1

Basics

Our purpose in this chapter will be to establish notation and terminology. The reader should already be acquainted with most of the concepts discussed and thus might wish to skim the chapter or skip ahead, returning if clarification is needed.

1.1 Smooth Functions

The set of real numbers will be denoted by $\mathbb{R}$. In this book, we will be concerned with questions of geometric analysis in an N-dimensional Euclidean space. That is, we will work in the space $\mathbb{R}^N$ of ordered N-tuples of real numbers. The *inner product* $x \cdot y$ of two elements $x, y \in \mathbb{R}^N$ is defined by setting

$$x \cdot y = \sum_{i=1}^{N} x_i y_i \, ,$$

where

$$x = (x_1, x_2, \ldots, x_N) \text{ and } y = (y_1, y_2, \ldots, y_N) \, .$$

Of course, the inner product is a symmetric, bilinear, positive definite function on $\mathbb{R}^N \times \mathbb{R}^N$. The *norm* of the element $x \in \mathbb{R}^N$, denoted by $|x|$, is defined by setting

$$|x| = \sqrt{x \cdot x} \, , \tag{1.1}$$

as we may since $x \cdot x$ is always nonnegative. The *standard orthonormal basis elements* for $\mathbb{R}^N$ will be denoted by $\mathbf{e}_i$, $i = 1, 2, \ldots, N$. Specifically, $\mathbf{e}_i$ is the vector with N entries, all of which are 0's except the ith entry, which is 1. For computational purposes, elements of $\mathbb{R}^N$ should be considered column vectors. Column vectors can waste space on the page, and so we sometimes take the liberty of using row vector notation, as we did above.

The *open ball of radius* $r > 0$ *centered at* x will be denoted by $\mathbb{B}(x, r)$ and is defined by setting

G.S. Krantz, R.H. Parks (eds.), *Geometric Integration Theory,* doi: 10.1007/978-0-8176-4679-0_1,

$$\mathbb{B}(x, r) = \{ y \in \mathbb{R}^N \ : \ |x - y| < r \} .$$

The *closed ball of radius* $r \geq 0$ *centered at* x will be denoted by $\overline{\mathbb{B}}(x, r)$ and is defined by setting

$$\overline{\mathbb{B}}(x, r) = \{ y \in \mathbb{R}^N \ : \ |x - y| \leq r \} .$$

The standard topology on the space $\mathbb{R}^N$ is defined by letting the *open sets* consist of all arbitrary unions of open balls. The *closed sets* are then defined to be the complements of the open sets. For any subset A of $\mathbb{R}^N$ (or of any topological space), there is a largest open set contained in A. That set, denoted by $\mathring{A}$, is called the *interior of* A. Similarly, A is contained in a smallest closed set containing A and that set, denoted by $\overline{A}$, is called the *closure of* A. The *topological boundary of* A, denoted by ∂A, is defined by setting

$$\partial A = \overline{A} \setminus \mathring{A} .$$

Remark 1.1.1.

(1) At this juncture, the only notion of boundary in sight is that of the topological boundary. Since later we shall be led to define another notion of boundary, we are taking care to emphasize that the present definition is the topological one. When it is clear from context that we are discussing the topological boundary, then we will refer simply to the "boundary of A."
(2) The notations $\mathring{A}$ and $\overline{A}$ for the interior and closure, respectively, of the set A are commonly used but are not universal. A variety of notations is used for the topological boundary of A, and ∂A is one of the more popular choices.

Let $U \subseteq \mathbb{R}^N$ be any open set. A function $f : U \to \mathbb{R}^M$ is said to be *continuously differentiable of order* k, or C^k, if f possesses all partial derivatives of order not exceeding k and all of those partial derivatives are continuous; we write $f \in C^k$ or $f \in C^k(U)$ if U is not clear from context. If the range of f is also not clear from context, then we write (for instance) $f \in C^k(U; \mathbb{R}^M)$. When $k = 1$, we simply say that f is *continuously differentiable*. The function f is said to be C^∞, or *infinitely differentiable*, provided that $f \in C^k$ for every positive k. The function f is said to be in C^ω, or *real analytic*, provided that it has a convergent power series expansion about each point of U. We direct the reader to [KPk 02] for matters related to real analytic functions. We also extend the preceding notation by using $f \in C^0$ to indicate that f is continuous.

The order of differentiability of a function is referred to as its *smoothness*. By a *smooth function*, one typically means an $f \in C^\infty$, but sometimes one may mean an $f \in C^k$, where k is an integer as large as turns out to be needed.

The *support* of a continuous function $f : U \to \mathbb{R}^M$, denoted by $\operatorname{supp} f$, is the closure of the set of points where $f \neq 0$. We will use C_c^k to denote the C^k functions with compact support; here k can be a nonnegative integer or ∞.

Let $\mathbb{Z}$ denote the integers, $\mathbb{Z}^+$ the nonnegative integers, and $\mathbb{N}$ the positive integers. A *multi-index* α is an element of $(\mathbb{Z}^+)^N$, the Cartesian product of N copies of $\mathbb{Z}^+$. If $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_N)$ is a multi-index and $x = (x_1, x_2, \ldots, x_N)$ is a point in $\mathbb{R}^N$, then we introduce the following standard notation:

$$x^\alpha \equiv (x_1)^{\alpha_1}(x_2)^{\alpha_2}\cdots(x_N)^{\alpha_N}\,,$$

$$|\alpha| \equiv \alpha_1+\alpha_2+\cdots+\alpha_N\,,$$

$$\frac{\partial^{|\alpha|}}{\partial x^\alpha} \equiv \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}}\frac{\partial^{\alpha_2}}{\partial x_2^{\alpha_2}}\cdots\frac{\partial^{\alpha_N}}{\partial x_N^{\alpha_N}}\,,$$

$$\alpha! \equiv (\alpha_1!)(\alpha_2!)\cdots(\alpha_N!)\,.$$

With this notation, a function f on U is C^k if $(\partial^{|\alpha|}/\partial x^\alpha)f$ exists and is continuous for all multi-indices α with $|\alpha| \le k$.

We will sometimes find it convenient to use the alternative notations

$$D_{x_i} f = \frac{\partial f}{\partial x_i} \quad\text{and}\quad D_{x_i\,x_j} f = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

for the partial derivatives of the function f (which may be a real-valued or vector-valued function).

Definition 1.1.2. If f is defined in a neighborhood of $p \in \mathbb{R}^N$, and if f takes values in $\mathbb{R}^M$, then we say that f is *differentiable* at p when there exists a linear function $Df(p) : \mathbb{R}^N \to \mathbb{R}^M$ such that

$$\lim_{x\to p} \frac{|f(x) - f(p) - \langle Df(p),\ x-p\rangle|}{|x-p|} = 0\,. \tag{1.2}$$

In case f is differentiable at p, we call $Df(p)$ the *differential* of f at p.

Advanced calculus tells us that if f is differentiable as in Definition 1.1.2, then the first partial derivatives of f exist and that we can evaluate the differential applied to the vector v using the equation

$$\langle Df(p),\ v\rangle = \sum_{i=1}^{N} v_i \frac{\partial f}{\partial x_i}(p) = \sum_{i=1}^{N} (\mathbf{e}_i \cdot v)\frac{\partial f}{\partial x_i}(p)\,, \tag{1.3}$$

where $v = \sum_{i=1}^{n} v_i \mathbf{e}_i$. The *Jacobian matrix*[1] of f at p is denoted by Jac f and is defined by

$$\operatorname{Jac} f \equiv \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(p) & \frac{\partial f_1}{\partial x_2}(p) & \cdots & \frac{\partial f_1}{\partial x_N}(p) \\ \frac{\partial f_2}{\partial x_1}(p) & \frac{\partial f_2}{\partial x_2}(p) & \cdots & \frac{\partial f_2}{\partial x_N}(p) \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_M}{\partial x_1}(p) & \frac{\partial f_M}{\partial x_2}(p) & \cdots & \frac{\partial f_M}{\partial x_N}(p) \end{pmatrix}.$$

[1] Carl Gustav Jacobi (1804–1851).

For $v \in \mathbb{R}^N$, we have

$$\langle Df(p),\ v\rangle = [\text{Jac}\, f]\, v\,, \tag{1.4}$$

where on the right-hand side of (1.4) the vector v is represented as a column vector and Jac f operates on v by matrix multiplication. Equation (1.4) is simply another way of writing (1.3). We will sometimes find it convenient to use the notation

$$D_v f(p) = \langle Df(p),\ v\rangle\,.$$

We will denote the collection of all M-by-N matrices with real entries by

$$\mathcal{M}_{M,N}\,.$$

The *Hilbert–Schmidt norm*[2] on $\mathcal{M}_{M,N}$ is defined by setting

$$\left|\,\left(a_{i,j}\right)\,\right| = \left(\sum_{i=1}^{M}\sum_{j=1}^{N}(a_{i,j})^2\right)^{1/2}$$

for $\left(a_{i,j}\right) \in \mathcal{M}_{M,N}$. The standard topology on $\mathcal{M}_{M,N}$ is that induced by the Hilbert–Schmidt norm. Of course, the mapping

$$\left(a_{i,j}\right) \longmapsto \sum_{i=1}^{M}\sum_{j=1}^{N} a_{i,j}\,\mathbf{e}_{i+(j-1)M}$$

from $\mathcal{M}_{M,N}$ to $\mathbb{R}^{MN}$ is a homeomorphism.

The function sending a point to its differential, when the differential exists, takes its values in the space of linear transformations from $\mathbb{R}^N$ to $\mathbb{R}^M$, a space often denoted by $\text{Hom}(\mathbb{R}^N, \mathbb{R}^M)$. The space $\text{Hom}(\mathbb{R}^N, \mathbb{R}^M)$ can be identified with $\mathcal{M}_{M,N}$ by representing each linear transformation by an $M \times N$ matrix. The Jacobian matrix provides that representation for the differential of a function.

The standard topology on $\text{Hom}(\mathbb{R}^N, \mathbb{R}^M)$ is that induced by the Hilbert–Schmidt norm on $\mathcal{M}_{M,N}$ and the identification of $\text{Hom}(\mathbb{R}^N, \mathbb{R}^M)$ with $\mathcal{M}_{M,N}$. On a finite-dimensional vector space, all norms induce the same topology, so, in particular, the same topology is given by the *mapping norm* on $\text{Hom}(\mathbb{R}^N, \mathbb{R}^M)$ defined by

$$\|L\| = \sup\{\ |L(v)|\ :\ v \in \mathbb{R}^N,\ |v| \le 1\ \}\,.$$

We see that $f : U \to \mathbb{R}^M$ is C^1 if and only if

$$p \longmapsto Df(p)$$

is a continuous mapping from U into $\text{Hom}(\mathbb{R}^N, \mathbb{R}^M)$.

[2] David Hilbert (1862–1943), Erhard Schmidt (1876–1959).

Definition 1.1.3. If $f \in C^k(U, \mathbb{R}^M)$, $k = 1, 2, \ldots$, we define the kth *differential* of f at p, denoted by $D^k f(p)$, to be the k-linear $\mathbb{R}^M$-valued function given by

$$\langle D^k f(p),\ (v_1, v_2, \ldots, v_k)\rangle = \sum_{i_1, i_2, \ldots, i_k = 1}^{N} \prod_{j=1}^{k} (\mathbf{e}_{i_j} \cdot v_j) \frac{\partial^k}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_k}} f(p)\,. \quad (1.5)$$

Note that in the case $k = 1$, equations (1.3) and (1.5) agree. Also note that the equality of mixed partial derivatives guarantees that $D^k f(p)$ is a symmetric function. The interested reader may consult [Fed 69, 1.9, 1.10, 3.1.11] to see the kth differential placed in the context of the symmetric algebra over a vector space.

Finally, note that in case $k > 1$, one can show inductively that (1.5) agrees with the value of the differential at p of the function

$$\langle D^{k-1} f(\cdot),\ (v_1, v_2, \ldots, v_{k-1})\rangle$$

applied to the vector v_k, that is,

$$\langle D^k f(p),\ (v_1, v_2, \ldots, v_k)\rangle = \langle D\, \langle D^{k-1} f(p),\ (v_1, v_2, \ldots, v_{k-1})\rangle,\ v_k\rangle$$

holds.

In case $M = 1$, one often identifies the differential of f with the *gradient vector* of f, denoted by grad f and defined by setting

$$\operatorname{grad} f = \sum_{i=1}^{N} \frac{\partial f}{\partial x_i}\, \mathbf{e}_i\,.$$

Similarly, the second differential of f can be identified with the *Hessian matrix*[3] of f, denoted by Hess (f) and defined by

$$\operatorname{Hess}(f) = \begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1\, \partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_1\, \partial x_N} \\ \dfrac{\partial^2 f}{\partial x_2\, \partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f}{\partial x_2\, \partial x_N} \\ \vdots & \vdots & & \vdots \\ \dfrac{\partial^2 f}{\partial x_N\, \partial x_1} & \dfrac{\partial^2 f}{\partial x_N\, \partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_N^2} \end{pmatrix}.$$

If f is suitably smooth, one has

$$v \cdot \operatorname{grad} f = \langle Df,\ v\rangle$$

and

$$v \cdot ([\operatorname{Hess}(f)]\, w) = \langle D^2 f,\ (v, w)\rangle\,,$$

for vectors v and w represented as columns and where $[\operatorname{Hess}(f)]\, w$ indicates matrix multiplication.

[3] Ludwig Otto Hesse (1811–1874).

1.2 Measures

Standard references for basic measure theory are [Fol 84], [Roy 88], and [Rud 87]. Since there are variations in terminology and notation among authors, we will briefly review measure theory. We shall *not* provide proofs of most statements, but instead refer the reader to [Fol 84], [Roy 88], and [Rud 87] for details.

Definition 1.2.1. Let X be a nonempty set.

(1) By a *measure* on X we mean a function μ defined on all subsets of X satisfying the conditions $\mu(\emptyset) = 0$, $A \subseteq B$ implies $\mu(A) \leq \mu(B)$, and

$$\mu\left(\bigcup_{A\in\mathcal{F}} A\right) \leq \sum_{A\in\mathcal{F}} \mu(A) \quad \text{if } \mathcal{F} \text{ is collection of subsets of } X \text{ with } \operatorname{card}(\mathcal{F}) \leq \aleph_0. \tag{1.6}$$

(2) If a set $A \subseteq X$ satisfies

$$\mu(E) = \mu(E \cap A) + \mu(E \setminus A) \quad \text{for all } E \subseteq X, \tag{1.7}$$

then we say that A is *μ-measurable*.

The condition (1.6) is called *countable subadditivity*. Since the empty union is the empty set and the empty sum is zero, countable subadditivity implies $\mu(\emptyset) = 0$. Nonetheless, it is worth emphasizing that $\mu(\emptyset) = 0$ must hold.

Proposition 1.2.2. *Let μ be a measure on the nonempty set X.*

(1) *If $\mu(A) = 0$, then A is μ-measurable.*
(2) *If A is μ-measurable and $B \subseteq X$, then*

$$\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)\,.$$

Definition 1.2.3. Let X be a nonempty set. By a *σ-algebra* on X is meant a family $\mathcal{M}$ of subsets of X such that

(1) $\emptyset \in \mathcal{M}$, $X \in \mathcal{M}$,
(2) $\mathcal{M}$ is closed under countable unions,
(3) $\mathcal{M}$ is closed under countable intersections, and
(4) $\mathcal{M}$ is closed under taking complements in X.

Theorem 1.2.4. *If μ is a measure on the nonempty set X, then the family of μ-measurable sets forms a σ-algebra.*

Theorem 1.2.5. *Let μ be a measure on the nonempty set X.*

(1) *If $\mathcal{F}$ is an at most countable family of pairwise disjoint μ-measurable sets, then*

$$\mu\left(\bigcup_{A\in\mathcal{F}} A\right) = \sum_{A\in\mathcal{F}} \mu(A)\,.$$

(2) *If* $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$ *is a nondecreasing family of* μ*-measurable sets, then*

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{i\to\infty} \mu(A_i)\,.$$

(3) *If* $B_1 \supseteq B_2 \supseteq B_3 \supseteq \cdots$ *is a nonincreasing family of* μ*-measurable sets and* $\mu(B_1) < \infty$*, then*

$$\mu\left(\bigcap_{i=1}^{\infty} B_i\right) = \lim_{i\to\infty} \mu(B_i)\,.$$

Remark 1.2.6. The conclusion (1) of Theorem 1.2.5 is called *countable additivity*. Many authors prefer the term *outer measure* for the countably subadditive functions we have called measures. Those authors define a measure to be a countably additive function on a σ-algebra. But if $\mathcal{M}$ is a σ-algebra and

$$m : \mathcal{M} \to \{t : 0 \le t \le \infty\}$$

is a countably additive function, then one can define $\mu(A)$ for any $A \subseteq X$ by setting

$$\mu(A) = \inf\{\, m(E) : A \subseteq E \in \mathcal{M}\,\}\,.$$

With μ so defined, we see that $\mu(A) = m(A)$ holds whenever $A \in \mathcal{M}$ and that every set in $\mathcal{M}$ is μ-measurable. Thus it is no loss of generality to assume from the outset that a measure is defined on all subsets of X. It should be stressed that even though the measure is defined on all subsets of X, some subsets of X will *not* be μ-measurable.

The notion of a regular measure, defined next, gives additional useful structure.

Definition 1.2.7. A measure μ on a nonempty set X is *regular* if for each set $A \subseteq X$ there exists a μ-measurable set B with $A \subseteq B$ and $\mu(A) = \mu(B)$.

One consequence of the additional structure available when one is working with a regular measure is given in the next lemma. The lemma is easily proved using the analogous result for μ-measurable sets, i.e., Theorem 1.2.5(2).

Lemma 1.2.8. *Let* μ *be a regular measure on the nonempty set* X*. If a sequence of subsets* $\{A_j\}$ *of* X *satisfies* $A_1 \subseteq A_2 \subseteq \cdots$*, then*

$$\mu\left(\bigcup_{j=1}^{\infty} A_j\right) = \lim_{j\to\infty} \mu(A_j)\,.$$

Definition 1.2.9. If X is a topological space, then the *Borel sets*[4] are the elements of the smallest σ-algebra containing the open sets.

[4] Émile Borel (1871–1956).

For a measure on a topological space, it is evident that the measurability of all the open sets implies the measurability of all the Borel sets, but it is typical for the Borel sets to be a proper subfamily of the measurable sets. For instance, the sets in $\mathbb{R}^N$ known as Suslin sets[5] or (especially in the descriptive set theory literature) as analytic sets are μ-measurable for measures μ of interest in geometric analysis. Any continuous image of a Borel set is a Suslin set, so every Borel set is ipso facto a Suslin set. Suslin sets are discussed in Section 1.7.

For the study of geometric analysis, the measures of interest always satisfy the following condition of Borel regularity.

Definition 1.2.10. Let μ be a measure on the topological space X. We say that μ is *Borel regular* if every open set is μ-measurable and if for each $A \subseteq X$, there exists a Borel set $B \subseteq X$ with $A \subseteq B$ and $\mu(A) = \mu(B)$.

Often we will be working in the more restrictive class of Radon measures[6] defined next.

Definition 1.2.11. Suppose μ is a measure on a locally compact Hausdorff space[7] X. We say that μ is a *Radon measure* if the following conditions hold:

(1) Every compact set has finite μ measure.
(2) Every open set is μ-measurable, and if $V \subseteq X$ is open, then

$$\mu(V) = \sup\{\, \mu(K) : K \text{ is compact and } K \subseteq V \,\}.$$

(3) For every $A \subseteq X$,

$$\mu(A) = \inf\{\, \mu(V) : V \text{ is open and } A \subseteq V \,\}.$$

Definition 1.2.12. Let X be a metric space with metric ϱ.

(1) For a set $A \subseteq X$, we define the *diameter of* A by setting

$$\operatorname{diam} A = \sup\{\, \varrho(x, y) : x, y \in A \,\}.$$

(2) For sets $A, B \subseteq X$, we define the *distance between* A *and* B by setting

$$\operatorname{dist}(A, B) = \inf\{\, \varrho(a, b) : a \in A,\ b \in B \,\}.$$

If A is the singleton set $\{a_0\}$, then we will abuse the notation by writing $\operatorname{dist}(a_0, B)$ instead of $\operatorname{dist}(\{a_0\}, B)$.

When one is working in a metric space, a convenient tool for verifying the measurability of the open sets is often provided by Carathéordory's criterion,[8] which we now introduce.

[5] Mikhail Yakovlevich Suslin (1894–1919).
[6] Johann Radon (1887–1956).
[7] Felix Hausdorff (1869–1942).
[8] Constantin Carathéodory (1873–1950).

Theorem 1.2.13 (Carathéodory's criterion). *Suppose μ is a measure on the metric space X. All open subsets of X are μ-measurable if and only if*

$$\mu(A) + \mu(B) \leq \mu(A \cup B) \tag{1.8}$$

holds whenever $A, B \subseteq X$ with $0 < \operatorname{dist}(A, B)$.

Proof. First, suppose all open subsets of X are μ-measurable and let $A, B \subseteq X$ with $0 < \operatorname{dist}(A, B)$ be given. Setting $d = \operatorname{dist}(A, B)$, we can define the open set

$$V = \{\, x \in X : \operatorname{dist}(x, A) < d/2 \,\}.$$

Since V is open, thus μ-measurable, we have

$$\mu(A \cup B) = \mu[(A \cup B) \cap V] + \mu[(A \cup B) \setminus V] = \mu(A) + \mu(B),$$

so (1.8) holds.

Conversely, let $V \subseteq X$ be open and suppose (1.8) holds whenever $A, B \subseteq X$ with $0 < \operatorname{dist}(A, B)$. Let $E \subseteq X$ be an arbitrary set. Without loss of generality, we may suppose that $\mu(E) < \infty$ holds. Using (1.8) inductively, we see that

$$\mu(E) \geq \sum_{i=1}^{n} \mu(\{x \in E : 1/(2i+1) \leq \operatorname{dist}(x, V) < 1/(2i)\})$$

and likewise,

$$\mu(E) \geq \sum_{i=1}^{n} \mu(\{x \in E : 1/(2i+2) \leq \operatorname{dist}(x, V) < 1/(2i+1)\}).$$

Since n was arbitrary, we conclude that

$$2\mu(E) \geq \sum_{i=1}^{\infty} \mu(\{x \in E : 1/(i+1) \leq \operatorname{dist}(x, V) < 1/i\}),$$

so

$$\begin{aligned} 0 &= \in_{n\to\infty} \sum_{i=n}^{\infty} \mu(\{x \in E : 1/(i+1) \leq \operatorname{dist}(x, V) < 1/i\}) \\ &\geq \mu(\{x \in E : 0 < \operatorname{dist}(x, V) < 1/n\}). \end{aligned}$$

Again using (1.8), we see that

$$\begin{aligned} \mu(E) &\geq \mu(E \cap V) + \mu(\{x \in E : 1/n \leq \operatorname{dist}(x, V)\}) \\ &\geq \mu(E \cap V) + \mu(E \setminus V) - \mu(\{x \in E : 0 < \operatorname{dist}(x, V) < 1/n\}), \end{aligned}$$

and letting $n \to \infty$, we obtain

$$\mu(E) \geq \mu(E \cap V) + \mu(E \setminus V).$$

Since $E \subseteq X$ was arbitrary, V is μ-measurable. □

1.2.1 Lebesgue Measure

To close out this section, we define Lebesgue measure[9] on $\mathbb{R}$. Other measures will be defined in Chapter 2.

Definition 1.2.14. For $A \subseteq \mathbb{R}$, the (one-dimensional) *Lebesgue measure* of A is denoted by $\mathcal{L}^1(A)$ and is defined by setting $\mathcal{L}^1(A)$ equal to

$$\inf \left\{ \sum_{I \in \mathcal{I}} \text{length}(I) : \mathcal{I} \text{ is a family of bounded open intervals, } A \subseteq \bigcup_{I \in \mathcal{I}} I \right\}. \quad (1.9)$$

Here, of course, if $I = (a, b)$ is an open interval, then $\text{length}(I) = b - a$.

It is easy to see that $\mathcal{L}^1$ is a measure, and it is easy to apply Carathéodory's criterion (by dividing long intervals into short intervals) to see that all open sets in the reals are $\mathcal{L}^1$ measurable. The purpose of the Lebesgue measure is to extend the notion of length to more general sets. It may not be obvious that the result of the construction agrees with the ordinary notion of length, so we confirm that fact next.

Lemma 1.2.15. *If a bounded, closed interval* $[a, b]$ *is contained in the union of finitely many nonempty, bounded, open intervals,* $(a_1, b_1), (a_2, b_2), \ldots, (a_n, b_n)$*, then it holds that*

$$b - a \le \sum_{i=1}^{n} (b_i - a_i) . \quad (1.10)$$

Proof. Noting that the result is obvious when $n = 1$, we argue by induction on n by supposing that the result holds for all bounded, closed intervals and all n less than or equal to the natural number N.

Consider

$$[a, b] \subseteq \bigcup_{i=1}^{N+1} (a_i, b_i) .$$

At least one of the intervals contains a, so by renumbering the intervals if need be, we may suppose $a \in (a_{N+1}, b_{N+1})$. Also, we may suppose $b_{N+1} < b$, because $b \le b_{N+1}$ would give us $b - a < b_{N+1} - a_{N+1}$.

We have

$$[b_{N+1}, b] \subseteq \bigcup_{i=1}^{N} (a_i, b_i) ,$$

and thus, by the induction hypothesis,

$$b - b_{N+1} \le \sum_{i=1}^{N} (b_i - a_i) ,$$

so

[9] Henri Léon Lebesgue (1875–1941).

$$b-a \le (b_{N+1}-a_{N+1})+(b-b_{N+1}) \le (b_{N+1}-a_{N+1})+\sum_{i=1}^{N}(b_i-a_i) = \sum_{i=1}^{N+1}(b_i-a_i)\,,$$

as required. □

Corollary 1.2.16. *The Lebesgue measure of the closed, bounded interval* $[a, b]$ *equals* $b - a$.

Proof. Clearly, we have $\mathcal{L}^1([a, b]) \le b - a$. To obtain the reverse inequality, we observe that, if $[a, b]$ is covered by a countable family of open intervals, then by compactness, $[a, b]$ is covered by finitely many of the open intervals. It then follows from the lemma that the sum of the lengths of the covering intervals exceeds $b - a$. □

Lebesgue measure is the unique translation-invariant measure on $\mathbb{R}$ that assigns measure 1 to the unit interval. The next example shows us that not every set is $\mathcal{L}^1$-measurable.

Example 1.2.17. Let $\mathbb{Q}$ denote the rational numbers. Notice that for each $a \in \mathbb{R}$, the set X_a defined by

$$X_a = \{\, a + q : q \in \mathbb{Q} \,\}$$

intersects the unit interval [0, 1]. Of course, if $a_1 - a_2$ is a rational number, then $X_{a_1} = X_{a_2}$, but also the converse is true: if $X_{a_1} = X_{a_2}$, then $a_1 - a_2 \in \mathbb{Q}$.

By the axiom of choice, there exists a set C such that

$$C \cap [0, 1] \cap X_a$$

has exactly one element for every $a \in \mathbb{R}$. By the way C is defined, the sets $C - q = \{\, c - q : c \in C \,\}$, $q \in [0, 1] \cap \mathbb{Q}$, must be pairwise disjoint. Because $\mathcal{L}^1$ is translation-invariant, all the sets $C - q$ have $\mathcal{L}^1$ measure equal to $\mathcal{L}^1(C)$, and if one of those sets is $\mathcal{L}^1$-measurable, then all of them are.

Now, if $t \in [0, 1]$, then there is $c \in [0, 1] \cap X_t$, that is, $c = t + q$ with $q \in \mathbb{Q}$. Equivalently, we can write $q = c - t$, so we see that $-1 \le q \le 1$ and $t \in C - q$. Thus we have

$$[0, 1] \subseteq \bigcup_{q \in [-1,1] \cap \mathbb{Q}} (C - q) \subseteq [-1, 2] \tag{1.11}$$

and the sets in the union are all pairwise disjoint.

If C were $\mathcal{L}^1$-measurable, then the left-hand containment in (1.11) would tell us that $\mathcal{L}^1(C) > 0$, while the right-hand containment would tell us that $\mathcal{L}^1(C) = 0$. Thus we have a contradiction. We conclude that C is not $\mathcal{L}^1$-measurable. □

The construction in the Example 1.2.17 is widely known. Less well known is the general fact that if μ is a Borel regular measure on a complete, separable metric space such that there are sets with positive, finite measure and with the property that no point has positive measure, then there must exist a set that is not μ-measurable (see [Fed 69, 2.2.4]).

The construction of nonmeasurable sets requires the use of the axiom of choice. In fact, Robert Solovay has used the method of forcing (originally developed by Paul Cohen (1934–2007)) to construct a model of set theory in which the axiom of choice is not valid and in which every set of reals is Lebesgue measurable (see [Sov 70]).

1.3 Integration

The definition of the integral in use in the mid 1800s was that given by Augustin-Louis Cauchy (1789–1850). Cauchy's definition is applicable to continuous integrands, and easily extends to piecewise continuous integrands, but does not afford more generality. This lack of generality in the definition of the definite integral compelled Bernhard Riemann (1826–1866) to clarify the notion of an integrable function for his investigation of the representation of functions by trigonometric series.

Recall that Riemann's definition of the integral of a function $f : [a, b] \to \mathbb{R}$ is based on the idea of partitioning the *domain* of the function into sub-intervals. This approach is mandated by the absence of a measure of the size of general subsets of the domain. Measure theory takes away that limitation and allows the definition of the integral to proceed by partitioning the domain via the inverse images of intervals in the *range*. While this change of the partitioning may seem minor, the consequences are far-reaching and have provided a theory that continues to serve us well.

1.3.1 Measurable Functions

Definition 1.3.1. Let μ be a measure on the nonempty set X.

(1) The term *μ-almost* can serve as an adjective or adverb in the following ways:
  (a) Let $\mathcal{P}(x)$ be a statement or formula that contains a free variable $x \in X$. We say that $\mathcal{P}(x)$ holds for *μ-almost every $x \in X$* if
$$\mu\Big(\{\, x \in X : \mathcal{P}(x) \text{ is false} \,\}\Big) = 0\,.$$
  If X is understood from context, then we simply say that $\mathcal{P}(x)$ holds *μ-almost everywhere*.
  (b) Two sets $A, B \subseteq X$ are *μ-almost equal* if their symmetric difference has μ-measure zero, i.e., $\mu\Big[(A \setminus B) \cup (B \setminus A)\Big] = 0$.
  (c) Two functions f and g, each defined for μ-almost every $x \in X$, are said to be *μ-almost equal* if $f(x) = g(x)$ holds for μ-almost every $x \in X$.

(2) Let Y be a topological space. By a *μ-measurable, Y-valued function* we mean a Y-valued function f defined for μ-almost every $x \in X$ such that the inverse image of any open subset U of Y is a μ-measurable subset of X, that is,
  (a) $f : D \subseteq X \to Y$,
  (b) $\mu(X \setminus D) = 0$, and
  (c) $f^{-1}(U)$ is μ-measurable whenever $U \subseteq Y$ is open.

Remark 1.3.2.

(1) For the purposes of measure and integration, two functions that are μ-almost equal are equivalent. This defines an equivalence relation.

(2) It is no loss of generality to assume that a μ-measurable function is defined at every point of X. In fact, suppose f is a μ-measurable, Y-valued function with domain D and let y_0 be any element of Y. We can define the μ-measurable

function $\widetilde{f} : X \to Y$ by setting $\widetilde{f} = f$ on D and $\widetilde{f}(x) = y_0$, for all $x \in X \setminus D$. Then f and $\widetilde{f}$ are μ-almost equal and $\widetilde{f}$ is defined at every point of X.

Next we state two classical theorems concerning measurable functions due to Egorov[10] and Luzin.[11]

Theorem 1.3.3 (Egorov's theorem). *Let μ be a measure on X and let $f_1, f_2, \ldots$ be real-valued, μ-measurable functions. If $A \subseteq X$ with $\mu(A) < \infty$,*

$$\lim_{n\to\infty} f_n(x) = g(x) \ \text{ exists for } \mu\text{-almost every } x \in A,$$

and $\epsilon > 0$, then there exists a μ-measurable set B, with $\mu(A \setminus B) < \epsilon$, such that f_n converges uniformly to g on B.

Theorem 1.3.4 (Luzin's theorem). *Let X be a metric space and let μ be a Borel regular measure on X. If $f : X \to \mathbb{R}$ is μ-measurable, $A \subseteq X$ is μ-measurable with $\mu(A) < \infty$, and $\epsilon > 0$, then there exists a closed set $C \subseteq A$, with $\mu(A \setminus C) < \epsilon$, such that f is continuous on C.*

One reason for the usefulness of the notion of a μ-measurable function is that the set of μ-measurable functions is closed under operations of interest in analysis (including limiting operations). This usefulness is further enhanced by using the extended real numbers, which we define next.

Definition 1.3.5. Often we will allow a function to take the values $+\infty = \infty$ and $-\infty$. To accommodate this generality, we define the *extended real numbers*

$$\overline{\mathbb{R}} = \mathbb{R} \cup \{\infty, -\infty\}.$$

The standard ordering on $\overline{\mathbb{R}}$ is defined by requiring

$$x \le y \ \text{ if and only if }$$
$$(x, y) \in \left(\{-\infty\} \times \overline{\mathbb{R}}\right) \bigcup \left(\mathbb{R} \times \{\infty\}\right) \bigcup \{(x, y) \in \mathbb{R} \times \mathbb{R} : x \le y\}.$$

The operation of addition is extended by requiring that it agree with values already defined for the real numbers, by demanding that the operation be commutative, and by assigning the values given in the following table:

$+$	$-\infty$	$x \in \mathbb{R}$	$+\infty$
$+\infty$	undefined	$+\infty$	$+\infty$
$-\infty$	$-\infty$	$-\infty$	undefined

The operation of multiplication is extended by requiring that it agree with values already defined for the real numbers, by demanding that the operation be commutative, and by assigning the values given in the following table:

[10] Dmitriĭ Fedorovich Egorov (1869–1931).

[11] Nikolai Nikolaevich Luzin (Nicolas Lusin) (1883–1950).

x	$-\infty \le x < 0$	0	$0 < x \le +\infty$
$+\infty$	$-\infty$	undefined	$+\infty$
$-\infty$	$+\infty$	undefined	$-\infty$

The topology on $\overline{\mathbb{R}}$ has as a basis the finite open intervals and the intervals of the form $[-\infty, a)$ and $(a, \infty]$ for $a \in \mathbb{R}$.

The extension of each arithmetic operation given above is maximal subject to the requirement that the operation remain continuous. Nonetheless, when defining integrals, it is convenient to extend the above definitions by adopting the convention that

$$0 \cdot \infty = 0 \cdot (-\infty) = 0 .$$

Theorem 1.3.6. *Let μ be a measure on the nonempty set X.*

(1) *If f and g are μ-measurable, extended-real-valued functions and if $f + g$ (respectively, fg) is defined μ-almost everywhere, then $f + g$ (respectively, fg) is μ-measurable.*
(2) *If f and g are μ-measurable, extended-real-valued functions, then the functions* $\max\{f, g\}$ *and* $\min\{f, g\}$ *are μ-measurable.*
(3) *If $f_1, f_2, \ldots$ are μ-measurable, extended-real-valued functions, then the functions* $\limsup_{n\to\infty} f_n$ *and* $\liminf_{n\to\infty} f_n$ *are μ-measurable.*

1.3.2 The Integral

Definition 1.3.7. For a function $f : X \to \overline{\mathbb{R}}$ we define the *positive part* of f to be the function $f^+ : X \to [0, \infty]$ defined by setting

$$f^+(x) = \begin{cases} f(x) & \text{if } f(x) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Similarly, the *negative part* of f is denoted by f^- and is defined by setting

$$f^-(x) = \begin{cases} f(x) & \text{if } f(x) < 0, \\ 0 & \text{otherwise.} \end{cases}$$

Definition 1.3.8.

(1) The *characteristic function of* $S \subseteq X$ is the function with domain X defined, for $x \in X$, by setting

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

(2) By a *simple function* is meant a linear combination of characteristic functions of subsets of X; that is, f is a simple function if it can be written in the form

$$f = \sum_{i=1}^{n} a_i \, \chi_{A_i} , \tag{1.12}$$

where the numbers a_i can be real or complex, but only finite values are allowed (that is, $a_i \neq \pm\infty$).

The nonnegative, μ-measurable, simple functions are of particular interest for integration theory.

Lemma 1.3.9. *Let μ be a measure on the nonempty set X. If $f : X \to [0, \infty]$ is μ-measurable, then there exists a sequence of μ-measurable, simple functions $h_n : X \to [0, \infty]$, $n = 1, 2, \ldots$, such that*

(1) $0 \le h_1 \le h_2 \le \cdots \le f$, *and*
(2) $\lim_{n\to\infty} h_n = f(x)$, *for all $x \in X$.*

Proof. We can set

$$h_n = n\,\chi_{B_n} + \sum_{i=1}^{n2^n-1} i \cdot 2^{-n}\,\chi_{A_i}\,,$$

where $B_n = f^{-1}\Big([n, \infty]\Big)$, and

$$A_i = f^{-1}\Big([i \cdot 2^{-n},\ (i+1)\cdot 2^{-n})\Big), \quad i = 1, 2, \ldots, n2^n - 1\,. \qquad \square$$

Definition 1.3.10. Let μ be a measure on the nonempty set X. If $f : X \to \overline{\mathbb{R}}$ is μ-measurable, then the *integral of f with respect to μ* or, more simply, the *μ-integral of f* (or, more simply yet, the *integral of f* when the measure is clear from context) is denoted by

$$\int f\,d\mu = \int_X f(x)\,d\mu(x)$$

and is defined as follows:

(1) In case f is a nonnegative, simple function written as in (1.12) with each A_i μ-measurable, we set

$$\int f\,d\mu = \sum_{i=1}^{n} a_i\,\mu(A_i)\,. \tag{1.13}$$

(2) In case f is a nonnegative function, we set

$$\int f\,d\mu = \sup\left\{\int h\,d\mu : 0 \le h \le f,\ h \text{ simple, } \mu\text{-measurable}\right\}. \tag{1.14}$$

(3) In case at least one of $\int f^+\,d\mu$ and $\int f^-\,d\mu$ is finite, so that

$$\int f^+\,d\mu - \int f^-\,d\mu$$

is defined, we set

$$\int f\,d\mu = \int f^+\,d\mu - \int f^-\,d\mu\,. \tag{1.15}$$

(4) In case both $\int f^+ \, d\mu$ and $\int f^- \, d\mu$ are infinite, the quantity $\int f \, d\mu$ is undefined.

Definition 1.3.11.

(1) To integrate f over a subset A of X, we multiply f by the characteristic function of A, that is,

$$\int_A f \, d\mu = \int f \cdot \chi_A \, d\mu \, .$$

(2) The definition of $\int f \, d\mu$ extends to complex-valued, respectively $\mathbb{R}^N$-valued, functions by separating f into real and imaginary parts, respectively components, and combining the resulting real-valued integrals using linearity.

(3) If $\int |f| \, d\mu$ is finite, then we say that f is *μ-integrable* (or simply *integrable* if the measure μ is clear from context). In particular, f is μ-integrable if and only if $|f|$ is μ-integrable.

Remark 1.3.12.

(1) By a *Lebesgue integrable* function is meant an $\mathcal{L}^1$-integrable function in the terminology of Definition 1.3.11(3).

(2) The theories of Riemann integration and Lebesgue integration are connected by the following theorem:

A bounded, real-valued function on a closed interval is Riemann integrable if and only if the set of points at which the function is discontinuous has Lebesgue measure zero.

We will not prove this result. A proof can be found in [Fol 84, Theorem (2.28)].

(3) The reader should be aware that the terminology in [Fed 69] is different from that which we use: In [Fed 69] a function is said to be "μ integrable" if $\int f \, d\mu$ is defined, the values $+\infty$ and $-\infty$ being allowed, and "μ summable" if $\int |f| \, d\mu$ is finite.

The following basic facts hold for integration of nonnegative functions.

Theorem 1.3.13. *Let μ be a measure on the nonempty set X. Suppose $f, g : X \to [0, \infty]$ are μ-measurable.*

(1) *If $A \subseteq X$ is μ-measurable, and $f(x) = 0$ holds for μ-almost all $x \in A$, then*

$$\int_A f \, d\mu = 0 \, .$$

(2) *If $A \subseteq X$ is μ-measurable and $\mu(A) = 0$, then*

$$\int_A f \, d\mu = 0 \, .$$

(3) *If* $0 \le c < \infty$, *then*

$$\int (c \cdot f)\, d\mu = c \int f\, d\mu\,.$$

(4) *If* $f \le g$, *then*

$$\int f\, d\mu \le \int g\, d\mu\,.$$

(5) *If* $A \subseteq B \subseteq X$ *are* μ*-measurable, then*

$$\int_A f\, d\mu \le \int_B f\, d\mu\,.$$

Proof. Conclusions (1)–(4) are immediate from the definitions, and conclusion (5) follows from (4). □

Of course, it is essential that the equation $\int (f+g)\, d\mu = \int f\, d\mu + \int g\, d\mu$ hold. Unfortunately, this equation is not an immediate consequence of the definition. To prove it we need the next lemma, which is a weak form of Lebesgue's monotone convergence theorem.

Lemma 1.3.14. *Let* μ *be a measure on the nonempty set* X. *If* $f : X \to [0, \infty]$ *is* μ*-measurable and* $0 \le h_1 \le h_2 \le \cdots \le f$ *is a sequence of simple,* μ*-measurable functions with* $\lim_{n\to\infty} h_n = f$, *then*

$$\lim_{n\to\infty} \int h_n\, d\mu = \int f\, d\mu\,.$$

Proof. The inequality $\lim_{n\to\infty} \int h_n\, d\mu \le \int f\, d\mu$ is immediate from the definition of the integral.

To obtain the reverse inequality, let ℓ be an arbitrary simple, μ-measurable function with $0 \le \ell \le f$ and write

$$\ell = \sum_{i=1}^{k} a_i\, \chi_{A_i}\,,$$

where each A_i is μ-measurable. Let $c \in (0, 1)$ also be arbitrary.

For each $m \in \mathbb{N}$, set

$$E_m = \{\, x : c \cdot \ell(x) \le h_m(x) \,\} \text{ and } \ell_m = c \cdot \ell \cdot \chi_{E_m}\,.$$

For $m \le n$, we have $\ell_m \le h_n$, so applying Theorem 1.3.13(4), we obtain

$$\int \ell_m\, d\mu \le \lim_{n\to\infty} \int h_n\, d\mu\,.$$

Finally, we note that for each $i = 1, 2, \ldots, k$, the sets $A_i \cap E_m$ increase to A_i as $m \to \infty$, so $\mu(A_i) = \lim_{m\to\infty} \mu(A_i \cap E_m)$ and thus

$$c \int \ell\, d\mu = \int c \cdot \ell\, d\mu = \lim_{m\to\infty} \int \ell_m\, d\mu \le \lim_{n\to\infty} \int h_n\, d\mu\,.$$

The result follows from the arbitrariness of ℓ and c. □

Theorem 1.3.15. *Let μ be a measure on the nonempty set X. If $f, g : X \to [0, \infty]$ are μ-measurable, then*

$$\int (f + g)\, d\mu = \int f\, d\mu + \int g\, d\mu\,.$$

Proof. The result clearly holds if f and g are simple functions, and the general case then follows from Lemmas 1.3.9 and 1.3.14. □

Corollary 1.3.16. *The μ-integrable functions form a vector space, and the μ-integral is a linear functional on the space of μ-integrable functions.*

The decisive results for integration theory are Fatou's lemma[12] and the monotone and dominated convergence theorems of Lebesgue (see any of [Fol 84], [Roy 88], and [Rud 87]). In the development outlined above, it is easiest first to prove Lebesgue's monotone convergence theorem, arguing as in the proof of Lemma 1.3.14. Then one uses the monotone convergence theorem to prove Fatou's lemma and the dominated convergence theorem. We state these results next.

Theorem 1.3.17. *Let μ be a measure on the nonempty set X.*

(1) **[Fatou's lemma]** *If $f_1, f_2, \ldots$ are nonnegative μ-measurable functions, then*

$$\liminf_{n\to\infty} \int_X f_n\, d\mu \geq \int_X \liminf_{n\to\infty} f_n\ d\mu\,.$$

(2) **[Lebesgue's monotone convergence theorem]** *If $f_1 \leq f_2 \leq \cdots$ are nonnegative μ-measurable functions, then*

$$\lim_{n\to\infty} \int_X f_n\, d\mu = \int_X \lim_{n\to\infty} f_n\ d\mu\,.$$

(3) **[Lebesgue's dominated convergence theorem]** *Let $f_1, f_2, \ldots$ be complex-valued μ-measurable functions that converge μ-almost everywhere to f. If there exists a nonnegative μ-measurable function g such that*

$$\sup_n |f_n(x)| \leq g(x) \ \ \text{and} \ \ \int_X g\, d\mu < \infty\,,$$

then

$$\lim_{n\to\infty} \int_X |f_n - f|\, d\mu = 0 \ \ \text{and} \ \ \lim_{n\to\infty} \int_X f_n\, d\mu = \int_X f\, d\mu\,.$$

One of the beauties of measure theory is that we deal in analysis almost exclusively with measurable functions and sets, and the ordinary operations of analysis would never cause us to leave the realm of measurable functions and sets. However, in geometric measure theory it is occasionally necessary to deal with functions that either are nonmeasurable or are not known a priori to be measurable. In such situations, it is convenient to have a notion of upper and lower integral.

---
12 Pierre Joseph Louis Fatou (1878–1929).

Definition 1.3.18. Let μ be a measure on the nonempty set X and let $f : X \to [0, \infty]$ be defined μ-almost everywhere. We denote the *upper μ-integral of f* by

$$\overline{\int} f \, d\mu$$

and define it by setting

$$\overline{\int} f \, d\mu = \inf \left\{ \int \psi \, d\mu : 0 \leq f \leq \psi \ \text{ and } \psi \text{ is } \mu\text{-measurable} \right\}.$$

Similarly, the *lower μ-integral of f* is denoted by

$$\underline{\int} f \, d\mu$$

and defined by setting

$$\underline{\int} f \, d\mu = \sup \left\{ \int \phi \, d\mu : 0 \leq \phi \leq f \ \text{ and } \phi \text{ is } \mu\text{-measurable} \right\}.$$

Lemma 1.3.19. *If μ is a measure on the nonempty set X and $f, g : X \to [0, \infty]$ are defined μ-almost everywhere, then the following hold:*

(1) $\underline{\int} f \, d\mu \leq \overline{\int} f \, d\mu$,

(2) *if* $f \leq g$, *then* $\underline{\int} f \, d\mu \leq \underline{\int} g \, d\mu$ *and* $\overline{\int} f \, d\mu \leq \overline{\int} g \, d\mu$,

(3) *if f is μ-measurable, then* $\underline{\int} f \, d\mu = \int f \, d\mu = \overline{\int} f \, d\mu$,

(4) *if* $0 \leq c$, *then* $\underline{\int} cf \, d\mu = c \underline{\int} f \, d\mu$ *and* $\overline{\int} cf \, d\mu = c \overline{\int} f \, d\mu$,

(5) $\underline{\int} f \, d\mu + \underline{\int} g \, d\mu \leq \underline{\int} (f + g) \, d\mu$ *and* $\overline{\int} (f + g) \, d\mu \leq \overline{\int} f \, d\mu + \overline{\int} g \, d\mu$.

The lemma follows easily from the definitions.

Proposition 1.3.20. *Suppose $f : X \to [0, \infty]$ satisfies $\overline{\int} f \, d\mu < \infty$. For such a function,*

$$\underline{\int} f \, d\mu = \overline{\int} f \, d\mu$$

holds if and only if f is μ-measurable.

Proof. Suppose the upper and lower μ-integrals of f are equal. Choose sequences of μ-measurable functions $g_1 \leq g_2 \leq \cdots \leq f$ and $h_1 \geq h_2 \geq \cdots \geq f$ with

$$\lim_{n\to\infty} \int g_n \, d\mu = \underline{\int} f \, d\mu = \overline{\int} f \, d\mu = \lim_{n\to\infty} \int h_n \, d\mu \, .$$

Then $g = \lim_{n\to\infty} g_n$ and $h = \lim_{n\to\infty} h_n$ are μ-measurable with $g \le f \le h$. Since, by Lebesgue's dominated convergence theorem, the μ-integrals of g and h are equal, we see that g and h must be μ-almost equal to each other, and thus μ-almost equal to f. □

1.3.3 Lebesgue Spaces

Definition 1.3.21. Fix $1 \le p \le \infty$. Let μ be a measure on the nonempty set X. The *Lebesgue space* $L^p(\mu)$ (or simply L^p if the choice of the measure is clear from context) is the vector space of μ-measurable, complex-valued functions satisfying

$$\|f\|_p < \infty \, ,$$

where $\|f\|_p$ is defined by setting

$$\|f\|_p = \begin{cases} \left(\int |f|^p \, d\mu \right)^{1/p}, & \text{if } p < \infty, \\ \inf \left\{ t : \mu\Big(X \cap \{ x : |f(x)| > t \} \Big) = 0 \right\}, & \text{if } p = \infty. \end{cases}$$

The elements of L^p are called L^p *functions*. Of course, the L^1 functions are just the μ-integrable functions. The L^2 functions are also called *square integrable functions*, and, for $1 \le p < \infty$, the L^p functions are also called *p-integrable functions*.

Remark 1.3.22.

(1) A frequently used tool in analysis is Hölder's inequality[13]

$$\int fg \, d\mu \le \|f\|_p \, \|g\|_q \, ,$$

where f and g are μ-measurable, $1 < p < \infty$, and $1/p + 1/q = 1$. We note that Hölder's inequality is also valid when the integrals are replaced by upper integrals. The proof of this generalization makes use of Lemma 1.3.19(2)5).

(2) The functional $\|\cdot\|_p$ is called the L^p*-norm*. In the cases $p = 1$ and $p = \infty$, it is easy to verify that the L^p-norm is, in fact, a norm, but for the case $1 < p < \infty$, this fact is a consequence of Minkowski's inequality[14]

$$\|f + g\|_p \le \|f\|_p + \|g\|_p \, .$$

(3) Much of the importance of the Lebesgue spaces stems from the discovery that L^p, $1 \le p < \infty$, is a complete metric space. This result is sometimes (for instance in [Roy 88]) called the Riesz–Fischer theorem.[15]

---

[13] Otto Ludwig Hölder (1859–1937).

[14] Hermann Minkowski (1864–1909).

[15] Frigyes Riesz (1880–1956), Ernst Sigismund Fischer (1875–1954).

1.3.4 Product Measures and the Fubini–Tonelli Theorem

Definition 1.3.23. Let μ be a measure on the nonempty set X and let ν be a measure on the nonempty set Y. The *Cartesian product of the measures* μ *and* ν is denoted $\mu \times \nu$ and is defined by setting

$$(\mu \times \nu)(E) = \inf \left\{ \sum_{i=1}^{\infty} \mu(A_i) \cdot \nu(B_i) : E \subseteq \bigcup_{i=1}^{\infty} A_i \times B_i, \right.$$
$$A_i \subseteq X \text{ is } \mu\text{-measurable, for } i = 1, 2, \ldots,$$
$$\left. B_i \subseteq Y \text{ is } \nu\text{-measurable, for } i = 1, 2, \ldots \right\}. \qquad (1.16)$$

It is immediately verified that $\mu \times \nu$ is a measure on $X \times Y$. Clearly the inequality

$$(\mu \times \nu)(A \times B) \leq \mu(A) \cdot \nu(B)$$

holds whenever $A \subseteq X$ is μ-measurable and $B \subseteq Y$ is ν-measurable. The product measure $\mu \times \nu$ is the largest measure satisfying that condition.

One of the main concerns in using product measures is justifying the interchange of the order of integration in a multiple integral. The next example illustrates a situation in which the order of integration in a double integral cannot be interchanged.

Example 1.3.24. The *counting measure on* X is defined by setting

$$\mu(E) = \begin{cases} \operatorname{card}(E) & \text{if } E \text{ is finite,} \\ \infty & \text{otherwise,} \end{cases}$$

for $E \subseteq X$. If ν is another measure on X for which $0 < \nu(X)$ and $\nu(\{x\}) = 0$ for each $x \in X$, and if $f : X \times X \to [0, \infty]$ is the characteristic function of the diagonal, that is,

$$f(x_1, x_2) = \begin{cases} 1 & \text{if } x_1 = x_2, \\ 0 & \text{otherwise,} \end{cases}$$

then

$$\int \left(\int f(x_1, x_2)\, d\mu(x_1) \right) d\nu(x_2) = \int 1\, d\nu = \nu(X) > 0\,,$$

but

$$\int \left(\int f(x_1, x_2)\, d\nu(x_2) \right) d\mu(x_1) = \int 0\, d\mu = 0\,.$$

□

To avoid the phenomenon in the preceding example we introduce a definition.

Definition 1.3.25. Let μ be a measure on the nonempty set X. We say that μ is σ*-finite* if X can be written as a countable union of μ-measurable sets each having finite μ measure.

The main facts about product measures, which often *do* allow the interchange of the order of integration, are stated in the next theorem. We refer the reader to any of [Fol 84], [Roy 88], and [Rud 87].

Theorem 1.3.26. *Let μ be a σ-finite measure on the nonempty set X and let ν be a σ-finite measure on the nonempty set Y.*

(1) *If $A \subseteq X$ is μ-measurable and $B \subseteq Y$ is ν-measurable, then $A \times B$ is $(\mu \times \nu)$-measurable and*

$$(\mu \times \nu)(A \times B) = \mu(A) \cdot \nu(B)\,.$$

(2) **(Tonelli's[16] theorem)** *If $f : X \times Y \to [0, \infty]$ is $(\mu \times \nu)$-measurable, then*

$$g(x) = \int f(x, y)\, d\nu(y) \tag{1.17}$$

defines a μ-measurable function on X,

$$h(y) = \int f(x, y)\, d\mu(x) \tag{1.18}$$

defines a ν-measurable function on Y, and

$$\begin{aligned} \int f\, d(\mu \times \nu) &= \int \left(\int f(x, y)\, d\mu(x) \right) d\nu(y) \\ &= \int \left(\int f(x, y)\, d\nu(y) \right) d\mu(x)\,. \end{aligned} \tag{1.19}$$

(3) **(Fubini's[17] theorem)** *If f is $(\mu \times \nu)$-integrable, then*
 (a) *$\phi(x) \equiv f(x, y)$ is μ-integrable for ν-almost every $y \in Y$,*
 (b) *$\psi(y) \equiv f(x, y)$ is ν-integrable for μ-almost every $x \in X$,*
 (c) *$g(x)$ defined by* (1.17) *is a μ-integrable function on X,*
 (d) *$h(y)$ defined by* (1.18) *is a ν-integrable function on Y, and*
 (e) *equation* (1.19) *holds.*

Definition 1.3.27. The *N-dimensional Lebesgue measure on $\mathbb{R}^N$*, denoted by $\mathcal{L}^N$, is defined inductively by setting $\mathcal{L}^N = \mathcal{L}^{N-1} \times \mathcal{L}^1$.

1.4 The Exterior Algebra

In an introductory vector calculus course, a vector is typically described as representing a direction and a magnitude, that is, an oriented line and a length. When later an oriented plane and an area in that plane are to be represented, a direction orthogonal to the plane and a length equal to the desired area are often used. This last device is viable only for $(N-1)$-dimensional oriented planes in N-dimensional space, because the complementary dimension must be 1. For the general case of an

[16] Leonida Tonelli (1885–1946).
[17] Guido Fubini (1879–1943).

oriented m-dimensional plane and an m-dimensional area in $\mathbb{R}^N$, some new idea must be invoked.

The straightforward way to represent an oriented m-dimensional plane in $\mathbb{R}^N$ is to specify an ordered m-tuple of independent vectors parallel to (contained in) the plane. To simultaneously represent an m-dimensional area in that plane, choose the vectors so that the m-dimensional area of the parallelepiped they determine equals that given m-dimensional area. Of course, a given oriented m-dimensional plane and m-dimensional area can be represented equally well by many different ordered m-tuples of vectors, and identifying any two such ordered m-tuples introduces an equivalence relation on the ordered m-tuples of vectors. To facilitate computation and understanding, the equivalence classes of ordered m-tuples are overlaid with a vector space structure. The result is the alternating algebra of m-vectors in $\mathbb{R}^N$. We now proceed to a formal definition.

Definition 1.4.1.

(1) Define an equivalence relation $\sim$ on

$$\left(\mathbb{R}^N\right)^m = \underbrace{\mathbb{R}^N \times \mathbb{R}^N \times \cdots \times \mathbb{R}^N}_{m \text{ factors}}$$

by requiring, for all $\alpha \in \mathbb{R}$ and $1 \le i < j \le m$,

(a)

$(u_1, \ldots, \alpha\, u_i, \ldots, u_j, \ldots, u_m) \sim (u_1, \ldots, u_i, \ldots, \alpha\, u_j, \ldots, u_m)$,

(b)

$(u_1, \ldots, u_i, \ldots, u_j, \ldots, u_m) \sim (u_1, \ldots, u_i + \alpha\, u_j, \ldots, u_j, \ldots, u_m)$,

(c)

$(u_1, \ldots, u_i, \ldots, u_j, \ldots, u_m) \sim (u_1, \ldots, -u_j, \ldots, u_i, \ldots, u_m)$,

and extending the resulting relation to be symmetric and transitive.

(2) The equivalence class of $(u_1, u_2, \ldots, \ldots, u_m)$ under $\sim$ is denoted by $u_1 \wedge u_2 \wedge \cdots \wedge u_m$. We call $u_1 \wedge u_2 \wedge \cdots \wedge u_m$ a *simple m-vector*.

(3) On the vector space of formal linear combinations of simple m-vectors, we define the equivalence relation $\approx$ by extending the relation defined by requiring

(a) $\alpha(u_1 \wedge u_2 \wedge \cdots \wedge u_m) \approx (\alpha u_1) \wedge u_2 \wedge \cdots \wedge u_m$,

(b) $(u_1 \wedge u_2 \wedge \cdots \wedge u_m) + (v_1 \wedge u_2 \wedge \cdots \wedge u_m) \approx (u_1 + v_1) \wedge u_2 \wedge \cdots \wedge u_m$.

(4) The equivalence classes of formal linear combinations of simple m-vectors under the relation $\approx$ are the *m-vectors in* $\mathbb{R}^N$. The vector space of m-vectors in $\mathbb{R}^N$ is denoted by $\bigwedge_m(\mathbb{R}^N)$.

(5) The *exterior algebra of* $\mathbb{R}^N$, denoted by $\bigwedge_*(\mathbb{R}^N)$, is the direct sum of the $\bigwedge_m(\mathbb{R}^N)$ together with the *exterior multiplication* defined by linearly extending the definition

$$(u_1 \wedge u_2 \wedge \cdots \wedge u_\ell) \wedge (v_1 \wedge v_2 \wedge \cdots \wedge v_m) = u_1 \wedge u_2 \wedge \cdots \wedge u_\ell \wedge v_1 \wedge v_2 \wedge \cdots \wedge v_m\,.$$

Remark 1.4.2.

(1) When $m = 1$, Definition 1.4.1(1) is vacuous, so $\bigwedge_1 (\mathbb{R}^N)$ is isomorphic to, and will be identified with, $\mathbb{R}^N$. If the vectors $u_1, u_2, \ldots, u_m$ are linearly dependent, then $u_1 \wedge u_2 \wedge \cdots \wedge u_m$ is the additive identity in $\bigwedge_m (\mathbb{R}^N)$, so we write $u_1 \wedge u_2 \wedge \cdots \wedge u_m = 0$. Consequently, when $N < m$, $\bigwedge_m (\mathbb{R}^N)$ is the trivial vector space containing only 0.

(2) As an exercise, the reader should convince himself that $\mathbf{e}_1 \wedge \mathbf{e}_2 + \mathbf{e}_3 \wedge \mathbf{e}_4 \in \bigwedge_2 (\mathbb{R}^4)$ is not a simple 2-vector.

For a nontrivial simple m-vector $u_1 \wedge u_2 \wedge \cdots \wedge u_m$ in $\mathbb{R}^N$, the *associated subspace* is that subspace spanned by the vectors $u_1, u_2, \ldots, u_m$. It is evident from Definition 1.4.1(1) that if $u_1 \wedge u_2 \wedge \cdots \wedge u_m = \pm v_1 \wedge v_2 \wedge \cdots \wedge v_m$, then their associated subspaces are equal. We assert that if $u_1 \wedge u_2 \wedge \cdots \wedge u_m = \pm v_1 \wedge v_2 \wedge \cdots \wedge v_m$, then also the m-dimensional area of the parallelepiped determined by $u_1, u_2, \ldots, u_m$ is equal to the m-dimensional area of the parallelepiped determined by $v_1, v_2, \ldots, v_m$. To see this last fact, we need the next proposition, which gives us a way to compute the m-dimensional areas in question. The proof is based on [Por 96].

Proposition 1.4.3. *Let $u_1, u_2, \ldots, u_m$ be vectors in $\mathbb{R}^N$. Then the parallelepiped determined by those vectors has m-dimensional area*

$$\sqrt{\det\left(U^{\mathrm{t}}\, U\right)}, \tag{1.20}$$

where U is the $N \times m$ matrix with $u_1, u_2, \ldots, u_m$ as its columns.

Proof. If the vectors $u_1, u_2, \ldots, u_m$ are pairwise orthogonal, then the result is immediate. Thus we will reduce the general case to this special case.

Notice that Cavalieri's principle[18] shows us that adding a multiple of u_j to another vector u_i, $i \neq j$, does not change the m-dimensional area of the parallelepiped determined by the vectors. But also notice that such an operation on the vectors u_i is equivalent to multiplying U on the right by an $m \times m$ triangular matrix with 1's on the diagonal. The Gram–Schmidt orthogonalization procedure[19] is effected by a sequence of operations of precisely this type. Thus we see that there is an upper triangular matrix A with 1's on the diagonal such that UA has orthogonal columns and the columns of UA determine a parallelepiped with the same m-dimensional area as the parallelepiped determined by $u_1, u_2, \ldots, u_m$. Since the columns of UA are orthogonal, we know that $\sqrt{\det\left((UA)^{\mathrm{t}}\,(UA)\right)}$ equals the m-dimensional area of the parallelepiped determined by its columns, and thus equals the m-dimensional area of the parallelepiped determined by $u_1, u_2, \ldots, u_m$. Finally, we compute

$$\begin{aligned} \det\left((UA)^{\mathrm{t}}\,(UA)\right) &= \det\left(A^{\mathrm{t}}\, U^{\mathrm{t}}\, U\, A\right) \\ &= \det\left(A^{\mathrm{t}}\right)\, \det\left(U^{\mathrm{t}}\, U\right) \det(A) \\ &= \det\left(U^{\mathrm{t}}\, U\right). \end{aligned}$$

□

[18] Bonaventura Francesco Cavalieri (1598–1647).

[19] Jørgen Pedersen Gram (1850–1916).

Corollary 1.4.4. *If $u_1, u_2, \ldots, u_m$ and $v_1, v_2, \ldots, v_m$ are vectors in $\mathbb{R}^N$ with*

$$u_1 \wedge u_2 \wedge \cdots \wedge u_m = \pm v_1 \wedge v_2 \wedge \cdots \wedge v_m\,,$$

then the m-dimensional area of the parallelepiped determined by the vectors $u_1, u_2, \ldots, u_m$ equals the m-dimensional area of the parallelepiped determined by the vectors $v_1, v_2, \ldots, v_m$.

Proof. We consider the m-tuples of vectors on the left-hand and right-hand sides of Definition 1.4.1(1a,b,c). Let U_l be the matrix whose columns are the vectors on the left-hand side and let U_r be the matrix whose columns are the vectors on the right-hand side. For (a), we have $U_r = U_l A$, where A is the $m \times m$ diagonal matrix with $1/\alpha$ in the ith column and α in the jth column. For (b), we have $U_r = U_l A$, where A is an $m \times m$ triangular matrix with 1's on the diagonal. For (c), we have $U_r = U_l A$, where A is an $m \times m$ permutation matrix with one of its 1's replaced by -1. In all three cases, $\det(A) = \pm 1$, and the result follows. □

For computational purposes, it is often convenient to use the basis

$$\mathbf{e}_{i_1} \wedge \mathbf{e}_{i_2} \wedge \cdots \wedge \mathbf{e}_{i_m}\,, \quad 1 \le i_1 < i_2 < \cdots < i_m \le N\,, \tag{1.21}$$

for $\bigwedge_m(\mathbb{R}^N)$. Specifying that the m-vectors in (1.21) are orthonormal induces the *standard inner product on* $\bigwedge_m(\mathbb{R}^N)$. The *exterior product* (sometimes called the *wedge product*)

$$\wedge : \textstyle\bigwedge_\ell(\mathbb{R}^N) \times \bigwedge_m(\mathbb{R}^N) \to \bigwedge_{\ell+m}(\mathbb{R}^N)$$

is an anticommutative, multilinear multiplication. Any linear $F : \mathbb{R}^N \to \mathbb{R}^P$ extends to a linear map $F_m : \bigwedge_m(\mathbb{R}^N) \to \bigwedge_m(\mathbb{R}^P)$ by defining

$$F_m(u_1 \wedge u_2 \wedge \cdots \wedge u_m) = F(u_1) \wedge F(u_2) \wedge \cdots \wedge F(u_m)\,.$$

1.5 The Generalized Pythagorean Theorem

The generalized Pythagorean theorem (Theorem 1.5.2 below) tells us that for a figure Σ lying in an m-dimensional affine subspace of $\mathbb{R}^N$, the square of the m-dimensional area of Σ equals the sum of the squares of the m-dimensional areas of the orthogonal projections of Σ onto all possible coordinate m-planes. For conceptual simplicity, we will restrict our attention to polyhedral figures Σ. We consider a few instances of this theorem:

- If $m = 1$ and Σ is a line segment, then the generalized Pythagorean theorem tells us that the square of the length of the segment is the sum of the squares of the lengths in each of the coordinate directions; that is, we recover the usual Pythagorean theorem.

- Suppose Σ is the parallelepiped generated by the m vectors $u_1, \ldots, u_m$ and U is the matrix whose columns are $u_1, \ldots, u_m$. Then the (signed) m-dimensional area of each projection of Σ onto a coordinate m-plane is given by an m-by-m minor determinant of U. Proposition 1.4.3 tells us that the m-dimensional area of Σ equals $\sqrt{\det(U^{\mathtt{t}}\, U)}$. Thus the generalized Pythagorean theorem implies—and, in fact, is equivalent to—the nontrivial fact that

$$\det(U^{\mathtt{t}}\, U) = \sum_{\lambda} [\det(U_\lambda)]^2 \tag{1.22}$$

  holds, where in (1.22) the summation extends over all $\lambda = \{i_1, \ldots, i_m\} \subseteq \{1, \ldots, N\}$ and where for each such λ, U_λ is the m-by-m submatrix whose rows are the rows numbered $i_1, \ldots, i_m$ in U.
- If Σ is an m-dimensional simplex in $\mathbb{R}^N$, then Σ automatically lies in an m-dimensional affine subspace of $\mathbb{R}^N$, and the generalized Pythagorean theorem applies to Σ. Figure 1.1 illustrates this situation when Σ is a triangle in $\mathbb{R}^3$. We have used A to denote the area of the triangle and A_{ij} to denote the area of the projection of the triangle onto the (x_i, x_j)-coordinate plane.

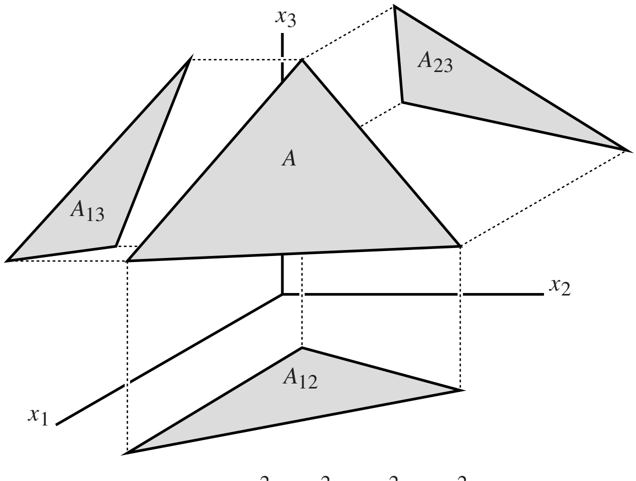

Fig. 1.1. $A^2 = A_{12}^2 + A_{13}^2 + A_{23}^2$.

In this section, we will give a geometrical proof of the generalized Pythagorean theorem. In particular, the proof will make no use of determinants. The main computation in the proof is made by applying the divergence theorem of advanced calculus to a constant vector field, while our other primary tool is the fact that the m-dimensional area of a figure is unchanged when the figure is mapped by an isometry.

Notation 1.5.1.

(1) Any m-dimensional polyhedral figure can be written as the union of m-dimensional simplices that intersect only in their boundaries. Thus, to prove the generalized Pythagorean theorem, it is sufficient to prove it when $\Sigma \subseteq \mathbb{R}^N$ is an m-simplex. Accordingly we will assume throughout the remainder of this section that Σ is the m-dimensional simplex determined by the $m+1$ points $u_0, \ldots, u_m$.
(2) We will denote the m-dimensional area of Σ by A.
(3) If $\lambda \subseteq \{1, 2, \ldots, N\}$ and $\mathrm{card}(\lambda) = K$, then $\Pi_\lambda : \mathbb{R}^N \to \mathbb{R}^K$ will be the orthogonal projection given by

$$\Pi_\lambda(x_1, x_2, \ldots, x_N) = (x_{i_1}, x_{i_2}, \ldots, x_{i_K}),$$

where $\lambda = \{i_1, i_2, \ldots, i_K\}$ and $i_1 < i_2 < \cdots < i_K$. We will need only the two cases $K = m$ and $K = 2$.
(4) If $\lambda \subseteq \{1, 2, \ldots, N\}$ and $\mathrm{card}(\lambda) = m$, let A_λ denote the m-dimensional area of $\Pi_\lambda(\Sigma)$. We will sometimes abuse this notation (as we did in Figure 1.1) by writing $A_{i_1,i_2,\ldots,i_K}$ instead of the more pedantic $A_{\{i_1,i_2,\ldots,i_K\}}$.
(5) Since a set $\lambda \subseteq \{1, 2, \ldots, m+1\}$ with $\mathrm{card}(\lambda) = m$ is most easily described by the one element it omits, we will write

$$A_{\hat{\imath}} = A_{1,\ldots,i-1,i+1,\ldots,m+1}.$$

Using the notation given above, we can state our result as follows:

Theorem 1.5.2 (Generalized Pythagorean theorem). *If Σ is an m-dimensional simplex in $\mathbb{R}^N$, then it holds that*

$$A^2 = \sum_{\substack{\lambda \subseteq \{1,\ldots,N\} \\ \mathrm{card}\,(\lambda)=m}} A_\lambda^2 . \tag{1.23}$$

Note that if $N = m$, the theorem is trivial. We first give a proof of the theorem in the case $N = m + 1$.

The Codimension-One Case, $N = m + 1$

Our proof for the case $N = m + 1$ will be based on an application of the divergence theorem.

Proposition 1.5.3. *Let Σ be an m-simplex in $\mathbb{R}^{m+1}$ with m-dimensional area A. Let $\mathbf{n}_0$ be a unit vector normal to Σ. Then*

$$A\,|\mathbf{n}_0 \cdot \mathbf{e}_i| = A_{\hat{\imath}}$$

holds for $i = 1, \ldots, m + 1$.

Proof. We may assume for convenience that $i = m + 1$. If $\mathbf{n}_0 \cdot \mathbf{e}_{m+1} = 0$, then the result is trivial, so we also may assume that $\mathbf{n}_0 \cdot \mathbf{e}_{m+1} > 0$, i.e., $\mathbf{n}_0$ points "up."

By translating Σ if necessary, we may assume that all the coordinates of all the points in Σ are positive. Consider the closed polyhedral cylinder $\mathcal{C}$ made up of the line segments connecting each point of Σ with its projection on the $(x_1, \ldots, x_m)$-coordinate hyperplane; that is,

$$\mathcal{C} = \left\{ (1-t)\,x + t\,\Pi_{1,\ldots,m}(x) : x \in \Sigma,\ 0 \leq t \leq 1 \right\}$$

(Figure 1.2 illustrates $\mathcal{C}$ in the case $m = 2$). It will be convenient to call Σ the "top" of $\mathcal{C}$ and to call $B \equiv \Pi_{1,\ldots,m}(\Sigma)$ the "bottom" of $\mathcal{C}$.

Note that except on the top and bottom of $\mathcal{C}$, the outward unit normal to $\partial\mathcal{C}$ is orthogonal to $\mathbf{e}_{m+1}$. On the top of $\mathcal{C}$ the outward unit normal to $\mathcal{C}$ equals $\mathbf{n}_0$, and on the bottom of $\mathcal{C}$ the outward unit normal to $\mathcal{C}$ equals $-\mathbf{e}_{m+1}$ (see Figure 1.2).

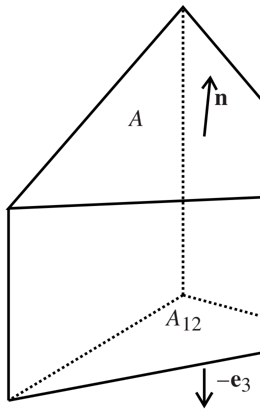

Fig. 1.2. Applying the divergence theorem.

The divergence theorem tells us that if $\mathbf{w}$ is a C^1 vector field on $\mathcal{C}$, then

$$\int_{\partial\mathcal{C}} \mathbf{w} \cdot \mathbf{n}\, d\sigma = \int_{\mathcal{C}} \operatorname{div} \mathbf{w}\, dV$$

holds, where $\mathbf{n}$ is the outward unit normal vector to $\partial\mathcal{C}$, $d\sigma$ is the element of m-dimensional area on $\partial\mathcal{C}$, and dV is the element of $(m+1)$-dimensional volume in $\mathcal{C}$.

Applying the divergence theorem to the constant vector field $\mathbf{w} \equiv \mathbf{e}_{m+1}$ on $\mathcal{C}$, we obtain

$$0 = \int_{\mathcal{C}} \operatorname{div} \mathbf{w}\, dV = \int_{\partial\mathcal{C}} \mathbf{w} \cdot \mathbf{n}\, d\sigma = A\, \mathbf{n}_0 \cdot \mathbf{e}_{m+1} - A_{\widehat{m+1}},$$

and the result follows. □

Corollary 1.5.4. *The generalized Pythagorean theorem holds when $N = m + 1$.*

Proof. Let $\mathbf{n}_0$ be a unit vector normal to $\Sigma \subseteq \mathbb{R}^{m+1}$. Since $\mathbf{n}_0$ is a unit vector, Proposition 1.5.3 gives us

$$A^2 = A^2 \sum_{i=1}^{m+1} (\mathbf{n}_0 \cdot \mathbf{e}_i)^2 = \sum_{i=1}^{m+1} A^2 (\mathbf{n}_0 \cdot \mathbf{e}_i)^2 = \sum_{i=1}^{m+1} A_{\hat{\imath}}^2 . \qquad \square$$

The Higher Codimension Case, $N \geq m + 2$

Definition 1.5.5. By a *coordinate-plane rotation of* $\mathbb{R}^N$ we will mean a linear transformation that for some $i < j$, rotates the (x_i, x_j)-plane while leaving the remaining $(N - 2)$ coordinates unchanged. We will call x_i and x_j the *rotated coordinates*.

Our strategy for completing the proof of the generalized Pythagorean theorem is to show that the result holds for Σ if and only if it holds for the image of Σ under a coordinate-plane rotation. We then show that a sequence of coordinate-plane rotations of Σ will move Σ into an m-dimensional plane parallel to a coordinate m-plane—a situation in which the generalized Pythagorean theorem holds trivially.

Notation 1.5.6.

(1) Suppose $F : \mathbb{R}^N \to \mathbb{R}^N$ is a linear transformation. We set

$$\widetilde{\Sigma} = F(\Sigma) .$$

For $\lambda \subseteq \{1, 2, \ldots, N\}$ with $\operatorname{card}(\lambda) = m$, $\widetilde{A}_\lambda$ will denote the m-dimensional area of $\Pi_\lambda(\widetilde{\Sigma})$. Similarly, when $N = m + 1$, we will use the notation $\widetilde{A}_{\hat{\imath}}$.

(2) For each positive integer K, we let $\mathbf{I}_{\mathbb{R}^K}$ be the identity map on $\mathbb{R}^K$.

Lemma 1.5.7. *Let* $F = \mathcal{R} \times \mathbf{I}_{\mathbb{R}^{N-2}}$, *where* $\mathcal{R} : \mathbb{R}^2 \to \mathbb{R}^2$ *is a rotation. Suppose* $\lambda \subseteq \{1, 2, \ldots, N\}$ *with* $\operatorname{card}(\lambda) = m$. *If*

$$\textit{either} \quad \{1, 2\} \cap \lambda = \emptyset \quad \textit{or} \quad \{1, 2\} \cap \lambda = \{1, 2\} ,$$

then $A_\lambda = \widetilde{A}_\lambda$.

Proof. When $\{1, 2\} \cap \lambda = \emptyset$ holds, we have

$$\Pi_\lambda(\Sigma) = \Pi_\lambda(\widetilde{\Sigma}) ,$$

so the result is trivial in this case.

Now suppose that $\{1, 2\} \subseteq \lambda$. Then we have

$$\Pi_\lambda \circ F = \Pi_\lambda \circ (\mathcal{R} \times \mathbf{I}_{\mathbb{R}^{N-2}}) = (\mathcal{R} \times \mathbf{I}_{\mathbb{R}^{m-2}}) \circ \Pi_\lambda ,$$

and the result follows because $\mathcal{R} \times \mathbf{I}_{\mathbb{R}^{m-2}}$ is an isometry. $\square$

In Lemma 1.5.7, we considered projections Π_λ such that λ either included the indices of both rotated coordinates or omitted the indices of both rotated coordinates. In contrast, the m-dimensional area of the projection is *not preserved* when λ includes *exactly one* of the indices of the rotated coordinates. But we do have the next result.

Lemma 1.5.8. *Let* $F = \mathcal{R} \times \mathbf{I}_{\mathbb{R}^{N-2}}$, *where* $\mathcal{R} : \mathbb{R}^2 \to \mathbb{R}^2$ *is a rotation. If* $\lambda' \subseteq \{3, 4, \ldots, N\}$ *with* $\operatorname{card}(\lambda') = m - 1$, *then*

$$A^2_{\{1\}\cup\lambda'} + A^2_{\{2\}\cup\lambda'} = \widetilde{A}^2_{\{1\}\cup\lambda'} + \widetilde{A}^2_{\{2\}\cup\lambda'} . \tag{1.24}$$

Proof. For notational convenience, suppose that

$$\lambda' = \{3, 4, \ldots, m+1\} .$$

Each summand in (1.24) is unchanged if Σ is replaced by its projection into $\mathbb{R}^{m+1}$, so we may and shall assume that $N = m + 1$.

We have already shown that the generalized Pythagorean theorem holds when $N = m + 1$, so we can apply that theorem to $\Sigma \subseteq \mathbb{R}^{m+1}$ and to $\widetilde{\Sigma} \subseteq \mathbb{R}^{m+1}$. Using also the fact that $A = \widetilde{A}$ (which holds because F is an isometry), we obtain

$$\sum_{i=1}^{m+1} A^2_{\widehat{i}} = A^2 = \widetilde{A}^2 = \sum_{i=1}^{m+1} \widetilde{A}^2_{\widehat{i}} .$$

Observe that

$$\sum_{i=1}^{m+1} A^2_{\widehat{i}} = A^2_{\lambda'\cup\{1\}} + A^2_{\lambda'\cup\{2\}} + \sum_{\substack{\lambda''\subseteq\lambda' \\ \operatorname{card}(\lambda'')=m-2}} A^2_{\lambda''\cup\{1,2\}}$$

and, likewise, that

$$\sum_{i=1}^{m+1} \widetilde{A}^2_{\widehat{i}} = \widetilde{A}^2_{\lambda'\cup\{1\}} + \widetilde{A}^2_{\lambda'\cup\{2\}} + \sum_{\substack{\lambda''\subseteq\lambda' \\ \operatorname{card}(\lambda'')=m-2}} \widetilde{A}^2_{\lambda''\cup\{1,2\}} .$$

Lemma 1.5.7 tells us that for each $\lambda'' \subseteq \lambda'$ with $\operatorname{card}(\lambda'') = m - 2$,

$$A_{\lambda''\cup\{1,2\}} = \widetilde{A}_{\lambda''\cup\{1,2\}}$$

holds, so the result follows. □

In Lemmas 1.5.7 and 1.5.8, we considered a rotation $\mathcal{R}$ in the (x_1, x_2)-plane merely for convenience of notation. By relabeling coordinates, we see that the following result holds.

Proposition 1.5.9. *Suppose* $F : \mathbb{R}^N \to \mathbb{R}^N$ *rotates the* (x_i, x_j)*-plane while leaving all the other coordinates unchanged (here* $i < j$*).*

(1) *If* $\lambda \subseteq \{1, 2, \ldots, N\}$ *with* $\operatorname{card}(\lambda) = m$ *and if*

$$\text{either} \quad \{i, j\} \cap \lambda = \emptyset \quad \text{or} \quad \{i, j\} \cap \lambda = \{i, j\} ,$$

then $A_\lambda = \widetilde{A}_\lambda$.

(2) *If* $\lambda' \subseteq \{1, 2, \ldots, N\}$ *with* $\mathrm{card}(\lambda') = m - 1$ *and if*

$$\{i, j\} \cap \lambda' = \emptyset ,$$

then

$$A^2_{\{i\}\cup\lambda'} + A^2_{\{j\}\cup\lambda'} = \widetilde{A}^2_{\{i\}\cup\lambda'} + \widetilde{A}^2_{\{j\}\cup\lambda'} .$$

In the next result, we show that the generalized Pythagorean theorem holds for Σ if and only if it holds for the image of Σ under a coordinate-plane rotation.

Corollary 1.5.10. *If* $F : \mathbb{R}^N \to \mathbb{R}^N$ *rotates the* (x_i, x_j)*-plane while leaving all the other coordinates unchanged (here* $i < j$*), then we have* $A = \widetilde{A}$ *and*

$$\sum_{\substack{\lambda\subseteq\{1,\ldots,N\}\\ \mathrm{card}\,(\lambda)=m}} A^2_\lambda = \sum_{\substack{\lambda\subseteq\{1,\ldots,N\}\\ \mathrm{card}\,(\lambda)=m}} \widetilde{A}^2_\lambda .$$

Consequently, the generalized Pythagorean theorem holds for Σ *if and only if it holds for* $\widetilde{\Sigma}$.

Proof. Observe that

$$\sum_{\substack{\lambda\subseteq\{1,\ldots,N\}\\ \mathrm{card}\,(\lambda)=m}} A^2_\lambda = \sum_{\substack{\lambda\subseteq\{1,\ldots,N\}\\ \mathrm{card}\,(\lambda)=m,\ \lambda\cap\{i,j\}=\emptyset}} A^2_\lambda + \sum_{\substack{\lambda\subseteq\{1,\ldots,N\}\\ \mathrm{card}\,(\lambda)=m,\ \lambda\cap\{i,j\}=\{i,j\}}} A^2_\lambda$$
$$+ \sum_{\substack{\lambda'\subseteq\{1,\ldots,N\}\\ \mathrm{card}\,(\lambda')=m-1,\ \lambda'\cap\{i,j\}=\emptyset}} \left(A^2_{\lambda'\cup\{i\}} + A^2_{\lambda'\cup\{j\}}\right)$$

and, likewise, that

$$\sum_{\substack{\lambda\subseteq\{1,\ldots,N\}\\ \mathrm{card}\,(\lambda)=m}} \widetilde{A}^2_\lambda = \sum_{\substack{\lambda\subseteq\{1,\ldots,N\}\\ \mathrm{card}\,(\lambda)=m,\ \lambda\cap\{i,j\}=\emptyset}} \widetilde{A}^2_\lambda + \sum_{\substack{\lambda\subseteq\{1,\ldots,N\}\\ \mathrm{card}\,(\lambda)=m,\ \lambda\cap\{i,j\}=\{i,j\}}} \widetilde{A}^2_\lambda$$
$$+ \sum_{\substack{\lambda'\subseteq\{1,\ldots,N\}\\ \mathrm{card}\,(\lambda')=m-1,\ \lambda'\cap\{i,j\}=\emptyset}} \left(\widetilde{A}^2_{\lambda'\cup\{i\}} + A^2_{\lambda'\cup\{j\}}\right) .$$

The result now follows from Proposition 1.5.9. □

Proof of the Generalized Pythagorean Theorem. By translating Σ if necessary, we may suppose that u_0 coincides with the origin. Let us also introduce the notation

$$u_i = (u_{i,1},\ u_{i,2}, \ldots,\ u_{i,N}) .$$

By Corollary 1.5.10, it suffices to prove the generalized Pythagorean theorem for the image of Σ after a sequence of coordinate-plane rotations. In fact, we will show that there exists a sequence of coordinate-plane rotations such that the resulting image of Σ is contained in the $(x_1, \ldots, x_m)$-coordinate plane. Since the generalized Pythagorean theorem holds trivially for a simplex lying in an m-dimensional coordinate plane, it follows that the generalized Pythagorean theorem holds for the originally given Σ.

• **The first sequence of coordinate-plane rotations.** We begin with the rotation $\mathcal{R}$ of the (x_1, x_2)-plane that maps $\Pi_{\{1,2\}}(u_1) = (u_{1,1},\ u_{1,2})$ to $(t, 0)$, where $t = (u_{1,1}^2 + u_{1,2}^2)^{1/2}$. When the coordinate-plane rotation $\mathcal{R} \times \mathbf{I}_{\mathbb{R}^{N-2}}$ is applied to Σ and Σ is replaced by its image—without changing notation—we obtain

$$u_1 = (u_{1,1}, 0,\ u_{1,3}, \ldots,\ u_{1,N})\,.$$

The second coordinate-plane rotation will rotate the (x_1, x_3)-plane so that $\Pi_{\{1,3\}}(u_1) = (u_{1,1},\ u_{1,3})$ is mapped to $(t, 0)$, where $t = (u_{1,1}^2 + u_{1,3}^2)^{1/2}$. After again replacing Σ by its image—still without changing notation—we obtain

$$u_1 = (u_{1,1}, 0, 0,\ u_{1,4}, \ldots,\ u_{1,N})\,.$$

After a total of $N - 1$ coordinate-plane rotations and replacements, we obtain

$$u_1 = (u_{1,1}, 0, 0, \ldots, 0)\,. \tag{1.25}$$

From now on, x_1 will *not* be one of the rotated coordinates in any of the coordinate-plane rotations we use. Consequently, (1.25) will continue to hold.

• **The $(i + 1)$st sequence of coordinate-plane rotations.** Suppose that we have

$$\begin{aligned} u_1 &= (u_{1,1}, 0, 0, \ldots, 0, 0, \ldots, 0)\,,\\ u_2 &= (u_{2,1}, u_{2,2}, 0, \ldots, 0, 0, \ldots, 0)\,,\\ &\;\vdots\\ u_i &= (u_{i,1}, u_{i,2}, u_{i,3}, \ldots, u_{i,i}, 0, \ldots, 0)\,. \end{aligned} \tag{1.26}$$

In particular, observe that (1.26) implies that the points $u_1, u_2, \ldots, u_i$ all lie in the $(x_1, x_2, \ldots, x_i)$-coordinate plane.

Arguing inductively, we will show that we can obtain (1.26) with $i = m$. Note that when $i = 1$, (1.26) is the same as (1.25).

Our next coordinate-plane rotation will rotate the (x_{i+1}, x_{i+2})-plane so that $\Pi_{\{i+1,i+2\}}(u_{i+1}) = (u_{i+1,i+1},\ u_{i+1,i+2})$ is mapped to $(t, 0)$, where $t = (u_{i+1,i+1}^2 + u_{i+1,i+2}^2)^{1/2}$. Then we obtain

$$u_{i+1} = (u_{i+1,1},\ u_{i+1,2}, \ldots,\ u_{i+1,i+1}, 0,\ u_{i+1,i+3}, \ldots,\ u_{i+1,N})\,.$$

Continuing in that fashion, we see that after a total of $N - i - 1$ coordinate-plane rotations, we obtain

$$u_{i+1} = (u_{i+1,1},\ u_{i+1,2}, \ldots,\ u_{i+1,i+1}, 0, 0, \ldots, 0)\,.$$

Since none of the coordinates $x_1, x_2, \ldots, x_i$ have been rotated coordinates for any of the coordinate-plane rotations we have used, the values of those coordinates will have remained unchanged. Thus we now have (1.26) with i replaced by $i + 1$.

Arguing as above for $i = 1, 2, \ldots, m-1$, we see that—including the first sequence of coordinate-plane rotations—after a grand total of $(N-1) + \sum_{i=1}^{m-1}(N-i-1) = (m/2)\,(2N-m-1)$ coordinate-plane rotations, we obtain (1.26) with i replaced by m. Thus we see that the image of Σ lies in the $(x_1, \ldots, x_m)$-coordinate plane, as desired. □

Remark 1.5.11. In [Bar 96], the reader will find a proof of the usual Pythagorean theorem via dimensional analysis. E. Thomann has conjectured (private communication) that the generalized Pythagorean theorem also might be provable via a dimensional analysis argument.

1.6 The Hausdorff Distance and Steiner Symmetrization

Consider the collection $\mathcal{P}(\mathbb{R}^N)$ of all subsets of $\mathbb{R}^N$. It is often useful, especially in geometric applications, to have a metric on $\mathcal{P}(\mathbb{R}^N)$. In this section we address methods for achieving this end. In Definition 1.2.12, we defined $\operatorname{dist}(S, T)$ for subsets S, T of a metric space; unfortunately, this function need not satisfy the triangle inequality. Also, in practice, $\mathcal{P}(\mathbb{R}^N)$ (the entire power set of $\mathbb{R}^N$) is probably too large a collection of objects to have a reasonable and useful metric topology (see [Dug 66, Section IX.9] for several characterizations of metrizability). With these considerations in mind, we shall restrict attention to the collection of nonempty, *bounded* subsets of $\mathbb{R}^N$.

Definition 1.6.1. Let S and T be nonempty, bounded subsets of $\mathbb{R}^N$. We set

$$\operatorname{HD}(S, T) = \max\left\{ \sup_{s\in S} \operatorname{dist}(s, T),\ \sup_{t\in T} \operatorname{dist}(S, t) \right\}. \tag{1.27}$$

This function is called the *Hausdorff distance*.

Notice that $\operatorname{HD}(S, T) = \operatorname{HD}(\overline{S}, T) = \operatorname{HD}(S, \overline{T}) = \operatorname{HD}(\overline{S}, \overline{T})$, so we further restrict our attention to the collection of nonempty sets that are both closed and bounded (i.e., compact) subsets of $\mathbb{R}^N$. For convenience, in this section, we will use $\mathcal{B}$ to denote the collection of nonempty, compact subsets of $\mathbb{R}^N$.

In Figure 1.3, if we let d denote the distance from a point on the left to the line segment on the right, then every point in the line segment is within distance $\sqrt{d^2 + (\epsilon/2)^2}$ of one of the points on the left—and that bound is sharp. Thus we see that $\operatorname{HD}(S, T) = \sqrt{d^2 + (\epsilon/2)^2}$.

Lemma 1.6.2. *Let $S, T \in \mathcal{B}$. Then there are points $s \in S$ and $t \in T$ such that* $\operatorname{HD}(S, T) = |s - t|$.

We leave the proof as an exercise for the reader (see Figure 1.4).

Proposition 1.6.3. *The function* HD *is a metric on $\mathcal{B}$.*

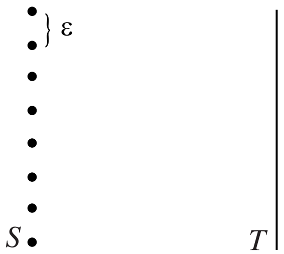

Fig. 1.3. The Hausdorff distance.

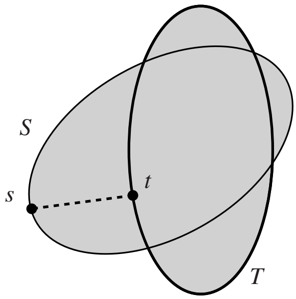

Fig. 1.4. Points that realize the Hausdorff distance.

Proof. Clearly HD ≥ 0, and if $S = T$, then $\text{HD}(S, T) = 0$.

Conversely, if $\text{HD}(S, T) = 0$ then let $s \in S$. By definition, there are points $t_j \in T$ such that $|s - t_j| \to 0$. Since T is compact, we may select a subsequence $\{t_{j_k}\}$ such that $t_{j_k} \to s$. Again, since T is compact, we then conclude that $s \in T$. Hence $S \subseteq T$. Similar reasoning shows that $T \subseteq S$. Hence $S = T$.

Finally, we come to the triangle inequality. Let $S, T, U \in \mathcal{B}$. Let $s \in S,\ t \in T,\ u \in U$. Then we have

$$
\begin{aligned}
|s - u| &\leq |s - t| + |t - u| \\
&\Downarrow \\
\text{dist}(S, u) &\leq |s - t| + |t - u| \\
&\Downarrow \\
\text{dist}(S, u) &\leq \text{dist}(S, t) + |t - u| \\
&\Downarrow \\
\text{dist}(S, u) &\leq \text{HD}(S, T) + |t - u| \\
&\Downarrow \\
\text{dist}(S, u) &\leq \text{HD}(S, T) + \text{dist}(T, u) \\
&\Downarrow \\
\text{dist}(S, u) &\leq \text{HD}(S, T) + \sup_{u \in U} \text{dist}(T, u)
\end{aligned}
$$

$$\Downarrow$$

$$\sup_{u\in U} \operatorname{dist}(S,u) \le \mathrm{HD}\,(S,T) + \sup_{u\in U} \operatorname{dist}(T,u).$$

By symmetry, we have

$$\sup_{s\in S} \operatorname{dist}(U,s) \le \mathrm{HD}\,(U,T) + \sup_{s\in S} \operatorname{dist}(T,s)$$

and thus

$$\begin{aligned}\max\{ \sup_{u\in U} \operatorname{dist}(S,u)\ ,\ \sup_{s\in S} \operatorname{dist}(U,s)\ \} \\ \le \max\Big\{\mathrm{HD}\,(S,T) + \sup_{u\in U} \operatorname{dist}(T,u)\ ,\ \mathrm{HD}\,(U,T) + \sup_{s\in S} \operatorname{dist}(T,s)\Big\}.\end{aligned}$$

We conclude that

$$\mathrm{HD}\,(U,S) \le \mathrm{HD}\,(U,T) + \mathrm{HD}\,(T,S).$$ □

There are fundamental questions concerning completeness, compactness, etc. that we need to ask about any metric space.

Theorem 1.6.4. *The metric space* $(\mathcal{B}, \mathrm{HD}\,)$ *is complete.*

Proof. Let $\{S_j\}$ be a Cauchy sequence in the metric space $(\mathcal{B}, \mathrm{HD}\,)$. We seek an element $S \in \mathcal{B}$ such that $S_j \to S$.

Elementary estimates, as in any metric space, show that the elements S_j are all contained in a common ball $B(0, R)$. We set S equal to

$$\bigcap_{j=1}^{\infty} \left(\overline{\bigcup_{\ell=j}^{\infty} S_\ell} \right).$$

Then S is nonempty, closed, and bounded, so it is an element of $\mathcal{B}$.

To see that $S_j \to S$, select $\epsilon > 0$. Choose J large enough so that if $j, k \ge J$ then $\mathrm{HD}\,(S_j, S_k) < \epsilon$. For $m > J$ set $T_m = \cup_{\ell=J}^m S_\ell$. Then it follows from the definition, and from Proposition 1.6.3, that $\mathrm{HD}\,(S_J, T_m) < \epsilon$ for every $m > J$. Therefore, with $U_p = \overline{\cup_{\ell=p}^{\infty} S_\ell}$ for every $p > J$, it follows that $\mathrm{HD}\,(S_J, U_p) \le \epsilon$.

We conclude that $\mathrm{HD}\,(S_J, \cap_{p=J+1}^K U_p) \le \epsilon$. Hence, by the continuity of the distance, $\mathrm{HD}\,(S_J, S) \le \epsilon$. That is what we wished to prove. □

As a corollary of the proof of Theorem 1.6.4 we obtain the following:

Corollary 1.6.5. *Let* $\{S_j\}$ *be a sequence of elements of* $\mathcal{B}$. *Suppose that* $S_j \to S$ *in the Hausdorff metric. Then*

$$\mathcal{L}^n(S) \ge \limsup_{j\to\infty} \mathcal{L}^n(S_j)\,.$$

The next theorem informs us of a seminal fact regarding the Hausdorff distance topology.

Theorem 1.6.6. *The set of nonempty compact subsets of* $\mathbb{R}^N$ *with the Hausdorff distance topology is boundedly compact, i.e., any bounded sequence has a subsequence that converges to a compact set.*

Proof. Let $A_1, A_2, \ldots$ be a bounded sequence in the Hausdorff distance. We may assume without loss of generality that each A_i is a subset of the closed unit N-cube, C_0. For each integer $k \geq 1$, subdivide the unit N-cube into 2^{kN} congruent subcubes of side length 2^{-k}; denote that collection of 2^{kN} subcubes by $\mathcal{S}_k$.

We will use an inductive construction and a diagonalization argument. Let $A_{0,i} = A_i$ for $i = 1, 2, \ldots$. For each $k \geq 1$, the sequence $A_{k,i}$, $i = 1, 2, \ldots$, will be a subsequence of the preceding sequence $A_{k-1,i}$, $i = 1, 2, \ldots$. Also, we will construct sets $C_0 \supseteq C_1 \supseteq \cdots$ inductively. Each C_k will be a union of a set of cubes in $\mathcal{S}_k$. The first set in this sequence is the unit cube C_0 itself. For each $k = 0, 1, \ldots$, the sequence $A_{k,i}$, $i = 1, 2, \ldots$, and the set C_k are to have the properties that

$$D \cap A_{k,i} \neq \emptyset \text{ holds for } i = 1, 2, \ldots \quad \text{whenever } D \in \mathcal{S}_k \text{ is one of the cubes forming } C_k, \tag{1.28}$$

and

$$A_{k,i} \subseteq C_k \quad \text{holds for all sufficiently large } i. \tag{1.29}$$

It is clear that (1.28) and (1.29) are satisfied when $k = 0$.

Assume $A_{k-1,i}$, $i = 1, 2, \ldots$, and C_{k-1} have been defined so that

$$D \cap A_{k-1,i} \neq \emptyset \text{ holds for } i = 1, 2, \ldots \quad \text{whenever } D \in \mathcal{S}_{k-1} \text{ is one of the cubes forming } C_{k-1},$$

and

$$A_{k-1,i} \subseteq C_{k-1} \quad \text{holds for all sufficiently large } i.$$

We let $\mathcal{C}_k$ be the collection of cubes in $\mathcal{S}_k$ that are subsets of C_{k-1} (here we are effectively subdividing the cubes that form C_{k-1}). A subcollection, $\mathcal{C} \subseteq \mathcal{C}_k$, will be called *admissible* if there are infinitely many i for which

$$D \cap A_{k-1,i} \neq \emptyset \quad \text{holds for all } D \in \mathcal{C}. \tag{1.30}$$

Let C_k be the union of a maximal admissible collection of subcubes, which is immediately seen to exist because $\mathcal{C}_k$ is finite. Let $A_{k,1}, A_{k,2}, \ldots$ be the subsequence of $A_{k-1,1}, A_{k-1,2}, \ldots$ consisting of those $A_{k-1,i}$ for which (1.30) is true. Observe that $A_{k,i} \subseteq C_k$ holds for sufficiently large i; otherwise, there is another subcube that could be added to the maximal collection while maintaining admissibility.

We set

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

and claim that C is the limit in the Hausdorff distance of $A_{k,k}$ as $k \to \infty$. Of course, C is nonempty by the finite intersection property. Let $\epsilon > 0$ be given. Clearly we can find an index k_0 such that

$$C_{k_0} \subseteq \{x : \operatorname{dist}(x, C) < \epsilon\}.$$

There is a number i_0 such that for $i \geq i_0$ we have

$$A_{k_0,i} \subseteq C_{k_0} \subseteq \{x : \operatorname{dist}(x, C) < \epsilon\}.$$

So, for $k \geq k_0 + i_0$, we know that

$$A_{k,k} \subseteq \{x : \operatorname{dist}(x, C) < \epsilon\}$$

holds. We let $k_1 \geq k_0 + i_0$ be such that

$$\sqrt{N}\, 2^{-k_1} < \epsilon.$$

Let $c \in C$ be arbitrary. Then $c \in C_{k_1}$, so there is some cube, D, of side length 2^{-k_1} containing c and for which

$$D \cap A_{k_1,i} \neq \emptyset$$

holds for all i. But then if $k \geq k_1$, we have $D \cap A_{k,k} \neq \emptyset$, so

$$\operatorname{dist}(c, A_{k,k}) \leq \sqrt{N}\, s^{-k} < \epsilon.$$

It follows that HD $(C, A_{k,k}) < \epsilon$ holds for all $k \geq k_1$. □

Next we give two more useful facts about the Hausdorff distance topology.

Definition 1.6.7. A subset C of a vector space is *convex* if for $x, y \in C$ and $0 \leq t \leq 1$ we have

$$(1 - t)\, x + t\, y \in C\, .$$

Proposition 1.6.8. *Let $\mathcal{C}$ be the collection of all closed, bounded, convex sets in $\mathbb{R}^N$. Then $\mathcal{C}$ is a closed subset of the metric space* $(\mathcal{B}, \text{HD}\,)$.

Proof. There are several amusing ways to prove this assertion. One is by contradiction. If $\{S_j\}$ is a convergent sequence in $\mathcal{C}$, then let $S \in B$ be its limit. If S does not lie in $\mathcal{C}$ then S is not convex. Thus there is a segment ℓ with endpoints lying in S but with some interior point p not in S.

Let $\epsilon > 0$ be selected so that the open ball $U(p, \epsilon)$ does not lie in S. Let a, b be the endpoints of ℓ. Choose j so large that HD $(S_j, S) < \epsilon/2$. For such j, there exist points $a_j, b_j \in S_j$ such that $|a_j - a| < \epsilon/3$ and $|b_j - b| < \epsilon/3$. But then each point $c_j(t) \equiv (1 - t)a_j + tb_j$ has distance less than $\epsilon/3$ from $c(t) \equiv (1 - t)a + tb$, $0 \leq t \leq 1$. In particular, there is a point p_j on the line segment ℓ_j connecting a_j to b_j such that $|p_j - p| < \epsilon/3$. Noting that p_j must lie in S_j, we see that we have contradicted our statement about $U(p, \epsilon)$. Therefore S must be convex. □

Proposition 1.6.9. *Let $\{S_j\}$ be a sequence of elements of $\mathcal{B}$, each of which is connected. Suppose that $S_j \to S$ in the Hausdorff metric. Then S must be connected.*

Proof. Suppose not. Then S is disconnected. So we may write $S = A \cup B$ with each of A and B closed and nonempty and $A \cap B = \emptyset$. Then there is a number $\eta > 0$ such that if $a \in A$ and $b \in B$ then $|a - b| > \eta$.

Choose j so large that HD $(S_j, S) < \eta/3$. Define

$$A_j = \{s \in S_j : \text{dist}(s, A) \leq \eta/3\} \quad \text{and} \quad B_j = \{s \in S_j : \text{dist}(s, B) \leq \eta/3\}.$$

Clearly $A_j \cap B_j = \emptyset$ and A_j, B_j are closed and nonempty. Moreover, $A_j \cup B_j = S_j$. That contradicts the connectedness of S_j and completes the proof. □

Remark 1.6.10. It is certainly possible to have totally disconnected sets E_j, $j = 1, 2, \ldots$, such that $E_j \to E$ as $j \to \infty$ and E is connected (exercise).

Now we turn to a new arena in which the Hausdorff distance is applicable.

Definition 1.6.11. Let V be an $(N-1)$-dimensional vector subspace of $\mathbb{R}^N$. *Steiner symmetrization*[20] with respect to V is the operation that associates with each bounded subset T of $\mathbb{R}^N$ the subset $\widetilde{T}$ of $\mathbb{R}^N$ having the property that for each straight line ℓ perpendicular to V, $\ell \cap \widetilde{T}$ is a closed line segment with center in V or is empty and the conditions

$$\mathcal{L}^1(\ell \cap \widetilde{T}) = \mathcal{L}^1(\ell \cap T) \tag{1.31}$$

and

$$\ell \cap \widetilde{T} = \emptyset \quad \text{if and only if} \quad \ell \cap T = \emptyset$$

hold, where in (1.31), $\mathcal{L}^1$ means the Lebesgue measure resulting from isometrically identifying the line ℓ with $\mathbb{R}$.

In Figure 1.5, B is the Steiner symmetrization of A with respect to the line L.

Steiner used symmetrization to give a proof of the isoperimetric theorem that he presented to the Berlin Academy of Science in 1836 (see [Str 36]). The results in the remainder of this section document a number of aspects of the behavior of Steiner symmetrization.

Proposition 1.6.12. *If T is a bounded $\mathcal{L}^N$-measurable subset of $\mathbb{R}^N$ and if S is obtained from T by Steiner symmetrization, then S is $\mathcal{L}^N$-measurable and*

$$\mathcal{L}^N(T) = \mathcal{L}^N(S).$$

Proof. This is a consequence of Fubini's theorem. □

Lemma 1.6.13. *Fix $0 < M < \infty$. If A and $A_1, A_2, \ldots$ are closed subsets of $\mathbb{R}^N \cap \overline{\mathbb{B}}(0, M)$ such that*

[20] Jakob Steiner (1796–1863).

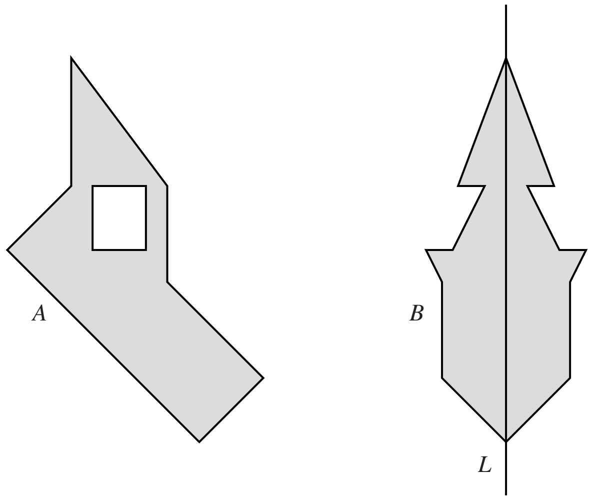

Fig. 1.5. Steiner symmetrization.

$$\bigcap_{i_0=1}^{\infty} \overline{\left[\bigcup_{i=i_0}^{\infty} A_i \right]} \subseteq A,$$

then

$$\limsup_i \mathcal{L}^N(A_i) \le \mathcal{L}^N(A).$$

Proof. Let $\epsilon > 0$ be arbitrary. Then there exists an open set U with $A \subseteq U$ and

$$\mathcal{L}^N(U) \le \mathcal{L}^N(A) + \epsilon.$$

A routine argument shows that for all sufficiently large i, $A_i \subseteq U$. It follows that

$$\limsup_i \mathcal{L}^N(A_i) \le \mathcal{L}^N(U),$$

and the fact that ϵ was arbitrary implies the lemma. □

Proposition 1.6.14. *If T is a compact subset of $\mathbb{R}^N$ and if S is obtained from T by Steiner symmetrization, then S is compact.*

Proof. Let V be an $(N-1)$-dimensional vector subspace of $\mathbb{R}^N$, and suppose that S is the result of Steiner symmetrization of T with respect to V. It is clear that the boundedness of T implies the boundedness of S. To see that S is closed, consider any sequence of points $p_1, p_2, \ldots$ in S that converges to some point p. Each p_i lies in a line ℓ_i perpendicular to V, and we know that

$$\operatorname{dist}(p_i, V) \le \frac{1}{2}\mathcal{L}^1(\ell_i \cap S) = \frac{1}{2}\mathcal{L}^1(\ell_i \cap T).$$

We also know that the line perpendicular to V and containing p must be the limit of the sequence of lines $\ell_1, \ell_2, \ldots$. Further, we know that

$$\operatorname{dist}(p, V) = \lim_{i\to\infty} \operatorname{dist}(p_i, V).$$

The inequality

$$\limsup_i \mathcal{L}^1(\ell_i \cap T) \le \mathcal{L}^1(\ell \cap T) \tag{1.32}$$

would allow us to conclude that

$$\operatorname{dist}(p, V) = \lim_{i\to\infty} \operatorname{dist}(p_i, V) \le \frac{1}{2} \limsup_{i\to\infty} \mathcal{L}^1(\ell_i \cap T) \le \frac{1}{2}\mathcal{L}^1(\ell \cap T),$$

and thus that $p \in S$.

To obtain the inequality (1.32), we let q_i be the vector parallel to V that translates ℓ_i to ℓ, and we apply Lemma 1.6.13, with N replaced by 1 and with ℓ identified with $\mathbb{R}$, to the sets $A_i = \tau_{q_i}(\ell_i \cap T)$, which are the translates of the sets $\ell_i \cap T$. We can take $A = \ell \cap T$, because T is closed. □

Proposition 1.6.15. *If T is a bounded, convex subset of $\mathbb{R}^N$ and S is obtained from T by Steiner symmetrization, then S is also a convex set.*

Proof. Let V be an $(N-1)$-dimensional vector subspace of $\mathbb{R}^N$, and suppose that S is the result of Steiner symmetrization of T with respect to V. Let x and y be two points of S. We let x' and y' denote the points obtained from x and y by reflection through the hyperplane V. Also, let ℓ_x and ℓ_y denote the lines perpendicular to V and passing through the points x and y, respectively. By the definition of Steiner symmetrization and the convexity of T, we see that $\ell_x \cap T$ must contain a line segment, say from p_x to q_x, of length at least $\operatorname{dist}(x, x')$. Likewise, $\ell_y \cap T$ contains a line segment from p_y to q_y of length at least $\operatorname{dist}(y, y')$. The convex hull of the four points p_x, q_x, p_y, q_y is a trapezoid, Q, which is a subset of T.

We claim that the trapezoid, Q', that is the convex hull of x, x', y, y' must be contained in S. Let x'' be the point of intersection of ℓ_x and V. Similarly, define y'' to be the intersection of ℓ_y and V. For any $0 \le \tau \le 1$, the line ℓ'' perpendicular to V and passing through

$$(1-\tau)x'' + \tau y''$$

intersects the trapezoid $Q \subseteq T$ in a line segment of length

$$d_1 = (1-\tau)\operatorname{dist}(p_x, q_x) + \tau \operatorname{dist}(p_y, q_y),$$

and it intersects the trapezoid Q' in a line segment, centered about V, of length

$$d_2 = (1-\tau)\operatorname{dist}(x, x') + \tau \operatorname{dist}(y, y').$$

But S must contain a closed line segment of ℓ'', centered about V, of length at least d_1. Since d_1 at least as large as d_2,

$$\ell'' \cap Q' \subseteq \ell'' \cap S.$$

Since the choice of $0 \le \tau \le 1$ was arbitrary, we conclude that $Q' \subseteq S$. In particular, the line segment from x to y is contained in Q' and thus in S. □

The power of Steiner symmetrization obtains from the following theorem.

Theorem 1.6.16. *Suppose that $\mathcal{C}$ is a nonempty family of nonempty compact subsets of $\mathbb{R}^N$ that is closed in the Hausdorff distance topology and that is closed under the operation of Steiner symmetrization with respect to any $(N-1)$-dimensional vector subspace of $\mathbb{R}^N$. Then $\mathcal{C}$ contains a closed ball (possibly of radius 0) centered at the origin.*

Proof. Let $\mathcal{C}$ be such a family of compact subsets of $\mathbb{R}^N$ and set

$$r = \inf\{s : \text{ there exists } T \in \mathcal{C} \text{ with } T \subseteq \overline{\mathbb{B}}(0, s)\}.$$

If $r = 0$, we are done, so we may assume $r > 0$. By Theorem 1.6.6, any uniformly bounded family of nonempty compact sets is compact in the Hausdorff distance topology, so we can suppose there exists a $T \in \mathcal{C}$ with $T \subseteq \overline{\mathbb{B}}(0, r)$.

We claim that $T = \overline{\mathbb{B}}(0, r)$. If not, then there exist $p \in \overline{\mathbb{B}}(0, r)$ and $\epsilon > 0$ such that $T \subseteq \overline{\mathbb{B}}(0, r) \setminus \mathbb{B}(p, \epsilon)$. Suppose T_1 is the result of Steiner symmetrization of T with respect to any arbitrarily chosen $(N-1)$-dimensional vector subspace V. Let ℓ be the line perpendicular to V and passing through p. For any line ℓ' parallel to ℓ and at distance less than ϵ from ℓ, the Lebesgue measure of the intersection of ℓ' with T must be strictly less than the length of the intersection of ℓ' with $\overline{\mathbb{B}}(0, r)$, so the intersection of ℓ' with $\partial\overline{\mathbb{B}}(0, r)$ is not in T_1. We conclude that if p_1 is either one of the points of intersection of the sphere of radius r about the origin with the line ℓ, then

$$\mathbb{B}(p_1, \epsilon) \cap \partial\overline{\mathbb{B}}(0, r) \cap T_1 = \emptyset.$$

Choose a finite set of distinct additional points $p_2, p_3, \ldots, p_k$ such that

$$\partial\overline{\mathbb{B}}(0, r) \subseteq \bigcup_{i=1}^{k} \mathbb{B}(p_i, \epsilon).$$

For $i = 1, 2, \ldots, k-1$, let T_{i+1} be the result of Steiner symmetrization of T_i with respect to the $(N-1)$-dimensional vector subspace perpendicular to the line through p_i and p_{i+1}. By the lemma it follows that

$$\mathbb{B}(p_i, \epsilon) \cap \partial\overline{\mathbb{B}}(0, r) \cap T_j = \emptyset$$

holds for $i \le j \le k$. Thus we have

$$T_k \cap \partial\overline{\mathbb{B}}(0, r) = \emptyset,$$

so

$$T_k \subseteq \overline{\mathbb{B}}(0, s)$$

holds for some $s < r$, a contradiction. □

1.7 Borel and Suslin Sets

In this section, we discuss the Borel and Suslin sets. The goal of the section is to show that for all reasonable measures on Euclidean space, the continuous images of Borel sets are measurable sets (Corollary 1.7.19). This result is based on three facts: every Borel set is a Suslin set (Theorem 1.7.9), the continuous image of a Suslin set is a Suslin set (Theorem 1.7.12), and all Suslin sets are measurable (Corollary 1.7.18).

To put it as briefly as possible, the Suslin sets in $\mathbb{R}^N$ are the sets obtained as the orthogonal projections of Borel sets in $\mathbb{R}^{N+M}$. The history of Suslin sets is of some interest. In [Leb 05] (on page 191) Lebesgue had claimed that every projection of a Borel set is again a Borel set—Lebesgue even gave what he believed was a proof. It was Suslin (see [Sus 17]) who showed that, in fact, the Borel sets form a proper subfamily within the Suslin sets, and consequently, there exists a Borel set whose orthogonal projection is *not* a Borel set. While it is clearly of interest to know that there exists a Suslin set that is not a Borel set, we will not prove or use that result. We refer the interested reader to [Fed 69, 2.2.11], [Hau 62, Section 33], or [Jec 78, Section 39].

Construction of the Borel Sets

In Section 1.2 we defined the Borel sets in a topological space to be the members of the smallest σ-algebra that includes all the open sets. The virtue of this definition is its efficiency, but the price we pay for that efficiency is the absence of a mechanism for constructing the Borel sets. In this section, we will provide that constructive definition of the Borel sets.

For definiteness we work on $\mathbb{R}^N$. We will use transfinite induction over the smallest uncountable ordinal ω_1 (see Appendix A.1 for a brief introduction to transfinite induction) to define families of sets Σ^0_α and Π^0_α, for $\alpha < \omega_1$. For us, the superscript 0's are superfluous, but we include them since they are typically used in descriptive set theory.

Definition 1.7.1. Set

$$\Sigma^0_1 = \text{the family of all open sets in } \mathbb{R}^N,$$

$$\Pi^0_1 = \text{the family of all closed sets in } \mathbb{R}^N.$$

If $\alpha < \omega_1$, and Σ^0_β and Π^0_β have been defined for all $\beta < \alpha$, then set

$$\Sigma^0_\alpha = \text{the family of sets of the form}$$

$$A = \bigcup_{i=1}^{\infty} A_i, \quad \text{where each } A_i \in \Pi^0_\beta \text{ for some } \beta < \alpha, \tag{1.33}$$

$$\Pi^0_\alpha = \text{the family of sets of the form } \mathbb{R}^N \setminus A \text{ for } A \in \Sigma^0_\alpha. \tag{1.34}$$

Since the complement of a union is the intersection of the complements, we see that we can also write

Π^0_α = the family of sets of the form

$$A = \bigcap_{i=1}^{\infty} A_i, \quad \text{where each } A_i \in \Sigma^0_\beta \text{ for some } \beta < \alpha. \tag{1.35}$$

By transfinite induction over ω_1, we see that for $\alpha < \omega_1$, all the elements of Σ^0_α and Π^0_α are Borel sets.

Lemma 1.7.2. *If* $1 \le \beta < \alpha < \omega_1$, *then*

$$\Sigma^0_\beta \subseteq \Pi^0_\alpha\,, \qquad \Pi^0_\beta \subseteq \Sigma^0_\alpha\,, \qquad \Sigma^0_\beta \subseteq \Sigma^0_\alpha\,, \qquad \Pi^0_\beta \subseteq \Pi^0_\alpha$$

hold.

Proof. By (1.33) and (1.35), we see that $\Sigma^0_\beta \subseteq \Pi^0_\alpha$ and $\Pi^0_\beta \subseteq \Sigma^0_\alpha$ hold whenever $1 \le \beta < \alpha < \omega_1$.

Every open set in Euclidean space is a countable union of closed sets, so $\Sigma^0_1 \subseteq \Sigma^0_2$ holds. Consequently, we also have $\Pi^0_1 \subseteq \Pi^0_2$. Since $\Sigma^0_1 \subseteq \Pi^0_2 \subseteq \Sigma^0_\alpha$ holds whenever $2 < \alpha$ and since $\Pi^0_1 \subseteq \Sigma^0_2$ holds, we have $\Sigma^0_1 \subseteq \Sigma^0_\alpha$ and $\Pi^0_1 \subseteq \Pi^0_\alpha$ for all $1 < \alpha < \omega_1$.

Fix $1 \le \beta < \alpha < \omega_1$. Suppose $\Sigma^0_\gamma \subseteq \Sigma^0_\alpha$ and $\Pi^0_\gamma \subseteq \Pi^0_\alpha$ hold whenever $\gamma < \beta$. Any set $A \in \Sigma^0_\beta$ must be of the form $A = \cup_{i=1}^{\infty} A_i$ with each $A_i \in \Pi^0_\gamma$ for some $\gamma < \beta$. Then since $\beta < \alpha$, we see that $A \in \Sigma^0_\alpha$. Thus $\Sigma^0_\beta \subseteq \Sigma^0_\alpha$. Similarly, we have $\Pi^0_\beta \subseteq \Pi^0_\alpha$. □

Corollary 1.7.3. *We have*

$$\bigcup_{\alpha<\omega_1} \Sigma^0_\alpha = \bigcup_{\alpha<\omega_1} \Pi^0_\alpha\,. \tag{1.36}$$

Theorem 1.7.4. *The family of sets in* (1.36) *is the* σ*-algebra of Borel subsets of* $\mathbb{R}^N$.

Proof. Let $\mathcal{B}$ denote the family of sets in (1.36). To see that $\mathcal{B}$ is closed under countable unions, suppose we are given $A_1, A_2, \ldots$ in $\mathcal{B}$. Considering the left-hand side of (1.36), we see that for each i, there is $\alpha_i < \omega_1$ such that $A_i \in \Sigma^0_{\alpha_i}$. Since the sequence $\alpha_1, \alpha_2, \ldots$ is countable, but ω_1 is uncountable, there is $\alpha^* < \omega_1$ with $\alpha_i < \alpha^*$ for all i (see Lemma A.1.4). We conclude that $\cup_{i=1}^{\infty} A_i \in \Sigma_{\alpha^*}$. Thus $\mathcal{B}$ is closed under countable unions. We argue similarly to see that $\mathcal{B}$ is closed under countable intersections and complements. □

Because in the definition of Π^0_α, equation (1.34) can be replaced by (1.35), Theorem 1.7.4 has the following corollary.

Corollary 1.7.5. *The family of Borel sets in* $\mathbb{R}^N$ *is the smallest family of sets containing the open sets that is closed under countable unions and countable intersections. Likewise, the family of Borel sets in* $\mathbb{R}^N$ *is the smallest family of sets, containing the closed sets, that is closed under countable unions and countable intersections.*

Suslin Sets

Recall that the positive integers are denoted by $\mathbb{N}$. We let $\widetilde{\mathcal{N}}$ denote the set of all finite sequences of positive integers and we let $\mathcal{N}$ denote the set of all infinite sequences of positive integers, so

$$\widetilde{\mathcal{N}} = \{\, (n_1, n_2, \ldots, n_k) : k \in \mathbb{N},\ n_i \in \mathbb{N} \text{ for } i = 1, 2, \ldots, k \,\},$$

$$\mathcal{N} = \{\, (n_1, n_2, \ldots) : n_i \in \mathbb{N} \text{ for } i = 1, 2, \ldots \,\}.$$

Definition 1.7.6. Let $\mathcal{M}$ be a collection of subsets of a set X. Suppose that there is a set $M_{n_1,n_2,\ldots,n_k} \in \mathcal{M}$ associated with every finite sequence of positive integers. We can represent this relation as a function $\nu : \widetilde{\mathcal{N}} \to \mathcal{M}$ defined by

$$(n_1, n_2, \ldots, n_k) \overset{\nu}{\longmapsto} M_{n_1,n_2,\ldots,n_k}\,.$$

Such a function ν is called a *determining system in* $\mathcal{M}$. Associated with the determining system ν is the set called the *nucleus of* ν denoted by $\mathrm{N}\,(\nu)$ and defined by

$$\mathrm{N}\,(\nu) = \bigcup_{\substack{n \in \mathcal{N} \\ n=(n_1,n_2,\ldots)}} \left(M_{n_1} \cap M_{n_1,n_2} \cap \cdots \cap M_{n_1,n_2,\ldots,n_k} \cap \cdots \right).$$

Suslin's operation (A) is the function that when applied to the argument ν produces the result $\mathrm{N}\,(\nu)$. We will say that $\mathrm{N}\,(\nu)$ is a *Suslin set generated by* $\mathcal{M}$. The family of all Suslin sets generated by $\mathcal{M}$ will be denoted by $\mathcal{M}_{(\mathrm{A})}$.

By the *Suslin sets in a topological space* we mean the Suslin sets generated by the family of closed sets.

Since $\mathcal{N}$ has the same cardinality as the real numbers, we see that the nucleus is formed by an *uncountable* union of countable intersections of sets in $\mathcal{M}$. We might expect that operation (A) could be extremely powerful, but at the outset it is not immediately clear what can be done with the operation. The next proposition tells us that operation (A) is at least as powerful as those used to form the Borel sets.

Proposition 1.7.7. *Suppose $A_1, A_2, \ldots \in \mathcal{M}$, then there exist determining systems ν_U and ν_I such that*

$$\mathrm{N}\,(\nu_U) = \bigcup_{i=1}^{\infty} A_i \quad \textit{and} \quad \mathrm{N}\,(\nu_I) = \bigcap_{i=1}^{\infty} A_i\,.$$

Proof. Define ν_U and ν_I by

$$(n_1, n_2, \ldots, n_k) \overset{\nu_U}{\longmapsto} A_{n_1}\,,$$

$$(n_1, n_2, \ldots, n_k) \overset{\nu_I}{\longmapsto} A_k\,.$$

It is easy to see that ν_U and ν_I have the desired properties. □

The next theorem that tells us that repeated applications of operation (A) produce nothing that cannot be produced with only one application of the operation.

Theorem 1.7.8. *If $\mathcal{M}$ is a family of sets, if $\emptyset \in \mathcal{M}$, and if $\mathcal{M}_{(\mathrm{A})}$ is the family of Suslin sets generated by $\mathcal{M}$, then any Suslin set generated by $\mathcal{M}_{(\mathrm{A})}$ is already an element of $\mathcal{M}_{(\mathrm{A})}$. Symbolically, we have*

$$\left(\mathcal{M}_{(\mathrm{A})}\right)_{(\mathrm{A})} = \mathcal{M}_{(\mathrm{A})} .$$

Proof. Let

$$(n_1, n_2, \ldots, n_k) \stackrel{\nu}{\longmapsto} M_{n_1,n_2,\ldots,n_k} \in \mathcal{M}_{(\mathrm{A})}$$

be a determining system in $\mathcal{M}_{(\mathrm{A})}$. For each $n_1, n_2, \ldots, n_k \in \widetilde{\mathcal{N}}$, the set $M_{n_1,n_2,\ldots,n_k}$ must itself be the nucleus of a determining system $\nu_{n_1,n_2,\ldots,n_k}$ in $\mathcal{M}$; that is,

$$(q_1, q_2, \ldots, q_\ell) \xmapsto{\nu_{n_1,n_2,\ldots,n_k}} M^{q_1,q_2,\ldots,q_\ell}_{n_1,n_2,\ldots,n_k} \in \mathcal{M} ,$$

$$M_{n_1,n_2,\ldots,n_k} = \bigcup_{\substack{q\in\mathcal{N} \\ q=(q_1,q_2,\ldots)}} \left(M^{q_1}_{n_1,n_2,\ldots,n_k} \cap M^{q_1,q_2}_{n_1,n_2,\ldots,n_k} \cap \cdots \cap M^{q_1,q_2,\ldots,q_\ell}_{n_1,n_2,\ldots,n_k} \cap \cdots\right) ,$$

$$\mathrm{N}(\nu) = \bigcup_{\substack{n\in\mathcal{N} \\ n=(n_1,n_2,\ldots)}} \left(M_{n_1} \cap M_{n_1,n_2} \cap \cdots \cap M_{n_1,n_2,\ldots,n_k} \cap \cdots\right) .$$

We can rewrite $\mathrm{N}(\nu)$ as the union of the sets

$$\begin{array}{ccccc} & \left(M^{q^1_1}_{n_1}\right. & \cap\ M^{q^1_1,q^1_2}_{n_1} & \cap \cdots \cap\ M^{q^1_1,q^1_2,\ldots,q^1_\ell}_{n_1} & \left.\cap \cdots\right) \\ \bigcap & \left(M^{q^2_1}_{n_1,n_2}\right. & \cap\ M^{q^2_1,q^2_2}_{n_1,n_2} & \cap \cdots \cap\ M^{q^2_1,q^2_2,\ldots,q^2_\ell}_{n_1,n_2} & \left.\cap \cdots\right) \\ \vdots & \vdots & \vdots & \vdots & \\ \bigcap & \left(M^{q^k_1}_{n_1,n_2,\ldots,n_k}\right. & \cap\ M^{q^k_1,q^k_2}_{n_1,n_2,\ldots,n_k} & \cap \cdots \cap\ M^{q^k_1,q^k_2,\ldots,q^k_\ell}_{n_1,n_2,\ldots,n_k} & \left.\cap \cdots\right) \\ \vdots & \vdots & \vdots & \vdots & . \end{array} \tag{1.37}$$

Notice that the set in the kth row and ℓth column of (1.37) is indexed by k subscripts and ℓ superscripts. The choices of the subscripts and superscripts are constrained by the following requirements:

$$\left.\begin{array}{l} \text{in any row, the list of subscripts is constant,} \\ \text{in any row, the list of superscripts grows by concatenation,} \\ \text{in any column, the list of subscripts grows by concatenation.} \end{array}\right\} \tag{1.38}$$

Let the prime numbers in increasing numerical order be given by the list

$$p_1, \; p_2, \; p_3, \ldots .$$

We can use the list of primes to encode the information concerning the number of subscripts, the number of superscripts, and their values as follows: Set

$$m = p_1^k \cdot p_2^\ell \cdot p_3^{n_1} \cdot p_4^{n_2} \; \cdots \; p_{k+2}^{n_k} \cdot p_{k+3}^{q_1^k} \cdot p_{k+4}^{q_2^k} \; \cdots \; p_{\ell+k+2}^{q_\ell^k} . \tag{1.39}$$

Given a positive integer m, the unique factorization of m into prime powers determines whether m is of the form (1.39). Certainly not every positive integer m is of the form (1.39), nor is every sequence of positive integers $m_1, m_2, \ldots$ consistent with the conditions (1.38), even if the individual numbers m_i are of the form (1.39). But it is true that any sequence of sets in (1.37) will give rise to a sequence of positive integers $m_1, m_2, \ldots$ of the form (1.39) that satisfies the conditions (1.38).

We now define the determining system

$$(m_1, m_2, \ldots, m_k) \stackrel{\sigma}{\longmapsto} S_{m_1, m_2, \ldots, m_k} .$$

For each positive integer m, if m is of the form (1.39), then the numbers k, ℓ, $n_1, n_2, \ldots, n_k$, $q_1^k, q_2^k, \ldots, q_\ell^k$ are uniquely determined by (1.39). So we can make the definition

$$T_m = \begin{cases} S_{n_1, n_2, \ldots, n_k}^{q_1^k, q_2^k, \ldots, q_\ell^k} & \text{if } m \text{ is of the form (1.39)} , \\ \emptyset & \text{otherwise.} \end{cases}$$

Then, for the sequence of positive integers $m_1, m_2, \ldots,$ set

$$S_{m_1, m_2, \ldots, m_k} = \begin{cases} T_{m_1} \cap T_{m_2} \cap \cdots \cap T_{m_k} & \text{if (1.38) is not violated,} \\ \emptyset & \text{otherwise.} \end{cases}$$

For $m = (m_1, m_2, \ldots) \in \mathcal{N}$, the set

$$S_{m_1} \cap S_{m_1, m_2} \cap \cdots \cap S_{m_1, m_2, \ldots, m_k} \cap \cdots$$

is either one of the sets in (1.37) or the empty set. By construction, every set in (1.37) gives rise to a sequence $m = (m_1, m_2, \ldots) \in \mathcal{N}$ such that

$$S_{m_1} \cap S_{m_1, m_2} \cap \cdots \cap S_{m_1, m_2, \ldots, m_k} \cap \cdots$$

equals that set in (1.37). Thus we have $\mathrm{N}(\nu) = \mathrm{N}(\sigma)$. □

Theorem 1.7.9. *Every Borel set in* $\mathbb{R}^N$ *is a Suslin set.*

Proof. By Proposition 1.7.7 and Theorem 1.7.8, the collection of Suslin sets is closed under countable unions and countable intersections. Thus by Corollary 1.7.5, the collection of Suslin sets contains all the Borel sets. □

Continuous Images of Suslin Sets

Suppose $f : X \to Y$ is a function from a set X to a set Y. The inverse image of a union of sets equals the union of the inverse images, and likewise the inverse image of an intersection of sets equals the intersection of the inverse images. Images of sets under functions are not as well behaved as inverse images; nonetheless, we do have the following result—which is easily verified.

Proposition 1.7.10. *Let $f : X \to Y$.*

(1) *For $\{A_\alpha\}_{\alpha\in I}$ a collection of subsets of X, $f\left(\bigcup_{\alpha\in I} A_\alpha\right) = \bigcup_{\alpha\in I} f(A_\alpha)$ holds.*
(2) *For $X \supseteq A_1 \supseteq A_2 \supseteq \cdots$, $f\left(\bigcap_{i=1}^{\infty} A_i\right) \subseteq \bigcap_{i=1}^{\infty} f(A_i)$ holds and strict inclusion is possible.*

To obtain an equality for images of intersections, we need to look at continuous functions and decreasing sequences of compact sets.

Proposition 1.7.11. *Let X and Y be topological spaces and let $f : X \to Y$ be continuous. If X is sequentially compact, $X \supseteq C_1 \supseteq C_2 \supseteq \cdots$, and if each C_i is a closed subset of X, then $f\left(\bigcap_{i=1}^{\infty} C_i\right) = \bigcap_{i=1}^{\infty} f(C_i)$.*

Proof. By Proposition 1.7.10, we need only show that $\bigcap_{i=1}^{\infty} f(C_i) \subseteq f\left(\bigcap_{i=1}^{\infty} C_i\right)$, so suppose $y \in \bigcap_{i=1}^{\infty} f(C_i)$.

For each i, there is $x_i \in C_i$ with $f(x_i) = y$, and because the sets C_i are decreasing, we have $x_j \in C_i$ whenever $j \ge i$.

Set $x_{0,j} = x_j$ for $j = 1, 2, \ldots$. Since C_1 is sequentially compact, there is a convergent subsequence $\{x_{1,j}\}_{j=1}^{\infty}$ of $\{x_{0,j}\}_{j=1}^{\infty}$. Arguing inductively, suppose $1 \le i$ and that we have already constructed a convergent sequence $\{x_{i,j}\}_{j=1}^{\infty}$ that is a subsequence of $\{x_{h,j}\}_{j=1}^{\infty}$, for $0 \le h \le i-1$, and is such that every $x_{i,j}$ is a point of C_i, for $j = 1, 2, \ldots$. Since $\{x_{i,j}\}_{j=1}^{\infty}$ is a subsequence of the original sequence $\{x_{0,j}\}_{j=1}^{\infty}$, there is a j_* such that $x_{i,j} \in C_{i+1}$ holds for all j with $j_* \le j$. Since C_{i+1} is sequentially compact, we can select a convergent subsequence $\{x_{i+1,j}\}_{j=1}^{\infty}$ of $\{x_{i,j}\}_{j=j_*}^{\infty}$, and thus satisfy the induction hypothesis.

By construction, the sequence $\{x_{j,j}\}_{j=1}^{\infty}$ is convergent. Hence we have $\lim_{j\to\infty} x_{j,j} \in \bigcap_{i=1}^{\infty} C_i$, $f\left(\lim_{j\to\infty} x_{j,j}\right) = \lim_{j=1}^{\infty} f\left(x_{j,j}\right) = y$, and thus we have shown that $y \in \bigcap_{i=1}^{\infty} C_i$. □

Theorem 1.7.12. *If $f : \mathbb{R}^N \to \mathbb{R}^M$ is continuous and $S \subseteq \mathbb{R}^N$ is a Suslin set, then $f(S)$ is a Suslin subset of $\mathbb{R}^M$.*

Proof. Since any closed subset of $\mathbb{R}^N$ is a countable union of compact sets, we see that if $\mathcal{K}$ is the collection of compact subsets of $\mathbb{R}^N$, then $\mathcal{K}_{(\mathrm{A})}$ is the collection of Suslin sets.

Let $S \subseteq \mathbb{R}^N$ be a Suslin set, and let ν be a determining system in $\mathcal{K}$ such that $S = \mathrm{N}(\nu)$. Since any finite intersection of compact sets is compact, we see that the determining system $(n_1, n_2, \ldots, n_k) \overset{\nu}{\longmapsto} K_{n_1,n_2,\ldots,n_k}$ has the same nucleus as the determining system $(n_1, n_2, \ldots, n_k) \overset{\widetilde{\nu}}{\longmapsto} H_{n_1,n_2,\ldots,n_k}$ in $\mathcal{K}$ given by

$$H_{n_1,n_2,\ldots,n_k} = K_{n_1} \cap K_{n_1,n_2} \cap \cdots \cap K_{n_1,n_2,\ldots,n_k}\,.$$

Because the sets $\left\{H_{n_1,n_2,\ldots,n_k}\right\}_{k=1}^{\infty}$ form a decreasing sequence of compact sets, we can apply Propositions 1.7.10 and 1.7.11 to conclude that

$$
\begin{aligned}
f(S) &= f[\mathrm{N}(\nu)] = f[\mathrm{N}(\widetilde{\nu})] \\
&= f\left[\bigcup_{\substack{n\in\mathcal{N}\\ n=(n_1,n_2,\dots)}} \left(H_{n_1}\cap H_{n_1,n_2}\cap\cdots\cap H_{n_1,n_2,\dots,n_k}\cap\cdots\right)\right] \\
&= \bigcup_{\substack{n\in\mathcal{N}\\ n=(n_1,n_2,\dots)}} \Big(f(H_{n_1})\cap f(H_{n_1,n_2})\cap\cdots\cap f(H_{n_1,n_2,\dots,n_k})\cap\cdots\Big),
\end{aligned}
$$

and so we see that $f(S)$ is a Suslin set in $\mathbb{R}^M$. □

Measurability of Suslin Sets

In order to prove that the Suslin sets are measurable, we need to introduce some additional structures similar to the nucleus of a determining system.

Definition 1.7.13. Let $(n_1, n_2, \dots, n_k) \overset{\nu}{\longmapsto} A_{n_1,n_2,\dots,n_k}$ be given. Let $h_1, h_2, \dots, h_s$ be a finite sequence of positive integers. We define the following sets:

$$
\mathrm{N}^{h_1,h_2,\dots,h_s}(\nu) = \bigcup_{\substack{(n_1,n_2,\dots)\in\mathcal{N}\\ n_i\le h_i,\ 1\le i\le s}} A_{n_1}\cap A_{n_1,n_2}\cap\cdots\cap A_{n_1,n_2,\dots,n_k}\cap\cdots, \tag{1.40}
$$

$$
\mathrm{N}_{h_1,h_2,\dots,h_s}(\nu) = \bigcup_{n_1=1}^{h_1}\bigcup_{n_2=1}^{h_2}\cdots\bigcup_{n_s=1}^{h_s} A_{n_1}\cap A_{n_1,n_2}\cap\cdots\cap A_{n_1,n_2,\dots,n_s}. \tag{1.41}
$$

The next proposition follows immediately from the definition.

Proposition 1.7.14. *Let* $(n_1, n_2, \dots, n_k) \overset{\nu}{\longmapsto} A_{n_1,n_2,\dots,n_k}$ *be given. We have*

$$
\mathrm{N}^1(\nu) \subseteq \cdots \subseteq \mathrm{N}^h(\nu) \subseteq \mathrm{N}^{h+1}(\nu) \subseteq \cdots,
$$

$$
\mathrm{N}(\nu) = \bigcup_{k=1}^{\infty} \mathrm{N}^k(\nu),
$$

$$
\mathrm{N}^{h_1,\dots,h_s,1}(\nu) \subseteq \cdots \subseteq \mathrm{N}^{h_1,\dots,h_s,k}(\nu) \subseteq \mathrm{N}^{h_1,\dots,h_s,k+1}(\nu) \subseteq \cdots,
$$

$$
\mathrm{N}^{h_1,\dots,h_s}(\nu) = \bigcup_{k=1}^{\infty} \mathrm{N}^{h_1,\dots,h_s,k}(\nu).
$$

Corollary 1.7.15. *If* μ *is a regular measure on the nonempty set* X *and* ν *is a determining system in any family of subsets of* X *and if* E *is any subset of* X*, then*

$$
\lim_{k\to\infty} \mu\Big[E\cap \mathrm{N}^k(\nu)\Big] = \mu\Big[E\cap \mathrm{N}(\nu)\Big],
$$

$$
\lim_{k\to\infty} \mu\Big[E\cap \mathrm{N}^{h_1,h_2,\dots,h_s,k}(\nu)\Big] = \mu\Big[E\cap \mathrm{N}^{h_1,h_2,\dots,h_s}(\nu)\Big].
$$

Proof. Recall that Lemma 1.2.8 tells us that for a regular measure the measure of the union of an increasing sequence of sets is the limit of the measures of the sets, so the result follows immediately from Proposition 1.7.14. □

We will need the following lemma.

Lemma 1.7.16. *Let* $(n_1, n_2, \ldots, n_k) \overset{\nu}{\longmapsto} A_{n_1,n_2,\ldots,n_k}$ *and* $(h_1, h_2, \ldots) \in \mathcal{N}$ *be given. Then we have*

$$\mathrm{N}_{h_1}(\nu) \cap \mathrm{N}_{h_1,h_2}(\nu) \cap \cdots \cap \mathrm{N}_{h_1,h_2,\ldots,h_s}(\nu) \cap \cdots \subseteq \mathrm{N}(\nu)\,. \tag{1.42}$$

Proof. Fix a point x belonging to the left-hand side of (1.42).

First we claim that there exists a positive integer $n_1^0 \le h_1$ such that for every k with $2 \le k$, there exist $n_2, n_3, \ldots, n_k$ with $n_i \le h_i$, for $2 \le i \le k$, and with

$$x \in A_{n_1^0} \cap A_{n_1^0,n_2} \cap \cdots \cap A_{n_1^0,n_2,\ldots,n_k}\,.$$

To verify this, suppose it were not true. Then for each index $n_1 \le h_1$ there would be exist a positive integer $k(n_1)$ such that

$$x \notin A_{n_1} \cap A_{n_1,n_2} \cap \cdots \cap A_{n_1,n_2,\ldots,n_{k(n_1)}}$$

whenever $n_i \le h_i$ for $i = 2, 3, \ldots, k(n_1)$.

Setting $K(1) = \max\{\, k(1),\ k(2), \ldots,\ k(h_1)\,\}$, we see that

$$x \notin \bigcup_{n_1=1}^{h_1} \bigcup_{n_2=1}^{h_2} \cdots \bigcup_{n_{K(1)}=1}^{h_{K(1)}} A_{n_1} \cap A_{n_1,n_2} \cap \cdots \cap A_{n_1,n_2,\ldots,n_{K(1)}}\,,$$

which contradicts our assumption that x is an element of the left-hand side of (1.42).

Arguing inductively, suppose we have selected positive integers $n_1^0, n_2^0, \ldots, n_s^0$ satisfying

$$\left.\begin{array}{l} n_1^0 \le h_1,\, n_2^0 \le h_2, \ldots, n_s^0 \le h_s, \\ \text{for every } k \text{ with } s+1 \le k, \text{ there exist } n_{s+1}, n_{s+2}, \ldots, n_k \\ \qquad \text{with } n_i \le h_i, \text{ for } s+1 \le i \le k, \text{ and with} \\ \qquad x \in A_{n_1^0} \cap A_{n_1^0,n_2^0} \cap \cdots \cap A_{n_1^0,n_2^0,\ldots,n_s^0,n_{s+1},n_{s+2},\ldots,n_k}\,. \end{array}\right\} \tag{1.43}$$

We claim that we can select $n_{s+1}^0 \le h_{s+1}$ so that (1.43) holds with s replaced by $s+1$. If no such n_{s+1}^0 existed, then for each index $n_{s+1} \le h_{s+1}$ there would exist a positive integer $k(n_{s+1})$ such that

$$x \notin A_{n_1^0} \cap A_{n_1^0,n_2^0} \cap \cdots \cap A_{n_1^0,n_2^0,\ldots,n_s^0,n_{s+1},n_{s+2},\ldots,n_{k(n_{s+1})}}$$

whenever $n_i \le h_i$ for $i = s+1, s+2, \ldots, k(n_{s+1})$.

Setting $K(s+1) = \max\{\, k(1),\ k(2), \ldots,\ k(h_{s+1})\,\}$, we see that

$$x \notin \bigcup_{n_1=1}^{h_1} \bigcup_{n_2=1}^{h_2} \cdots \bigcup_{n_{K(s+1)}=1}^{h_{K(s+1)}} A_{n_1} \cap A_{n_1,n_2} \cap \cdots \cap A_{n_1,n_2,\ldots,n_{K(s+1)}},$$

which contradicts our assumption that x is an element of the left-hand side of (1.42).

Thus there exists an infinite sequence $n_1^0 \le h_1, n_2^0 \le h_2, \ldots$ such that

$$x \in A_{n_1^0} \cap A_{n_1^0,n_2^0} \cap \cdots \cap A_{n_1^0,n_2^0,\ldots,n_k^0} \cap \cdots ;$$

hence $x \in \mathrm{N}(\nu)$. □

Theorem 1.7.17. *Let μ be a regular measure on the nonempty set X, and let $\mathcal{M}$ be the collection of μ-measurable subsets of X. If ν is a determining system in $\mathcal{M}$, then $\mathrm{N}(\nu)$ is μ-measurable.*

Proof. Let the determining system ν be $(n_1, n_2, \ldots, n_k) \overset{\nu}{\longmapsto} M_{n_1,n_2,\ldots,n_k}$, and set $A = \mathrm{N}(\nu)$. We need to show that for any set $E \subseteq X$, we have

$$\mu(E \cap A) + \mu(E \setminus A) \le \mu(E).$$

We may assume that $\mu(E) < \infty$. Let $\epsilon > 0$ be arbitrary.

Using Corollary 1.7.15, we can inductively define a sequence of positive integers $h_1, h_2, \ldots$ such that

$$\mu\Big[C \cap \mathrm{N}^{h_1}(\nu)\Big] \ge \mu\Big[E \cap \mathrm{N}(\nu)\Big] - \epsilon/2$$

and

$$\mu\Big[C \cap \mathrm{N}^{h_1,h_2,\ldots,h_k}(\nu)\Big] \ge \mu\Big[E \cap \mathrm{N}^{h_1,h_2,\ldots,h_{k-1}}(\nu)\Big] - \epsilon/2^k.$$

We have $\mathrm{N}^{h_1,h_2,\ldots,h_k}(\nu) \subseteq \mathrm{N}_{h_1,h_2,\ldots,h_k}(\nu)$, so

$$\mu\Big[E \cap \mathrm{N}_{h_1,h_2,\ldots,h_k}(\nu)\Big] \ge \mu\Big[E \cap \mathrm{N}^{h_1,h_2,\ldots,h_k}(\nu)\Big] \ge \mu(E \cap \mathrm{N}(\nu)) - \epsilon$$

holds, and thus, since $\mathrm{N}_{h_1,h_2,\ldots,h_k}(\nu)$ is μ-measurable,

$$\begin{aligned} \mu(E) &= \mu\Big[E \cap \mathrm{N}_{h_1,h_2,\ldots,h_k}(\nu)\Big] + \mu\Big[E \setminus \mathrm{N}_{h_1,h_2,\ldots,h_k}(\nu)\Big] \\ &\ge \mu\Big[E \cap \mathrm{N}(\nu)\Big] + \mu\Big[E \setminus \mathrm{N}_{h_1,h_2,\ldots,h_k}(\nu)\Big] - \epsilon. \end{aligned}$$

Now the sequence of sets $\Big\{\mathrm{N}_{h_1,h_2,\ldots,h_k}(\nu)\Big\}_{k=1,2,\ldots}$ is descending, and by Lemma 1.7.16 its limit is a subset of $\mathrm{N}(\nu)$. Consequently the sequence $\Big\{X \setminus \mathrm{N}_{h_1,h_2,\ldots,h_k}\Big\}_{k=1,2,\ldots}$ is ascending and its limit contains the set $X \setminus \mathrm{N}(\nu)$. Hence

$$\lim_{k\to\infty} \mu\Big[E \setminus \mathrm{N}_{h_1,h_2,\ldots,h_k}(\nu)\Big] = \mu\left[E \setminus \bigcup_{k=1}^{\infty} \mathrm{N}_{h_1,h_2,\ldots,h_k}(\nu)\right] \ge \mu\Big[E \setminus \mathrm{N}(\nu)\Big],$$

so

$$\mu(E) \geq \mu\Big[E \cap \mathrm{N}(\nu)\Big] + \mu\Big[E \setminus \mathrm{N}(\nu)\Big] - \epsilon \,,$$

and the result follows since ϵ is an arbitrary positive number. □

Corollary 1.7.18. *If μ is a Borel regular measure on the topological space X, then all the Suslin sets in X are μ-measurable.*

Corollary 1.7.19. *If $f : \mathbb{R}^N \to \mathbb{R}^M$ is continuous, μ is a Borel regular measure on $\mathbb{R}^M$, and $S \subseteq \mathbb{R}^N$ is a Suslin set, then $f(S)$ is μ-measurable.*

Remark 1.7.20. The particular properties of Euclidean space required for Corollary 1.7.19 are that every open set is a countable union of closed sets and that every closed set is a countable union of compact sets.