices have integer coordinates. In other words, we consider the natural grid of lines parallel to the axes, that is, the grid generated by the lattice $\mathbb{Z}^d$. We shall also use the grids formed by cubes of side length 2^{-N} obtained by successively bisecting the original grid.

We either accept or reject cubes in the initial grid as part of $\mathcal{Q}$ according to the following rule: if Q is entirely contained in $\mathcal{O}$ then we accept Q; if Q intersects both $\mathcal{O}$ and $\mathcal{O}^c$ then we tentatively accept it; and if Q is entirely contained in $\mathcal{O}^c$ then we reject it.

As a second step, we bisect the tentatively accepted cubes into 2^d cubes with side length $1/2$. We then repeat our procedure, by accepting the smaller cubes if they are completely contained in $\mathcal{O}$, tentatively accepting them if they intersect both $\mathcal{O}$ and $\mathcal{O}^c$, and rejecting them if they are contained in $\mathcal{O}^c$. Figure 3 illustrates these steps for an open set in $\mathbb{R}^2$.

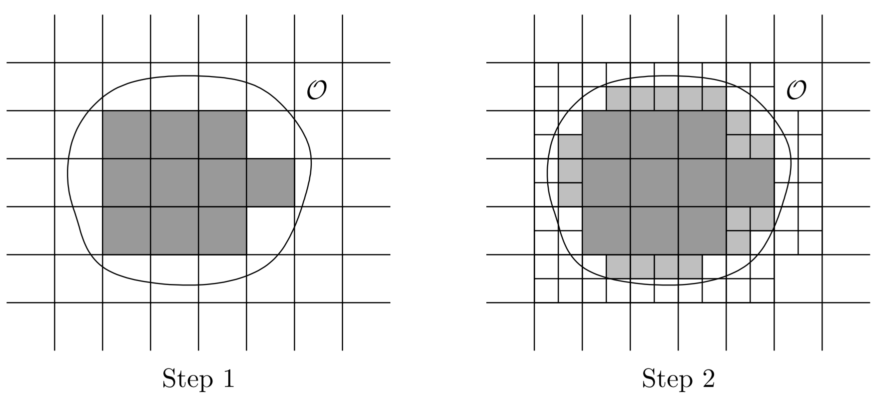

Figure 3. Decomposition of $\mathcal{O}$ into almost disjoint cubes

This procedure is then repeated indefinitely, and (by construction) the resulting collection $\mathcal{Q}$ of all accepted cubes is countable and consists of almost disjoint cubes. To see why their union is all of $\mathcal{O}$, we note that given $x \in \mathcal{O}$ there exists a cube of side length 2^{-N} (obtained from successive bisections of the original grid) that contains x and that is entirely contained in $\mathcal{O}$. Either this cube has been accepted, or it is contained in a cube that has been previously accepted. This shows that the union of all cubes in $\mathcal{Q}$ covers $\mathcal{O}$.

Once again, if $\mathcal{O} = \bigcup_{j=1}^{\infty} R_j$ where the rectangles R_j are almost disjoint, it is reasonable to assign to $\mathcal{O}$ the measure $\sum_{j=1}^{\infty} |R_j|$. This is natural since the volume of the boundary of each rectangle should be 0, and the overlap of the rectangles should not contribute to the volume of $\mathcal{O}$. We note, however, that the above decomposition into cubes is not unique, and it is not immediate that the sum is independent of this decomposition. So in $\mathbb{R}^d$, with $d \geq 2$, the notion of volume or area, even for open sets, is more subtle.

The general theory developed in the next section actually yields a notion of volume that is consistent with the decompositions of open sets of the previous two theorems, and applies to all dimensions. Before we come to that, we discuss an important example in $\mathbb{R}$.

The Cantor set

The Cantor set plays a prominent role in set theory and in analysis in general. It and its variants provide a rich source of enlightening examples.

We begin with the closed unit interval $C_0 = [0, 1]$ and let C_1 denote the set obtained from deleting the middle third open interval from $[0, 1]$, that is,

$$C_1 = [0, 1/3] \cup [2/3, 1].$$

Next, we repeat this procedure for each sub-interval of C_1; that is, we delete the middle third open interval. At the second stage we get

$$C_2 = [0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1].$$

We repeat this process for each sub-interval of C_2, and so on (Figure 4).

This procedure yields a sequence C_k, $k = 0, 1, 2, \ldots$ of compact sets with

$$C_0 \supset C_1 \supset C_2 \supset \cdots \supset C_k \supset C_{k+1} \supset \cdots .$$

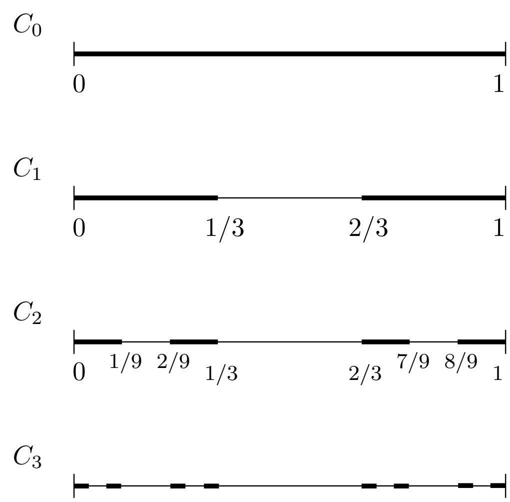

Figure 4. Construction of the Cantor set

The **Cantor set** $\mathcal{C}$ is by definition the intersection of all C_k's:

$$\mathcal{C} = \bigcap_{k=0}^{\infty} C_k.$$

The set $\mathcal{C}$ is not empty, since all end-points of the intervals in C_k (all k) belong to $\mathcal{C}$.

Despite its simple construction, the Cantor set enjoys many interesting topological and analytical properties. For instance, $\mathcal{C}$ is closed and bounded, hence compact. Also, $\mathcal{C}$ is totally disconnected: given any $x, y \in \mathcal{C}$ there exists $z \notin \mathcal{C}$ that lies between x and y. Finally, $\mathcal{C}$ is perfect: it has no isolated points (Exercise 1).

Next, we turn our attention to the question of determining the "size" of $\mathcal{C}$. This is a delicate problem, one that may be approached from different angles depending on the notion of size we adopt. For instance, in terms of cardinality the Cantor set is rather large: it is not countable. Since it can be mapped to the interval $[0, 1]$, the Cantor set has the cardinality of the continuum (Exercise 2).

However, from the point of view of "length" the size of $\mathcal{C}$ is small. Roughly speaking, the Cantor set has length zero, and this follows from the following intuitive argument: the set $\mathcal{C}$ is covered by sets C_k whose lengths go to zero. Indeed, C_k is a disjoint union of 2^k intervals of length

3^{-k}, making the total length of C_k equal to $(2/3)^k$. But $\mathcal{C} \subset C_k$ for all k, and $(2/3)^k \to 0$ as k tends to infinity. We shall define a notion of measure and make this argument precise in the next section.

2 The exterior measure

The notion of exterior measure is the first of two important concepts needed to develop a theory of measure. We begin with the definition and basic properties of exterior measure. Loosely speaking, the exterior measure m_* assigns to *any* subset of $\mathbb{R}^d$ a first notion of size; various examples show that this notion coincides with our earlier intuition. However, the exterior measure lacks the desirable property of additivity when taking the union of disjoint sets. We remedy this problem in the next section, where we discuss in detail the other key concept of measure theory, the notion of measurable sets.

The exterior measure, as the name indicates, attempts to describe the volume of a set E by approximating it from the outside. The set E is covered by cubes, and if the covering gets finer, with fewer cubes overlapping, the volume of E should be close to the sum of the volumes of the cubes.

The precise definition is as follows: if E is *any* subset of $\mathbb{R}^d$, the **exterior measure**[1] of E is

$$m_*(E) = \inf \sum_{j=1}^{\infty} |Q_j|, \tag{1}$$

where the infimum is taken over all countable coverings $E \subset \bigcup_{j=1}^{\infty} Q_j$ by closed cubes. The exterior measure is always non-negative but could be infinite, so that in general we have $0 \leq m_*(E) \leq \infty$, and therefore takes values in the extended positive numbers.

We make some preliminary remarks about the definition of the exterior measure given by (1).

(i) It is important to note that it would not suffice to allow *finite* sums in the definition of $m_*(E)$. The quantity that would be obtained if one considered only coverings of E by finite unions of cubes is in general larger than $m_*(E)$. (See Exercise 14.)

(ii) One can, however, replace the coverings by cubes, with coverings by rectangles; or with coverings by balls. That the former alternative

[1] Some authors use the term **outer measure** instead of exterior measure.

yields the same exterior measure is quite direct. (See Exercise 15.) The equivalence with the latter is more subtle. (See Exercise 26 in Chapter 3.)

We begin our investigation of this new notion by providing examples of sets whose exterior measures can be calculated, and we check that the latter matches our intuitive idea of volume (length in one dimension, area in two dimensions, etc.)

EXAMPLE 1. The exterior measure of a point is zero. This is clear once we observe that a point is a cube with volume zero, and which covers itself. Of course the exterior measure of the empty set is also zero.

EXAMPLE 2. The exterior measure of a closed cube is equal to its volume. Indeed, suppose Q is a closed cube in $\mathbb{R}^d$. Since Q covers itself, we must have $m_*(Q) \leq |Q|$. Therefore, it suffices to prove the reverse inequality.

We consider an arbitrary covering $Q \subset \bigcup_{j=1}^{\infty} Q_j$ by cubes, and note that it suffices to prove that

$$|Q| \leq \sum_{j=1}^{\infty} |Q_j|. \tag{2}$$

For a fixed $\epsilon > 0$ we choose for each j an open cube S_j which contains Q_j, and such that $|S_j| \leq (1+\epsilon)|Q_j|$. From the open covering $\bigcup_{j=1}^{\infty} S_j$ of the compact set Q, we may select a finite subcovering which, after possibly renumbering the rectangles, we may write as $Q \subset \bigcup_{j=1}^{N} S_j$. Taking the closure of the cubes S_j, we may apply Lemma 1.2 to conclude that $|Q| \leq \sum_{j=1}^{N} |S_j|$. Consequently,

$$|Q| \leq (1+\epsilon) \sum_{j=1}^{N} |Q_j| \leq (1+\epsilon) \sum_{j=1}^{\infty} |Q_j|.$$

Since ϵ is arbitrary, we find that the inequality (2) holds; thus $|Q| \leq m_*(Q)$, as desired.

EXAMPLE 3. If Q is an open cube, the result $m_*(Q) = |Q|$ still holds. Since Q is covered by its closure $\overline{Q}$, and $|\overline{Q}| = |Q|$, we immediately see that $m_*(Q) \leq |Q|$. To prove the reverse inequality, we note that if Q_0 is a closed cube contained in Q, then $m_*(Q_0) \leq m_*(Q)$, since any covering of Q by a countable number of closed cubes is also a covering of Q_0 (see Observation 1 below). Hence $|Q_0| \leq m_*(Q)$, and since we can choose Q_0 with a volume as close as we wish to $|Q|$, we must have $|Q| \leq m_*(Q)$.

EXAMPLE 4. The exterior measure of a rectangle R is equal to its volume. Indeed, arguing as in Example 2, we see that $|R| \le m_*(R)$. To obtain the reverse inequality, consider a grid in $\mathbb{R}^d$ formed by cubes of side length $1/k$. Then, if $\mathcal{Q}$ consists of the (finite) collection of all cubes entirely contained in R, and $\mathcal{Q}'$ the (finite) collection of all cubes that intersect the complement of R, we first note that $R \subset \bigcup_{Q \in (\mathcal{Q} \cup \mathcal{Q}')} Q$. Also, a simple argument yields

$$\sum_{Q \in \mathcal{Q}} |Q| \le |R|.$$

Moreover, there are $O(k^{d-1})$ cubes[2] in $\mathcal{Q}'$, and these cubes have volume k^{-d}, so that $\sum_{Q \in \mathcal{Q}'} |Q| = O(1/k)$. Hence

$$\sum_{Q \in (\mathcal{Q} \cup \mathcal{Q}')} |Q| \le |R| + O(1/k),$$

and letting k tend to infinity yields $m_*(R) \le |R|$, as desired.

EXAMPLE 5. The exterior measure of $\mathbb{R}^d$ is infinite. This follows from the fact that any covering of $\mathbb{R}^d$ is also a covering of any cube $Q \subset \mathbb{R}^d$, hence $|Q| \le m_*(\mathbb{R}^d)$. Since Q can have arbitrarily large volume, we must have $m_*(\mathbb{R}^d) = \infty$.

EXAMPLE 6. The Cantor set $\mathcal{C}$ has exterior measure 0. From the construction of $\mathcal{C}$, we know that $\mathcal{C} \subset C_k$, where each C_k is a disjoint union of 2^k closed intervals, each of length 3^{-k}. Consequently, $m_*(\mathcal{C}) \le (2/3)^k$ for all k, hence $m_*(\mathcal{C}) = 0$.

Properties of the exterior measure

The previous examples and comments provide some intuition underlying the definition of exterior measure. Here, we turn to the further study of m_* and prove five properties of exterior measure that are needed in what follows.

First, we record the following remark that is immediate from the definition of m_*:

[2] We remind the reader of the notation $f(x) = O(g(x))$, which means that $|f(x)| \le C|g(x)|$ for some constant C and all x in a given range. In this particular example, there are fewer than Ck^{d-1} cubes in question, as $k \to \infty$.

- For every $\epsilon > 0$, there exists a covering $E \subset \bigcup_{j=1}^{\infty} Q_j$ with

$$\sum_{j=1}^{\infty} m_*(Q_j) \leq m_*(E) + \epsilon.$$

The relevant properties of exterior measure are listed in a series of observations.

Observation 1 (Monotonicity) *If $E_1 \subset E_2$, then $m_*(E_1) \leq m_*(E_2)$.*

This follows once we observe that any covering of E_2 by a countable collection of cubes is also a covering of E_1.

In particular, monotonicity implies that every bounded subset of $\mathbb{R}^d$ has finite exterior measure.

Observation 2 (Countable sub-additivity) *If $E = \bigcup_{j=1}^{\infty} E_j$, then $m_*(E) \leq \sum_{j=1}^{\infty} m_*(E_j)$.*

First, we may assume that each $m_*(E_j) < \infty$, for otherwise the inequality clearly holds. For any $\epsilon > 0$, the definition of the exterior measure yields for each j a covering $E_j \subset \bigcup_{k=1}^{\infty} Q_{k,j}$ by closed cubes with

$$\sum_{k=1}^{\infty} |Q_{k,j}| \leq m_*(E_j) + \frac{\epsilon}{2^j}.$$

Then, $E \subset \bigcup_{j,k=1}^{\infty} Q_{k,j}$ is a covering of E by closed cubes, and therefore

$$\begin{aligned} m_*(E) \leq \sum_{j,k} |Q_{k,j}| &= \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} |Q_{k,j}| \\ &\leq \sum_{j=1}^{\infty} \left(m_*(E_j) + \frac{\epsilon}{2^j} \right) \\ &= \sum_{j=1}^{\infty} m_*(E_j) + \epsilon. \end{aligned}$$

Since this holds true for every $\epsilon > 0$, the second observation is proved.

Observation 3 *If $E \subset \mathbb{R}^d$, then $m_*(E) = \inf m_*(\mathcal{O})$, where the infimum is taken over all open sets $\mathcal{O}$ containing E.*

By monotonicity, it is clear that the inequality $m_*(E) \le \inf m_*(\mathcal{O})$ holds. For the reverse inequality, let $\epsilon > 0$ and choose cubes Q_j such that $E \subset \bigcup_{j=1}^\infty Q_j$, with

$$\sum_{j=1}^\infty |Q_j| \le m_*(E) + \frac{\epsilon}{2}.$$

Let Q_j^0 denote an open cube containing Q_j, and such that $|Q_j^0| \le |Q_j| + \epsilon/2^{j+1}$. Then $\mathcal{O} = \bigcup_{j=1}^\infty Q_j^0$ is open, and by Observation 2

$$\begin{aligned} m_*(\mathcal{O}) \le \sum_{j=1}^\infty m_*(Q_j^0) &= \sum_{j=1}^\infty |Q_j^0| \\ &\le \sum_{j=1}^\infty \left(|Q_j| + \frac{\epsilon}{2^{j+1}} \right) \\ &\le \sum_{j=1}^\infty |Q_j| + \frac{\epsilon}{2} \\ &\le m_*(E) + \epsilon. \end{aligned}$$

Hence $\inf m_*(\mathcal{O}) \le m_*(E)$, as was to be shown.

Observation 4 *If $E = E_1 \cup E_2$, and $d(E_1, E_2) > 0$, then*

$$m_*(E) = m_*(E_1) + m_*(E_2).$$

By Observation 2, we already know that $m_*(E) \le m_*(E_1) + m_*(E_2)$, so it suffices to prove the reverse inequality. To this end, we first select δ such that $d(E_1, E_2) > \delta > 0$. Next, we choose a covering $E \subset \bigcup_{j=1}^\infty Q_j$ by closed cubes, with $\sum_{j=1}^\infty |Q_j| \le m_*(E) + \epsilon$. We may, after subdividing the cubes Q_j, assume that each Q_j has a diameter less than δ. In this case, each Q_j can intersect at most one of the two sets E_1 or E_2. If we denote by J_1 and J_2 the sets of those indices j for which Q_j intersects E_1 and E_2, respectively, then $J_1 \cap J_2$ is empty, and we have

$$E_1 \subset \bigcup_{j \in J_1} Q_j \quad \text{as well as} \quad E_2 \subset \bigcup_{j \in J_2} Q_j.$$

Therefore,

$$\begin{aligned} m_*(E_1) + m_*(E_2) &\leq \sum_{j \in J_1} |Q_j| + \sum_{j \in J_2} |Q_j| \\ &\leq \sum_{j=1}^{\infty} |Q_j| \\ &\leq m_*(E) + \epsilon. \end{aligned}$$

Since ϵ is arbitrary, the proof of Observation 4 is complete.

Observation 5 *If a set E is the countable union of almost disjoint cubes $E = \bigcup_{j=1}^{\infty} Q_j$, then*

$$m_*(E) = \sum_{j=1}^{\infty} |Q_j|.$$

Let $\tilde{Q}_j$ denote a cube strictly contained in Q_j such that $|Q_j| \leq |\tilde{Q}_j| + \epsilon/2^j$, where ϵ is arbitrary but fixed. Then, for every N, the cubes $\tilde{Q}_1, \tilde{Q}_2, \ldots, \tilde{Q}_N$ are disjoint, hence at a finite distance from one another, and repeated applications of Observation 4 imply

$$m_*\left(\bigcup_{j=1}^{N} \tilde{Q}_j\right) = \sum_{j=1}^{N} |\tilde{Q}_j| \geq \sum_{j=1}^{N} \left(|Q_j| - \epsilon/2^j\right).$$

Since $\bigcup_{j=1}^{N} \tilde{Q}_j \subset E$, we conclude that for every integer N,

$$m_*(E) \geq \sum_{j=1}^{N} |Q_j| - \epsilon.$$

In the limit as N tends to infinity we deduce $\sum_{j=1}^{\infty} |Q_j| \leq m_*(E) + \epsilon$ for every $\epsilon > 0$, hence $\sum_{j=1}^{\infty} |Q_j| \leq m_*(E)$. Therefore, combined with Observation 2, our result proves that we have equality.

This last property shows that if a set can be decomposed into almost disjoint cubes, its exterior measure equals the sum of the volumes of the cubes. In particular, by Theorem 1.4 we see that the exterior measure of an open set equals the sum of the volumes of the cubes in a decomposition, and this coincides with our initial guess. Moreover, this also yields a proof that the sum is independent of the decomposition.

One can see from this that the volumes of simple sets that are calculated by elementary calculus agree with their exterior measure. This assertion can be proved most easily once we have developed the requisite tools in integration theory. (See Chapter 2.) In particular, we can then verify that the exterior measure of a ball (either open or closed) equals its volume.

Despite observations 4 and 5, one *cannot* conclude in general that if $E_1 \cup E_2$ is a disjoint union of subsets of $\mathbb{R}^d$, then

$$m_*(E_1 \cup E_2) = m_*(E_1) + m_*(E_2). \tag{3}$$

In fact (3) holds when the sets in question are not highly irregular or "pathological" but are measurable in the sense described below.

3 Measurable sets and the Lebesgue measure

The notion of measurability isolates a collection of subsets in $\mathbb{R}^d$ for which the exterior measure satisfies all our desired properties, including additivity (and in fact countable additivity) for disjoint unions of sets.

There are a number of different ways of defining measurability, but these all turn out to be equivalent. Probably the simplest and most intuitive is the following: A subset E of $\mathbb{R}^d$ is **Lebesgue measurable**, or simply **measurable**, if for any $\epsilon > 0$ there exists an open set $\mathcal{O}$ with $E \subset \mathcal{O}$ and

$$m_*(\mathcal{O} - E) \leq \epsilon.$$

This should be compared to Observation 3, which holds for *all* sets E.

If E is measurable, we define its **Lebesgue measure** (or **measure**) $m(E)$ by

$$m(E) = m_*(E).$$

Clearly, the Lebesgue measure inherits all the features contained in Observations 1 - 5 of the exterior measure.

Immediately from the definition, we find:

Property 1 *Every open set in $\mathbb{R}^d$ is measurable.*

Our immediate goal now is to gather various further properties of measurable sets. In particular, we shall prove that the collection of measurable sets behave well under the various operations of set theory: countable unions, countable intersections, and complements.

Property 2 *If $m_*(E) = 0$, then E is measurable. In particular, if F is a subset of a set of exterior measure 0, then F is measurable.*

By Observation 3 of the exterior measure, for every $\epsilon > 0$ there exists an open set $\mathcal{O}$ with $E \subset \mathcal{O}$ and $m_*(\mathcal{O}) \leq \epsilon$. Since $(\mathcal{O} - E) \subset \mathcal{O}$, monotonicity implies $m_*(\mathcal{O} - E) \leq \epsilon$, as desired.

As a consequence of this property, we deduce that the Cantor set $\mathcal{C}$ in Example 6 is measurable and has measure 0.

Property 3 *A countable union of measurable sets is measurable.*

Suppose $E = \bigcup_{j=1}^\infty E_j$, where each E_j is measurable. Given $\epsilon > 0$, we may choose for each j an open set $\mathcal{O}_j$ with $E_j \subset \mathcal{O}_j$ and $m_*(\mathcal{O}_j - E_j) \leq \epsilon/2^j$. Then the union $\mathcal{O} = \bigcup_{j=1}^\infty \mathcal{O}_j$ is open, $E \subset \mathcal{O}$, and $(\mathcal{O} - E) \subset \bigcup_{j=1}^\infty (\mathcal{O}_j - E_j)$, so monotonicity and sub-additivity of the exterior measure imply

$$m_*(\mathcal{O} - E) \leq \sum_{j=1}^\infty m_*(\mathcal{O}_j - E_j) \leq \epsilon.$$

Property 4 *Closed sets are measurable.*

First, we observe that it suffices to prove that compact sets are measurable. Indeed, any closed set F can be written as the union of compact sets, say $F = \bigcup_{k=1}^\infty F \cap B_k$, where B_k denotes the closed ball of radius k centered at the origin; then Property 3 applies.

So, suppose F is compact (so that in particular $m_*(F) < \infty$), and let $\epsilon > 0$. By Observation 3 we can select an open set $\mathcal{O}$ with $F \subset \mathcal{O}$ and $m_*(\mathcal{O}) \leq m_*(F) + \epsilon$. Since F is closed, the difference $\mathcal{O} - F$ is open, and by Theorem 1.4 we may write this difference as a countable union of almost disjoint cubes

$$\mathcal{O} - F = \bigcup_{j=1}^\infty Q_j.$$

For a fixed N, the finite union $K = \bigcup_{j=1}^N Q_j$ is compact; therefore $d(K, F) > 0$ (we isolate this little fact in a lemma below). Since $(K \cup F) \subset \mathcal{O}$, Observations 1, 4, and 5 of the exterior measure imply

$$\begin{aligned} m_*(\mathcal{O}) &\geq m_*(F) + m_*(K) \\ &= m_*(F) + \sum_{j=1}^N m_*(Q_j). \end{aligned}$$

Hence $\sum_{j=1}^{N} m_*(Q_j) \leq m_*(\mathcal{O}) - m_*(F) \leq \epsilon$, and this also holds in the limit as N tends to infinity. Invoking the sub-additivity property of the exterior measure finally yields

$$m_*(\mathcal{O} - F) \leq \sum_{j=1}^{\infty} m_*(Q_j) \leq \epsilon,$$

as desired.

We digress briefly to complete the above argument by proving the following.

Lemma 3.1 *If F is closed, K is compact, and these sets are disjoint, then $d(F, K) > 0$.*

Proof. Since F is closed, for each point $x \in K$, there exists $\delta_x > 0$ so that $d(x, F) > 3\delta_x$. Since $\bigcup_{x \in K} B_{2\delta_x}(x)$ covers K, and K is compact, we may find a subcover, which we denote by $\bigcup_{j=1}^{N} B_{2\delta_j}(x_j)$. If we let $\delta = \min(\delta_1, \ldots, \delta_N)$, then we must have $d(K, F) \geq \delta > 0$. Indeed, if $x \in K$ and $y \in F$, then for some j we have $|x_j - x| \leq 2\delta_j$, and by construction $|y - x_j| \geq 3\delta_j$. Therefore

$$|y - x| \geq |y - x_j| - |x_j - x| \geq 3\delta_j - 2\delta_j \geq \delta,$$

and the lemma is proved.

Property 5 *The complement of a measurable set is measurable.*

If E is measurable, then for every positive integer n we may choose an open set $\mathcal{O}_n$ with $E \subset \mathcal{O}_n$ and $m_*(\mathcal{O}_n - E) \leq 1/n$. The complement $\mathcal{O}_n^c$ is closed, hence measurable, which implies that the union $S = \bigcup_{n=1}^{\infty} \mathcal{O}_n^c$ is also measurable by Property 3. Now we simply note that $S \subset E^c$, and

$$(E^c - S) \subset (\mathcal{O}_n - E),$$

such that $m_*(E^c - S) \leq 1/n$ for all n. Therefore, $m_*(E^c - S) = 0$, and $E^c - S$ is measurable by Property 2. Therefore E^c is measurable since it is the union of two measurable sets, namely S and $(E^c - S)$.

Property 6 *A countable intersection of measurable sets is measurable.*

This follows from Properties 3 and 5, since

$$\bigcap_{j=1}^{\infty} E_j = \left(\bigcup_{j=1}^{\infty} E_j^c \right)^c.$$

In conclusion, we find that the family of measurable sets is closed under the familiar operations of set theory. We point out that we have shown more than simply closure with respect to finite unions and intersections: we have proved that the collection of measurable sets is closed under *countable* unions and intersections. This passage from finite operations to infinite ones is crucial in the context of analysis. We emphasize, however, that the operations of *uncountable* unions or intersections are not permissible when dealing with measurable sets!

Theorem 3.2 *If $E_1, E_2, \ldots,$ are disjoint measurable sets, and $E = \bigcup_{j=1}^{\infty} E_j$, then*

$$m(E) = \sum_{j=1}^{\infty} m(E_j).$$

Proof. First, we assume further that each E_j is bounded. Then, for each j, by applying the definition of measurability to E_j^c, we can choose a closed subset F_j of E_j with $m_*(E_j - F_j) \leq \epsilon/2^j$. For each fixed N, the sets $F_1, \ldots, F_N$ are compact and disjoint, so that $m\left(\bigcup_{j=1}^{N} F_j\right) = \sum_{j=1}^{N} m(F_j)$. Since $\bigcup_{j=1}^{N} F_j \subset E$, we must have

$$m(E) \geq \sum_{j=1}^{N} m(F_j) \geq \sum_{j=1}^{N} m(E_j) - \epsilon.$$

Letting N tend to infinity, since ϵ was arbitrary we find that

$$m(E) \geq \sum_{j=1}^{\infty} m(E_j).$$

Since the reverse inequality always holds (by sub-additivity in Observation 2), this concludes the proof when each E_j is bounded.

In the general case, we select any sequence of cubes $\{Q_k\}_{k=1}^{\infty}$ that increases to $\mathbb{R}^d$, in the sense that $Q_k \subset Q_{k+1}$ for all $k \geq 1$ and $\bigcup_{k=1}^{\infty} Q_k = \mathbb{R}^d$. We then let $S_1 = Q_1$ and $S_k = Q_k - Q_{k-1}$ for $k \geq 2$. If we define measurable sets by $E_{j,k} = E_j \cap S_k$, then

$$E = \bigcup_{j,k} E_{j,k}.$$

The union above is disjoint and every $E_{j,k}$ is bounded. Moreover $E_j = \bigcup_{k=1}^{\infty} E_{j,k}$, and this union is also disjoint. Putting these facts together, and using what has already been proved, we obtain

$$m(E) = \sum_{j,k} m(E_{j,k}) = \sum_j \sum_k m(E_{j,k}) = \sum_j m(E_j),$$

as claimed.

With this, the countable additivity of the Lebesgue measure on measurable sets has been established. This result provides the necessary connection between the following:

- our primitive notion of volume given by the exterior measure,
- the more refined idea of measurable sets, and
- the countably infinite operations allowed on these sets.

We make two definitions to state succinctly some further consequences.

If $E_1, E_2, \dots$ is a countable collection of subsets of $\mathbb{R}^d$ that increases to E in the sense that $E_k \subset E_{k+1}$ for all k, and $E = \bigcup_{k=1}^\infty E_k$, then we write $E_k \nearrow E$.

Similarly, if $E_1, E_2, \dots$ decreases to E in the sense that $E_k \supset E_{k+1}$ for all k, and $E = \bigcap_{k=1}^\infty E_k$, we write $E_k \searrow E$.

Corollary 3.3 *Suppose $E_1, E_2, \dots$ are measurable subsets of $\mathbb{R}^d$.*

(i) *If $E_k \nearrow E$, then $m(E) = \lim_{N\to\infty} m(E_N)$.*

(ii) *If $E_k \searrow E$ and $m(E_k) < \infty$ for some k, then*

$$m(E) = \lim_{N\to\infty} m(E_N).$$

Proof. For the first part, let $G_1 = E_1$, $G_2 = E_2 - E_1$, and in general $G_k = E_k - E_{k-1}$ for $k \geq 2$. By their construction, the sets G_k are measurable, disjoint, and $E = \bigcup_{k=1}^\infty G_k$. Hence

$$m(E) = \sum_{k=1}^\infty m(G_k) = \lim_{N\to\infty} \sum_{k=1}^N m(G_k) = \lim_{N\to\infty} m\left(\bigcup_{k=1}^N G_k\right),$$

and since $\bigcup_{k=1}^N G_k = E_N$ we get the desired limit.

For the second part, we may clearly assume that $m(E_1) < \infty$. Let $G_k = E_k - E_{k+1}$ for each k, so that

$$E_1 = E \cup \bigcup_{k=1}^\infty G_k$$

is a disjoint union of measurable sets. As a result, we find that

$$\begin{aligned} m(E_1) &= m(E) + \lim_{N\to\infty} \sum_{k=1}^{N-1} (m(E_k) - m(E_{k+1})) \\ &= m(E) + m(E_1) - \lim_{N\to\infty} m(E_N). \end{aligned}$$

Hence, since $m(E_1) < \infty$, we see that $m(E) = \lim_{N\to\infty} m(E_N)$, and the proof is complete.

The reader should note that the second conclusion may fail without the assumption that $m(E_k) < \infty$ for some k. This is shown by the simple example when $E_n = (n, \infty) \subset \mathbb{R}$, for all n.

What follows provides an important geometric and analytic insight into the nature of measurable sets, in terms of their relation to open and closed sets. Its thrust is that, in effect, an arbitrary measurable set can be well approximated by the open sets that contain it, and alternatively, by the closed sets it contains.

Theorem 3.4 *Suppose E is a measurable subset of $\mathbb{R}^d$. Then, for every $\epsilon > 0$:*

(i) *There exists an open set $\mathcal{O}$ with $E \subset \mathcal{O}$ and $m(\mathcal{O} - E) \leq \epsilon$.*

(ii) *There exists a closed set F with $F \subset E$ and $m(E - F) \leq \epsilon$.*

(iii) *If $m(E)$ is finite, there exists a compact set K with $K \subset E$ and $m(E - K) \leq \epsilon$.*

(iv) *If $m(E)$ is finite, there exists a finite union $F = \bigcup_{j=1}^N Q_j$ of closed cubes such that*

$$m(E \triangle F) \leq \epsilon.$$

The notation $E \triangle F$ stands for the **symmetric difference** between the sets E and F, defined by $E \triangle F = (E - F) \cup (F - E)$, which consists of those points that belong to only one of the two sets E or F.

Proof. Part (i) is just the definition of measurability. For the second part, we know that E^c is measurable, so there exists an open set $\mathcal{O}$ with $E^c \subset \mathcal{O}$ and $m(\mathcal{O} - E^c) \leq \epsilon$. If we let $F = \mathcal{O}^c$, then F is closed, $F \subset E$, and $E - F = \mathcal{O} - E^c$. Hence $m(E - F) \leq \epsilon$ as desired.

For (iii), we first pick a closed set F so that $F \subset E$ and $m(E - F) \leq \epsilon/2$. For each n, we let B_n denote the ball centered at the origin of radius

n, and define compact sets $K_n = F \cap B_n$. Then $E - K_n$ is a sequence of measurable sets that decreases to $E - F$, and since $m(E) < \infty$, we conclude that for all large n one has $m(E - K_n) \leq \epsilon$.

For the last part, choose a family of closed cubes $\{Q_j\}_{j=1}^{\infty}$ so that

$$E \subset \bigcup_{j=1}^{\infty} Q_j \quad \text{and} \quad \sum_{j=1}^{\infty} |Q_j| \leq m(E) + \epsilon/2.$$

Since $m(E) < \infty$, the series converges and there exists $N > 0$ such that $\sum_{j=N+1}^{\infty} |Q_j| < \epsilon/2$. If $F = \bigcup_{j=1}^{N} Q_j$, then

$$\begin{aligned} m(E \triangle F) &= m(E - F) + m(F - E) \\ &\leq m\left(\bigcup_{j=N+1}^{\infty} Q_j\right) + m\left(\bigcup_{j=1}^{\infty} Q_j - E\right) \\ &\leq \sum_{j=N+1}^{\infty} |Q_j| + \sum_{j=1}^{\infty} |Q_j| - m(E) \\ &\leq \epsilon. \end{aligned}$$

Invariance properties of Lebesgue measure

A crucial property of Lebesgue measure in $\mathbb{R}^d$ is its translation-invariance, which can be stated as follows: if E is a measurable set and $h \in \mathbb{R}^d$, then the set $E_h = E + h = \{x + h : x \in E\}$ is also measurable, and $m(E + h) = m(E)$. With the observation that this holds for the special case when E is a cube, one passes to the exterior measure of arbitrary sets E, and sees from the definition of m_* given in Section 2 that $m_*(E_h) = m_*(E)$. To prove the measurability of E_h under the assumption that E is measurable, we note that if $\mathcal{O}$ is open, $\mathcal{O} \supset E$, and $m_*(\mathcal{O} - E) < \epsilon$, then $\mathcal{O}_h$ is open, $\mathcal{O}_h \supset E_h$, and $m_*(\mathcal{O}_h - E_h) < \epsilon$.

In the same way one can prove the relative dilation-invariance of Lebesgue measure. Suppose $\delta > 0$, and denote by δE the set $\{\delta x : x \in E\}$. We can then assert that δE is measurable whenever E is, and $m(\delta E) = \delta^d m(E)$. One can also easily see that Lebesgue measure is reflection-invariant. That is, whenever E is measurable, so is $-E = \{-x : x \in E\}$ and $m(-E) = m(E)$.

Other invariance properties of Lebesgue measure are in Exercise 7 and 8, and Problem 4 of Chapter 2.

σ-algebras and Borel sets

A **σ-algebra** of sets is a collection of subsets of $\mathbb{R}^d$ that is closed under countable unions, countable intersections, and complements.

The collection of all subsets of $\mathbb{R}^d$ is of course a σ-algebra. A more interesting and relevant example consists of all measurable sets in $\mathbb{R}^d$, which we have just shown also forms a σ-algebra.

Another σ-algebra, which plays a vital role in analysis, is the **Borel σ-algebra** in $\mathbb{R}^d$, denoted by $\mathcal{B}_{\mathbb{R}^d}$, which by definition is the smallest σ-algebra that contains all open sets. Elements of this σ-algebra are called **Borel sets**.

The definition of the Borel σ-algebra will be meaningful once we have defined the term "smallest," and shown that such a σ-algebra exists and is unique. The term "smallest" means that if $\mathcal{S}$ is any σ-algebra that contains all open sets in $\mathbb{R}^d$, then necessarily $\mathcal{B}_{\mathbb{R}^d} \subset \mathcal{S}$. Since we observe that any intersection (not necessarily countable) of σ-algebras is again a σ-algebra, we may define $\mathcal{B}_{\mathbb{R}^d}$ as the intersection of all σ-algebras that contain the open sets. This shows the existence and uniqueness of the Borel σ-algebra.

Since open sets are measurable, we conclude that the Borel σ-algebra is contained in the σ-algebra of measurable sets. Naturally, we may ask if this inclusion is strict: do there exist Lebesgue measurable sets which are not Borel sets? The answer is "yes." (See Exercise 35.)

From the point of view of the Borel sets, the Lebesgue sets arise as the **completion** of the σ-algebra of Borel sets, that is, by adjoining all subsets of Borel sets of measure zero. This is an immediate consequence of Corollary 3.5 below.

Starting with the open and closed sets, which are the simplest Borel sets, one could try to list the Borel sets in order of their complexity. Next in order would come countable intersections of open sets; such sets are called $\boldsymbol{G_\delta}$ sets. Alternatively, one could consider their complements, the countable union of closed sets, called the $\boldsymbol{F_\sigma}$ sets.[3]

Corollary 3.5 *A subset E of $\mathbb{R}^d$ is measurable*

(i) *if and only if E differs from a G_δ by a set of measure zero,*

(ii) *if and only if E differs from an F_σ by a set of measure zero.*

Proof. Clearly E is measurable whenever it satisfies either (i) or (ii), since the F_σ, G_δ, and sets of measure zero are measurable.

[3]The terminology G_δ comes from German "Gebiete" and "Durschnitt"; F_σ comes from French "fermé" and "somme."

Conversely, if E is measurable, then for each integer $n \geq 1$ we may select an open set $\mathcal{O}_n$ that contains E, and such that $m(\mathcal{O}_n - E) \leq 1/n$. Then $S = \bigcap_{n=1}^{\infty} \mathcal{O}_n$ is a G_δ that contains E, and $(S - E) \subset (\mathcal{O}_n - E)$ for all n. Therefore $m(S - E) \leq 1/n$ for all n; hence $S - E$ has exterior measure zero, and is therefore measurable.

For the second implication, we simply apply part (ii) of Theorem 3.4 with $\epsilon = 1/n$, and take the union of the resulting closed sets.

Construction of a non-measurable set

Are all subsets of $\mathbb{R}^d$ measurable? In this section, we answer this question when $d = 1$ by constructing a subset of $\mathbb{R}$ which is *not* measurable.[4] This justifies the conclusion that a satisfactory theory of measure cannot encompass *all* subsets of $\mathbb{R}$.

The construction of a non-measurable set $\mathcal{N}$ uses the axiom of choice, and rests on a simple equivalence relation among real numbers in $[0, 1]$.

We write $x \sim y$ whenever $x - y$ is rational, and note that this is an equivalence relation since the following properties hold:

- $x \sim x$ for every $x \in [0, 1]$
- if $x \sim y$, then $y \sim x$
- if $x \sim y$ and $y \sim z$, then $x \sim z$.

Two equivalence classes either are disjoint or coincide, and $[0, 1]$ is the disjoint union of all equivalence classes, which we write as

$$[0, 1] = \bigcup_{\alpha} \mathcal{E}_\alpha.$$

Now we construct the set $\mathcal{N}$ by choosing exactly one element x_α from each $\mathcal{E}_\alpha$, and setting $\mathcal{N} = \{x_\alpha\}$. This (seemingly obvious) step requires further comment, which we postpone until after the proof of the following theorem.

Theorem 3.6 *The set $\mathcal{N}$ is not measurable.*

The proof is by contradiction, so we assume that $\mathcal{N}$ is measurable. Let $\{r_k\}_{k=1}^{\infty}$ be an enumeration of all the rationals in $[-1, 1]$, and consider the translates

$$\mathcal{N}_k = \mathcal{N} + r_k.$$

---

[4] The existence of such a set in $\mathbb{R}$ implies the existence of corresponding non-measurable subsets of $\mathbb{R}^d$ for each d, as a consequence of Proposition 3.4 in the next chapter.

We claim that the sets $\mathcal{N}_k$ are disjoint, and

$$[0,1] \subset \bigcup_{k=1}^{\infty} \mathcal{N}_k \subset [-1,2]. \tag{4}$$

To see why these sets are disjoint, suppose that the intersection $\mathcal{N}_k \cap \mathcal{N}_{k'}$ is non-empty. Then there exist rationals $r_k \neq r'_k$ and α and β with $x_\alpha + r_k = x_\beta + r_{k'}$; hence

$$x_\alpha - x_\beta = r_{k'} - r_k.$$

Consequently $\alpha \neq \beta$ and $x_\alpha - x_\beta$ is rational; hence $x_\alpha \sim x_\beta$, which contradicts the fact that $\mathcal{N}$ contains only *one* representative of each equivalence class.

The second inclusion is straightforward since each $\mathcal{N}_k$ is contained in $[-1,2]$ by construction. Finally, if $x \in [0,1]$, then $x \sim x_\alpha$ for some α, and therefore $x - x_\alpha = r_k$ for some k. Hence $x \in \mathcal{N}_k$, and the first inclusion holds.

Now we may conclude the proof of the theorem. If $\mathcal{N}$ were measurable, then so would be $\mathcal{N}_k$ for all k, and since the union $\bigcup_{k=1}^{\infty} \mathcal{N}_k$ is disjoint, the inclusions in (4) yield

$$1 \leq \sum_{k=1}^{\infty} m(\mathcal{N}_k) \leq 3.$$

Since $\mathcal{N}_k$ is a translate of $\mathcal{N}$, we must have $m(\mathcal{N}_k) = m(\mathcal{N})$ for all k. Consequently,

$$1 \leq \sum_{k=1}^{\infty} m(\mathcal{N}) \leq 3.$$

This is the desired contradiction, since neither $m(\mathcal{N}) = 0$ nor $m(\mathcal{N}) > 0$ is possible.

Axiom of choice

That the construction of the set $\mathcal{N}$ is possible is based on the following general proposition.

- Suppose E is a set and $\{E_\alpha\}$ is a collection of non-empty subsets of E. (The indexing set of α's is not assumed to be countable.) Then there is a function $\alpha \mapsto x_\alpha$ (a "choice function") such that $x_\alpha \in E_\alpha$, for all α.

In this general form this assertion is known as the **axiom of choice**. This axiom occurs (at least implicitly) in many proofs in mathematics, but because of its seeming intuitive self-evidence, its significance was not at first understood. The initial realization of the importance of this axiom was in its use to prove a famous assertion of Cantor, the **well-ordering principle**. This proposition (sometimes referred to as "transfinite induction") can be formulated as follows.

A set E is **linearly ordered** if there is a binary relation $\leq$ such that:

(a) $x \leq x$ for all $x \in E$.

(b) If $x, y \in E$ are distinct, then either $x \leq y$ or $y \leq x$ (but not both).

(c) If $x \leq y$ and $y \leq z$, then $x \leq z$.

We say that a set E can be **well-ordered** if it can be linearly ordered in such a way that *every* non-empty subset $A \subset E$ has a smallest element in that ordering (that is, an element $x_0 \in A$ such that $x_0 \leq x$ for any other $x \in A$).

A simple example of a well-ordered set is $\mathbb{Z}^+$, the positive integers with their usual ordering. The fact that $\mathbb{Z}^+$ is well-ordered is an essential part of the usual (finite) induction principle. More generally, the well-ordering principle states:

- Any set E can be well-ordered.

It is in fact nearly obvious that the well-ordering principle implies the axiom of choice: if we well-order E, we can choose x_α to be the smallest element in E_α, and in this way we have constructed the required choice function. It is also true, but not as easy to show, that the converse implication holds, namely that the axiom of choice implies the well-ordering principle. (See Problem 6 for another equivalent formulation of the Axiom of Choice.)

We shall follow the common practice of assuming the axiom of choice (and hence the validity of the well-ordering principle).[5] However, we should point out that while the axiom of choice seems self-evident the well-ordering principle leads quickly to some baffling conclusions: one only needs to spend a little time trying to imagine what a well-ordering of the reals might look like!

[5]It can be proved that in an appropriate formulation of the axioms of set theory, the axiom of choice is independent of the other axioms; thus we are free to accept its validity.

4 Measurable functions

With the notion of measurable sets in hand, we now turn our attention to the objects that lie at the heart of integration theory: measurable functions.

The starting point is the notion of a **characteristic function** of a set E, which is defined by

$$\chi_E(x) = \begin{cases} 1 & \text{if } x \in E, \\ 0 & \text{if } x \notin E. \end{cases}$$

The next step is to pass to the functions that are the building blocks of integration theory. For the Riemann integral it is in effect the class of **step functions**, with each given as a finite sum

$$f = \sum_{k=1}^{N} a_k \chi_{R_k}, \tag{5}$$

where each R_k is a rectangle, and the a_k are constants.

However, for the Lebesgue integral we need a more general notion, as we shall see in the next chapter. A **simple function** is a finite sum

$$f = \sum_{k=1}^{N} a_k \chi_{E_k} \tag{6}$$

where each E_k is a measurable set of finite measure, and the a_k are constants.

4.1 Definition and basic properties

We begin by considering only real-valued functions f on $\mathbb{R}^d$, which we allow to take on the infinite values $+\infty$ and $-\infty$, so that $f(x)$ belongs to the extended real numbers

$$-\infty \leq f(x) \leq \infty.$$

We shall say that f is **finite-valued** if $-\infty < f(x) < \infty$ for all x. In the theory that follows, and the many applications of it, we shall almost always find ourselves in situations where a function takes on infinite values on at most a set of measure zero.

A function f defined on a measurable subset E of $\mathbb{R}^d$ is **measurable**, if for all $a \in \mathbb{R}$, the set

$$f^{-1}([-\infty, a)) = \{x \in E : f(x) < a\}$$

is measurable. To simplify our notation, we shall often denote the set $\{x \in E : f(x) < a\}$ simply by $\{f < a\}$ whenever no confusion is possible.

First, we note that there are many equivalent definitions of measurable functions. For example, we may require instead that the inverse image of closed intervals be measurable. Indeed, to prove that f is measurable if and only if $\{x : f(x) \leq a\} = \{f \leq a\}$ is measurable for every a, we note that in one direction, one has

$$\{f \leq a\} = \bigcap_{k=1}^{\infty} \{f < a + 1/k\},$$

and recall that the countable intersection of measurable sets is measurable. For the other direction, we observe that

$$\{f < a\} = \bigcup_{k=1}^{\infty} \{f \leq a - 1/k\}.$$

Similarly, f is measurable if and only if $\{f \geq a\}$ (or $\{f > a\}$) is measurable for every a. In the first case this is immediate from our definition and the fact that $\{f \geq a\}$ is the complement of $\{f < a\}$, and in the second case this follows from what we have just proved and the fact that $\{f \leq a\} = \{f > a\}^c$. A simple consequence is that $-f$ is measurable whenever f is measurable.

In the same way, one can show that if f is finite-valued, then it is measurable if and only if the sets $\{a < f < b\}$ are measurable for every $a, b \in \mathbb{R}$. Similar conclusions hold for whichever combination of strict or weak inequalities one chooses. For example, if f is finite-valued, then it is measurable if and only if $\{a \leq f < b\}$ for all $a, b \in \mathbb{R}$. By the same arguments one sees the following:

Property 1 *The finite-valued function f is measurable if and only if $f^{-1}(\mathcal{O})$ is measurable for every open set $\mathcal{O}$, and if and only if $f^{-1}(F)$ is measurable for every closed set F.*

Note that this property also applies to extended-valued functions, if we make the additional hypothesis that both $f^{-1}(\infty)$ and $f^{-1}(-\infty)$ are measurable sets.

Property 2 *If f is continuous on $\mathbb{R}^d$, then f is measurable. If f is measurable and finite-valued, and Φ is continuous, then $\Phi \circ f$ is measurable.*

In fact, Φ is continuous, so $\Phi^{-1}((-\infty, a))$ is an open set $\mathcal{O}$, and hence $(\Phi \circ f)^{-1}((-\infty, a)) = f^{-1}(\mathcal{O})$ is measurable.

It should be noted, however, that in general it is not true that $f \circ \Phi$ is measurable whenever f is measurable and Φ is continuous. See Exercise 35.

Property 3 *Suppose $\{f_n\}_{n=1}^{\infty}$ is a sequence of measurable functions. Then*

$$\sup_n f_n(x), \qquad \inf_n f_n(x), \qquad \limsup_{n\to\infty}, f_n(x) \qquad \textit{and} \qquad \liminf_{n\to\infty} f_n(x)$$

are measurable.

Proving that $\sup_n f_n$ is measurable requires noting that $\{\sup_n f_n > a\} = \bigcup_n \{f_n > a\}$. This also yields the result for $\inf_n f_n(x)$, since this quantity equals $-\sup_n(-f_n(x))$.

The result for the limsup and liminf also follows from the two observations

$$\limsup_{n\to\infty} f_n(x) = \inf_k\{\sup_{n\geq k} f_n\} \qquad \text{and} \qquad \liminf_{n\to\infty} f_n(x) = \sup_k\{\inf_{n\geq k} f_n\}.$$

Property 4 *If $\{f_n\}_{n=1}^{\infty}$ is a collection of measurable functions, and*

$$\lim_{n\to\infty} f_n(x) = f(x),$$

then f is measurable.

Since $f(x) = \limsup_{n\to\infty} f_n(x) = \liminf_{n\to\infty} f_n(x)$, this property is a consequence of property 3.

Property 5 *If f and g are measurable, then*

(i) *The integer powers f^k, $k \geq 1$ are measurable.*

(ii) *$f + g$ and fg are measurable if both f and g are finite-valued.*

For (i) we simply note that if k is odd, then $\{f^k > a\} = \{f > a^{1/k}\}$, and if k is even and $a \geq 0$, then $\{f^k > a\} = \{f > a^{1/k}\} \cup \{f < -a^{1/k}\}$.

For (ii), we first see that $f + g$ is measurable because

$$\{f + g > a\} = \bigcup_{r\in\mathbb{Q}} \{f > a - r\} \cap \{g > r\},$$

with $\mathbb{Q}$ denoting the rationals.

Finally, fg is measurable because of the previous results and the fact that

$$fg = \frac{1}{4}[(f+g)^2 + (f-g)^2].$$

We shall say that two functions f and g defined on a set E are equal **almost everywhere**, and write

$$f(x) = g(x) \qquad \text{a.e. } x \in E,$$

if the set $\{x \in E : f(x) \neq g(x)\}$ has measure zero. We sometimes abbreviate this by saying that $f = g$ a.e. More generally, a property or statement is said to hold almost everywhere (a.e.) if it is true except on a set of measure zero.

One sees easily that if f is measurable and $f = g$ a.e., then g is measurable. This follows at once from the fact that $\{f < a\}$ and $\{g < a\}$ differ by a set of measure zero. Moreover, all the properties above can be relaxed to conditions holding almost everywhere. For instance, if $\{f_n\}_{n=1}^{\infty}$ is a collection of measurable functions, and

$$\lim_{n\to\infty} f_n(x) = f(x) \qquad \text{a.e.},$$

then f is measurable.

Note that if f and g are defined almost everywhere on a measurable subset $E \subset \mathbb{R}^d$, then the functions $f + g$ and fg can only be defined on the intersection of the domains of f and g. Since the union of two sets of measure zero has again measure zero, $f + g$ is defined almost everywhere on E. We summarize this discussion as follows.

Property 6 *Suppose f is measurable, and $f(x) = g(x)$ for a.e. x. Then g is measurable.*

In this light, Property 5 (ii) also holds when f and g are finite-valued almost everywhere.

4.2 Approximation by simple functions or step functions

The theorems in this section are all of the same nature and provide further insight in the structure of measurable functions. We begin by approximating pointwise, non-negative measurable functions by simple functions.

Theorem 4.1 *Suppose f is a non-negative measurable function on $\mathbb{R}^d$. Then there exists an increasing sequence of non-negative simple functions $\{\varphi_k\}_{k=1}^{\infty}$ that converges pointwise to f, namely,*

$$\varphi_k(x) \leq \varphi_{k+1}(x) \quad \textit{and} \quad \lim_{k\to\infty} \varphi_k(x) = f(x), \ \textit{for all } x.$$

Proof. We begin first with a truncation. For $k \geq 1$, let Q_k denote the cube centered at the origin and of side length k. Then we define

$$F_k(x) = \begin{cases} f(x) & \text{if } x \in Q_k \text{ and } f(x) \leq k, \\ k & \text{if } x \in Q_k \text{ and } f(x) > k, \\ 0 & \text{otherwise.} \end{cases}$$

Then, $F_k(x) \to f(x)$ as k tends to infinity for all x. Now, we partition the range of F_k, namely $[0, k]$, as follows. For fixed $k, j \geq 1$, we define

$$E_{\ell,j} = \left\{ x \in Q_k : \frac{\ell}{j} < F_k(x) \leq \frac{\ell+1}{j} \right\}, \qquad \text{for } 0 \leq \ell < kj.$$

Then we may form

$$F_{k,j}(x) = \sum_{\ell} \frac{\ell}{j}\, \chi_{E_{\ell,j}}(x).$$

Each $F_{k,j}$ is a simple function that satisfies $0 \leq F_k(x) - F_{k,j}(x) \leq 1/j$ for all x. If we now choose $j = k$, and let $\varphi_k = F_{k,k}$, then we see that $0 \leq F_k(x) - \varphi_k(x) \leq 1/k$ for all x, and $\{\varphi_k\}$ satisfies all the desired properties.

Note that the result holds for non-negative functions that are extended-valued, if the limit $+\infty$ is allowed. We now drop the assumption that f is non-negative, and also allow the extended limit $-\infty$.

Theorem 4.2 *Suppose f is measurable on $\mathbb{R}^d$. Then there exists a sequence of simple functions $\{\varphi_k\}_{k=1}^{\infty}$ that satisfies*

$$|\varphi_k(x)| \leq |\varphi_{k+1}(x)| \quad \textit{and} \quad \lim_{k\to\infty} \varphi_k(x) = f(x), \ \textit{for all } x.$$

In particular, we have $|\varphi_k(x)| \leq |f(x)|$ for all x and k.

Proof. We use the following decomposition of the function f: $f(x) = f^+(x) - f^-(x)$, where

$$f^+(x) = \max(f(x), 0) \quad \text{and} \quad f^-(x) = \max(-f(x), 0).$$

Since both f^+ and f^- are non-negative, the previous theorem yields two increasing sequences of non-negative simple functions $\{\varphi_k^{(1)}(x)\}_{k=1}^\infty$ and $\{\varphi_k^{(2)}(x)\}_{k=1}^\infty$ which converge pointwise to f^+ and f^-, respectively. Then, if we let

$$\varphi_k(x) = \varphi_k^{(1)}(x) - \varphi_k^{(2)}(x),$$

we see that $\varphi_k(x)$ converges to $f(x)$ for all x. Finally, the sequence $\{|\varphi_k|\}$ is increasing because the definition of f^+, f^- and the properties of $\varphi_k^{(1)}$ and $\varphi_k^{(2)}$ imply that

$$|\varphi_k(x)| = \varphi_k^{(1)}(x) + \varphi_k^{(2)}(x).$$

We may now go one step further, and approximate by step functions. Here, in general, the convergence may hold only almost everywhere.

Theorem 4.3 *Suppose f is measurable on $\mathbb{R}^d$. Then there exists a sequence of step functions $\{\psi_k\}_{k=1}^\infty$ that converges pointwise to $f(x)$ for almost every x.*

Proof. By the previous result, it suffices to show that if E is a measurable set with finite measure, then $f = \chi_E$ can be approximated by step functions. To this end, we recall part (iv) of Theorem 3.4, which states that for every ϵ there exist cubes $Q_1, \ldots, Q_N$ such that $m(E \triangle \bigcup_{j=1}^N Q_j) \leq \epsilon$. By considering the grid formed by extending the sides of these cubes, we see that there exist almost disjoint rectangles $\tilde{R}_1, \ldots, \tilde{R}_M$ such that $\bigcup_{j=1}^N Q_j = \bigcup_{j=1}^M \tilde{R}_j$. By taking rectangles R_j contained in $\tilde{R}_j$, and slightly smaller in size, we find a collection of disjoint rectangles that satisfy $m(E \triangle \bigcup_{j=1}^M R_j) \leq 2\epsilon$. Therefore

$$f(x) = \sum_{j=1}^M \chi_{R_j}(x),$$

except possibly on a set of measure $\leq 2\epsilon$. Consequently, for every $k \geq 1$, there exists a step function $\psi_k(x)$ such that if

$$E_k = \{x : \ f(x) \neq \psi_k(x)\},$$

then $m(E_k) \leq 2^{-k}$. If we let $F_K = \bigcup_{j=K+1}^\infty E_j$ and $F = \bigcap_{K=1}^\infty F_K$, then $m(F) = 0$ since $m(F_K) \leq 2^{-K}$, and $\psi_k(x) \to f(x)$ for all x in the complement of F, which is the desired result.

4.3 Littlewood's three principles

Although the notions of measurable sets and measurable functions represent new tools, we should not overlook their relation to the older concepts they replaced. Littlewood aptly summarized these connections in the form of three principles that provide a useful intuitive guide in the initial study of the theory.

(i) Every set is nearly a finite union of intervals.

(ii) Every function is nearly continuous.

(iii) Every convergent sequence is nearly uniformly convergent.

The sets and functions referred to above are of course assumed to be measurable. The catch is in the word "nearly," which has to be understood appropriately in each context. A precise version of the first principle appears in part (iv) of Theorem 3.4. An exact formulation of the third principle is given in the following important result.

Theorem 4.4 (Egorov) *Suppose $\{f_k\}_{k=1}^{\infty}$ is a sequence of measurable functions defined on a measurable set E with $m(E) < \infty$, and assume that $f_k \to f$ a.e on E. Given $\epsilon > 0$, we can find a closed set $A_\epsilon \subset E$ such that $m(E - A_\epsilon) \le \epsilon$ and $f_k \to f$ uniformly on A_ϵ.*

Proof. We may assume without loss of generality that $f_k(x) \to f(x)$ for every $x \in E$. For each pair of non-negative integers n and k, let

$$E_k^n = \{x \in E : |f_j(x) - f(x)| < 1/n, \quad \text{for all } j > k\}.$$

Now fix n and note that $E_k^n \subset E_{k+1}^n$, and $E_k^n \nearrow E$ as k tends to infinity. By Corollary 3.3, we find that there exists k_n such that $m(E - E_{k_n}^n) < 1/2^n$. By construction, we then have

$$|f_j(x) - f(x)| < 1/n \qquad \text{whenever } j > k_n \text{ and } x \in E_{k_n}^n.$$

We choose N so that $\sum_{n=N}^{\infty} 2^{-n} < \epsilon/2$, and let

$$\tilde{A}_\epsilon = \bigcap_{n \ge N} E_{k_n}^n.$$

We first observe that

$$m(E - \tilde{A}_\epsilon) \le \sum_{n=N}^{\infty} m(E - E_{k_n}^n) < \epsilon/2.$$

Next, if $\delta > 0$, we choose $n \geq N$ such that $1/n < \delta$, and note that $x \in \tilde{A}_\epsilon$ implies $x \in E^n_{k_n}$. We see therefore that $|f_j(x) - f(x)| < \delta$ whenever $j > k_n$. Hence f_k converges uniformly to f on $\tilde{A}_\epsilon$.

Finally, using Theorem 3.4 choose a closed subset $A_\epsilon \subset \tilde{A}_\epsilon$ with $m(\tilde{A}_\epsilon - A_\epsilon) < \epsilon/2$. As a result, we have $m(E - A_\epsilon) < \epsilon$ and the theorem is proved.

The next theorem attests to the validity of the second of Littlewood's principle.

Theorem 4.5 (Lusin) *Suppose f is measurable and finite valued on E with E of finite measure. Then for every $\epsilon > 0$ there exists a closed set F_ϵ, with*

$$F_\epsilon \subset E, \quad \text{and} \quad m(E - F_\epsilon) \leq \epsilon$$

and such that $f|_{F_\epsilon}$ is continuous.

By $f|_{F_\epsilon}$ we mean the restriction of f to the set F_ϵ. The conclusion of the theorem states that if f is viewed as a function defined only on F_ϵ, then f is continuous. However, the theorem does *not* make the stronger assertion that the function f defined on E is continuous at the points of F_ϵ.

Proof. Let f_n be a sequence of step functions so that $f_n \to f$ a.e. Then we may find sets E_n so that $m(E_n) < 1/2^n$ and f_n is continuous outside E_n. By Egorov's theorem, we may find a set $A_{\epsilon/3}$ on which $f_n \to f$ uniformly and $m(E - A_{\epsilon/3}) \leq \epsilon/3$. Then we consider

$$F' = A_{\epsilon/3} - \bigcup_{n \geq N} E_n$$

for N so large that $\sum_{n \geq N} 1/2^n < \epsilon/3$. Now for every $n \geq N$ the function f_n is continuous on F'; thus f (being the uniform limit of $\{f_n\}$) is also continuous on F'. To finish the proof, we merely need to approximate the set F' by a closed set $F_\epsilon \subset F'$ such that $m(F' - F_\epsilon) < \epsilon/3$.

5* The Brunn-Minkowski inequality

Since addition and multiplication by scalars are basic features of vector spaces, it is not surprising that properties of these operations arise in a fundamental way in the theory of Lebesgue measure on $\mathbb{R}^d$. We have already discussed in this connection the translation-invariance and relative dilation-invariance of Lebesgue measure. Here we come to the study of the sum of two measurable sets A and B, defined as

$$A + B = \{x \in \mathbb{R}^d : \; x = x' + x'' \quad \text{with } x' \in A \text{ and } x'' \in B\}.$$

This notion is of importance in a number of questions, in particular in the theory of convex sets; we shall apply it to the isoperimetric problem in Chapter 3.

In this regard the first (admittedly vague) question we can pose is whether one can establish any general estimate for the measure of $A + B$ in terms of the measures of A and B (assuming that these three sets are measurable). We can see easily that it is not possible to obtain an upper bound for $m(A + B)$ in terms of $m(A)$ and $m(B)$. Indeed, simple examples show that we may have $m(A) = m(B) = 0$ while $m(A + B) > 0$. (See Exercise 20.)

In the converse direction one might ask for a general estimate of the form

$$m(A + B)^\alpha \geq c_\alpha \left(m(A)^\alpha + m(B)^\alpha\right),$$

where α is a positive number and the constant c_α is independent of A and B. Clearly, the best one can hope for is $c_\alpha = 1$. The role of the exponent α can be understood by considering **convex sets**. Such sets A are defined by the property that whenever x and y are in A then the line segment joining them, $\{xt + y(1 - t) : \; 0 \leq t \leq 1\}$, also belongs to A. If we recall the definition $\lambda A = \{\lambda x, \; x \in A\}$ for $\lambda > 0$, we note that whenever A is convex, then $A + \lambda A = (1 + \lambda)A$. However, $m((1 + \lambda)A) = (1 + \lambda)^d m(A)$, and thus the presumed inequality can hold only if $(1 + \lambda)^{d\alpha} \geq 1 + \lambda^{d\alpha}$, for all $\lambda > 0$. Now

$$(a + b)^\gamma \geq a^\gamma + b^\gamma \qquad \text{if } \gamma \geq 1 \text{ and } a, b \geq 0, \tag{7}$$

while the reverse inequality holds if $0 \leq \gamma \leq 1$. (See Exercise 38.) This yields $\alpha \geq 1/d$. Moreover, (7) shows that the inequality with the exponent $1/d$ implies the corresponding inequality with $\alpha \geq 1/d$, and so we are naturally led to the inequality

$$m(A + B)^{1/d} \geq m(A)^{1/d} + m(B)^{1/d}. \tag{8}$$

Before proceeding with the proof of (8), we need to mention a technical impediment that arises. While we may assume that A and B are measurable, it does not necessarily follow that then $A + B$ is measurable. (See Exercise 13 in the next chapter.) However it is easily seen that this difficulty does not occur when, for example, A and B are closed sets, or when one of them is open. (See Exercise 19.)

With the above considerations in mind we can state the main result.

Theorem 5.1 *Suppose A and B are measurable sets in $\mathbb{R}^d$ and their sum $A + B$ is also measurable. Then the inequality* (8) *holds.*

Let us first check (8) when A and B are rectangles with side lengths $\{a_j\}_{j=1}^d$ and $\{b_j\}_{j=1}^d$, respectively. Then (8) becomes

$$\left(\prod_{j=1}^{d}(a_j + b_j)\right)^{1/d} \geq \left(\prod_{j=1}^{d} a_j\right)^{1/d} + \left(\prod_{j=1}^{d} b_j\right)^{1/d}, \tag{9}$$

which by homogeneity we can reduce to the special case where $a_j + b_j = 1$ for each j. In fact, notice that if we replace a_j, b_j by $\lambda_j a_j, \lambda_j b_j$, with $\lambda_j > 0$, then both sides of (9) are multiplied by $(\lambda_1\lambda_2 \cdots \lambda_d)^{1/d}$. We then need only choose $\lambda_j = (a_j + b_j)^{-1}$. With this reduction, the inequality (9) is an immediate consequence of the arithmetic-geometric inequality (Exercise 39)

$$\frac{1}{d}\sum_{j=1}^{d} x_j \geq \left(\prod_{j=1}^{d} x_j\right)^{1/d}, \qquad \text{for all } x_j \geq 0:$$

we add the two inequalities that result when we set $x_j = a_j$ and $x_j = b_j$, respectively.

We next turn to the case when each A and B are the union of finitely many rectangles whose interiors are disjoint. We shall prove (8) in this case by induction on the total number of rectangles in A and B. We denote this number by n. Here it is important to note that the desired inequality is unchanged when we translate A and B independently. In fact, replacing A by $A + h$ and B by $B + h'$ replaces $A + B$ by $A + B + h + h'$, and thus the corresponding measures remain the same. We now choose a pair of disjoint rectangles R_1 and R_2 in the collection making up A, and we note that they can be separated by a coordinate hyperplane. Thus we may assume that for some j, after translation by an appropriate h, R_1 lies in $A_- = A \cap \{x_j \leq 0\}$, and R_2 in $A_+ = A \cap \{0 \leq x_j\}$. Observe also that both A_+ and A_- contain at least one less rectangle than A does, and $A = A_- \cup A_+$.

We next translate B so that $B_- = B \cap \{x_j \leq 0\}$ and $B_+ = B \cap \{x_j \geq 0\}$ satisfy

$$\frac{m(B_\pm)}{m(B)} = \frac{m(A_\pm)}{m(A)}.$$

However, $A+B \supset (A_+ + B_+) \cup (A_- + B_-)$, and the union on the right is essentially disjoint, since the two parts lie in different half-spaces. Moreover, the total number of rectangles in either A_+ and B_+, or A_- and B_- is also less than n. Thus the induction hypothesis applies and

$$\begin{aligned}
m(A+B) &\geq m(A_+ + B_+) + m(A_- + B_-) \\
&\geq \left(m(A_+)^{1/d} + m(B_+)^{1/d}\right)^d + \left(m(A_-)^{1/d} + m(B_-)^{1/d}\right)^d \\
&= m(A_+)\left[1 + \left(\frac{m(B)}{m(A)}\right)^{1/d}\right]^d + m(A_-)\left[1 + \left(\frac{m(B)}{m(A)}\right)^{1/d}\right]^d \\
&= \left(m(A)^{1/d} + m(B)^{1/d}\right)^d,
\end{aligned}$$

which gives the desired inequality (8) when A and B are both finite unions of rectangles with disjoint interiors.

Next, this quickly implies the result when A and B are open sets of finite measure. Indeed, by Theorem 1.4, for any $\epsilon > 0$ we can find unions of almost disjoint rectangles A_ϵ and B_ϵ, such that $A_\epsilon \subset A$, $B_\epsilon \subset B$, with $m(A) \leq m(A_\epsilon) + \epsilon$ and $m(B) \leq m(B_\epsilon) + \epsilon$. Since $A + B \supset A_\epsilon + B_\epsilon$, the inequality (8) for A_ϵ and B_ϵ and a passage to a limit gives the desired result. From this, we can pass to the case where A and B are arbitrary compact sets, by noting first that $A+B$ is then compact, and that if we define $A^\epsilon = \{x : d(x, A) < \epsilon\}$, then A^ϵ are open, and $A^\epsilon \searrow A$ as $\epsilon \to 0$. With similar definitions for B^ϵ and $(A+B)^\epsilon$, we observe also that $A + B \subset A^\epsilon + B^\epsilon \subset (A+B)^{2\epsilon}$. Hence, letting $\epsilon \to 0$, we see that (8) for A^ϵ and B^ϵ implies the desired result for A and B. The general case, in which we assume that A, B, and $A+B$ are measurable, then follows by approximating A and B from inside by compact sets, as in (iii) of Theorem 3.4.

6 Exercises

1. Prove that the Cantor set $\mathcal{C}$ constructed in the text is totally disconnected and perfect. In other words, given two distinct points $x, y \in \mathcal{C}$, there is a point $z \notin \mathcal{C}$ that lies between x and y, and yet $\mathcal{C}$ has no isolated points.

[Hint: If $x, y \in \mathcal{C}$ and $|x - y| > 1/3^k$, then x and y belong to two different intervals in C_k. Also, given any $x \in \mathcal{C}$ there is an end-point y_k of some interval in C_k that satisfies $x \neq y_k$ and $|x - y_k| \leq 1/3^k$.]

2. The Cantor set $\mathcal{C}$ can also be described in terms of ternary expansions.

(a) Every number in $[0,1]$ has a ternary expansion

$$x = \sum_{k=1}^{\infty} a_k 3^{-k}, \qquad \text{where } a_k = 0, 1, \text{ or } 2.$$

Note that this decomposition is not unique since, for example, $1/3 = \sum_{k=2}^{\infty} 2/3^k$. Prove that $x \in \mathcal{C}$ if and only if x has a representation as above where every a_k is either 0 or 2.

(b) The **Cantor-Lebesgue function** is defined on $\mathcal{C}$ by

$$F(x) = \sum_{k=1}^{\infty} \frac{b_k}{2^k} \qquad \text{if } x = \textstyle\sum_{k=1}^{\infty} a_k 3^{-k}, \text{ where } b_k = a_k/2.$$

In this definition, we choose the expansion of x in which $a_k = 0$ or 2.

Show that F is well defined and continuous on $\mathcal{C}$, and moreover $F(0) = 0$ as well as $F(1) = 1$.

(c) Prove that $F : \mathcal{C} \to [0,1]$ is surjective, that is, for every $y \in [0,1]$ there exists $x \in \mathcal{C}$ such that $F(x) = y$.

(d) One can also extend F to be a continuous function on $[0,1]$ as follows. Note that if (a,b) is an open interval of the complement of $\mathcal{C}$, then $F(a) = F(b)$. Hence we may define F to have the constant value $F(a)$ in that interval.

A geometrical construction of F is described in Chapter 3.

3. Cantor sets of constant dissection. Consider the unit interval $[0,1]$, and let ξ be a fixed real number with $0 < \xi < 1$ (the case $\xi = 1/3$ corresponds to the Cantor set $\mathcal{C}$ in the text).

In stage 1 of the construction, remove the centrally situated open interval in $[0,1]$ of length ξ. In stage 2, remove two central intervals each of relative length ξ, one in each of the remaining intervals after stage 1, and so on.

Let $\mathcal{C}_\xi$ denote the set which remains after applying the above procedure indefinitely.[6]

(a) Prove that the complement of $\mathcal{C}_\xi$ in $[0,1]$ is the union of open intervals of total length equal to 1.

(b) Show directly that $m_*(\mathcal{C}_\xi) = 0$.

[Hint: After the k^{th} stage, show that the remaining set has total length $= (1-\xi)^k$.]

4. Cantor-like sets. Construct a closed set $\hat{\mathcal{C}}$ so that at the k^{th} stage of the construction one removes 2^{k-1} centrally situated open intervals each of length ℓ_k, with

$$\ell_1 + 2\ell_2 + \cdots + 2^{k-1}\ell_k < 1.$$

[6] The set we call $\mathcal{C}_\xi$ is sometimes denoted by $\mathcal{C}_{\frac{1-\xi}{2}}$.

(a) If ℓ_j are chosen small enough, then $\sum_{k=1}^{\infty} 2^{k-1}\ell_k < 1$. In this case, show that $m(\hat{\mathcal{C}}) > 0$, and in fact, $m(\hat{\mathcal{C}}) = 1 - \sum_{k=1}^{\infty} 2^{k-1}\ell_k$.

(b) Show that if $x \in \hat{\mathcal{C}}$, then there exists a sequence of points $\{x_n\}_{n=1}^{\infty}$ such that $x_n \notin \hat{\mathcal{C}}$, yet $x_n \to x$ and $x_n \in I_n$, where I_n is a sub-interval in the complement of $\hat{\mathcal{C}}$ with $|I_n| \to 0$.

(c) Prove as a consequence that $\hat{\mathcal{C}}$ is perfect, and contains no open interval.

(d) Show also that $\hat{\mathcal{C}}$ is uncountable.

5. Suppose E is a given set, and $\mathcal{O}_n$ is the open set:

$$\mathcal{O}_n = \{x : \ d(x, E) < 1/n\}.$$

Show:

(a) If E is compact, then $m(E) = \lim_{n\to\infty} m(\mathcal{O}_n)$.

(b) However, the conclusion in (a) may be false for E closed and unbounded; or E open and bounded.

6. Using translations and dilations, prove the following: Let B be a ball in $\mathbb{R}^d$ of radius r. Then $m(B) = v_d r^d$, where $v_d = m(B_1)$, and B_1 is the unit ball, $B_1 = \{x \in \mathbb{R}^d : \ |x| < 1\}$.

A calculation of the constant v_d is postponed until Exercise 14 in the next chapter.

7. If $\delta = (\delta_1, \ldots, \delta_d)$ is a d-tuple of positive numbers $\delta_i > 0$, and E is a subset of $\mathbb{R}^d$, we define δE by

$$\delta E = \{(\delta_1 x_1, \ldots, \delta_d x_d) : \ \text{where } (x_1, \ldots, x_d) \in E\}.$$

Prove that δE is measurable whenever E is measurable, and

$$m(\delta E) = \delta_1 \cdots \delta_d m(E).$$

8. Suppose L is a linear transformation of $\mathbb{R}^d$. Show that if E is a measurable subset of $\mathbb{R}^d$, then so is $L(E)$, by proceeding as follows:

(a) Note that if E is compact, so is $L(E)$. Hence if E is an F_σ set, so is $L(E)$.

(b) Because L automatically satisfies the inequality

$$|L(x) - L(x')| \leq M|x - x'|$$

for some M, we can see that L maps any cube of side length ℓ into a cube of side length $c_d M\ell$, with $c_d = 2\sqrt{d}$. Now if $m(E) = 0$, there is a collection of cubes $\{Q_j\}$ such that $E \subset \bigcup_j Q_j$, and $\sum_j m(Q_j) < \epsilon$. Thus $m_*(L(E)) \leq c'\epsilon$, and hence $m(L(E)) = 0$. Finally, use Corollary 3.5.

One can show that $m(L(E)) = |\det L|\, m(E)$; see Problem 4 in the next chapter.

9. Give an example of an open set $\mathcal{O}$ with the following property: the boundary of the closure of $\mathcal{O}$ has positive Lebesgue measure.

[Hint: Consider the set obtained by taking the union of open intervals which are deleted at the odd steps in the construction of a Cantor-like set.]

10. This exercise provides a construction of a decreasing sequence of positive continuous functions on the interval $[0,1]$, whose pointwise limit is *not* Riemann integrable.

Let $\hat{\mathcal{C}}$ denote a Cantor-like set obtained from the construction detailed in Exercise 4, so that in particular $m(\hat{\mathcal{C}}) > 0$. Let F_1 denote a piecewise-linear and continuous function on $[0,1]$, with $F_1 = 1$ in the complement of the first interval removed in the construction of $\hat{\mathcal{C}}$, $F_1 = 0$ at the center of this interval, and $0 \leq F_1(x) \leq 1$ for all x. Similarly, construct $F_2 = 1$ in the complement of the intervals in stage two of the construction of $\hat{\mathcal{C}}$, with $F_2 = 0$ at the center of these intervals, and $0 \leq F_2 \leq 1$. Continuing this way, let $f_n = F_1 \cdot F_2 \cdots F_n$ (see Figure 5).

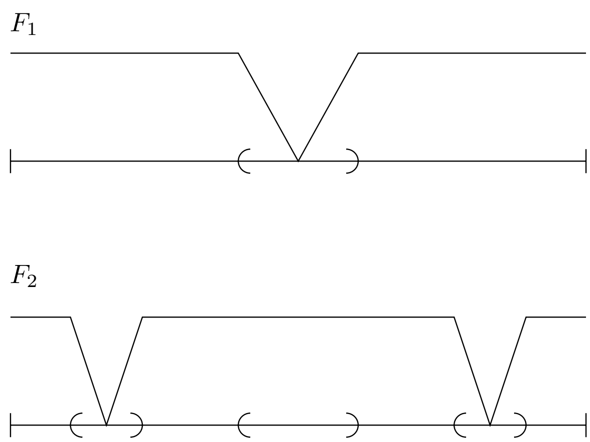

Figure 5. Construction of $\{F_n\}$ in Exercise 10

Prove the following:

(a) For all $n \geq 1$ and all $x \in [0,1]$, one has $0 \leq f_n(x) \leq 1$ and $f_n(x) \geq f_{n+1}(x)$. Therefore, $f_n(x)$ converges to a limit as $n \to \infty$ which we denote by $f(x)$.

(b) The function f is discontinuous at every point of $\hat{\mathcal{C}}$.

[Hint: Note that $f(x) = 1$ if $x \in \hat{\mathcal{C}}$, and find a sequence of points $\{x_n\}$ so that $x_n \to x$ and $f(x_n) = 0$.]

Now $\int f_n(x)\, dx$ is decreasing, hence $\int f_n$ converges. However, a bounded function is Riemann integrable if and only if its set of discontinuities has measure zero.

(The proof of this fact, which is given in the Appendix of Book I, is outlined in Problem 4.) Since f is discontinuous on a set of positive measure, we find that f is not Riemann integrable.

11. Let A be the subset of $[0, 1]$ which consists of all numbers which do not have the digit 4 appearing in their decimal expansion. Find $m(A)$.

12. Theorem 1.3 states that every open set in $\mathbb{R}$ is the disjoint union of open intervals. The analogue in $\mathbb{R}^d$, $d \geq 2$, is generally false. Prove the following:

(a) An open disc in $\mathbb{R}^2$ is not the disjoint union of open rectangles.

[Hint: What happens to the boundary of any of these rectangles?]

(b) An open connected set Ω is the disjoint union of open rectangles if and only if Ω is itself an open rectangle.

13. The following deals with G_δ and F_σ sets.

(a) Show that a closed set is a G_δ and an open set an F_σ.

[Hint: If F is closed, consider $\mathcal{O}_n = \{x : d(x, F) < 1/n\}$.]

(b) Give an example of an F_σ which is not a G_δ.

[Hint: This is more difficult; let F be a denumerable set that is dense.]

(c) Give an example of a Borel set which is not a G_δ nor an F_σ.

14. The purpose of this exercise is to show that covering by a *finite* number of intervals will not suffice in the definition of the outer measure m_*.

The **outer Jordan content** $J_*(E)$ of a set E in $\mathbb{R}$ is defined by

$$J_*(E) = \inf \sum_{j=1}^{N} |I_j|,$$

where the inf is taken over every *finite* covering $E \subset \bigcup_{j=1}^{N} I_j$, by intervals I_j.

(a) Prove that $J_*(E) = J_*(\overline{E})$ for every set E (here $\overline{E}$ denotes the closure of E).

(b) Exhibit a countable subset $E \subset [0, 1]$ such that $J_*(E) = 1$ while $m_*(E) = 0$.

15. At the start of the theory, one might define the outer measure by taking coverings by rectangles instead of cubes. More precisely, we define

$$m_*^{\mathcal{R}}(E) = \inf \sum_{j=1}^{\infty} |R_j|,$$

where the inf is now taken over all countable coverings $E \subset \bigcup_{j=1}^{\infty} R_j$ by (closed) rectangles.

Show that this approach gives rise to the same theory of measure developed in the text, by proving that $m_*(E) = m_*^{\mathcal{R}}(E)$ for every subset E of $\mathbb{R}^d$.

[Hint: Use Lemma 1.1.]

16. The Borel-Cantelli lemma. Suppose $\{E_k\}_{k=1}^{\infty}$ is a countable family of measuable subsets of $\mathbb{R}^d$ and that

$$\sum_{k=1}^{\infty} m(E_k) < \infty.$$

Let

$$\begin{aligned} E &= \{x \in \mathbb{R}^d : x \in E_k, \text{ for infinitely many } k\} \\ &= \limsup_{k \to \infty}(E_k). \end{aligned}$$

(a) Show that E is measurable.

(b) Prove $m(E) = 0$.

[Hint: Write $E = \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} E_k$.]

17. Let $\{f_n\}$ be a sequence of measurable functions on $[0,1]$ with $|f_n(x)| < \infty$ for a.e x. Show that there exists a sequence c_n of positive real numbers such that

$$\frac{f_n(x)}{c_n} \to 0 \qquad \text{a.e. } x$$

[Hint: Pick c_n such that $m(\{x : |f_n(x)/c_n| > 1/n\}) < 2^{-n}$, and apply the Borel-Cantelli lemma.]

18. Prove the following assertion: Every measurable function is the limit a.e. of a sequence of continuous functions.

19. Here are some observations regarding the set operation $A + B$.

(a) Show that if either A and B is open, then $A + B$ is open.

(b) Show that if A and B are closed, then $A + B$ is measurable.

(c) Show, however, that $A + B$ might not be closed even though A and B are closed.

[Hint: For (b) show that $A + B$ is an F_σ set.]

20. Show that there exist closed sets A and B with $m(A) = m(B) = 0$, but $m(A + B) > 0$:

(a) In $\mathbb{R}$, let $A = \mathcal{C}$ (the Cantor set), $B = \mathcal{C}/2$. Note that $A + B \supset [0, 1]$.

(b) In $\mathbb{R}^2$, observe that if $A = I \times \{0\}$ and $B = \{0\} \times I$ (where $I = [0, 1]$), then $A + B = I \times I$.

21. Prove that there is a continuous function that maps a Lebesgue measurable set to a non-measurable set.

[Hint: Consider a non-measurable subset of $[0, 1]$, and its inverse image in $\mathcal{C}$ by the function F in Exercise 2.]

22. Let $\chi_{[0,1]}$ be the characteristic function of $[0, 1]$. Show that there is no everywhere continuous function f on $\mathbb{R}$ such that

$$f(x) = \chi_{[0,1]}(x) \qquad \text{almost everywhere.}$$

23. Suppose $f(x, y)$ is a function on $\mathbb{R}^2$ that is separately continuous: for each fixed variable, f is continuous in the other variable. Prove that f is measurable on $\mathbb{R}^2$.

[Hint: Approximate f in the variable x by piecewise-linear functions f_n so that $f_n \to f$ pointwise.]

24. Does there exist an enumeration $\{r_n\}_{n=1}^{\infty}$ of the rationals, such that the complement of

$$\bigcup_{n=1}^{\infty} \left(r_n - \frac{1}{n}, r_n + \frac{1}{n} \right)$$

in $\mathbb{R}$ is non-empty?

[Hint: Find an enumeration where the only rationals outside of a fixed bounded interval take the form r_n, with $n = m^2$ for some integer m.]

25. An alternative definition of measurability is as follows: E is measurable if for every $\epsilon > 0$ there is a *closed* set F contained in E with $m_*(E - F) < \epsilon$. Show that this definition is equivalent with the one given in the text.

26. Suppose $A \subset E \subset B$, where A and B are measurable sets of finite measure. Prove that if $m(A) = m(B)$, then E is measurable.

27. Suppose E_1 and E_2 are a pair of compact sets in $\mathbb{R}^d$ with $E_1 \subset E_2$, and let $a = m(E_1)$ and $b = m(E_2)$. Prove that for any c with $a < c < b$, there is a compact set E with $E_1 \subset E \subset E_2$ and $m(E) = c$.

[Hint: As an example, if $d = 1$ and E is a measurable subset of $[0, 1]$, consider $m(E \cap [0, t])$ as a function of t.]

28. Let E be a subset of $\mathbb{R}$ with $m_*(E) > 0$. Prove that for each $0 < \alpha < 1$, there exists an open interval I so that

$$m_*(E \cap I) \geq \alpha\, m_*(I).$$

Loosely speaking, this estimate shows that E contains almost a whole interval.

[Hint: Choose an open set $\mathcal{O}$ that contains E, and such that $m_*(E) \geq \alpha\, m_*(\mathcal{O})$. Write $\mathcal{O}$ as the countable union of disjoint open intervals, and show that one of these intervals must satisfy the desired property.]

29. Suppose E is a measurable subset of $\mathbb{R}$ with $m(E) > 0$. Prove that the **difference set** of E, which is defined by

$$\{z \in \mathbb{R} : \ z = x - y \text{ for some } x, y \in E\},$$

contains an open interval centered at the origin.

If E contains an interval, then the conclusion is straightforward. In general, one may rely on Exercise 28.

[Hint: Indeed, by Exercise 28, there exists an open interval I so that $m(E \cap I) \geq (9/10)\, m(I)$. If we denote $E \cap I$ by E_0, and suppose that the difference set of E_0 does not contain an open interval around the origin, then for arbitrarily small a the sets E_0, and $E_0 + a$ are disjoint. From the fact that $(E_0 \cup (E_0 + a)) \subset (I \cup (I + a))$ we get a contradiction, since the left-hand side has measure $2m(E_0)$, while the right-hand side has measure only slightly larger than $m(I)$.]

A more general formulation of this result is as follows.

30. If E and F are measurable, and $m(E) > 0$, $m(F) > 0$, prove that

$$E + F = \{x + y : x \in E,\ x \in F\}$$

contains an interval.

31. The result in Exercise 29 provides an alternate proof of the non-measurability of the set $\mathcal{N}$ studied in the text. In fact, we may also prove the non-measurability of a set in $\mathbb{R}$ that is very closely related to the set $\mathcal{N}$.

Given two real numbers x and y, we shall write as before that $x \sim y$ whenever the difference $x - y$ is rational. Let $\mathcal{N}^*$ denote a set that consists of one element in each equivalence class of $\sim$. Prove that $\mathcal{N}^*$ is non-measurable by using the result in Exercise 29.

[Hint: If $\mathcal{N}^*$ is measurable, then so are its translates $\mathcal{N}_n^* = \mathcal{N}^* + r_n$, where $\{r_n\}_{n=1}^\infty$ is an enumeration of $\mathbb{Q}$. How does this imply that $m(\mathcal{N}^*) > 0$? Can the difference set of $\mathcal{N}^*$ contain an open interval centered at the origin?]

32. Let $\mathcal{N}$ denote the non-measurable subset of $I = [0, 1]$ constructed at the end of Section 3.

(a) Prove that if E is a measurable subset of $\mathcal{N}$, then $m(E) = 0$.

(b) If G is a subset of $\mathbb{R}$ with $m_*(G) > 0$, prove that a subset of G is non-measurable.

[Hint: For (a) use the translates of E by the rationals.]

33. Let $\mathcal{N}$ denote the non-measurable set constructed in the text. Recall from the exercise above that measurable subsets of $\mathcal{N}$ have measure zero.

Show that the set $\mathcal{N}^c = I - \mathcal{N}$ satisfies $m_*(\mathcal{N}^c) = 1$, and conclude that if $E_1 = \mathcal{N}$ and $E_2 = \mathcal{N}^c$, then

$$m_*(E_1) + m_*(E_2) \neq m_*(E_1 \cup E_2),$$

although E_1 and E_2 are disjoint.

[Hint: To prove that $m_*(\mathcal{N}^c) = 1$, argue by contradiction and pick a measurable set U such that $U \subset I$, $\mathcal{N}^c \subset U$ and $m_*(U) < 1 - \epsilon$.]

34. Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be any two Cantor sets (constructed in Exercise 3). Show that there exists a function $F : [0,1] \to [0,1]$ with the following properties:

(i) F is continuous and bijective,

(ii) F is monotonically increasing,

(iii) F maps $\mathcal{C}_1$ surjectively onto $\mathcal{C}_2$.

[Hint: Copy the construction of the standard Cantor-Lebesgue function.]

35. Give an example of a measurable function f and a continuous function Φ so that $f \circ \Phi$ is non-measurable.

[Hint: Let $\Phi : \mathcal{C}_1 \to \mathcal{C}_2$ as in Exercise 34, with $m(\mathcal{C}_1) > 0$ and $m(\mathcal{C}_2) = 0$. Let $N \subset \mathcal{C}_1$ be non-measurable, and take $f = \chi_{\Phi(N)}$.]

Use the construction in the hint to show that there exists a Lebesgue measurable set that is not a Borel set.

36. This exercise provides an example of a measurable function f on $[0,1]$ such that every function g equivalent to f (in the sense that f and g differ only on a set of measure zero) is discontinuous at *every* point.

(a) Construct a measurable set $E \subset [0,1]$ such that for any non-empty open sub-interval I in $[0,1]$, both sets $E \cap I$ and $E^c \cap I$ have positive measure.

(b) Show that $f = \chi_E$ has the property that whenever $g(x) = f(x)$ a.e x, then g must be discontinuous at every point in $[0,1]$.

[Hint: For the first part, consider a Cantor-like set of positive measure, and add in each of the intervals that are omitted in the first step of its construction, another Cantor-like set. Continue this procedure indefinitely.]

37. Suppose Γ is a curve $y = f(x)$ in $\mathbb{R}^2$, where f is continuous. Show that $m(\Gamma) = 0$.

[Hint: Cover Γ by rectangles, using the uniform continuity of f.]

38. Prove that $(a+b)^\gamma \geq a^\gamma + b^\gamma$ whenever $\gamma \geq 1$ and $a, b \geq 0$. Also, show that the reverse inequality holds when $0 \leq \gamma \leq 1$.

[Hint: Integrate the inequality between $(a+t)^{\gamma-1}$ and $t^{\gamma-1}$ from 0 to b.]

39. Establish the inequality

$$\frac{x_1 + \cdots + x_d}{d} \geq (x_1 \cdots x_d)^{1/d} \qquad \text{for all } x_j \geq 0,\ j = 1, \ldots, d \tag{10}$$

by using backward induction as follows:

(a) The inequality is true whenever d is a power of 2 ($d = 2^k$, $k \geq 1$).

(b) If (10) holds for some integer $d \geq 2$, then it must hold for $d-1$, that is, one has $(y_1 + \cdots + y_{d-1})/(d-1) \geq (y_1 \cdots y_{d-1})^{1/(d-1)}$ for all $y_j \geq 0$, with $j = 1, \ldots, d-1$.

[Hint: For (a), if $k \geq 2$, write $(x_1 + \cdots + x_{2^k})/2^k$ as $(A+B)/2$, where $A = (x_1 + \cdots + x_{2^{k-1}})/2^{k-1}$, and apply the inequality when $d = 2$. For (b), apply the inequality to $x_1 = y_1, \ldots, x_{d-1} = y_{d-1}$ and $x_d = (y_1 + \cdots + y_{d-1})/(d-1)$.]

7 Problems

1. Given an irrational x, one can show (using the pigeon-hole principle, for example) that there exists infinitely many fractions p/q, with relatively prime integers p and q such that

$$\left| x - \frac{p}{q} \right| \leq \frac{1}{q^2}.$$

However, prove that the set of those $x \in \mathbb{R}$ such that there exist infinitely many fractions p/q, with relatively prime integers p and q such that

$$\left| x - \frac{p}{q} \right| \leq \frac{1}{q^3} \qquad (\text{or } \leq 1/q^{2+\epsilon}),$$

is a set of measure zero.

[Hint: Use the Borel-Cantelli lemma.]

2. Any open set Ω can be written as the union of closed cubes, so that $\Omega = \bigcup Q_j$ with the following properties

(i) The Q_j's have disjoint interiors.

(ii) $d(Q_j, \Omega^c) \approx$ side length of Q_j. This means that there are positive constants c and C so that $c \leq d(Q_j, \Omega^c)/\ell(Q_j) \leq C$, where $\ell(Q_j)$ denotes the side length of Q_j.

3. Find an example of a measurable subset C of $[0,1]$ such that $m(C) = 0$, yet the difference set of C contains a non-trivial interval centered at the origin. Compare with the result in Exercise 29.

[Hint: Pick the Cantor set $C = \mathcal{C}$. For a fixed $a \in [-1,1]$, consider the line $y = x + a$ in the plane, and copy the construction of the Cantor set, but in the cube $Q = [0,1] \times [0,1]$. First, remove all but four closed cubes of side length $1/3$, one at each corner of Q; then, repeat this procedure in each of the remaining cubes (see Figure 6). The resulting set is sometimes called a Cantor dust. Use the property of nested compact sets to show that the line intersects this Cantor dust.]

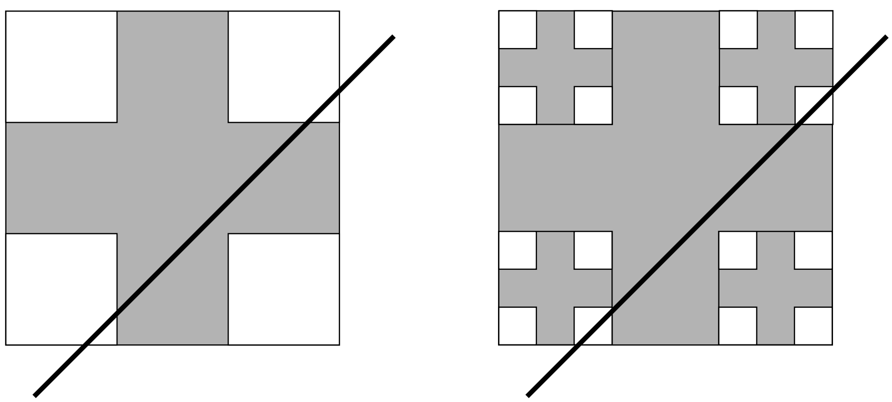

Figure 6. Construction of the Cantor dust

4. Complete the following outline to prove that a bounded function on an interval $[a,b]$ is Riemann integrable if and only if its set of discontinuities has measure zero. This argument is given in detail in the appendix to Book I.

Let f be a *bounded* function on a compact interval J, and let $I(c,r)$ denote the open interval centered at c of radius $r > 0$. Let $\text{osc}(f,c,r) = \sup |f(x) - f(y)|$, where the supremum is taken over all $x, y \in J \cap I(c,r)$, and define the oscillation of f at c by $\text{osc}(f,c) = \lim_{r\to 0} \text{osc}(f,c,r)$. Clearly, f is continuous at $c \in J$ if and only if $\text{osc}(f,c) = 0$.

Prove the following assertions:

(a) For every $\epsilon > 0$, the set of points c in J such that $\text{osc}(f,c) \geq \epsilon$ is compact.

(b) If the set of discontinuities of f has measure 0, then f is Riemann integrable.

[Hint: Given $\epsilon > 0$ let $A_\epsilon = \{c \in J : \text{osc}(f,c) \geq \epsilon\}$. Cover A_ϵ by a finite number of open intervals whose total length is $\leq \epsilon$. Select an appropriate partition of J and estimate the difference between the upper and lower sums of f over this partition.]

(c) Conversely, if f is Riemann integrable on J, then its set of discontinuities has measure 0.

[Hint: The set of discontinuities of f is contained in $\bigcup_n A_{1/n}$. Choose a partition P such that $U(f,P) - L(f,P) < \epsilon/n$. Show that the total length of the intervals in P whose interior intersect $A_{1/n}$ is $\leq \epsilon$.]

5. Suppose E is measurable with $m(E) < \infty$, and

$$E = E_1 \cup E_2, \qquad E_1 \cap E_2 = \emptyset.$$

If $m(E) = m_*(E_1) + m_*(E_2)$, then E_1 and E_2 are measurable.

In particular, if $E \subset Q$, where Q is a finite cube, then E is measurable if and only if $m(Q) = m_*(E) + m_*(Q - E)$.

6.* The fact that the axiom of choice and the well-ordering principle are equivalent is a consequence of the following considerations.

One begins by defining a partial ordering on a set E to be a binary relation $\leq$ on the set E that satisfies:

(i) $x \leq x$ for all $x \in E$.

(ii) If $x \leq y$ and $y \leq x$, then $x = y$.

(iii) If $x \leq y$ and $y \leq z$, then $x \leq z$.

If in addition $x \leq y$ or $y \leq x$ whenever $x, y \in E$, then $\leq$ is a linear ordering of E.

The axiom of choice and the well-ordering principle are then logically equivalent to the Hausdorff maximal principle:

Every non-empty partially ordered set has a (non-empty) maximal linearly ordered subset.

In other words, if E is partially ordered by $\leq$, then E contains a non-empty subset F which is linearly ordered by $\leq$ and such that if F is contained in a set G also linearly ordered by $\leq$, then $F = G$.

An application of the Hausdorff maximal principle to the collection of all well-orderings of subsets of E implies the well-ordering principle for E. However, the proof that the axiom of choice implies the Hausdorff maximal principle is more complicated.

7.* Consider the curve $\Gamma = \{y = f(x)\}$ in $\mathbb{R}^2$, $0 \leq x \leq 1$. Assume that f is twice continuously differentiable in $0 \leq x \leq 1$. Then show that $m(\Gamma + \Gamma) > 0$ if and only if $\Gamma + \Gamma$ contains an open set, if and only if f is not linear.

8.* Suppose A and B are open sets of finite positive measure. Then we have equality in the Brunn-Minkowski inequality (8) if and only if A and B are convex and similar, that is, there are a $\delta > 0$ and an $h \in \mathbb{R}^d$ such that

$$A = \delta B + h.$$