1 Introduction and Preliminaries

This chapter serves two purposes. The first purpose is to prepare the reader for a more systematic development in later chapters of the methods of real analysis through some introductory accounts of a few specific topics. The second purpose is, in view of the possible situation where some readers might not be conversant with basic concepts in elementary abstract analysis, to acquaint them with the fundamentals of abstract analysis. Nevertheless, readers are assumed to have some basic training in rigorous analysis as usually offered by courses in advanced calculus, and to have some acquaintance with the rudiments of linear algebra.

Throughout the book, the field of real numbers and that of complex numbers are denoted, respectively, by $\mathbb{R}$ and $\mathbb{C}$, while the set of all positive integers and the set of all integers are denoted by $\mathbb{N}$ and $\mathbb{Z}$ respectively.

The standard set-theoretical terminology is assumed; but terminology and notations regarding mappings will now be briefly recalled. If T is a mapping from a set A into a set B (expressed by $T : A \to B$), $T(a)$ denotes the element in B which is associated with $a \in A$ under the mapping T; for a subset **S** of **A**, the set $\{T(x) : x \in S\}$ is denoted by TS and is called the **image of *S* under *T***; thus $T\{a\} = \{T(a)\}$. $T(a)$ is sometimes simply written as Ta if no confusion is possible, and at times, an element a of a set and the set $\{a\}$ consisting of an element are not clearly distinguished as different objects. For example, Ta and $T\{a\}$ may not be distinguished and Ta is also called the image of a under T. A mapping $T : A \to B$ is said to be **one-to-one** or **injective** if $Ta = Ta'$ leads to $a = a'$, and is said to be **surjective** if $TA = B$; T is **bijective** if it is both injective and surjective. If $TA = B$, T is also referred to as a mapping from A onto B. Mappings are also called maps. Synonyms for maps are operators and transformations. As usual, a map from a set into $\mathbb{R}$ or $\mathbb{C}$ is called a function.

Some convenient notations for operations on sets are now introduced. Regarding a family $\mathcal{F} = \{A_\alpha\}_{\alpha \in I}$ of sets indexed by an index set I, the union $\bigcup_{\alpha \in I} A_\alpha$ is also expressed by $\bigcup \mathcal{F}$; if A and B are sets in a vector space and α a scalar, the set $\{x + y : x \in A, y \in B\}$ is denoted by $A + B$, and the set $\{\alpha x : x \in A\}$ by αA.

1.1 Summability of systems of real numbers

Summability of systems of real numbers is a special case in the theory of integration, to be treated in Chapter 2, but it reveals many essential points of the theory.

For a set S, the family of all nonempty finite subsets of S will be denoted by $F(S)$. Consider now a system $\{c_\alpha\}_{\alpha\in I}$ of real numbers indexed by an index set I. The system $\{c_\alpha\}_{\alpha\in I}$ will be denoted simply by $\{c_\alpha\}$ if the index set I is assumed either explicitly or implicitly. The system is said to be **summable** if there is $\ell \in \mathbb{R}$, such that for any $\varepsilon > 0$ there is $A \in F(I)$, with the property that whenever $B \in F(I)$ and $B \supset A$, then

$$\left|\sum_{\alpha\in B} c_\alpha - \ell\right| < \varepsilon. \tag{1.1}$$

Exercise 1.1.1 Show that if ℓ in the preceding definition exists, then it is unique.

If $\{c_\alpha\}$ is summable, the uniquely determined ℓ in the above definition is called the sum of $\{c_\alpha\}$ and is denoted by $\sum_{\alpha\in I} c_\alpha$.

Before we go further it is worthwhile remarking that the convergence of the series $\sum_{n=1}^{\infty} c_n$ depends on the order $1 < 2 < 3 < \cdots$ and $\sum_{n\in\mathbb{N}} c_n$, if it exists, does not depend on how $\mathbb{N}$ is ordered. Hence $\sum_{n\in\mathbb{N}} c_n$ may not exist while $\sum_{n=1}^{\infty} c_n$ exists. We will come back to this remark in Exercise 1.1.5.

Theorem 1.1.1 *If $\{c_\alpha^{(1)}\}_{\alpha\in I}$ and $\{c_\alpha^{(2)}\}_{\alpha\in I}$ are summable, then so is $\{ac_\alpha^{(1)} + bc_\alpha^{(2)}\}_{\alpha\in I}$ for fixed real numbers a and b, and*

$$\sum_{\alpha\in I}\left(ac_\alpha^{(1)} + bc_\alpha^{(2)}\right) = a\sum_{\alpha\in I} c_\alpha^{(1)} + b\sum_{\alpha\in I} c_\alpha^{(2)}.$$

Proof We may assume that $|a| + |b| > 0$, and for convenience put $\sum_{\alpha\in I} c_\alpha^{(1)} = l_1$, $\sum_{\alpha\in I} c_\alpha^{(2)} = l_2$. Let $\varepsilon > 0$ be given, there are A_1 and A_2 in $F(I)$ such that when B_1, B_2 are in $F(I)$ with $B_1 \supset A_1$, $B_2 \supset A_2$, we have $|\sum_{\alpha\in B_1} c_\alpha^{(1)} - l_1| < \frac{\varepsilon}{|a|+|b|}$ and $|\sum_{\alpha\in B_2} c_\alpha^{(2)} - l_2| < \frac{\varepsilon}{|a|+|b|}$. Choose now $A = A_1 \cup A_2$, then for $B \in F(I)$ with $B \supset A$, we have $|\sum_{\alpha\in B}(ac_\alpha^{(1)} + bc_\alpha^{(2)}) - (al_1 + bl_2)| \leq |a||\sum_{\alpha\in B} c_\alpha^{(1)} - l_1| + |b||\sum_{\alpha\in B} c_\alpha^{(2)} - l_2| < \frac{|a|\varepsilon}{|a|+|b|} + \frac{|b|\varepsilon}{|a|+|b|} = \varepsilon$. This shows that $\{ac_\alpha^{(1)} + bc_\alpha^{(2)}\}$ is summable and $\sum_{\alpha\in I}(ac_\alpha^{(1)} + bc_\alpha^{(2)}) = al_1 + bl_2$. ■

Theorem 1.1.2 *If $c_\alpha \geq 0\ \forall\alpha \in I$, then $\{c_\alpha\}$ is summable if and only if*

$$\left\{\sum_{\alpha\in A} c_\alpha : A \in F(I)\right\} \tag{1.2}$$

is bounded.

Proof That boundedness of (1.2) is necessary for $\{c_\alpha\}$ to be summable is left as an exercise. Now we show that boundedness of (1.2) is sufficient for $\{c_\alpha\}$ to be summable. Let ℓ be the least upper bound of $\{\sum_{\alpha\in A} c_\alpha : A \in F(I)\}$; for any $\varepsilon > 0$ there is $A \in F(I)$ such that

$$0 \leq \ell - \sum_{\alpha\in A} c_\alpha < \varepsilon. \tag{1.3}$$

Let now $B \in F(I)$ and $B \supset A$, then

$$\left|\sum_{\alpha\in B} c_\alpha - \ell\right| = \ell - \sum_{\alpha\in B} c_\alpha \leq \ell - \sum_{\alpha\in A} c_\alpha < \varepsilon.$$ ■

We note before moving on that if a subset S of $\mathbb{R}$ is bounded from above, then the least upper bound of S exists uniquely and is denoted by $\sup S$; similarly, if S is bounded from below, then the greatest lower bound exists uniquely and is denoted by $\inf S$. If $S = \{s_\alpha : \alpha \in I\}$, then $\inf S$ and $\sup S$ are also expressed, respectively, by $\inf_{\alpha\in I} s_\alpha$ and $\sup_{\alpha\in I} s_\alpha$.

Exercise 1.1.2 Show that boundedness of (1.2) is necessary for $\{c_\alpha\}$ to be summable.

Because of Theorem 1.1.2, if $\{c_\alpha\}$ is a system of nonnegative real numbers and is not summable, then we write $\sum_{\alpha\in I} c_\alpha = +\infty$. Hence, $\sum_{\alpha\in I} c_\alpha$ always has a meaning if $\{c_\alpha\}$ is a system of nonnegative numbers.

Theorem 1.1.3 (Cauchy criterion) *A system $\{c_\alpha\}$ is summable if and only if for any $\varepsilon > 0$ there is $A \in F(I)$, such that $|\sum_{\alpha\in B} c_\alpha| < \varepsilon$ whenever $B \in F(I)$ and $A \cap B = \emptyset$.*

Proof Sufficiency: Choose $A \in F(I)$ such that $|\sum_{\alpha\in B} c_\alpha| < 1$ for $B \in F(I)$, satisfying $A \cap B = \emptyset$, then obviously if $B \in F(I)$ with $B \cap A = \emptyset$, we have $\sum_{\alpha\in B} c_\alpha^+ < 1$, where $c_\alpha^+ = c_\alpha$ or 0 according to whether $c_\alpha \geq 0$ or < 0. Now, for $B \in F(I)$, we have

$$\sum_{\alpha\in B} c_\alpha^+ = \sum_{\alpha\in B\cap A} c_\alpha^+ + \sum_{\alpha\in B\setminus A} c_\alpha^+ < \sum_{\alpha\in A} c_\alpha^+ + 1,$$

i.e., $\{\sum_{\alpha\in B} c_\alpha^+ : B \in F(I)\}$ is bounded; hence by Theorem 1.1.2 $\{c_\alpha^+\}$ is summable.

Similarly $\{c_\alpha^-\}$ is summable, where $c_\alpha^- = -c_\alpha$ or 0 according to whether $c_\alpha \leq 0$ or > 0. Now $c_\alpha = c_\alpha^+ - c_\alpha^-$, hence $\{c_\alpha\}$ is summable by Theorem (1.1).

The necessary part is left for the reader to verify. ■

Exercise 1.1.3 Suppose that $\{c_\alpha\}_{\alpha\in I}$ is summable and that J is a nonempty subset of I. Show that (i) $\{c_\alpha\}_{\alpha\in J}$ is summable, and (ii) $\sum_{\alpha\in I} c_\alpha = \sum_{\alpha\in J} c_\alpha + \sum_{\alpha\in I\setminus J} c_\alpha$.

Exercise 1.1.4 Show that $\{c_\alpha\}$ is summable if and only if $\{|c_\alpha|\}$ is summable; show also that $\{c_\alpha\}$ is summable if and only if

$$\left\{\left|\sum_{\alpha\in A} c_\alpha\right| : A \in F(I)\right\}$$

is bounded.

Exercise 1.1.5 Show that $\{c_\alpha\}_{\alpha\in\mathbb{N}}$ is summable if and only if the series $\sum_{\alpha=1}^{\infty} c_\alpha$ is absolutely convergent. Show also that $\sum_{\alpha\in\mathbb{N}} c_\alpha = \sum_{\alpha=1}^{\infty} c_\alpha$ if $\{c_\alpha\}_{\alpha\in\mathbb{N}}$ is summable.

Exercise 1.1.6 Show that $\{c_\alpha\}_{\alpha\in I}$ is summable if and and only if (i) $\{\alpha \in I : c_\alpha \neq 0\}$ is finite or countable; and (ii) if $\{\alpha \in I : c_\alpha \neq 0\} = \{\alpha_1, \alpha_2, \ldots\}$ is infinite; then the series $\sum_{k=1}^{\infty} c_{\alpha_k}$ converges absolutely.

Exercise 1.1.7 Suppose that for each $n = 1, 2, 3, \ldots$, there is $A_n \in F(I)$, with the property that for each $A \in F(I)$, there is a positive integer N such that $A \subset A_n$ for all $n \geq N$. Show that if $\{c_\alpha\}_{\alpha\in I}$ is summable, then

$$\sum_{\alpha\in I} c_\alpha = \lim_{n\to\infty} \sum_{\alpha\in A_n} c_\alpha.$$

Give an example to show that it is possible that $\lim_{n\to\infty} \sum_{\alpha\in A_n} c_\alpha$ exists and is finite, but $\{c_\alpha\}$ is not summable.

Example 1.1.1 Suppose that $I = \bigcup_{n\in\mathbb{N}} I_n$, where I_n's are pairwise disjoint. Let $\{c_\alpha\}_{\alpha\in I}$ be summable, then $\sum_{\alpha\in I} c_\alpha = \sum_{n\in\mathbb{N}} (\sum_{\alpha\in I_n} c_\alpha)$. By Exercise 1.1.4, we may assume that $c_\alpha \geq 0$ for all $\alpha \in I$. It follows from $\sum_{\alpha\in I} c_\alpha = \sup\{\sum_{\alpha\in A} c_\alpha : A \in F(I)\}$ that $\sum_{\alpha\in I} c_\alpha \leq \sum_{n\in\mathbb{N}} (\sum_{\alpha\in I_n} c_\alpha)$. It remains to be seen that $\sum_{\alpha\in I} c_\alpha \geq \sum_{n\in\mathbb{N}} (\sum_{\alpha\in I_n} c_\alpha)$. Let $k \in \mathbb{N}$ and $\varepsilon > 0$. For each $n = 1, \ldots, k$, there is a finite set $A_n \subset I_n$ such that $\sum_{\alpha\in I_n} c_\alpha < \sum_{\alpha\in A_n} c_\alpha + \frac{\varepsilon}{k}$. Then, if we put $B_k = \bigcup_{n=1}^{k} A_n$, we have $\sum_{\alpha\in I} c_\alpha \geq \sum_{\alpha\in B_k} c_\alpha > \sum_{n=1}^{k} (\sum_{\alpha\in I_n} c_\alpha - \frac{\varepsilon}{k}) = \sum_{n=1}^{k} (\sum_{\alpha\in I_n} c_\alpha) - \varepsilon$; since $\varepsilon > 0$ is arbitrary, $\sum_{\alpha\in I} c_\alpha \geq \sum_{n=1}^{k} (\sum_{\alpha\in I_n} c_\alpha)$ for each $k \in \mathbb{N}$. Now let $k \to \infty$ to obtain $\sum_{\alpha\in I} c_\alpha \geq \sum_{n\in\mathbb{N}} (\sum_{\alpha\in I_n} c_\alpha)$. Observe from the proof that $\{\sum_{\alpha\in I_n} c_\alpha\}_{n\in\mathbb{N}}$ is summable.

We shall recognize in Example 2.3.3 that summability considered in this section is the integrability with respect to the counting measure on I.

1.2 Double series

Let $I = \mathbb{N} \times \mathbb{N} = \{(i, j) : i, j = 1, 2, \ldots\}$ and write c_{ij} for $c_{(i,j)}$. When the summability of the system $\{c_{ij}\}$ is in question, the system $\{c_{ij}\}$ is referred to as a **double series** and is denoted by $\sum c_{ij}$. Hence the double series $\sum c_{ij}$ is summable if $\{c_{ij}\} = \{c_{(i,j)}\}$ is summable, and $\sum_{(i,j)\in I} c_{ij}$ is called the sum of the double series $\sum c_{ij}$.

For a double sequence $\{a_{mn}\}$, we say that $\lim_{m,n\to\infty} a_{mn} = \ell$, if for any $\varepsilon > 0$ there is a positive integer N such that $|a_{mn} - \ell| < \varepsilon$ whenever $m, n \geq N$.

Theorem 1.2.1 *If the double series $\sum c_{ij}$ is summable, then*

$$\sum_{(i,j)\in I} c_{ij} = \lim_{m,n\to\infty} \sum_{j=1}^{n} \sum_{i=1}^{m} c_{ij} = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} c_{ij} = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} c_{ij}.$$

Proof We show first that $\sum_{(i,j)\in I} c_{ij} = \lim_{n,m\to\infty} \sum_{j=1}^{n} \sum_{i=1}^{m} c_{ij}$. Let $\ell = \sum_{(i,j)\in I} c_{ij}$. Given $\varepsilon > 0$, there is $A \in F(I)$ such that

$$\left| \sum_{(i,j)\in B} c_{ij} - \ell \right| < \varepsilon$$

whenever $B \in F(I)$ and $B \supset A$. Let $N = \max\{i \vee j : (i,j) \in A\}$, where $i \vee j$ is the larger of i and j. For $n, m \geq N$, let $B_{mn} = \{(i,j) \in I : 1 \leq i \leq m,\ 1 \leq j \leq n\}$, then $B_{mn} \in F(I)$ and $B_{mn} \supset A$, hence

$$\left| \sum_{j=1}^{n} \sum_{i=1}^{m} c_{ij} - \ell \right| = \left| \sum_{(i,j)\in B_{mn}} c_{ij} - \ell \right| < \varepsilon.$$

This means that $\ell = \lim_{m,n\to\infty} \sum_{j=1}^{n} \sum_{i=1}^{m} c_{ij}$.

Since $\sum_{(i,j)\in I} c_{ij} = \sum_{(i,j)\in I} c_{ij}^{+} - \sum_{(i,j)\in I} c_{ij}^{-}$, in the remaining part of the proof, we may assume that $c_{ij} \geq 0$ for all $(i,j) \in I$. Observe then that

$$\ell = \sup_{n,m\geq 1} \sum_{j=1}^{n} \sum_{i=1}^{m} c_{ij}.$$

Hence,

$$\ell \geq \lim_{m\to\infty} \left(\sum_{j=1}^{n} \sum_{i=1}^{m} c_{ij} \right) = \sum_{j=1}^{n} \sum_{i=1}^{\infty} c_{ij}$$

for each n and consequently

$$\ell \geq \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} c_{ij}.$$

On the other hand,

$$\begin{aligned} \ell &= \sup_{n,m\geq 1} \sum_{j=1}^{n} \sum_{i=1}^{m} c_{ij} \leq \sup_{n\geq 1} \left(\sum_{j=1}^{n} \sum_{i=1}^{\infty} c_{ij} \right) = \lim_{n\to\infty} \left(\sum_{j=1}^{n} \sum_{i=1}^{\infty} c_{ij} \right) \\ &= \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} c_{ij}. \end{aligned}$$

We have shown that $\ell = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} c_{ij}$; similarly,

$$\ell = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} c_{ij}. \qquad \blacksquare$$

Example 1.2.1 If $\{a_n\}_{n\in\mathbb{N}}$ and $\{b_n\}_{n\in\mathbb{N}}$ are summable, then the double series $\sum a_n b_m$ is summable and $\sum_{(n,m)\in\mathbb{N}\times\mathbb{N}} a_n b_m = (\sum_{n\in\mathbb{N}} a_n)(\sum_{m\in\mathbb{N}} b_m)$. That $\sum a_n b_m$ is summable follows from Exercise 1.1.4 and the observation that $\{\sum_{(n,m)\in A} |a_n b_m| : A \in F(\mathbb{N}\times\mathbb{N})\}$ is bounded from above by $(\sum_{n\in\mathbb{N}} |a_n|)\cdot(\sum_{m\in\mathbb{N}} |b_m|)$. Then, by Theorem 1.2.1, $\sum_{(n,m)\in\mathbb{N}\times\mathbb{N}} a_n b_m = \sum_{n\in\mathbb{N}}\sum_{m\in\mathbb{N}} a_n b_m = (\sum_{n\in\mathbb{N}} a_n)(\sum_{m\in\mathbb{N}} b_m)$. For $k \geq 2$ in $\mathbb{N}$, put $A_k = \{(n,m) \in \mathbb{N}\times\mathbb{N} : n+m=k\}$; then $\sum_{(n,m)\in\mathbb{N}\times\mathbb{N}} a_n b_m = \sum_{\substack{k\in\mathbb{N}\\ k\geq 2}} (\sum_{(n,m)\in A_k} a_{nm})$ from Example 1.1.1. The system $\{\sum_{(n,m)\in A_k} a_n b_m\}_{k\geq 2}$ is called the product of $\{a_n\}$ and $\{b_n\}$; we have shown that the sum of the product is the product of the sums.

The following exercise complements Theorem 1.2.1.

Exercise 1.2.1 Copy the proof of Theorem 1.2.1 to show that if $c_{ij} \geq 0$ for all i and j in $\mathbb{N}$, then the conclusion of Theorem 1.2.1 still holds, even if $\sum_{(i,j)\in I} c_{ij} = \infty$ (recall that for a system $\{c_\alpha\}$ of nonnegative numbers, $\sum_\alpha c_\alpha = \infty$ means that $\{c_\alpha\}$ is not summable).

Remark For i, j in $\mathbb{N}$, let

$$c_{ij} = \begin{cases} 1 & \text{if } i = j; \\ -1 & \text{if } j = i+1; \\ 0 & \text{otherwise,} \end{cases}$$

then $\sum c_{ij}$ is not summable and $0 = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} c_{ij} \neq \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} c_{ij} = 1$.

1.3 Coin tossing

A pair of symbols H and T, associated, respectively, with nonnegative numbers p and q such that $p + q = 1$ is called a **Bernoulli trial** and is denoted by $B(p, q)$. A Bernoulli trial $B(p, q)$ is a mathematical model for the tossing of a coin, of which heads occur with probability p and tails turn out with probability q; this explains the symbols H and T. In particular, $B(\frac{1}{2}, \frac{1}{2})$ models the tossing of a fair coin.

In this section, we consider the first step towards construction of a mathematical model for a sequence of tossing of a fair coin. For convenience, we replace H and T by 1 and 0 in this order; then an infinite sequence $\omega = (\omega_1, \omega_2, \ldots, \omega_k, \ldots)$ of 0's and 1's represents a realization of a sequence of coin tossing. Let

$$\Omega = \{0,1\}^\infty := \{\omega = (\omega_k),\ \omega_k = 0 \text{ or } 1 \text{ for each } k\},$$

where we adopt the usual convention of expressing an infinite sequence $(\omega_1, \ldots, \omega_k, \ldots)$ by (ω_k) with the understanding that ω_k is the entry at the k-th position of the sequence. In terminology of probability theory, elements in Ω are called **sample points** of a sequence of coin tossings and Ω is called the **sample space** of the sequence of tossings. Subsets of Ω will often be referred to as **events**. Now for $n \in \mathbb{N}$, let

$$\Omega_n = \{0, 1\}^n := \{(\varepsilon_1, \ldots, \varepsilon_n) : \varepsilon_j \in \{0, 1\},\ j = 1, \ldots, n\},$$

and for $(\varepsilon_1, \ldots, \varepsilon_n) \in \Omega_n$, call the set

$$E(\varepsilon_1, \ldots, \varepsilon_n) = \{\omega = (\omega_k) \in \Omega : \omega_k = \varepsilon_k,\ k = 1, \ldots, n\}$$

an **elementary cylinder**; but if n is to be emphasized, it is called an **elementary cylinder of rank n**. A finite union of elementary cylinders is called a **cylinder** in Ω. Since intersection of two elementary cylinders is either empty or an elementary cylinder, every cylinder in Ω can be expressed as a disjoint union of elementary cylinders; in fact, if Z is a cylinder in Ω, there is $n \in \mathbb{N}$ and $H \subset \Omega_n$ such that

$$Z = \bigcup\{E(\varepsilon_1, \ldots, \varepsilon_n) : (\varepsilon_1, \ldots, \varepsilon_n) \in H\},$$

of which one notes that $E(\varepsilon_1, \ldots, \varepsilon_n)$'s are mutually disjoint. Of course, a cylinder Z can be expressed as above in many ways. We denote by $\mathcal{Q}$ the family of all cylinders in Ω. Since $\Omega = E(0) \cup E(1)$, $\Omega \in \mathcal{Q}$; $\emptyset$ is also in $\mathcal{Q}$, because it is the union of an empty family of elementary cylinders.

Exercise 1.3.1 Show that $\mathcal{Q}$ is an algebra of subsets of Ω, in the sense that $\mathcal{Q}$ satisfies the following conditions: (i) $\Omega \in \mathcal{Q}$; (ii) if $Z \in \mathcal{Q}$, then $Z^c = \Omega \backslash Z$ is in $\mathcal{Q}$; and (iii) if Z_1, Z_2 are in $\mathcal{Q}$, then $Z_1 \cup Z_2$ is in $\mathcal{Q}$.

For an event Z in $\mathcal{Q}$, we define its probability $P(Z)$ as follows. First, for an elementary cylinder $C = E(\varepsilon_1, \ldots, \varepsilon_n)$, define $P(C) = (\frac{1}{2})^n$; intuitively, this definition of $P(C)$ means that we consider the modeling of a sequence of independent tossing of a fair coin. Now if $Z \in \mathcal{Q}$ is given by

$$Z = \bigcup\{E(\varepsilon_1, \ldots, \varepsilon_n) : (\varepsilon_1, \ldots, \varepsilon_n) \in H\},$$

where $H \subset \Omega_n$, then define

$$P(Z) = \sum_{(\varepsilon_1, \ldots, \varepsilon_n) \in H} P(E(\varepsilon_1, \ldots, \varepsilon_n)) = \#H \cdot 2^{-n},$$

where $\#H$ is the number of elements in H. We claim that $P(Z)$ is well defined. Actually if Z is also given by

$$Z = \bigcup\{E(\varepsilon_1, \ldots, \varepsilon_m) : (\varepsilon_1, \ldots, \varepsilon_m) \in H'\},$$

where $H' \subset \Omega_m$, then (assuming $m \geq n$) $H' = \{(\varepsilon_1, \ldots, \varepsilon_m) \in \Omega_m : (\varepsilon_1, \ldots, \varepsilon_n) \in H\}$ and therefore $\#H' = \#H \cdot 2^{m-n}$; consequently

$$\sum_{(\varepsilon_1,\ldots,\varepsilon_m)\in H'} P(E(\varepsilon_1, \ldots, \varepsilon_m)) = \#H' \cdot 2^{-m} = \#H \cdot 2^{m-n} \cdot 2^{-m}$$

$$= \#H \cdot 2^{-n} = \sum_{(\varepsilon_1,\ldots,\varepsilon_n)\in H} P(E(\varepsilon_1, \ldots, \varepsilon_n)),$$

implying that the definition of $P(Z)$ is independent of how Z is expressed as a finite disjoint union of elementary cylinders of a given rank. We complete the definition of P by letting $P(\emptyset) = 0$. Note that $P(\Omega) = 1$.

Exercise 1.3.2

(i) Show that P is additive on $\mathcal{Q}$, i.e. $P(Z_1 \cup Z_2) = P(Z_1) + P(Z_2)$ if Z_1, Z_2 are disjoint elements of $\mathcal{Q}$.

(ii) For $k \in \mathbb{N}$ and $\varepsilon \in \{0, 1\}$, put $E_\varepsilon^k = \{\omega \in \Omega : \omega_k = \varepsilon\}$. Show that

$$P(E_{\varepsilon_1}^{k_1} \cap \cdots \cap E_{\varepsilon_n}^{k_n}) = \prod_{j=1}^{n} P(E_{\varepsilon_j}^{k_j}) = 2^{-n}$$

for any finite sequence $k_1 < k_2 < \cdots < k_n$ in $\mathbb{N}$.

From now on we write $d_j(\omega) = \omega_j, j = 1, 2, \ldots,$ if $\omega = (\omega_1, \omega_2, \ldots) \in \Omega$; and for each n define a function S_n on Ω by

$$S_n(\omega) = \sum_{j=1}^{n} d_j(\omega).$$

Exercise 1.3.3 Show that, for each $k = 0, 1, 2, \ldots, n$, the set $\{S_n = k\} := \{\omega \in \Omega : S_n(\omega) = k\}$ is in $\mathcal{Q}$ and

$$P(\{S_n = k\}) = \binom{n}{k} \frac{1}{2^n},$$

where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

For a given realization ω of a sequence of independent coin tossing, $S_n(\omega)$ is the number of heads that appear in the first n tosses and $\frac{S_n(\omega)}{n}$ measures the relative frequency of appearance of heads in the first n tosses. Let

$$E = \left\{ \omega \in \Omega : \lim_{n\to\infty} \frac{S_n(\omega)}{n} = \frac{1}{2} \right\};$$

E is easily seen to be not in $\mathcal{Q}$. Nevertheless, we expect that P can be extended to be defined on a larger family of sets than $\mathcal{Q}$ in such a way that $P(A)$ can be interpreted as the probability of event A, and such that $P(E)$ is defined with value 1. We expect $P(E) = 1$, because this is what a fair coin is accounted for intuitively. Discussion of the subject matter of this section will be continued in Example 1.7.1, Example 2.1.1, Example 3.4.6, and Example 7.5.2; and eventually we shall answer positively to this expectation in the paragraph following Corollary 7.5.3.

1.4 Metric spaces and normed vector spaces

The usefulness of the concept of continuity has already surfaced in elementary analysis of functions defined on an interval. This section considers a structure on a set which allows one to speak of "nearness" for elements in the set, so that a concept of continuity can be defined for functions defined on the set, parallel to that for functions defined on an interval of the real line. We shall not treat the most general situation; instead, we consider the situation where an abstract concept of distance can be defined between elements of the set, because this situation abounds sufficiently for our purposes later. When the set considered is a vector space, it is natural to consider the case where the distance defined and the linear structure of the set mingle well, as in the case of a real line or Euclidean plane. This leads to the concept of normed vector spaces.

Let M be a nonempty set and let $\rho : M \times M \to [0, +\infty)$ satisfy (i) $\rho(x,y) = \rho(y,x) \geq 0$ for all $x, y \in M$ and $\rho(x,y) = 0$ if and only if $x = y$; (ii) $\rho(x,z) \leq \rho(x,y) + \rho(y,z)$ for all x, y, and z in M. Such a ρ is then called a **metric** on M, and (M, ρ) is called a **metric space**. Usually we say that M is a metric space with metric ρ, or simply that M is a metric space when a certain metric ρ is explicitly or implicitly implied. For a nonempty subset S of M the restriction of ρ to $S \times S$ is a metric on S which will also be denoted by ρ. The metric space (S, ρ) is called a subspace of (M, ρ) and ρ is called the metric on S inherited from M. Unless stated otherwise, if S is a subset of a metric space M, S is equipped with the metric inherited from M. For a nonempty subset A of M, the **diameter** of A, denoted diam A, is defined by

$$\operatorname{diam} A := \sup_{x,y \in A} \rho(x,y);$$

while $\operatorname{diam} A = 0$ if $A = \emptyset$.

A subset A of M is said to be **bounded** if $\operatorname{diam} A < \infty$. In other words, A is bounded if $\{\rho(x, x_0) : x \in A\}$ is a bounded set in $\mathbb{R}$ for every $x_0 \in M$.

Elements of a metric space are often called **points** of the space.

Example 1.4.1 Let $M = \mathbb{R}^n$ and for $x, y \in \mathbb{R}^n$ let $\rho(x,y) = |x - y|$, where $|x| = (\sum_{i=1}^n x_i^2)^{\frac{1}{2}}$ if $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$. To show that ρ is a metric on $\mathbb{R}^n$ we first establish the well-known **Schwarz inequality**: $|x \cdot y| \leq |x||y|$ if $x, y \in \mathbb{R}^n$, where, for $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ in $\mathbb{R}^n$, $x \cdot y := \sum_{i=1}^n x_i y_i$ is called the inner product of x and y. For this purpose we note first that for $x \in \mathbb{R}^n$, $|x|^2 = x \cdot x$ and that we may assume that $x \neq 0$ and $y \neq 0$, hence $|x| > 0$ and $|y| > 0$. For $t \in \mathbb{R}$, we have

$$\begin{aligned} 0 \leq |x + ty|^2 &= (x + ty) \cdot (x + ty) = |x|^2 + 2t(x \cdot y) + t^2|y|^2 \\ &= \big(|x| + t|y|\big)^2 + 2t\big(x \cdot y - |x||y|\big), \end{aligned}$$

from which by taking $t = -|x|/|y|$ we obtain $x \cdot y \leq |x||y|$. Then $|x \cdot y| \leq |x||y|$ follows, because $-(x \cdot y) \leq |x|| - y| = |x||y|$. Now for x, y, and z in $\mathbb{R}^n$, we have

$$\begin{aligned} \rho(x,z)^2 &= |x - z|^2 = |x - y + y - z|^2 = |x - y|^2 + 2(x - y) \cdot (y - z) + |y - z|^2 \\ &\leq |x - y|^2 + 2|x - y||y - z| + |y - z|^2 = \big(|x - y| + |y - z|\big)^2 \\ &= \big[\rho(x,y) + \rho(y,z)\big]^2, \end{aligned}$$

i.e.

$$\rho(x,z) \leq \rho(x,y) + \rho(y,z).$$

Hence $\mathbb{R}^n$ is a metric space with metric ρ defined above. This metric is called the **Euclidean metric** on $\mathbb{R}^n$. Unless stated otherwise, $\mathbb{R}^n$ is considered as a metric space with this metric, then $\mathbb{R}^n$ is called the n-dimensional Euclidean space.

Similarly, $\mathbb{C}^n$ is a metric space, with the metric ρ defined by $\rho(\zeta, \eta) = (\sum_{j=1}^n |\zeta_j - \eta_j|^2)^{1/2}$ for $\zeta = (\zeta_1, \ldots, \zeta_n)$ and $\eta = (\eta_1, \ldots, \eta_2)$ in $\mathbb{C}^n$. $\mathbb{C}^n$ with this metric is called the **n-dimensional unitary space**. This follows, as in the case of the Euclidean metric for $\mathbb{R}^n$, from the Schwarz inequality $|\zeta \cdot \eta| \leq |\zeta||\eta|$ for ζ, η in $\mathbb{C}^n$, where $\zeta \cdot \eta = \sum_{j=1}^n \zeta_j \bar{\eta}_j$ and $|\zeta| = (\sum_{j=1}^n |\zeta_j|^2)^{\frac{1}{2}}$. As before, if $t \in \mathbb{R}$, we have

$$\begin{aligned} 0 \leq |\zeta + t\eta|^2 &= (\zeta + t\eta) \cdot (\zeta + t\eta) = |\zeta|^2 + 2t \operatorname{Re} \zeta \cdot \eta + t^2|\eta|^2 \\ &= (|\zeta| + t|\eta|)^2 + 2t\{\operatorname{Re} \zeta \cdot \eta - |\zeta||\eta|\}, \end{aligned}$$

from which we infer that $\operatorname{Re} \zeta \cdot \eta \leq |\zeta||\eta|$ by choosing $t = -|\zeta||\eta|^{-1}$ if $\eta \neq 0$. Then, $|\zeta \cdot \eta| \leq |\zeta||\eta|$ follows from replacing ζ by $e^{-i\theta}\zeta$ if $\zeta \cdot \eta = |\zeta \cdot \eta|e^{i\theta}$. Note that for a complex number α, $\bar{\alpha}$ denotes the conjugate of α, while $\operatorname{Re} \alpha$ denotes the real part of α.

Example 1.4.2 For a closed finite interval $[a, b]$ in $\mathbb{R}$, let $C[a, b]$ denote the space of all real-valued continuous functions defined on $[a, b]$. For $f, g \in C[a, b]$, let $\rho(f, g) = \max_{a \leq t \leq b} |f(t) - g(t)|$. It is easily verified that $C[a, b]$ is a metric space with metric ρ so defined. Unless stated otherwise, $C[a, b]$ is equipped with this metric, which is often referred to as the **uniform metric** on $C[a, b]$. $C[a, b]$ is also used to denote the space of all complex-valued continuous functions on $[a, b]$ with metric defined similarly. When $C[a, b]$ denotes the latter space, it shall be explicitly indicated.

Exercise 1.4.1 Show that $\mathbb{R}^n$ is also a metric space, with metric ρ defined by $\rho(x, y) = \max_{1 \leq i \leq n} |x_i - y_i|$ if $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$.

A map from $\mathbb{N}$, the set of all positive integers, to a set M is called a **sequence** in M or a **sequence of elements** of M. Such a sequence will be denoted by $\{x_n\}$, where x_n

is the image of the positive integer n under the mapping. If $\{x_n\}$ is a sequence in M, then $\{x_{n_k}\}$ is called a subsequence of $\{x_n\}$ if $n_1 < n_2 < \cdots < n_k < \cdots$ is a subsequence of $\{n\}$. A sequence $\{x_n\}$ in a metric space M is said to converge to $x \in M$ if for any $\varepsilon > 0$ there is $n_0 \in \mathbb{N}$ such that $\rho(x_n, x) < \varepsilon$ whenever $n \geq n_0$. Since x is uniquely determined, x is called the **limit** of $\{x_n\}$ and is denoted by $\lim_{n\to\infty} x_n$. That $x = \lim_{n\to\infty} x_n$ is often expressed by $x_n \to x$. If $\lim_{n\to\infty} x_n$ exists, then we say that $\{x_n\}$ converges in M and $\{x_n\}$ is referred to as a **convergent sequence**. A sequence $\{x_n\}$ in M is usually expressed by $\{x_n\} \subset M$ by abuse of notation, and therefore $\{x_n\}$ also denotes the range of the sequence $\{x_n\}$. A sequence in M is said to be bounded if its range is bounded.

Example 1.4.3 $\{f_n\} \subset C[a, b]$ converges if and only if $f_n(x)$ converges uniformly for $x \in [a, b]$.

A sequence $\{x_n\} \subset M$ is called a **Cauchy sequence** if for any $\varepsilon > 0$, there is $n_0 \in \mathbb{N}$ such that $\rho(x_n, x_m) < \varepsilon$ whenever $n, m \geq n_0$. Clearly, a Cauchy sequence is bounded.

Exercise 1.4.2 Show that if $\{x_n\} \subset M$ converges, then $\{x_n\}$ is a Cauchy sequence.

Exercise 1.4.3 Let $\{x_n\}$ be a Cauchy sequence. Show that if $\{x_n\}$ has a convergent subsequence, then $\{x_n\}$ converges.

A metric space M is called ***complete*** if every Cauchy sequence in M converges in M.

Exercise 1.4.4 Show that both $\mathbb{R}^n$ and $C[a, b]$ are complete.

Exercise 1.4.5 If instead of the uniform metric we equip $C[a, b]$ with a new metric ρ', defined by

$$\rho'(f, g) = \int_a^b |f(t) - g(t)| dt$$

for f, g in $C[a, b]$, show that $C[a, b]$ is not complete when considered as a metric space with metric ρ'.

Exercise 1.4.6 Show that any nonempty set M can be considered as a complete metric space by defining $\rho(x, y) = 0$ or 1 depending on $x = y$ or $x \neq y$. Such a metric ρ is said to be discrete.

Let M_1, M_2 be metric spaces with metrics ρ_1 and ρ_2 respectively. A map $T : M_1 \to M_2$ is said to be **continuous at** $x \in M_1$ if for any $\varepsilon > 0$, there is $\delta > 0$ such that $\rho_2(T(x), T(y)) < \varepsilon$ whenever $\rho_1(x, y) < \delta$. If T is continuous at every point of M_1, then T is said to be continuous on M_1 and is called a **continuous map** from M_1 into M_2. A continuous map from a metric space M into $\mathbb{R}$ or $\mathbb{C}$ is called a **continuous function** on M and is generically denoted by f. The space of all continuous real(complex)-valued functions on a metric space M is denoted by $C(M)$; $C(M)$ is a real- or complex vector space depending on whether the functions in question are real- or complex-valued.

A point x of a set A in a metric space is called an **interior** point of A if there is $\varepsilon > 0$ such that $y \in A$ whenever $\rho(x, y) < \varepsilon$; the set of all interior points of A is denoted by $\mathring{A}$. A set G in a metric space M is said to be **open** if $\mathring{G} = G$. The complement of an open set is called a **closed** set. For $x \in M$ and $r > 0$, let $B_r(x) = \{y \in M : \rho(y,x) < r\}$ and $C_r(x) = \{y \in M : \rho(y,x) \leq r\}$. It is easily verified that $B_r(x)$ is an open set and $C_r(x)$ is a closed set. $B_r(x)$ ($C_r(x)$) is usually referred to as the **open (closed)** ball centered at x and with radius r. A point $x \in M$ is said to be **isolated** if $B_r(x) = \{x\}$ for some $r > 0$. A set $N \subset M$ is called a **neighborhood** of $x \in M$ if N contains an open set which contains x; similarly, if N contains an open set which contains a set A, then N is called a **neighborhood** of A. It is clear that a sequence $\{x_n\}$ in M converges to $x \in M$ if and only if, for any neighborhood N of x, there is $n_0 \in \mathbb{N}$ such that $x_n \in N$ whenever $n \geq n_0$. One notes that if x_0 is an isolated point of M, then any map T from M into any metric space is continuous at x_0.

Note that open sets depend on the metric ρ, and when ρ is to be emphasized, an open set in a metric space with metric ρ is more precisely said to be open w.r.t. ρ.

Exercise 1.4.7 Let M_1, M_2 be metric spaces and let $T : M_1 \to M_2$.

(i) Show that T is continuous at $x \in M_1$ if and only if, for any sequence $\{x_n\} \subset M_1$ with $\lim_{n\to\infty} x_n = x$, it holds that $\lim_{n\to\infty} T(x_n) = T(x)$ in M_2; also show that T is continuous at $x \in M_1$ if and only if, for every sequence $\{x_n\} \subset M_1$ with $\lim_{n\to\infty} x_n = x$, it holds that $\{x_n\}$ has a subsequence $\{x_{n_k}\}$ such that $\lim_{k\to\infty} T(x_{n_k}) = T(x)$.

(ii) Show that T is continuous at $x \in M_1$ if and only if, for any neighborhood N of $T(x)$ in M_2, the set $T^{-1}N = \{y \in M_1 : T(y) \in N\}$ is a neighborhood of x in M_1.

(iii) Show that T is continuous on M_1 if and only if for any open set $G_2 \subset M_2$, $T^{-1}G_2$ is an open subset of M_1.

Exercise 1.4.8 Let $\mathcal{T}$ be the family of all open subsets of a metric space M. Show that:

(i) $\emptyset$ and M are in $\mathcal{T}$;

(ii) $A, B \in \mathcal{T} \Rightarrow A \cap B \in \mathcal{T}$;

(iii) if $\{A_i\}_{i\in I} \subset \mathcal{T}$, then $\bigcup_{i\in I} A_i \in \mathcal{T}$, where I is any index set.

Suppose that (M_1, ρ_1) and (M_2, ρ_2) are metric spaces. Let $M_1 \times M_2 := \{(x,y) : x \in M_1, y \in M_2\}$ be the **Cartesian product** of M_1 and M_2; define a metric ρ on $M_1 \times M_2$ by

$$\rho((x,y),(x',y')) = \rho_1(x,x') + \rho_2(y,y')$$

for (x,y) and (x',y') in $M_1 \times M_2$. It is easily verified that ρ is actually a metric on $M_1 \times M_2$. With this metric ρ, $M_1 \times M_2$ is called the **product space** of M_1 and M_2 as metric space.

Exercise 1.4.9 Let $M_1 \times M_2$ be the product space of metric spaces M_1 and M_2.

(i) For $A \subset M_1$ and $B \subset M_2$, show that $A \times B$ is open in $M_1 \times M_2$ if and only if both A and B are open in M_1 and M_2 respectively.

(ii) Let G be an open set in $M_1 \times M_2$; show that $G_1 := \{x \in M_1 : (x,y) \in G$ for some y in $M_2\}$ and $G_2 := \{y \in M_2 : (x,y) \in G$ for some x in $M_1\}$ are open in M_1 and M_2 respectively.

Let $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$ and let E be a vector space over $\mathbb{K}$. Elements of $\mathbb{K}$ are called scalars. Suppose that for each $x \in E$, there is a nonnegative number $\|x\|$ associated with it so that:

(i) $\|x\| = 0$ if and only if x is the zero element of E;

(ii) $\|\alpha x\| = |\alpha|\|x\|$ for all $\alpha \in \mathbb{K}$ and $x \in E$;

(iii) $\|x + y\| \leq \|x\| + \|y\|$ for all x, y in E (**triangle inequality**).

Then E is called a **normed vector space** (abbreviated as **n.v.s.**) with **norm** $\|\cdot\|$, and $\|\cdot\|$ is called a **norm on** E.

If E is a n.v.s., for x, y in E, let

$$\rho(x, y) = \|x - y\|,$$

then ρ is a metric on E and is called the metric associated with norm $\|\cdot\|$. Unless stated otherwise, we always consider this metric for a n.v.s.. The n.v.s. E with norm $\|\cdot\|$ is denoted by $(E, \|\cdot\|)$ if the norm $\|\cdot\|$ is to be emphasized.

Lemma 1.4.1 *Suppose that E is a n.v.s. and $x_n \to x$ in E, then $\|x\| = \lim_{n\to\infty} \|x_n\|$. In other words, $\|\cdot\|$ is a continuous function on E.*

Proof The lemma follows from the following sequence of triangle inequalities:

$$\|x_n\| - \|x_n - x\| \leq \|x\| \leq \|x_n\| + \|x_n - x\|. \qquad \blacksquare$$

A *normed vector space* is called a **Banach space** if it is a complete metric space.

Both $\mathbb{R}^n$ and $C[a, b]$ are Banach spaces, with norms given respectively by $\|x\| = (\sum_{i=1}^n x_i^2)^{\frac{1}{2}}$ for $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and $\|f\| = \max_{a\leq t\leq b} |f(t)|$ for $f \in C[a, b]$. Similarly, the unitary space $\mathbb{C}^n$ is a Banach space with norm $\|z\| = (\sum_{j=1}^n |z_j|^2)^{\frac{1}{2}}$ for $z = (z_1, \ldots, z_n)$ in $\mathbb{C}^n$. The norms defined above for $\mathbb{R}^n$ and $\mathbb{C}^n$ are called respectively the **Euclidean norm** and the **unitary norm** and are denoted by $|\cdot|$ in both cases, in accordance with the notations introduced in Example 1.4.1; note that their associated metrics are the metrics introduced for $\mathbb{R}^n$ and C^n in Example 1.4.1. The norm defined for $C[a, b]$ is called the **uniform norm**; its associated metric is the uniform metric defined in Example 1.4.2.

A class of well-known Banach spaces, the l^p spaces, will be introduced in §1.6. This class of Banach spaces anticipates the important and more general class of L^p spaces treated in Section 2.7 and in Chapter 6.

In the remaining part of this section, linear maps from a normed vector space E into a normed vector space F over the same field $\mathbb{R}$ or $\mathbb{C}$ are considered. Recall that a map T from a vector space E into a vector space F over the same field is said to be **linear** if $T(\alpha x + \beta y) = \alpha T(x) + \beta T(y)$, for all x, y in E and all scalars α, β. Linear maps are more often called **linear transformations** or **linear operators**.

Exercise 1.4.10 Suppose that T is a linear transformation from E into F. Show that T is continuous on E if and only if it is continuous at one point.

Theorem 1.4.1 *Let T be a linear transformation from E into F, then T is continuous if and only if there is $C \geq 0$ such that*

$$\|Tx\| \leq C\|x\|$$

for all $x \in E$.

Proof If there is $C \geq 0$ such that $\|Tx\| \leq C\|x\|$ holds for all $x \in E$, then T is obviously continuous at $x = 0$ and hence by Exercise 1.4.10 is continuous on E.

Conversely, suppose that T is continuous on E, and is hence continuous at $x = 0$. There is then $\delta > 0$ such that if $\|x\| \leq \delta$, then $\|Tx\| \leq 1$. Let now $x \in E$ and $x \neq 0$, then $\left\| \frac{\delta}{\|x\|} x \right\| = \delta$, so $\left\| T\left(\frac{\delta}{\|x\|} x\right) \right\| \leq 1$. Thus $\|Tx\| \leq \frac{1}{\delta}\|x\|$. If we choose $C = \frac{1}{\delta}$, then $\|Tx\| \leq C\|x\|$ for all $x \in E$. ■

From this theorem it follows that if T is a continuous linear transformation from E into F, then

$$\|T\| := \sup_{x \in E, x \neq 0} \frac{\|Tx\|}{\|x\|} < +\infty,$$

and is the smallest C for which $\|Tx\| \leq C\|x\|$ for all $x \in E$. $\|T\|$ is called the norm of T. Of course, $\|T\|$ can be defined for any linear transformation T from E into F; then $\|Tx\| \leq \|T\|\|x\|$ holds always and T is continuous if and only if $\|T\| < +\infty$. Hence a continuous linear transformation is also called a **bounded** linear transformation.

Exercise 1.4.11 Show that $\|T\| = \sup_{x \in E, \|x\|=1} \|Tx\|$.

Exercise 1.4.12 Let $L(E, F)$ be the space of all bounded linear transformations from E into F. Show that it is a normed vector space with norm $\|T\|$ for $T \in L(E, F)$ as previously defined.

Remark Any linear map T from a Euclidean space $\mathbb{R}^n$ into a Euclidean space $\mathbb{R}^m$ is continuous. This follows from the representation of T by a matrix (a_{jk}), $1 \leq j \leq m$, $1 \leq k \leq n$, of real entries, in the sense that if $y = Tx$, then $y_j = \sum_{k=1}^n a_{jk}x_k$, $j = 1, \ldots, m$, where $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_m)$, by observing that

$$|y|^2 = \sum_{j=1}^m \left(\sum_{k=1}^n a_{jk}x_k \right)^2 \leq \left(\sum_{j=1}^m \sum_{k=1}^n a_{jk}^2 \right) |x|^2.$$

Theorem 1.4.2 *If F is a Banach space, then $L(E, F)$ is a Banach space.*

Proof Let $\{T_n\}$ be a Cauchy sequence in $L(E, F)$. Since

$$\|T_nx - T_mx\| = \|(T_n - T_m)x\| \leq \|T_n - T_m\| \cdot \|x\|,$$

$\{T_nx\}$ is a Cauchy sequence in F for each $x \in E$. Since F is complete, $\lim_{n\to\infty} T_nx$ exists. Put $Tx = \lim_{n\to\infty} T_nx$. T is obviously a linear transformation from E into F.

We claim now $T \in L(E, F)$. Since $\{T_n\}$ is Cauchy, $\|T_n\| \leq C$ for some $C > 0$, and for all n. Now, from Lemma 1.4.1,

$$\|Tx\| = \lim_{n\to\infty} \|T_nx\| \leq \left(\sup_n \|T_n\|\right)\|x\| \leq C\|x\|$$

for each $x \in E$. Hence T is a bounded linear transformation.

We show next, $\lim_{n\to\infty} \|T_n - T\| = 0$. Given $\varepsilon > 0$, there is n_0 such that $\|T_n - T_m\| < \varepsilon$ if $n, m \geq n_0$. Let $n \geq n_0$, we have

$$\begin{aligned}
\|T_n - T\| &= \sup_{x\in E, \|x\|=1} \|T_nx - Tx\| \\
&= \sup_{x\in E, \|x\|=1} \lim_{m\to\infty} \|T_nx - T_mx\| \\
&\leq \sup_{x\in E, \|x\|=1} \left(\sup_{m\geq n_0} \|T_n - T_m\|\right)\|x\| \\
&\leq \sup_{x\in E, \|x\|=1} \varepsilon\|x\| = \varepsilon;
\end{aligned}$$

this shows that $\lim_{n\to\infty} \|T_n - T\| = 0$, or $\lim_{n\to\infty} T_n = T$. Thus the sequence $\{T_n\}$ has a limit in $L(E, F)$. Therefore $L(E, F)$ is complete. ∎

$L(E, \mathbb{C})$, or $L(E, \mathbb{R})$, depending on whether E is a complex or a real vector space, is called the topological dual of E and is denoted by E^*; E^* is a Banach space. Elements of E^* are called bounded linear functionals on E.

When $E = F$, $L(E, F)$ is usually abbreviated to $L(E)$. For S, T in $L(E)$, $S \circ T$ is in $L(E)$ and $\|S \circ T\| \leq \|S\| \cdot \|T\|$, as follows directly from definitions. Usually, we shall denote $S \circ T$ by ST; then for S, T, and U in $L(E)$, $(ST)U = S(TU)$, and we may therefore denote TT by T^2, $(TT)T$ by T^3, $\ldots$ etc. for $T \in L(E)$ free of misinterpretation. Note that $\|T^k\| \leq \|T\|^k$ for $T \in L(E)$ and $k \in \mathbb{N}$. For convenience, we put $T^\circ = 1$, the identity map on E.

Exercise 1.4.13 Let S be a nonempty set and consider the vector space $B(S)$ of all bounded real(complex)-valued functions on S. Addition and multiplication by scalar in $B(S)$ are usual for functions. For $f \in B(S)$, let $\|f\| = \sup_{s\in S} |f(s)|$.

(i) Show that $(B(S), \|\cdot\|)$ is a Banach space.

(ii) For $a \in B(S)$, define $A : B(S) \to B(S)$ by $(Af)(s) = a(s)f(s)$, $s \in S$. Show that A is a bounded linear transformation from $B(S)$ into itself and that $\|A\| = \|a\|$.

Exercise 1.4.14 Consider $C[0,1]$ and let $g \in C[0,1]$. Define a linear functional ℓ on $C[0,1]$ by

$$\ell(f) = \int_0^1 f(x)g(x)dx.$$

Show that $\ell \in C[0,1]^*$ and $\|\ell\| = \int_0^1 |g(x)|dx$.

Exercise 1.4.15 Let g be a continuous function on $[0,1] \times [0,1]$ and for $f \in C[0,1]$, let the function Tf be defined by $Tf(x) = \int_0^1 g(x,y)f(y)dy$. Show that $T \in L(C[0,1])$ and $\|T\| = \max_{x\in[0,1]} \int_0^1 |g(x,y)|dy$.

We now consider a series of elements in a n.v.s. E. A symbol of the form $\sum_{k=1}^{\infty} x_k$ with each x_k in E is called a **series**. For each $n \in \mathbb{N}$, $\sum_{k=1}^{n} x_k$ is called the n-th partial **sum** of the series $\sum_{k=1}^{\infty} x_k$. If it happens that $\lim_{n\to\infty} \sum_{k=1}^{n} x_k$ exists in E, say x, then the series $\sum_{k=1}^{\infty} x_k$ is said to be convergent in E and x is called the **sum** of the series, $\sum_{k=1}^{\infty} x_k$, symbolically expressed by $x = \sum_{k=1}^{\infty} x_k$, i.e. when $\sum_{k=1}^{\infty} x_k$ converges, we attach a meaning to the symbol $\sum_{k=1}^{\infty} x_k$ by referring to it as $\lim_{n\to\infty} \sum_{k=1}^{n} x_k$, or the sum of the series.

Theorem 1.4.3 *Let $\{x_k\}$ be a sequence in a Banach space E such that $\sum_{k=1}^{\infty} \|x_k\| < \infty$. Then $\sum_{k=1}^{\infty} x_k$ converges in E.*

Proof For $n \in \mathbb{N}$, let $y_n = \sum_{k=1}^{n} x_k$. Then for $m > n$ in $\mathbb{N}$,

$$\|y_m - y_n\| = \left\| \sum_{k=n+1}^{m} x_k \right\| \leq \sum_{k=n+1}^{m} \|x_k\| \to 0$$

as $n \to \infty$. This means that $\{y_n\}$ is a Cauchy sequence in E, but the fact that E is complete implies that $\{y_n\}$ converges in E, i.e. $\lim_{n\to\infty} \sum_{k=1}^{n} x_k$ exists in E. ■

Exercise 1.4.16 Suppose that $\sum_{k=1}^{\infty} x_k$ is a convergent series in a n.v.s. E. Show that

$$\left\| \sum_{k=1}^{\infty} x_k \right\| \leq \sum_{k=1}^{\infty} \|x_k\|.$$

Exercise 1.4.17 Suppose that $\sum_{k=1}^{\infty} \alpha_k$ is a convergent series in $\mathbb{R}$.

(i) If x is an element of a n.v.s. E, show that $\sum_{k=1}^{\infty} \alpha_k x$ converges in E.

(ii) If $\{x_k\}$ is a bounded sequence in a Banach space E and $\sum_{k=1}^{\infty} \alpha_k$ is absolutely convergent, show that $\sum_{k=1}^{\infty} \alpha_k x_k$ converges in E.

The following example, which complements Theorem 1.4.3, illustrates a method to extract a convergent subsequence from a given sequence.

Example 1.4.4 If a series $\sum_{n=1}^{\infty} x_n$ in a n.v.s. E converges whenever $\sum_{n=1}^{\infty} \|x_n\| < \infty$, then E is a Banach space. To show this, let $\{y_n\}$ be a Cauchy sequence in E. Since $\{y_n\}$ is Cauchy, there is an increasing sequence $n_1 < n_2 < \cdots < n_k < \cdots$ in $\mathbb{N}$ such that $\|y_{n_{k+1}} - y_{n_k}\| < \frac{1}{k^2}$ for each k. Then $\sum_{k=1}^{\infty} \|y_{n_{k+1}} - y_{n_k}\| < \infty$ and hence

$\sum_{k=1}^{\infty}(y_{n_{k+1}} - y_{n_k})$ converges, which is equivalent to $\{y_{n_k}\}$ being a convergent sequence. We have shown that $\{y_n\}$ has a convergent subsequence; thus $\{y_n\}$ converges by Exercise 1.4.3 and E is therefore complete.

Remark We conclude this section with a remark on norms on a vector space E. Suppose that $\|\cdot\|'$ and $\|\cdot\|''$ are different norms on a vector space E, in general, $\|\cdot\|'$ and $\|\cdot\|''$ will generate different families of open sets; but a moment's reflection convinces us that $\|\cdot\|'$ and $\|\cdot\|''$generate the same family of open sets if and only if there is $c > 0$ such that

$$c\|x\|'' \leq \|x\|' \leq \frac{1}{c}\|x\|''$$

for all x in E (in this case $\|\cdot\|'$ and $\|\cdot\|''$ are said to be equivalent). We shall see in Proposition 1.7.2 that all norms on a finite-dimensional vector space are equivalent.

1.5 Semi-continuities

For real-valued functions, the fact that the real field $\mathbb{R}$ is ordered plays an important role in the analysis of functions. In particular, for real-valued functions defined on a metric space, lower semi-continuity and upper semi-continuity are useful concepts that owe their existence to $\mathbb{R}$ being ordered. Semi-continuities are our concern in this section. For a subset S of $\mathbb{R}$ we shall adopt the convention that $\inf S = \infty$ and $\sup S = -\infty$ if S is empty; and that $\inf S = -\infty$ if S is not bounded from below, while $\sup S = \infty$ if S is not bounded from above.

For a sequence x_n, $n = 1, 2, \ldots$, of real numbers, let

$$\liminf_{n\to\infty} x_n = \lim_{n\to\infty}\left(\inf_{k\geq n} x_k\right), \tag{1.4}$$

$$\limsup_{n\to\infty} x_n = \lim_{n\to\infty}\left(\sup_{k\geq n} x_k\right). \tag{1.5}$$

Notice that $\inf_{k\geq n} x_k$ is increasing and $\sup_{k\geq n} x_k$ is decreasing as n increases, hence both limits on the right-hand sides of (1.4) and (1.5) exist, although they may not be finite. Thus $\liminf_{n\to\infty} x_n$ and $\limsup_{n\to\infty} x_n$ always exist, and are called respectively the **inferior limit** and the **superior limit** of $\{x_n\}$. Clearly, $\liminf_{n\to\infty} x_n \leq \limsup_{n\to\infty} x_n$.

Exercise 1.5.1

(i) Show that $\lim_{n\to\infty} x_n$ exists if and only if $\liminf_{n\to\infty} x_n = \limsup_{n\to\infty} x_n$, and $\lim_{n\to\infty} x_n$ is the common value $\liminf_{n\to\infty} x_n = \limsup_{n\to\infty} x_n$ if it exists.

(ii) Show that $\liminf_{n\to\infty}(x_n + y_n) \geq \liminf_{n\to\infty} x_n + \liminf_{n\to\infty} y_n$ ($\limsup_{n\to\infty}(x_n + y_n) \leq \limsup_{n\to\infty} x_n + \limsup_{n\to\infty} y_n$), if $\liminf_{n\to\infty} x_n + \liminf_{n\to\infty} y_n$ ($\limsup_{n\to\infty} x_n + \limsup_{n\to\infty} y_n$) is meaningful. Note that $\alpha + \beta$ is meaningful if at least one of α and β is finite, or if both α and β are either ∞ or $-\infty$.

(iii) Show that $\liminf_{n\to\infty}(x_n + y_n) \le \liminf_{n\to\infty} x_n + \limsup_{n\to\infty} y_n$ if the right-hand side is meaningful and that $\limsup_{n\to\infty}(x_n + y_n) \ge \liminf_{n\to\infty} x_n + \limsup_{n\to\infty} y_n$ if the right-hand side is meaningful.

A real-valued function f defined on a metric space M with metric ρ is said to be **lower semi-continuous (upper semi-continuous)** at $x \in M$ if, for every sequence $\{x_n\}$ in M with $x = \lim_{n\to\infty} x_n$, $f(x) \le \liminf_{n\to\infty} f(x_n)$ $(f(x) \ge \limsup_{n\to\infty} f(x_n))$ holds. Lower semi-continuity and upper semi-continuity will often be abbreviated as l.s.c. and u.s.c. respectively. It is clear that a function f is l.s.c. (u.s.c.) at x if and only if for any given $\varepsilon > 0$ there is $\delta > 0$ such that $f(y) > f(x) - \varepsilon$ $(f(y) < f(x) + \varepsilon)$ if $\rho(y, x) < \delta$.

Exercise 1.5.2

(i) Show that f is lower semi-continuous (upper semi-continuous) at x if and only if

$$f(x) = \lim_{\delta\searrow 0}\left[\inf_{y\in M,\ \rho(x,y)<\delta} f(y)\right] \left(f(x) = \lim_{\delta\searrow 0}\left[\sup_{y\in M,\ \rho(x,y)<\delta} f(y)\right]\right);$$

(ii) show that f is continuous at x if and only if f is both lower semi-continuous and upper semi-continuous at x.

Because of the assertions of Exercise 1.5.2, if x is not an isolated point of M, we define $\liminf_{y\to x} f(y)$ and $\limsup_{y\to x} f(y)$ by

$$\liminf_{y\to x} f(y) = \lim_{\delta\searrow 0}\left[\inf_{y\in M,\ 0<\rho(x,y)<\delta} f(y)\right];$$

$$\limsup_{y\to x} f(y) = \lim_{\delta\searrow 0}\left[\sup_{y\in M,\ 0<\rho(x,y)<\delta} f(y)\right],$$

since $\inf_{y\in M,\ 0<\rho(x,y)<\delta} f(y)$ increases as δ decreases and $\sup_{y\in M,\ 0<\rho(x,y)<\delta} f(y)$ decreases as δ decreases, both $\liminf_{y\to x} f(y)$ and $\limsup_{y\to x} f(y)$ exist, although they may not be finite. If $\liminf_{y\to x} f(y) = \limsup_{y\to x} f(y)$, the common value is called the limit of $f(y)$ as $y \to x$ and is denoted by $\lim_{y\to x} f(y)$. Usually, $\lim_{y\to x} f(y)$ is simply called the limit of the function f at x. Note that $\liminf_{y\to x} f(y)$ and $\limsup_{y\to x} f(y)$ are defined if f is defined on a neighborhood of x with x excluded. If x is an isolated point of M and f is defined at x, then $\liminf_{y\to x} f(y) = \limsup_{y\to x} f(y) = \lim_{y\to x} f(y) = f(x)$ by definition.

Exercise 1.5.3

(i) Show that $\liminf_{y\to x} f(y) \le \limsup_{y\to x} f(y)$ and that f is continuous at x if and only if $\lim_{y\to x} f(y) = f(x)$.

(ii) Show that f is l.s.c. (u.s.c.) at x if and only if $f(x) \le \liminf_{y\to x} f(y)$ $(f(x) \ge \limsup_{y\to x} f(y))$.

If f is lower semi-continuous (upper semi-continuous) at every point of M, then f is said to be lower semi-continuous (upper semi-continuous) on M.

Exercise 1.5.4 Show that f is lower semi-continuous (upper semi-continuous) on M if and only if $\{x \in M : f(x) > \alpha\}$ ($\{x \in M : f(x) < \alpha\}$) is open for every $\alpha \in \mathbb{R}$.

Exercise 1.5.5 Let f_α, $\alpha \in I$, be a family of real-valued continuous functions defined on M and assume that $\sup_{\alpha\in I} f_\alpha(x)$ ($\inf_{\alpha\in I} f_\alpha(x)$) is finite for each $x \in M$; show that $\sup_{\alpha\in I} f_\alpha(x)$ ($\inf_{\alpha\in I} f(x)$) is lower (upper) semi-continuous on M.

Exercise 1.5.6 Suppose that f is a real-valued function defined on a metric space and assume that f is bounded from below on M, i.e. there is $c \in \mathbb{R}$ such that $f(z) \geq c$ for all $z \in M$. For each $k \in \mathbb{N}$ is defined a function f_k on M by

$$f_k(x) = \inf_{z\in M}\{f(z) + k\rho(x,z)\}, \quad x \in M.$$

(i) Show that $f_k(x)$ is finite for all $x \in M$ and

$$|f_k(x) - f_k(y)| \leq k\rho(x,y)$$

for all x, y in M.

(ii) Suppose that f is l.s.c. on M. Show that

$$f(x) = \lim_{k\to\infty} f_k(x), \quad x \in M.$$

(iii) Show that f is l.s.c. on M if and only if there is an increasing sequence $\{f_k\}$ of continuous functions on M such that

$$f(x) = \lim_{k\to\infty} f_k(x)$$

for all $x \in M$.

Exercise 1.5.7 A metric space M is called a compact space if every sequence in M has a subsequence which converges in M. Show that if f is lower semi–continuous (upper semi-continuous) on a compact metric space M, then f assumes its minimum (maximum) on M. (Hint: There is a sequence $\{x_n\}$ in M such that $\lim_{n\to\infty} f(x_n) = \inf_{x\in M} f(x)$)

1.6 The space $\ell^p(\mathbb{Z})$

The Banach spaces considered in this section are included in the more general class of L^p spaces, to be introduced in Section 2.7; but it is expedient to give a separate and direct treatment here without recourse to general theory of measure and integration.

Let $\mathbb{Z}$ be the set of all integers and consider the space L of all real-valued functions defined on $\mathbb{Z}$. With the usual definition of addition of functions and multiplication of a function by a scalar, L is a real vector space. For $f \in L$ and $j \in \mathbb{Z}$, if we denote $f(j)$ by f_j, then f can be identified with the two-way sequence $(f_j)_{j\in\mathbb{Z}}$ of real numbers and L is the space of all sequences $(a_j)_{j\in\mathbb{Z}}$ of real numbers. For $f \in L$ and $1 \le p \le \infty$, let

$$\|f\|_p = \begin{cases} \left(\sum_{j\in\mathbb{Z}} |f(j)|^p\right)^{\frac{1}{p}} & \text{if } p < \infty; \\ \sup_{j\in\mathbb{Z}} |f(j)| & \text{if } p = \infty. \end{cases}$$

Now consider the space $\ell^p(\mathbb{Z})$, $1 \le p \le \infty$, defined by

$$\ell^p(\mathbb{Z}) = \{f \in L : \|f\|_p < \infty\}.$$

Presently we shall prove that $\ell^p(\mathbb{Z})$ is a vector space and $\|\cdot\|_p$ is a norm on $\ell^p(\mathbb{Z})$, but for this purpose we first show an inequality which is a generalization of the Schwarz inequality and is called Hölder's inequality. Two extended real numbers $p, q \ge 1$ are called **conjugate** exponents if $\frac{1}{p} + \frac{1}{q} = 1$ ($\frac{1}{\infty} = 0$; for further arithmetic conventions regarding ∞ and $-\infty$, see the first paragraph of Section 2.2), while two nonnegative numbers α and β will be called a **convex** pair if $\alpha + \beta = 1$.

Lemma 1.6.1 *If α and β is a convex pair, then for any $0 \le \zeta, \eta < \infty$ the following inequality holds:*

$$\zeta^\alpha \eta^\beta \le \alpha\zeta + \beta\eta. \tag{1.6}$$

Proof We may assume that $0 < \alpha, \beta < 1$ and $\zeta, \eta > 0$.

Since $(1+x)^\alpha \le \alpha x + 1$, for $x \ge 0$, we have

$$y^\alpha \le \alpha y + \beta, \ \ y \ge 1. \tag{1.7}$$

Now either $\zeta\eta^{-1} \ge 1$ or $\zeta^{-1}\eta \ge 1$; if $\zeta\eta^{-1} \ge 1$, take $y = \zeta\eta^{-1}$ in (1.7), while if $\zeta^{-1}\eta \ge 1$, take $y = \zeta^{-1}\eta$ in (1.7) with α and β interchanged, then proceed to (1.6). ∎

Lemma 1.6.2 (**Hölder's inequality**) *If $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ are in $\mathbb{R}^n$, then for conjugate exponents p and q we have*

$$\sum_{j=1}^n |x_j y_j| \le \|x\|_p \|y\|_q.$$

Remark Since an element x of $\mathbb{R}^n$ can be identified with an element f of L by $f(1) = x_1, \dots, f(n) = x_n$, and $f(j) = 0$ for other j, $\|x\|_p$ is defined.

Proof of Lemma 1.6.2 It is clear that if one of p and q is ∞, the lemma is trivial, hence we suppose that $1 < p, q < \infty$. Since $\|x\|_p = 0$ if and only if $x = 0$, we may assume that $\|x\|_p > 0$ and $\|y\|_p > 0$. For $1 \le j \le n$, choose $\zeta = \left(\frac{|x_j|}{\|x\|_p}\right)^p$ and $\eta = \left(\frac{|y_j|}{\|y\|_q}\right)^q$ in Lemma 1.6.1. with $\alpha = \frac{1}{p}$ and $\beta = \frac{1}{q}$, then

$$\frac{|x_j y_j|}{\|x\|_p\|y\|_q} \le \frac{1}{p}\frac{|x_j|^p}{\|x\|_p^p} + \frac{1}{q}\frac{|y_j|^q}{\|y\|_q^q},$$

and consequently

$$\sum_{j=1}^{n}|x_j y_j| \le \|x\|_p\|y\|_q\left(\frac{1}{p}+\frac{1}{q}\right) = \|x\|_p\|y\|_q. \qquad \blacksquare$$

Exercise 1.6.1 Suppose that $\alpha > 0$ and $\beta > 0$ is a convex pair. Show that

$$\zeta^\alpha\eta^\beta = \alpha\zeta + \beta\eta,\ \zeta \ge 0,\ \eta \ge 0$$

if and only if $\zeta = \eta$.

We are now in a position to prove that $\ell^p(\mathbb{Z})$ is a vector space and $\|\cdot\|_p$ is a norm on $\ell^p(\mathbb{Z})$. That $\|f\|_p = 0$ if and only if $f = 0$ and that $\lambda f \in \ell^p(\mathbb{Z})$ and $\|\lambda f\|_p = |\lambda|\|f\|_p$ for $\lambda \in \mathbb{R}$ and $f \in \ell^p(\mathbb{Z})$ are obvious. It only remains to show that $\|f+g\|_p \le \|f\|_p + \|g\|_p$ for f, g in $\ell^p(\mathbb{Z})$. For this purpose, we may assume that $1 < p < \infty$ and $\|f+g\|_p > 0$. Under this assumption, there is $A \in F(\mathbb{Z})$ such that $\sum_{j\in A}|f(j)+g(j)|^p > 0$. For such A, we have

$$0 < \sum_{j\in A}|f(j)+g(j)|^p \le \sum_{j\in A}|f(j)+g(j)|^{p-1}\big(|f(j)|+|g(j)|\big),$$

from which, by using Hölder's inequality (see Lemma 1.6.2.), we have

$$\begin{aligned}
0 &< \sum_{j\in A}|f(j)+g(j)|^p \\
&\le \left(\sum_{j\in A}|f(j)+g(j)|^{(p-1)q}\right)^{\frac{1}{q}}\left\{\left(\sum_{j\in A}|f(j)|^p\right)^{\frac{1}{p}} + \left(\sum_{j\in A}|g(j)|^p\right)^{\frac{1}{p}}\right\} \\
&\le \left(\sum_{j\in A}|f(j)+g(j)|^p\right)^{\frac{1}{q}}\big(\|f\|_p + \|q\|_p\big),
\end{aligned}$$

and thus, on dividing the last sequence of inequalities by $\left(\sum_{j\in A}|f(j)+g(j)|^p\right)^{\frac{1}{q}}$, we obtain

$$\left(\sum_{j\in A}|f(j)+g(j)|^p\right)^{\frac{1}{p}} \le \|f\|_p + \|g\|_p. \tag{1.8}$$

Now observe that (1.8) holds for any $A \in F(\mathbb{Z})$. Taking the supremum on the left-hand side of (1.8) over $A \in F(\mathbb{Z})$, we see that $\|f+g\|_p \leq \|f\|_p + \|g\|_p$. Therefore, $\ell^p(\mathbb{Z})$ is a vector space and $\|\cdot\|_p$ is a norm on $\ell^p(\mathbb{Z})$. We shall always refer to $\ell^p(\mathbb{Z})$ as a normed vector space with this norm.

Exercise 1.6.2 Let $k_1 < \cdots < k_n$ be a finite sequence in $\mathbb{Z}$ of length n; define a map T from $\ell^p(\mathbb{Z})$ to the n-dimensional Euclidean $\mathbb{R}^n$ by

$$T(f) = \big(f(k_1), \ldots, f(k_n)\big), \; f \in \ell^p(\mathbb{Z}).$$

Show that T is continuous from $\ell^p(\mathbb{Z})$ onto $\mathbb{R}^n$ and that the image under T of any open set in $\ell^p(\mathbb{Z})$ is an open set in $\mathbb{R}^n$.

Exercise 1.6.3 Suppose $1 \leq p < \infty$; show that $|a_1 + \cdots + a_n|^p \leq n^{p-1} \sum_{j=1}^n |a_j|^p$ for $a_1, \ldots, a_n$ in $\mathbb{R}$.

Exercise 1.6.4 Let $f_1, f_2, \ldots, f_n, \ldots$ be a Cauchy sequence in $\ell^p(\mathbb{Z})$; show that $\lim_{n\to\infty} f_n(j)$ exists and is finite for every $j \in \mathbb{Z}$.

Exercise 1.6.5 Show that $\ell^\infty(\mathbb{Z})$ is a Banach space.

Theorem 1.6.1 *$\ell^p(\mathbb{Z})$ is a Banach space for $1 \leq p \leq \infty$.*

Proof The case $p = \infty$ is relatively easy and is left as an exercise (see Exercise 1.6.5). Consider now the case $1 \leq p < \infty$. Let $f_1, f_2, \ldots, f_n, \ldots$ be a Cauchy sequence in $\ell^p(\mathbb{Z})$, then $\lim_{n\to\infty} f_n(j)$ exists and is finite for each $j \in \mathbb{Z}$ (see Exercise 1.6.4), say $f(j) = \lim_{n\to\infty} f_n(j)$. We show first that $f \in \ell^p(\mathbb{Z})$. Since $f_1, f_2, \ldots, f_n, \ldots$ is a Cauchy sequence, it is necessarily bounded. Let $\|f_n\|_p \leq M$ for all n. There is $n_0 \in \mathbb{N}$ such that

$$\|f_n - f_m\|_p < 1, \;\; n, m \geq n_0.$$

Now fix $m \geq n_0$ and let $A \in F(\mathbb{Z})$, then

$$\begin{aligned}\sum_{j\in A} |f(j)|^p &= \lim_{n\to\infty} \sum_{j\in A} |f_n(j)|^p = \lim_{n\to\infty} \sum_{j\in A} |f_n(j) - f_m(j) + f_m(j)|^p \\ &\leq \limsup_{n\to\infty} \sum_{j\in A} \big\{|f_n(j) - f_m(j)| + |f_m(j)|\big\}^p,\end{aligned}$$

from which, by Exercise 1.6.3, we have

$$\begin{aligned}\sum_{j\in A} |f(j)|^p &\leq \limsup_{n\to\infty} 2^{p-1} \left\{ \sum_{j\in A} |f_n(j) - f_m(j)|^p + \sum_{j\in A} |f_m(j)|^p \right\} \\ &\leq 2^{p-1} \left\{ \limsup_{n\to\infty} \|f_n - f_m\|_p^p + \|f_m\|_p^p \right\} \\ &\leq 2^{p-1}\{1 + M^p\}.\end{aligned}$$

Thus,

$$\sum_{j\in\mathbb{Z}}|f(j)|^p = \sup_{A\in F(\mathbb{Z})}\sum_{j\in A}|f(j)|^p \leq 2^{p-1}(1+M^p) < \infty,$$

which shows $f \in \ell^p(\mathbb{Z})$. We now claim $\lim_{n\to\infty} f_n = f$ in $\ell^p(\mathbb{Z})$. Actually, given $\varepsilon > 0$, there is $N \in \mathbb{N}$ such that

$$\|f_n - f_m\|_p < \varepsilon, \ \ n, m \geq N.$$

Now, for $n \geq N$ and $A \in F(\mathbb{Z})$,

$$\begin{aligned}\sum_{j\in A}|f(j) - f_n(j)|^p &= \lim_{m\to\infty}\sum_{j\in A}|f_m(j) - f_n(j)|^p \\ &\leq \liminf_{m\to\infty}\|f_m - f_n\|_p^p \leq \varepsilon^p,\end{aligned}$$

which implies

$$\|f - f_n\|_p^p = \sup_{A\in F(\mathbb{Z})}\sum_{j\in A}|f(j) - f_n(j)|^p \leq \varepsilon^p,$$

or

$$\|f - f_n\|_p \leq \varepsilon, \ \ n \geq N.$$

In other words, $\lim_{n\to\infty} f_n = f$ in $\ell^p(\mathbb{Z})$. This shows that $\ell^p(\mathbb{Z})$ is complete and hence is a Banach space. ■

Exercise 1.6.6 Let f, g be in $\ell^1(\mathbb{Z})$.

(i) Show that $\{f(n-m)g(m)\}_{(n,m)\in\mathbb{Z}\times\mathbb{Z}}$ is summable and

$$\sum_{(n,m)\in\mathbb{Z}\times\mathbb{Z}} f(n-m)g(m) = \sum_{n\in\mathbb{Z}}\sum_{m\in\mathbb{Z}} f(n-m)g(m).$$

(ii) Define $f * g(n) = \sum_{m\in\mathbb{Z}} f(n-m)g(m)$, $n \in \mathbb{Z}$. Show that $f * g \in \ell^1(\mathbb{Z})$, $f * g = g * f$, and $\|f * g\|_1 \leq \|f\|_1\|g\|_1$.

Exercise 1.6.7 Suppose that $f \in l^p(\mathbb{Z})$ and $g \in l^1(\mathbb{Z})$. Show that $f * g$ can be defined similarly as in Exercise 1.6.6 (ii); then show that $f * g = g * f$, and

$$\|f * g\|_p \leq \|f\|_p\|g\|_1.$$

Remark For any nonempty set S and $1 \leq p \leq \infty$, the Banach space $\ell^p(S)$ can be defined in the same way that $\ell^p(\mathbb{Z})$ is defined. The first such space is the space $\ell^2(\mathbb{N})$ introduced by D. Hilbert in his study of the Fredholm theory of integral equations.

1.7 Compactness

This section is devoted to a study of compactness, introduced in Exercise 1.5.7. Existence of mathematical objects in analysis often involves arguments of compactness: for example, Exercise 1.5.7 guarantees that if f is a lower semi-continuous function defined on a compact metric space M, then there exists $x_0 \in M$ such that

$$f(x_0) = \min_{x \in M} f(x).$$

Recall from Exercise 1.5.7 that a metric space M is called a **compact** space if every sequence in M has a subsequence which converges in M. One observes readily that a compact metric space is necessarily complete. There is a characterization of compact metric spaces which is often useful. To prepare for the statement of such a characterization, we call a point x_0 of a metric space M a **limit point** of a set $A \subset M$ if every neighborhood of x_0 contains a point of A other than x_0.

Exercise 1.7.1 Let A be a subset of a metric space M.

(i) Show that a point x_0 is a limit point of A if and only if every neighborhood of x_0 contains infinitely many points of A;

(ii) show that A is closed if and only if it contains all its limit points. Infer in particular that a finite set is closed.

Theorem 1.7.1 *A metric space M is compact if and only if every infinite subset of M has a limit point.*

Proof Suppose first that M is compact and let A be an infinite subset of M. We shall show that A has a limit point. Since A is infinite, there is a sequence $\{x_n\}$ in A formed of mutually different points. As M is compact, $\{x_n\}$ has a subsequence $\{x_{n_k}\}$ which converges to $x \in M$. Since $\{x_{n_k}\}$ is formed of mutually different points in A and $x = \lim_{k\to\infty} x_{n_k}$, x is a limit point of A. We have shown that if M is compact, then every infinite subset of M has a limit point.

Next, suppose that every infinite subset of M has a limit point. Let us show that M is compact. Suppose that $\{x_n\}$ is a sequence in M. If the range of the sequence $\{x_n\}$ is a finite set, then $x_{n_1} = x_{n_2} = \cdots = x_{n_k} = \cdots$ for some subsequence $\{n_k\}$ of $\{n\}$, and hence the subsequence $\{x_{n_k}\}$ of $\{x_n\}$, being a constant sequence, converges. On the other hand, if the range of $\{x_n\}$ is infinite, then it has a limit point x. It is clear that x is the limit of a subsequence of $\{x_n\}$. Thus M is compact. ■

A subset K of a metric space is said to be compact if K is a compact metric space with metric inherited from M. From the **Bolzano–Weierstrass theorem**, which states that every bounded infinite subset of $\mathbb{R}$ has a limit point, it follows that every bounded closed subset of $\mathbb{R}$ is compact. Historically, the Bolzano–Weierstrass theorem is the genesis of the concept of compact sets.

Exercise 1.7.2 Suppose that $K_1 \supset K_2 \supset \cdots \supset K_n \supset K_{n+1} \supset \cdots$ is a decreasing sequence of nonempty compact sets in a metric space. Show that $\bigcap_n K_n \neq \emptyset$.

Exercise 1.7.3 Show that the Bolzano–Weierstrass theorem holds also for $\mathbb{R}^k$, $k \geq 2$ and then infer that every bounded closed subset of $\mathbb{R}^k$ is compact. Show also that every bounded closed set in the unitary space $\mathbb{C}^k$ is compact.

Exercise 1.7.4

(i) Show that compact subsets of a metric space are both bounded and closed.

(ii) Show that a subset of the Euclidean space $\mathbb{R}^k$ or of the unitary space $\mathbb{C}^n$ is compact if and only if it is both bounded and closed.

(iii) Let, for each $n \in \mathbb{Z}$, e_n be the element of $l^2(\mathbb{Z})$ (see Section 1.6) such that $e_n(j) = \delta_{nj}$, $j \in \mathbb{Z}$. Show that $\{e_n\}_{n\in\mathbb{Z}}$ is a bounded and closed subset of $l^2(\mathbb{Z})$, but it is not compact. Recall that δ_{nj} is the Kronecker delta, defined by $\delta_{nj} = 1$ or 0 according to whether $n = j$ or $n \neq j$.

Proposition 1.7.1 *If T is a continuous map from a metric space M_1 into a metric space M_2, then for every compact set K in M_1, TK is a compact set in M_2, i.e. continuous images of compact sets are compact.*

Proof Let K be a compact set in M_1; we may assume that K is nonempty. Suppose that $\{y_n\}$ is a sequence in TK; we have to show that $\{y_n\}$ has a subsequence which converges to an element in TK. For each $n \in \mathbb{N}$, pick $x_n \in K$ such that $y_n = Tx_n$. Since K is compact, $\{x_n\}$ has a subsequence $\{x_{n_k}\}$ such that $x_{n_k} \to x \in K$. Since T is continuous, $y_{n_k} = Tx_{n_k} \to Tx$. Thus the subsequence $\{y_{n_k}\}$ of $\{y_n\}$ converges to an element in TK. ■

An interesting consequence of Proposition 1.7.1 is the following proposition concerning norms on a finite-dimensional vector space.

Proposition 1.7.2 *If E is a finite-dimensional vector space, then any two norms $\|\cdot\|'$ and $\|\cdot\|''$ on E are equivalent, in the sense that there is $c > 0$ such that $c\|v\|'' \leq \|v\|' \leq \frac{1}{c}\|v\|''$ for all $v \in E$.*

Proof For definiteness we assume that E is a complex vector space. Let $n = \dim E$, and choose a basis $\{v_1, \ldots, v_n\}$ of E. Define a norm $\|\cdot\|$ on E by

$$\|v\| = \left\{ \sum_{j=1}^{n} |\alpha_j|^2 \right\}^{1/2}$$

if $v = \sum_{j=1}^n \alpha_j v_j$, where each $\alpha_j \in \mathbb{C}$. Let Γ be the set $\{v = \sum_{j=1}^n \alpha_j v_j : \sum_{j=1}^n |\alpha_j|^2 = 1\}$ in E. Define a map $T : \mathbb{C}^n \to E$ by

$$T(\zeta) = \sum_{j=1}^{n} \zeta_j v_j, \quad \zeta = (\zeta_1, \ldots, \zeta_n) \in \mathbb{C}^n.$$

From $\|T(\zeta) - T(\eta)\|' \leq \sum_{j=1}^{n} |\zeta_j - \eta_j| \|v_j\|' \leq \sqrt{n} \max_{1 \leq j \leq n} \|v_j\|' |\zeta - \eta|$, where $|\zeta - \eta|$ is the norm of $\zeta - \eta$ in the unitary space $\mathbb{C}^n$, it follows that T is continuous from the unitary space $\mathbb{C}^n$ into $(E, \|\cdot\|')$. Note that T is bijective. Since Γ is the image under T of the compact set $\{\zeta \in \mathbb{C}^n : \sum_{j=1}^{n} |\zeta_j|^2 = 1\}$ in $\mathbb{C}^n$, Γ is compact in $(E, \|\cdot\|')$, by Proposition 1.7.1. Now let $r = \inf_{v\in\Gamma} \|v\|'$ and observe that since Γ is compact in $(E, \|\cdot\|')$ and Γ does not contain the zero element of E, $r = \min_{v\in\Gamma} \|v\|' > 0$; in other words, $\|v\|' \geq r > 0$ for all v with $\|v\| = 1$. Now let $v \in E$, $v \neq 0$, then $\|\frac{v}{\|v\|}\|' \geq r$ or $r\|v\| \leq \|v\|'$. On the other hand, $\|v\|' \leq \sum_{j=1}^{n} |\alpha_j| \|v_j\|' \leq \sqrt{n}(\max_{1\leq j\leq n} \|v_j\|')\{\sum_{j=1}^{n} |\alpha_j|^2\}^{1/2} = \sqrt{n}(\max_{1\leq j\leq n} \|v_j\|')\|v\|$, or, if we let $\sqrt{n}(\max_{1\leq j\leq n} \|v_j\|') = R$, we have

$$\|v\|' \leq R\|v\|$$

for all $v \in E$ (note: we write $v = \sum_{j=1}^{n} \alpha_j v_j$ for $v \in E$). We choose then $c' > 0$ such that $c' \leq r$ and $\frac{1}{c'} \geq R$, then

$$c'\|v\| \leq \|v\|' \leq \frac{1}{c'}\|v\|, \quad v \in E.$$

Similarly, there is $c'' > 0$ such that

$$c''\|v\| \leq \|v\|'' \leq \frac{1}{c''}\|v\|, \quad v \in E.$$

Then, for $v \in E$,

$$c'c''\|v\|'' \leq c'\|v\| \leq \|v\|' \leq \frac{1}{c'}\|v\| \leq \frac{1}{c'c''}\|v\|'',$$

or

$$c\|v\|'' \leq \|v\|' \leq \frac{1}{c}\|v\|'',$$

where $c = c'c'' > 0$. ■

Corollary 1.7.1 *Finite-dimensional vector subspaces of a n.v.s. E are all closed.*

Proof For definiteness, assume that E is a real n.v.s. with norm $\|\cdot\|$. Consider any finite-dimensional vector subspace F of E, put n = dimension of F and choose a basis $\{v_1, \ldots, v_n\}$ of F. Define a new norm $\|\cdot\|'$ on F as follows: for $u = \sum_{j=1}^{n} \alpha_j v_j$ where $\alpha_1, \ldots, \alpha_n$ are real numbers, let $\|u\|' = (\sum_{j=1}^{n} \alpha_j^2)^{1/2}$. Clearly, $\|\cdot\|'$ is a norm on F. Let T be the linear map from the Euclidean space $\mathbb{R}^n$ onto F, defined by $Tx = \sum_{j=1}^{n} x_j v_j$ for $x = (x_1, \ldots, x_n)$. If we denote by $|\cdot|$ the Euclidean norm for $\mathbb{R}^n$, then $\|Tx\|' = |x|$. By Proposition 1.7.2, there is $c > 0$ such that $c\|u\|' \leq \|u\| \leq c^{-1}\|u\|'$

for $u \in F$; consequently, $\|Tx\| \le c^{-1}\|Tx\|' = c^{-1}|x|$ for $x \in \mathbb{R}^n$ and hence T is a continuous map from $\mathbb{R}^n$ into E. To show that F is closed in E, we have to show that if $\{u_k\}$ is a sequence in F which converges in E, then the limit is in F. Since $\{u_k\}$ converges, it is bounded, say $\|u_k\| \le A$ for all k for some $A > 0$. Now write $u_k = \sum_{j=1}^n \alpha_j^{(k)} v_j$ and put $\alpha^{(k)} = (\alpha_1^{(k)}, \ldots, \alpha_n^{(k)})$, then $u_k = T\alpha^{(k)}$ and $|\alpha^{(k)}| = \|u_k\|' \le c^{-1}\|u_k\| \le c^{-1}A$ for each k. Thus $\{u_k\}$ is contained in the image $K \subset F$ of the closed ball $\{x \in \mathbb{R}^n : |x| \le c^{-1}A\}$ under T. Since closed balls in $\mathbb{R}^n$ are compact, K is compact by Proposition 1.7.1 and is therefore closed in E. Now $\{u_k\} \subset K$ implies that its limit is in $K \subset F$. This shows that F is closed. ∎

Corollary 1.7.2 *Suppose that F is an affine subspace of $\mathbb{R}^n$, then for each $x \in \mathbb{R}^n$, there is unique y in F such that $|x - y| = \min_{z \in F} |x - z|$. Furthermore, y is characterized by the condition that $(x - y) \cdot (z - y) = 0$ for all $z \in F$.*

Proof We need only consider the case that F is a proper affine subspace of $\mathbb{R}^n$ and x is not in F. Since F is closed by Corollary 1.7.1, $\inf_{z \in F} |x - z| = l > 0$. Let $K = \{z \in F : |x - z| \le 2l\}$, then $\inf_{z \in F} |x - z| = \inf_{z \in K} |x - z|$; but, since K is compact, there is $y \in K$ such that $l = \min_{z \in F} |x - z| = \min_{z \in K} |x - z| = |x - y|$. Consider now $z \in F$ and let $f(t) = |x - y + t(z - y)|^2 = |x - y|^2 + 2t(x - y) \cdot (z - y) + t^2|z - y|^2$ for $t \in \mathbb{R}$. Since f assumes minimum l^2 at $t = 0$, $f'(0) = 2(x - y) \cdot (z - y) = 0$. Hence y satisfies the condition that $(x - y) \cdot (z - y) = 0$ for all $z \in F$; on the other hand, if $y \in F$ satisfies the condition that $(x - y) \cdot (z - y) = 0$ for all $z \in F$, then for any $z \in F$ we have $|x - z|^2 = |x - y + y - z|^2 = |x - y|^2 + 2(x - y) \cdot (y - z) + |y - z|^2 = |x - y|^2 + |y - z|^2 \ge |x - y|^2$, i.e. $|x - y| = \min_{z \in F} |x - z|$. Thus, we have shown that there is $y \in F$ such that $|x - y| = \min_{z \in F} |x - z|$ and that y is characterized by the condition that $(x - y) \cdot (z - y) = 0$ for all $z \in F$. It remains to show that y is unique. Let y and y' in F satisfy $|x - y| = |x - y'| = \min_{z \in F} |x - z|$, then

$$(x - y) \cdot (z - y) = 0, \quad (x - y') \cdot (z - y') = 0$$

for all z in F. Choose $z = y'$ and y respectively in these equalities; we have

$$(x - y) \cdot (y - y') = 0, \quad (x - y') \cdot (y - y') = 0;$$

subtract the first equality from the second; we have $(y - y') \cdot (y - y') = 0 = |y - y'|^2$, implying $y = y'$. ∎

The map $x \mapsto y$, as asserted by Corollary 1.7.2, is called the **orthogonal** projection from $\mathbb{R}^n$ onto F. If this map is denoted by P, then (1) $Px = x$ if and only if $x \in F$; (2) $P^2 = P$; and (3) $|Px - Px'| \le |x - x'|$. That (1) and (2) hold is fairly obvious. To see that (3) holds, observe firstly that

$$(x - x' - Px + Px') \cdot (Px - Px') = 0,$$

from which it follows that $|Px - Px'|^2 = (x - x') \cdot (Px - Px') \le |x - x'||Px - Px'|$ and hence (3) holds. It follows from (3) that P is a continuous map.

Remark If F is a vector subspace of $\mathbb{R}^n$, then

(i) P is actually a linear map, as follows easily from the characterization that $(x - Px) \cdot z = 0$ for all $z \in F$;

(ii) since $(x - Px) \cdot Px = 0$, $|x|^2 = |Px|^2 + |x - Px|^2$ for every $x \in \mathbb{R}^n$; this last equality is called the **Pythagoras relation**.

Proposition 1.7.3 *Suppose that T is an injective and continuous map from a compact metric space M_1 into a metric space M_2. Then $T^{-1} : TM_1 \to M_1$ is continuous.*

Proof Let $y \in TM_1$ and $\{y_n\}$ be a sequence in TM_1 with $y = \lim_{n\to\infty} y_n$. To show that T^{-1} is continuous at y, we have to show that $\{y_n\}$ has a subsequence $\{y_{n_k}\}$ such that $\lim_{k\to\infty} T^{-1}y_{n_k} = T^{-1}y$ (cf. Exercise 1.4.7 (i)). Let $x_n = T^{-1}y_n$. Since M_1 is compact, $\{x_n\}$ has a subsequence $\{x_{n_k}\}$ which converges to x in M_1. Now $y_{n_k} = Tx_{n_k} \to Tx$ entails that $Tx = y$ and hence $\lim_{k\to\infty} T^{-1}y_{n_k} = \lim_{k\to\infty} x_{n_k} = x = T^{-1}y$. ■

We shall presently give a useful characterization of compact sets in a complete metric space corresponding to the characterization of compact sets in $\mathbb{R}^k$ as bounded and closed sets (see Exercise 1.7.4 (ii)).

A finite family of open balls with radius $\varepsilon > 0$ in a metric space M is called an **ε-net** for a subset A of M if its union contains A. A set A in a metric space is said to be **totally bounded** if for any $\varepsilon > 0$ there is an ε-net for A.

Exercise 1.7.5

(i) Show that a set in $\mathbb{R}^n$ is totally bounded if and only if it is bounded.

(ii) Show that a set A in a metric space is totally bounded if and only if for any $\varepsilon > 0$ there is an ε-net for A whose balls have their centers in A.

Lemma 1.7.1 *A subset A of a metric space M is totally bounded if and only if every sequence in A has a Cauchy subsequence. In particular, compact sets are totally bounded.*

Proof Suppose that A is totally bounded and let $\{x_n\}$ be a sequence in A. There is a $\frac{1}{2}$-net for A and hence one of its balls contains a subsequence $\{x_n^{(1)}\}$ of $\{x_n\}$. After the sequence $\{x_n^{(1)}\}$ is chosen, we then choose a $\frac{1}{4}$-net for A. As before one of the balls of this $\frac{1}{4}$-net contains a subsequence $\{x_n^{(2)}\}$ of $\{x_n^{(1)}\}$. We proceed in this way to obtain a sequence of subsequences, $\{x_n^{(1)}\}, \{x_n^{(2)}\}, \ldots, \{x_n^{(k)}\}, \ldots$ of $\{x_n\}$, each of which is a subsequence of the preceding one, and for each k the sequence $\{x_n^{(k)}\}$ is contained in a ball of radius 2^{-k}. Now, $\{x_n^{(n)}\}$ is a subsequence of $\{x_n\}$. For each positive integer n_0, if $n > m \geq n_0$, both $x_n^{(n)}$ and $x_m^{(m)}$ are in a ball of radius 2^{-n_0}, hence $\rho(x_n^{(n)}, x_m^{(m)}) \leq 2^{-n_0+1}$, from which it follows that $\{x_n^{(n)}\}$ is a Cauchy sequence. Thus each sequence in A has a Cauchy subsequence.

Next, suppose that each sequence in A has a Cauchy subsequence. We are going to show that A is totally bounded. Suppose to the contrary that for some $\varepsilon_0 > 0$, no ε_0-net for A exists. Choose $x_1 \in A$, since $B_{\varepsilon_0}(x_1)$ does not cover A there is $x_2 \in A \backslash B_{\varepsilon_0}(x_1)$. Suppose that $x_1, \ldots, x_n$ in A have been chosen so that $\rho(x_i, x_j) \geq \varepsilon_0$ for

$i, j \leq n$ and $i \neq j$, then choose $x_{n+1} \in A \setminus \bigcup_{i=1}^{n} B_{\varepsilon_0}(x_i)$. Such an x_{n+1} exists because $\{B_{\varepsilon_0}(x_0), \ldots, B_{\varepsilon_0}(x_n)\}$ is not an ε_0-net for A. But then $\rho(x_i, x_j) \geq \varepsilon_0$ for $i, j \leq n+1$ and $i \neq j$. By mathematical induction we have thus exhibited a sequence $\{x_n\}$ in A such that $\rho(x_i, x_j) \geq \varepsilon_0$ when $i \neq j$. Such a sequence can not have a Cauchy subsequence, this contradicts our assumption about A. Thus A is totally bounded. ■

Theorem 1.7.2 *A subset K of a complete metric space M is compact if and only if K is closed and totally bounded.*

Proof Suppose that K is compact, then K is closed. Since each sequence in K has a convergent subsequence which is therefore Cauchy, Lemma 1.7.1 implies that K is totally bounded. Next, suppose K is closed and totally bounded and let $\{x_n\}$ be a sequence in K, then $\{x_n\}$ has a Cauchy subsequence $\{x_n'\}$ by Lemma 1.7.1. But since K is a closed subset of a complete metric space, it is complete and hence $\{x_n'\}$ converges in K. This shows that K is compact. ■

Let A be a subset of a metric space; the smallest closed set which contains A is called the **closure** of A and is denoted by $\overline{A}$. Obviously, $\overline{A}$ is the intersection of all those closed sets containing A. If $\overline{A} = M$, we say that A is dense in M, or that A is a dense subset of M. A metric space M is said to be **separable** if it contains a countable dense subset. A subset of a metric space is separable, if it is separable as a metric space; it is **precompact**, if its closure is compact.

Since the closure of a totally bounded set is totally bounded, Corollary 1.7.3 follows from Theorem 1.7.2 (see Exercise 1.7.6 and Exercise 1.7.7):

Corollary 1.7.3 *A set in a complete metric space is precompact if and only if it is totally bounded.*

Exercise 1.7.6 Show that the closure of a totally bounded set is totally bounded.

Exercise 1.7.7 Show that a set in a complete metric space is precompact if and only if it is totally bounded.

Exercise 1.7.8 Show that a totally bounded subset of a metric space is separable. In particular, a compact subset of a metric space is separable.

Example 1.7.1 (Sequence space) This example illustrates a method to construct a compact space from a sequence (M_k, ρ_k), $k = 1, 2, \ldots,$ of compact metric spaces with $diam M_k \leq C$ for all k. For such a sequence, put $M = \prod_{k=1}^{\infty} M_k = \{x = (x_1, \ldots, x_k, \ldots) : x_k \in M_k, k = 1, 2, \ldots\}$. We shall often denote $x = (x_1, \ldots, x_k, \ldots)$ by (x_k). For $x = (x_k), y = (y_k)$ in M, let

$$\rho(x, y) = \sum_{k=1}^{\infty} \frac{1}{k^2} \rho_k(x_k, y_k). \tag{1.9}$$

It is clear that ρ is a metric on M, and with this metric diam $M \leq 2C$. If $\{x^{(n)}\}_{n \in \mathbb{N}}$ is a sequence in M, and $x \in M$, then $\rho_k(x_k^{(n)}, x_k) \leq k^2 \rho(x^{(n)}, x)$ for each k, from which it follows that if $\lim_{n\to\infty} x^{(n)} = x$ in M, then $\lim_{n\to\infty} x_k^{(n)} = x_k$ in M_k for each k. Conversely, if $\lim_{n\to\infty} x_k^{(n)} = x_k$ for each k, we claim that $\lim_{n\to\infty} x^{(n)} = x$ in M. Let $\varepsilon > 0$ be given. There is $k_0 \in \mathbb{N}$ such that $\sum_{k=k_0+1}^{\infty} \frac{1}{k^2} \rho_k(x_k^{(n)}, x_k) \leq C \sum_{k=k_0+1}^{\infty} \frac{1}{k^2} < \frac{\varepsilon}{2}$. Now, since $\lim_{n\to\infty} \rho_k(x_k^{(n)}, x_k) = 0$ for $k = 1, \ldots, k_0$, there is $L \in \mathbb{N}$ such that $\rho_k(x_k^{(n)}, x_k) < \frac{\varepsilon}{4}$ for $k = 1, \ldots, k_0$, whenever $n \geq L$. Consequently, when $n \geq L$, we have

$$\rho(x^{(n)}, x) = \sum_{k=1}^{k_0} \frac{1}{k^2} \rho_k(x_k^{(n)}, x_k) + \sum_{k=k_0+1}^{\infty} \frac{1}{k^2} \rho_k(x_k^{(n)}, x_k) < \frac{\varepsilon}{4} \sum_{k=1}^{k_0} \frac{1}{k^2} + \frac{\varepsilon}{2} < \varepsilon;$$

this means $\lim_{n\to\infty} x^{(n)} = x$. Thus, we have shown that $\lim_{n\to\infty} x^{(n)} = x$ in M if and only if $\lim_{n\to\infty} x_k^{(n)} = x_k$ in M_k for each k. We show now that M is compact. Suppose that $\{x^{(n)}\}$ is a sequence in M; we have to show that $\{x^{(n)}\}$ has a subsequence which converges in M. We achieve this by the well-known **diagonalization procedure**. Since M_1 is compact $\{x_1^{(n)}\}$ has a subsequence $\{x_1^{(n_j^{(1)})}\}$ which converges in M_1 to, say, x_1; then $\{x_2^{(n_j^{(1)})}\}$ has a subsequence $\{x_2^{(n_j^{(2)})}\}$ which converges in M_2 to x_2; continuing in this fashion, we obtain an array of subsequences of $\{x^{(n)}\}$:

$$\begin{array}{l} x^{(n_1^{(1)})},\ x^{(n_2^{(1)})},\ \ldots,\ x^{(n_j^{(1)})},\ \ldots \\ x^{(n_1^{(2)})},\ x^{(n_2^{(2)})},\ \ldots,\ x^{(n_j^{(2)})},\ \ldots \\ \vdots \\ x^{(n_1^{(j)})},\ x^{(n_2^{(j)})},\ \ldots,\ x^{(n_j^{(j)})},\ \ldots \\ \vdots \end{array} \tag{1.10}$$

where each low contains the next one as a subsequence, and for each $k \in \mathbb{N}$,

$$\lim_{j\to\infty} x_k^{(n_j^{(k)})} = x_k \quad (\text{in } M_k). \tag{1.11}$$

Now, put $n_j = n_j^{(j)}$, $j = 1, 2, \ldots$. $\{x^{(n_j)}\}$ is a subsequence of $\{x^{(n)}\}$ formed of the diagonal elements of the array (1.10). Observe that $\{x^{(n_j)}\}_{j\geq k}$ is a subsequence of $\{x^{(n_j^{(k)})}\}$ for each k, therefore $\lim_{j\to\infty} x_k^{(n_j)} = x_k$ by (1.11) for each k, and consequently $\{x^{(n_j)}\}$ converges in M to (x_k), as we have shown previously in this example. We have shown that $\{x^{(n)}\}$ has a converging subsequence in M. Thus M is compact. In particular, if each M_k is a finite set with discrete metric (see Exercise 1.4.6), then M is compact with metric given by (1.9). We have encountered such a space $\Omega = \{0, 1\} \times \{0, 1\} \times \cdots$ in Section 1.3, of which one observes readily that each set in the algebra $\mathcal{Q}$ is a closed subset of Ω and is hence compact.

Remark In Example 1.7.1, the assumption that $\operatorname{diam} M_k \leq C$ for all k is not necessary, because, if we replace each ρ_k by $\rho_k' = (\operatorname{diam} M_k)^{-1} \rho_k$, then each (M_k, ρ_k') is compact and diam $M_k \leq 1$ w.r.t. the new metric ρ'_k. Hence from any sequence (M_k, ρ_k) of compact metric spaces, one can construct a compact sequence space as in Example 1.7.1.

Now we give a characterization of compact sets which is usually taken as the definition for compact sets in topological spaces.

A family $\{S_\alpha\}$ of subsets of a given set S is called a **covering** of a subset A of S if $A \subset \bigcup_\alpha S_\alpha$; then we also say that $\{S_\alpha\}$ covers A. If S is a metric space and each set S_α is open, $\{S_\alpha\}$ is called an **open covering** of A if it covers A. A subset A of a metric space is said to have the **finite covering property** if every open covering of A has a finite subfamily which covers A.

Lemma 1.7.2 *Let K be a compact subset of a metric space and suppose that $\{G_\alpha\}_{\alpha \in I}$ is an open covering of K, then there is $\delta > 0$, called a* **Lebesgue** *number of K relative to $\{G_\alpha\}$, such that any subset A of K with* diam $A \leq \delta$ *is contained in G_α for some $\alpha \in I$.*

Proof Suppose the contrary. Then for each $n \in \mathbb{N}$ there is a subset A_n of K with diam $A_n \leq \frac{1}{n}$ such that A_n is contained in no G_α. Then choose $x_n \in A_n$. Since K is compact, the sequence $\{x_n\}$ has a subsequence $\{x_{n_k}\}$ which converges to $x \in K$. Let $x \in G_{\alpha_0}$, $\alpha_0 \in I$, and choose $r > 0$ so that $B_r(x) \subset G_{\alpha_0}$. If k is sufficiently large, $\frac{1}{n_k} < \frac{r}{2}$ and $x_{n_k} \in B_{\frac{r}{2}}(x)$; consequently $A_{n_k} \subset B_r(x) \subset G_{\alpha_0}$. This contradicts the fact that A_{n_k} is contained in no G_α. The contradiction proves the lemma. ■

Theorem 1.7.3 *A subset K of a metric space M is compact if and only if K has the finite covering property.*

Proof Suppose first that K has the finite covering property. Consider a sequence $\{x_n\}$ in K; we shall show that $\{x_n\}$ has a subsequence which converges to a point in K. Suppose the contrary, then for each $x \in K$, there is an open ball B_x centered at x such that $x_n \in B_x$ for only finitely many n. $\{B_x\}_{x \in K}$ is an open covering of K, hence has a finite subfamily $\{B_1, \ldots, B_l\}$ which also covers K. Since $\bigcup_{j=1}^{l} B_j \supset K$ and $x_n \in B_j$ for only finitely many n for each j, $x_n \in K$ for only finitely many n, contradicting the fact that $\{x_n\}$ is a sequence in K. Thus $\{x_n\}$ has a subsequence which converges in K, showing that K is compact.

Next, suppose that K is compact. Let $\{G_\alpha\}$ be an open covering of K; we are going to show that $\{G_\alpha\}$ has a finite subfamily which also covers K. Choose a Lebesgue number $\delta > 0$ of K relative to $\{G_\alpha\}$ according to Lemma 1.7.2. Since K is totally bounded by Lemma 1.7.1, there is an $\frac{\delta}{2}$-net $\{B_1, \ldots, B_k\}$ containing K. For $j = 1, \ldots, k$, diam $K \cap B_j \leq \delta$ implies $K \cap B_j \subset G_{\alpha_j}$ for some α_j, and consequently $K \subset \bigcup_{j=1}^{k} G_{\alpha_j}$ i.e. $\{G_{\alpha_1}, \ldots, G_{\alpha_k}\}$ is a finite subfamily of $\{G_\alpha\}$ and it covers K. This shows that K has the finite covering property. ■

Corollary 1.7.4 (Finite intersection property) *Let $\{K_\alpha\}_{\alpha \in I}$ be a family of compact sets in a metric space M with the property that intersection of any finite subfamily of $\{K_\alpha\}$ is nonempty. Then $\bigcap_{\alpha \in I} K_\alpha \neq \emptyset$.*

Proof Suppose the contrary, that $\bigcap_{\alpha \in I} K_\alpha = \emptyset$. Choose and fix $\alpha_0 \in I$. Then for $x \in K_{\alpha_0}$, there is $\alpha_x \in I$ such that $x \in K^c_{\alpha_x}$; hence $\{K^c_\alpha\}_{\alpha \in I}$ is an open covering of K_{α_0}. There is therefore a finite set $\{\alpha_1, \ldots, \alpha_k\} \subset I$ such that $\bigcup_{j=1}^k K^c_{\alpha_j} \supset K_{\alpha_0}$, by Theorem 1.7.3; this last inclusion relation means that $K_{\alpha_0} \cap K_{\alpha_1} \cap \cdots \cap K_{\alpha_k}$ is empty, contradicting our assumption about the family $\{K_\alpha\}$. The contradiction proves the corollary. ■

Two applications of Theorem 1.7.3 will now be given; both concerned with the uniformity concept. Suppose that T is a map from a metric space M_1 with metric ρ_1 into a metric space M_2 with metric ρ_2. T is said to be **uniformly** continuous on M_1 if for any given $\varepsilon > 0$, there is $\delta > 0$ such that $\rho_2(Tx, Ty) < \varepsilon$ whenever x and y are in M_1 with $\rho_1(x, y) < \delta$. Obviously, if T is uniformly continuous on M_1, it is, *a fortiori*, continuous on M_1. A sequence $\{T_n\}$ of maps from M_1 into M_2 is said to converge pointwise to a map T from M_1 into M_2 if $Tx = \lim_{n \to \infty} T_n x$ for each $x \in M_1$; it is said to converge **uniformly** to T on M_1 if for any given $\varepsilon > 0$, there is $n_0 \in \mathbb{N}$ such that $\rho_2(T_n x, Tx) \leq \varepsilon$ for every $x \in M_1$ whenever $n \geq n_0$.

Theorem 1.7.4 *If T is a continuous map from a compact metric space M_1 into a metric space M_2, then T is uniformly continuous on M_1.*

Proof Let $\varepsilon > 0$ be given, and let $x \in M_1$. Since T is continuous at x, there is $\delta_x > 0$ such that $\rho_2(Ty, Tx) < \varepsilon/2$ if $\rho_1(y, x) < \delta_x$. Consider $\{B_{\frac{1}{2}\delta_x}(x)\}_{x \in M_1}$; it is an open covering of M_1; by Theorem 1.7.3, it contains a finite subfamily, say $\{B_{\frac{1}{2}\delta_{x_1}}(x_1), \ldots, B_{\frac{1}{2}\delta_{x_l}}(x_l)\}$, which also covers M_1. Choose $\delta = \frac{1}{2}\min\{\delta_{x_1}, \ldots, \delta_{x_l}\}$. Suppose now that $x, y \in M_1$ with $\rho_1(x, y) < \delta$, and let $x \in B_{\frac{1}{2}\delta_{x_j}}(x_j)$, $1 \leq j \leq l$. Then $\rho_1(y, x_j) \leq \rho_1(x, y) + \rho_1(x, x_j) < \delta + \frac{1}{2}\delta_{x_j} \leq \delta_{x_j}$, hence $\rho_2(Ty, Tx_j) < \frac{\varepsilon}{2}$; since $x \in B_{\frac{1}{2}\delta_{x_j}}(x_j)$, $\rho_2(Tx, Tx_j) < \frac{\varepsilon}{2}$. Therefore, $\rho_2(Tx, Ty) \leq \rho_2(Tx, Tx_j) + \rho_2(Tx_j, Ty) < \varepsilon$. This shows that T is uniformly continuous. ■

Theorem 1.7.5 (Dini) *Let $\{f_n\}$ be a sequence of real-valued continuous functions defined on a compact metric space M such that $f_1(x) \leq f_2(x) \leq \cdots \leq f_n(x) \leq \cdots$ and converges to a finite real number $f(x)$ for each $x \in M$. If, further, f is continuous on M, then the sequence $\{f_n\}$ converges uniformly to f on M.*

Proof Given $\varepsilon > 0$ and $x \in M$, there is $k_x \in \mathbb{N}$ such that $0 \leq f(x) - f_{k_x}(x) < \frac{\varepsilon}{3}$. Because both f and f_{k_x} are continuous, there is an open ball $B(x)$ centered at x such that $|f(y) - f(x)| < \frac{\varepsilon}{3}$ and $|f_{k_x}(y) - f_{k_x}(x)| < \frac{\varepsilon}{3}$ whenever $y \in B(x)$; as a consequence, we have

$$0 \leq f(y) - f_{k_x}(y) \leq |f(y) - f(x)| + |f(x) - f_{k_x}(x)| + |f_{k_x}(x) - f_{k_x}(y)| < \varepsilon$$

whenever $y \in B(x)$, or

$$0 \leq f(y) - f_k(y) < \varepsilon \tag{1.12}$$

whenever $y \in B(x)$ and $k \geq k_x$. Now $\{B(x) : x \in M\}$ is an open covering of M; by Theorem 1.7.3 it has a finite subfamily, say $\{B(x_1), \ldots, B(x_l)\}$, which also covers M. Let $k_0 = \max\{k_{x_1}, \ldots, k_{x_l}\}$; then for $y \in M$ and $k \geq k_0$, it follows from (1.12) that

$$0 \leq f(y) - f_k(y) < \varepsilon,$$

because $y \in B(x_j)$ for some $1 \leq j \leq l$ and $k \geq k_0 \geq k_{x_j}$. Thus the sequence $\{f_n\}$ converges to f uniformly on M. ■

We come now, in the final part of this section, to prove a historically important theorem characterizing precompact sets in the n.v.s. $C(X)$ of all continuous real(complex)-valued functions defined on a compact metric space X with norm given by

$$\|f\| = \sup_{x \in X} |f(x)| = \max_{x \in X} |f(x)|$$

for $f \in C(X)$, where $\sup_{x \in X} |f(x)| = \max_{x \in X} |f(x)|$ is a consequence of Exercise 1.5.7. Clearly, $C(X)$ is a n.v.s. with norm given as such. For a compact metric space X, the norm given previously on $C(X)$ is implicitly assumed without further notice. Actually $C(X)$ is a Banach space; to show this we need a lemma.

Lemma 1.7.3 *Let $\{f_n\}$ be a sequence of continuous functions defined on a metric space M. Suppose that $\{f_n\}$ converges uniformly to a function f on M, then f is continuous on M.*

Proof Let $x \in M$. We shall show that f is continuous at x. Given $\varepsilon > 0$, by the uniform convergence of $\{f_n\}$ to f on M there is $n_0 \in \mathbb{N}$ such that $|f_{n_0}(y) - f(y)| < \frac{\varepsilon}{3}$ for all y in M. Since f_{n_0} is continuous at x, there is $\delta > 0$ such that $|f_{n_0}(y) - f_{n_0}(x)| < \frac{\varepsilon}{3}$ whenever $\rho(x, y) < \delta$. Hence if $\rho(x, y) < \delta$, then

$$\begin{aligned} |f(y) - f(x)| &\leq |f_{n_0}(y) - f(y)| + |f_{n_0}(y) - f_{n_0}(x)| + |f_{n_0}(x) - f(x)| \\ &< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon, \end{aligned}$$

which shows that f is continuous at x. ■

Proposition 1.7.4 *$C(X)$ is a Banach space.*

Proof Let $\{f_n\}$ be a Cauchy sequence in $C(X)$; we have to show that $\{f_n\}$ converges in $C(X)$. Since $|f_n(x) - f_m(x)| \leq \|f_n - f_m\|$ for $x \in X$, $\{f_n(x)\}$ is a Cauchy sequence of scalars and hence converges to a scalar $f(x)$ for every x in X; thus as a sequence of functions, $\{f_n\}$ converges pointwise to a function f on X. Actually $\{f_n\}$ converges uniformly to f on X. Given $\varepsilon > 0$, there is $n_0 \in \mathbb{N}$ such that $\|f_n - f_m\| < \varepsilon$ whenever $n, m \geq n_0$, hence $|f_n(x) - f_m(x)| < \varepsilon$ for all x in X and $n, m \geq n_0$, and thus $|f_n(x) - f(x)| \leq \varepsilon$ for all x in X if $n \geq n_0$, by letting $m \to \infty$. It follows then from Lemma 1.7.3 that $f \in C(X)$. We claim finally that $\lim_{n \to \infty} \|f_n - f\| = 0$, i.e. $\{f_n\}$ converges to f in $C(X)$. To see this, for $\varepsilon > 0$ given choose $n_0 \in \mathbb{N}$ as above, then $|f_n(x) - f(x)| \leq \varepsilon$

for all $x \in X$ and $n \geq n_0$; this means that $\sup_{x\in X} |f_n(x) - f(x)| \leq \varepsilon$ when $n \geq n_0$, or $\|f_n - f\| \leq \varepsilon$ when $n \geq n_0$. Thus $\lim_{n\to\infty} \|f_n - f\| = 0$. ■

A family $\mathcal{F}$ of functions defined on a metric space M is called an **equicontinuous** family if for each given $\varepsilon > 0$ there is $\delta > 0$ such that whenever $\rho(x,y) < \delta$, then $|f(x) - f(y)| < \varepsilon$ for all $f \in \mathcal{F}$. Note that functions in an equicontinuous family are necessarily uniformly continuous.

The theorem that follows is not only historically important, but is also useful in the theory of differential equations.

Theorem 1.7.6 (Arzelà–Ascoli) *If X is a compact metric space, a subset K of $C(X)$ is precompact if and only if it is bounded in $C(X)$ and equicontinuous as a family of functions on X.*

Proof Suppose that K is precompact. Since $C(X)$ is complete, as asserted by Proposition 1.7.4, K is totally bounded by Corollary 1.7.3. Let $\varepsilon > 0$ and let $f_1, \ldots, f_n$ be the centers of an $\frac{\varepsilon}{3}$-net for K. Since $f_1, \ldots, f_n$ are uniformly continuous on X, by Theorem 1.7.4, there is $\delta > 0$ such that

$$|f_i(x) - f_i(y)| < \frac{\varepsilon}{3}$$

for $i = 1, \ldots, n$ when $\rho(x,y) < \delta$. Consider now $f \in K$ and choose $j \in \{1, \ldots, n\}$ so that

$$\sup_{x\in X} |f(x) - f_j(x)| < \frac{\varepsilon}{3};$$

such j exists because $f_1, \ldots, f_n$ are centers of an $\frac{\varepsilon}{3}$-net for K. Then if $\rho(x,y) < \delta$, we have

$$\begin{aligned} |f(x) - f(y)| &\leq |f(x) - f_j(x)| + |f_j(x) - f_j(y)| + |f_j(y) - f(y)| \\ &< \frac{2}{3}\varepsilon + |f_j(x) - f_j(y)| < \varepsilon, \end{aligned}$$

and therefor K is equicontinuous. Since K is totally bounded, it is bounded in $C(X)$.

Conversely, suppose that K is bounded in $C(X)$ and is equicontinuous as a family of functions on X. Let $\varepsilon > 0$. Choose $\delta > 0$ such that $|f(x) - f(y)| < \frac{\varepsilon}{4}$ for $f \in K$ when $f(x,y) < \delta$. As X is compact, there is a δ-net for X with centers $x_1, \ldots, x_n$. For simplicity's sake, in the argument that follows we assume that functions in $C(X)$ are real-valued; the corresponding argument when $C(X)$ consists of complex-valued functions will be clear. Since K is bounded in $C(X)$, there is $L > 0$ so that $|f(x)| \leq L$ for all $f \in K$ and all $x \in X$. Divide the interval $[-L, L]$ into k equal parts by the partition

$$y_0 = -L < y_1 < \cdots < y_k = L,$$

where k is chosen so that $|y_i - y_{i+1}| < \frac{\varepsilon}{4}$ for $i = 0, \ldots, k-1$. We say that an n-tuple $(y_{i_1}, \ldots, y_{i_n})$ of numbers $y_0, \ldots, y_k$ is admissible if for some $f \in K$ the following inequalities hold:

$$|f(x_j) - y_{i_j}| < \frac{\varepsilon}{4}, \quad j = 1, \ldots, n. \tag{1.13}$$

Clearly, for each $f \in K$ there is an n-tuple $(y_{i_1}, \ldots, y_{i_n})$ so that (1.13) holds. Hence the set Y of all admissible n-tuples is nonempty. Note that Y is finite. For each n-tuple $y = (y_{i_1}, \ldots, y_{i_n})$ in Y choose and fix an $f_y \in K$ so that (1.13) holds, with f replaced by f_y. Let now $f \in K$. Choose $y = (y_{i_1}, \ldots, y_{i_n})$ in Y such that (1.13) holds. For $x \in X$ choose x_j, $1 \le j \le n$, so that $\rho(x, x_j) < \delta$. Then

$$|f(x) - f_y(x)| \le |f(x) - f(x_j)| + |f(x_j) - y_{i_j}| + |y_{i_j} - f_y(x_j)| + |f_y(x_j) - f_y(x)|,$$

from which we infer that $\|f - f_y\| < \varepsilon$ from the fact that both f and f_y satisfy (1.13) as well as from the way $\delta > 0$ is chosen. Thus $\{B_\varepsilon(f_y) : y \in Y\}$ is an ε-net for K. We have shown that K is totally bounded. Hence K is precompact by Corollary 1.7.3. ∎

Example 1.7.2 Let $K = \{f \in C^1[0,1] : f(0) = a \text{ and } |f'| \le g\}$, where $a \in \mathbb{R}$ and g is a nonnegative continuous function on $[0,1]$. It is clear from Theorem 1.7.6 that K is a precompact set in $C[0,1]$.

1.8 Extension of continuous functions

We consider in this section the question of when a continuous real-valued function defined on a subset of a metric space can be extended continuously to the whole space.

Lemma 1.8.1 (Uryson) *Let A, B be nonempty disjoint closed sets in a metric space M, then there is a continuous function defined on M such that $0 \le f \le 1$, $f = 0$ on A, and $f = 1$ on B.*

Proof For a set $S \subset M$, the function $x \mapsto \rho(x, S) := \inf_{z \in S} \rho(x, z)$ is continuous on M. This follows from the obvious inequality

$$|\rho(x, S) - \rho(y, S)| \le \rho(x, y)$$

for x, y in M. Since A and B are disjoint closed sets, $\rho(x, A) + \rho(x, B) > 0$ for $x \in M$, we may then define $f : M \to \mathbb{R}$ by

$$f(x) = \frac{\rho(x, A)}{\rho(x, A) + \rho(x, B)}, \quad x \in M.$$

Clearly f is continuous and is the function to be sought. ∎

Corollary 1.8.1 *Let A and B be nonempty disjoint closed sets in a metric space M; then for any pair $\alpha < \beta$ of real numbers, there is a continuous function f defined on M such that $\alpha \leq f \leq \beta, f = \alpha$ on A, and $f = \beta$ on B.*

Exercise 1.8.1 Prove Corollary 1.8.1.

Theorem 1.8.1 (Tietze) *Suppose that g is a bounded continuous function defined on a closed set C in a metric space M, and let $\gamma = \sup_{x \in C} |g(x)|$. Then there is a continuous function f defined on M such that $f = g$ on C and $\sup_{x \in M} |f(x)| = \gamma$.*

Proof We may assume that $M \backslash C$ contains infinitely many points, because otherwise M consists only of points from C and a finite number of isolated points, in which case the theorem is trivial. Then we may pick any two points x_1 and x_2 outside C, define $g(x_1) = -\gamma$, $g(x_2) = \gamma$, and replace C by $C \cup \{x_1, x_2\}$. Thus we may assume that $\min_{x \in C} g(x) = -\gamma$ and $\max_{x \in C} g(x) = \gamma$.

Now let $A = \{x \in C : g(x) \leq -\frac{\gamma}{3}\}$, $B = \{x \in C : g(x) \geq \frac{\gamma}{3}\}$, then A and B are disjoint nonempty closed sets. By Corollary 1.8.1 there is a continuous function f_1 on M such that $|f_1| \leq \frac{\gamma}{3}, f_1 = -\frac{\gamma}{3}$ on A and $f_1 = \frac{\gamma}{3}$ on B. It is readily verified that $|g - f_1| \leq \frac{2}{3}\gamma$ on C. Note that $\min_{x \in C}\{g(x) - f_1(x)\} = -\frac{2}{3}\gamma$ and $\max_{x \in C}\{g(x) - f_1(x)\} = \frac{2}{3}\gamma$.

Repeat the argument of the last paragraph with g replaced by $g - f_1$ and γ by $\frac{2}{3}\gamma$; we obtain a continuous function f_2 on M such that $|f_2| \leq \frac{1}{3} \cdot \frac{2}{3}\gamma$ and $|g - f_1 - f_2| \leq (\frac{2}{3})^2\gamma$ on C. Continuing in this fashion, we obtain a sequence $\{f_n\}$ of continuous functions on M such that $|f_n| \leq \frac{1}{3}(\frac{2}{3})^{n-1}\gamma$ and $|g - \sum_{j=1}^{n} f_j| \leq (\frac{2}{3})^n\gamma$ on C. It follows then that $\sum_n f_n$ converges uniformly to a continuous function f on M and $f = g$ on C. Now, $|f| \leq \sum_{j=1}^{\infty} |f_j| \leq \sum_{j=1}^{\infty} \frac{1}{3}(\frac{2}{3})^{j-1}\gamma = \gamma$. ■

Remark The function g in Theorem 1.8.1 is usually called an **extension** of the function f, while f is called the **restriction** of g on C and is often denoted as $g|_C$.

1.9 Connectedness

A metric space M is said to be **connected** if any nonempty subset of M which is both open and closed is M itself. Obviously any discrete space cannot be connected except when it consists of only one point. A subset of a metric space M is called connected if it is connected as a metric space with its metric inherited from M.

Exercise 1.9.1 Show that a metric space M is connected if and only if it cannot be expressed as a disjoint union of two nonempty subsets, both of which are open.

Theorem 1.9.1 *A finite closed interval in $\mathbb{R}$ is connected.*

Proof Let the interval be $I = [a, b]$, $-\infty < a, b < \infty$. Suppose that I is not connected, then $I = A \cup B$, where $A \cap B = \emptyset$ and both A and B are nonempty open and closed in I. We may suppose $a \in A$. Since B is bounded below by a, $\inf B \in I$. Since B is closed in I, $\inf B \in B$ and hence cannot be in A, which implies $a < \inf B$. Thus $(a, \inf B) \subset A$, and $\inf B$ is a limit point of A, but that A is closed implies $\inf B$ is in A, a contradiction. ■

Exercise 1.9.2

(i) Modify the arguments in the proof of Theorem 1.9.1 to show that any interval in $\mathbb{R}$ is connected whether it is finite or infinite and whether it is closed, open, or half-open.

(ii) Show that a subset A of $\mathbb{R}$ is connected if and only if for any pair $x < y$ of elements in A, $[x, y] \subset A$. Conclude then that connected sets in $\mathbb{R}$ are intervals.

Exercise 1.9.3 Show that every open set in $\mathbb{R}$ is a disjoint union of at most countably many open intervals.

1.10 Locally compact spaces

An account of compact sets in a locally compact metric space will now be given in regard to construction of some useful continuous functions relating to compact sets.

A metric space X is called a **locally compact** space if every x in X has a compact neighborhood. Clearly, $\mathbb{R}^n$ with the Euclidean metric is a locally compact space. We observe the following two facts for a locally compact space X:

(i) If K is a compact subset of X, then K has a compact neighborhood.

(ii) If K is a compact subset of X and $x \in X \backslash K$, then K has a compact neighborhood W_x not containing x.

To see (i), consider the open covering $\{\overset{\circ}{U}_x\}_{x \in K}$, where U_x is a compact neighborhood of x, and extract from it a finite subcovering $\{\overset{\circ}{U}_{x_1}, \ldots, \overset{\circ}{U}_{x_k}\}$ of K; then $\bigcup_{j=1}^{k} U_{x_j}$ is a compact neighborhood of K. Now if $x \in X \backslash K$, put $\delta = \mathrm{dist}(x, K) > 0$; then $W_x = V \cap \{y \in X : \mathrm{dist}(y, K) \le \frac{1}{2}\delta\}$ is a compact neighborhood of K not containing x, where V is a compact neighborhood of K as asserted in (i); thus (ii) holds.

Lemma 1.10.1 *Suppose that K is a compact subset of a locally compact space X and is contained in an open set G. Then K has a compact neighborhood V contained in G.*

Proof Because of (i) we may assume that $X \backslash G \neq \emptyset$. For each $x \in X$ let W_x be a compact neighborhood of K not containing x, as in (ii), and consider the family $\mathcal{F} = \{W_x \cap G^c : x \in G^c\}$ of compact sets; Clearly, $\bigcap \mathcal{F} = \emptyset$ and by the finite intersection property (Corollary 1.7.4) there are $x_1, \ldots, x_k$ in G^c such that $\bigcap_{j=1}^{k} \{W_{x_j} \cap G^c\} = \left[\bigcap_{j=1}^{k} W_{x_j}\right] \cap G^c = \emptyset$. We infer then from the last set relation that $V = \bigcap_{j=1}^{k} W_{x_j}$ is a compact neighborhood of K contained in G. ■

Lemma 1.10.2 *Let $\mathcal{F} = \{G_1, \ldots, G_n\}$ be a finite open covering of a compact set K in a locally compact space X; then there are compact sets $K_1, \ldots, K_n$ in X such that $K_j \subset G_j$ for each $j = 1, \ldots, n$ and $K \subset \bigcup_{j=1}^{n} K_j$.*

Proof For $x \in K$, there is j, $1 \leq j \leq n$, such that $x \in G_j$; then Lemma 1.10.1 implies that x has a compact neighborhood $V_x \subset G_j$. Since $\{\overset{\circ}{V}_x : x \in K\}$ is an open covering of K, there are $x_1, \ldots, x_k$ in K such that $\bigcup_{j=1}^k \overset{\circ}{V}_{x_j} \supset K$. For each $j = 1, \ldots, n$, let $\mathcal{F}_j = \{V_{x_i} : V_{x_i} \subset G_j\}$ and put $K_j = \bigcup \mathcal{F}_j$; then K_j is a compact set $\subset G_j$ and $\bigcup_{j=1}^n K_j = \bigcup_{i=1}^k V_{x_i} \supset K$. ■

Remark In Lemma 1.10.2, some of the K_j's might be empty; but if $\mathcal{F}$ has the property that every one of its proper subfamily is not a covering of K, then each K_j is nonempty.

For a function f defined on a metric space X, we shall denote by **supp f** the closure of the set $\{x \in X : f(x) \neq 0\}$. If *suppf* (which is called the **support** of f) is compact, f is called a function with **compact support**. The family of all continuous functions with compact support in a metric space X is denoted by $C_c(X)$. Note that $C_c(X)$ is a real or complex vector space depending on whether real-valued or complex-valued functions are considered. For an open set G in a metric space X, the family of all continuous functions f on X with compact support such that $0 \leq f \leq 1$ and $\operatorname{supp} f \subset G$ is to be denoted by $U_c(G)$.

Corollary 1.10.1 *Suppose that K is a compact set contained in an open set G of a locally compact space X. Then there is f in $U_c(G)$ such that $f = 1$ on K.*

Proof K has a compact neighborhood V contained in G by Lemma 1.10.1; then K and $\overset{\circ}{V}{}^c$ are disjoint closed subsets of X. Using the Uryson lemma (Lemma 1.8.1), we find a continuous function f on X such that $0 \leq f \leq 1, f = 0$ on $\overset{\circ}{V}{}^c$, and $f = 1$ on K. Since $\operatorname{supp} f \subset V \subset G, f \in U_c(G)$. ■

Suppose now that K is a compact set in a metric space X and $\mathcal{F} = \{G_1, \ldots, G_n\}$ is a finite open covering of K, then a collection $\{u_1, \ldots, u_n\}$ of continuous functions is called a **partition of unity** of K **subordinate** to $\mathcal{F}$ if $u_j \in U_c(G_j)$ for each $j = 1, \ldots, n$ and $\sum_{j=1}^n u_j(x) = 1$ for all $x \in K$.

Theorem 1.10.1 (Partition of unity) *Suppose that K is a compact set in a locally compact metric space X and that $\mathcal{F}$ is a finite open covering of K. Then K has a partition of unity subordinate to $\mathcal{F}$.*

Proof Let $\mathcal{F} = \{G_1, \ldots, G_n\}$. There are compact sets $K_1, \ldots, K_n$ such that $K_j \subset G_j$ for each j and $K \subset \bigcup_{j=1}^n K_j$, by Lemma 1.10.2. For each $j = 1, \ldots, n$, it then follows from Corollary 1.10.1 that there is a $f_j \in U_c(G_j)$ such that $f_j = 1$ on K_j. Define functions $u_1, \ldots, u_n$ by

$$u_1 = f_1,\ u_2 = (1 - f_1)f_2, \ldots,\ u_n = (1 - f_1) \cdots (1 - f_{n-1})f_n.$$

Then, $u_j \in U_c(G_j)$, $j = 1, \dots, n$. Now

$$\sum_{j=1}^{n} u_j = 1 - (1 - f_1) \cdots (1 - f_n), \tag{1.14}$$

as can be verified from $u_1 = 1 - (1 - f_1)$, $u_1 + u_2 = 1 - (1 - f_1)(1 - f_2)$, and so on. If $x \in K$, then $x \in K_j$ for some j and therefore $(1 - f_1(x)) \cdots (1 - f_n(x)) = 0$; consequently $\sum_{j=1}^{n} u_j(x) = 1$, by (1.14). ■

2 A Glimpse of Measure and Integration

This chapter gives a quick but precise exposition of the essentials of measure and integration so that an overall view of the subject is provided at the outset.

Preliminaries on various types of families of sets and set functions defined on them are covered in the first section, for later use in this chapter as well as in subsequent chapters.

The important L^p spaces are also introduced in this chapter for the reader to have an early appreciation of the power of the basic convergence theorems, which, together with the Egoroff theorem, reveal convincingly the relevance of σ-additivity of measures.

2.1 Families of sets and set functions

Sets considered in this section are subsets of a given fixed set Ω, which is sometimes referred to as a **universal set**; the family of all subsets of Ω is called the **power set** of Ω and is denoted by 2^Ω. A function τ defined on a nonempty family Φ of subsets of Ω and taking complex or extended real values is called a **set function**. If the empty set $\phi \in \Phi$, we always require that $\tau(\phi) = 0$. But hereafter in this chapter a set function τ is always assumed to take only nonnegative extended real values; and it is said to be finite if $\tau(A)$ is finite for $A \in \Phi$, while it is **σ-finite** if there is a sequence $\{A_n\} \subset \Phi$ such that $\bigcup \Phi \subset \bigcup_n A_n$ and $\tau(A_n) < \infty$ for each n. A set function τ is **monotone** if $\tau(A) \leq \tau(B)$ for A, B in Φ with $A \subset B$. A monotone set function τ with domain Φ is said to be **continuous from below** at $A \in \Phi$, if for every increasing sequence $\{A_n\} \subset \Phi$ with $A = \bigcup_n A_n$ the equality $\tau(A) = \lim_{n\to\infty} \tau(A_n)$ holds. Note that since τ is monotone, $\lim_{n\to\infty} \tau(A_n)$ exists. The set function τ is **continuous from below on Φ** if it is continuous from below at every $A \in \Phi$. A set function with ϕ in its domain is called a **premeasure** on Ω.

A family $\mathcal{P}$ of subsets of Ω is called a **π-system** on Ω if $A \cap B \in \mathcal{P}$ whenever A and B are in $\mathcal{P}$. The families $\{(-\infty, \alpha] : \alpha \in \mathbb{R}\}$ and $\{(a, b) : -\infty < a \leq b < \infty\}$ are π-systems on $\mathbb{R}$.

A family $\mathcal{A}$ of subsets of Ω is called an **algebra** on Ω if

(a_1) $\Omega \in \mathcal{A}$;
(a_2) if $A \in \mathcal{A}$, then $A^c := \Omega \backslash A$ is in $\mathcal{A}$;
(a_3) $A \cup B \in \mathcal{A}$ whenever A and B are in $\mathcal{A}$.

It is readily seen that if $\{A_1, \ldots, A_n\}$ is any finite subfamily of an algebra $\mathcal{A}$, then $\bigcup_{j=1}^n A_j \in \mathcal{A}$, and consequently $\bigcap_{j=1}^n A_j \in \mathcal{A}$, because $(\bigcap_{j=1}^n A_j)^c = \bigcup_{j=1}^n A_j^c$. One also notes that if A, B are in $\mathcal{A}$, then $A \backslash B := A \cap B^c$ is in $\mathcal{A}$.

A family Σ of subsets of Ω is called a **σ-algebra** on Ω if it is an algebra on Ω and if $\{A_n\}$ is a sequence in Σ; then $\bigcup_n A_n \in \Sigma$. Since $(\bigcap_n A_n)^c = \bigcup_n A_n^c$, $\bigcap_n A_n \in \Sigma$ if $\{A_n\}$ is a sequence in a σ-algebra Σ.

A family $\mathcal{L}$ of subsets of Ω is called a **λ-system** on Ω if the following conditions hold for $\mathcal{L}$:

(λ_1) $\Omega \in \mathcal{L}$;
(λ_2) if $A \in \mathcal{L}$, then $A^c \in \mathcal{L}$;
(λ_3) if $\{A_n\}$ is a disjoint sequence in $\mathcal{L}$, then $\bigcup_n A_n \in \mathcal{L}$.

Observe that if $\mathcal{L}$ is a λ-system on Ω and if A, B are in $\mathcal{L}$ with $A \subset B$, then $B \backslash A \in \mathcal{L}$, because $B \backslash A = A^c \cap B = (A \cup B^c)^c$.

Π-systems, λ-systems, algebras, and σ-algebras on Ω will often be simply referred to as π-systems, λ-systems, algebras, and σ-algebras if Ω is clearly implied in a statement.

We state without proof a trivial lemma for later reference.

Lemma 2.1.1 *A family of subsets of Ω is a σ-algebra on Ω if and only if it is both a π-system and a λ-system on Ω.*

Since the intersection of any collection of λ-systems on Ω is a λ-system, for any family Φ of subsets of Ω the smallest λ-system on Ω containing Φ exists and is denoted by $\lambda(\Phi)$. Similarly, the smallest σ-algebra on Ω containing Φ exists and is denoted by $\sigma(\Phi)$. We note that $\lambda(\Phi) \subset \sigma(\Phi)$ always, because any σ-algebra is a λ-system.

A λ-system satisfies a set of conditions which is a little weaker than that for a σ-algebra; but it turns out that often the set of conditions for λ-systems is much easier to verify than that for σ-algebras. The following theorem was first discovered by W. Sierpinski, and has been shown to be very useful in probability theory by E.B. Dynkin. It is now often referred to as the **(π–λ) Theorem**.

Theorem 2.1.1 (π-λ Theorem) *If $\mathcal{P}$ is a π-system on Ω, then $\lambda(\mathcal{P}) = \sigma(\mathcal{P})$.*

Proof Let $\mathcal{L}_0 = \lambda(\mathcal{P})$. If $\mathcal{L}_0$ is a π-system, then $\mathcal{L}_0$ is a σ-algebra, by Lemma 2.1.1, consequently $\mathcal{L}_0 \supset \sigma(\mathcal{P})$; but since $\mathcal{L}_0 = \lambda(\mathcal{P}) \subset \sigma(\mathcal{P})$, we have $\lambda(\mathcal{P}) = \sigma(\mathcal{P})$. It remains therefore to show that $\mathcal{L}_0$ is a π-system. For $A \in \mathcal{L}_0$, let

$$\mathcal{L}_A = \{B \subset \Omega : A \cap B \in \mathcal{L}_0\}.$$

To show that $\mathcal{L}_0$ is a π-system is to show that $\mathcal{L}_A \supset \mathcal{L}_0$ for every $A \in \mathcal{L}_0$. Clearly, $\mathcal{L}_A$ is a λ-system. Observe then that if $B \in \mathcal{P}$, then $\mathcal{L}_B \supset \mathcal{P}$, since $\mathcal{P}$ is a π-system, and hence $\mathcal{L}_B$ is a λ-system containing $\mathcal{P}$. Therefore, $\mathcal{L}_B \supset \mathcal{L}_0$ if $B \in \mathcal{P}$, this means that $A \cap B \in \mathcal{L}_0$ if $A \in \mathcal{L}_0$ and $B \in \mathcal{P}$, or $\mathcal{L}_A \supset \mathcal{P}$ if $A \in \mathcal{L}_0$. Since $\mathcal{L}_A$ is a λ-system, we then have $\mathcal{L}_A \supset \mathcal{L}_0$ for $A \in \mathcal{L}_0$. Thus $\mathcal{L}_0$ is a π-system and the theorem is proved. ■

We reiterate that hereafter in this chapter set functions are assumed to take nonnegative extended real values.

We shall call a set function μ defined on an algebra $\mathcal{A}$ on Ω an **additive** set function if $\mu(A \cup B) = \mu(A) + \mu(B)$ whenever $A, B \in \mathcal{A}$ and $A \cap B = \phi$. Recall that $\mu(\phi) = 0$. An additive set function μ on an algebra $\mathcal{A}$ is $\boldsymbol{\sigma}$**-additive** if

$$\mu\left(\bigcup_n A_n\right) = \sum_n \mu(A_n),$$

whenever $\{A_n\}$ is a disjoint sequence in $\mathcal{A}$ with $\bigcup_n A_n \in \mathcal{A}$.

Exercise 2.1.1 Let μ be an additive set function defined on an algebra $\mathcal{A}$ on Ω.

(i) Show that μ is monotone.

(ii) Show that if $A_1, \ldots, A_n$ are in $\mathcal{A}$, then $\mu(\bigcup_{j=1}^n A_j) \le \sum_{j=1}^n \mu(A_j)$.

(iii) Show that μ is σ-additive if and only if μ is continuous from below on $\mathcal{A}$.

(iv) Show that if $\{A_j\}_{j=1}^\infty \subset \mathcal{A}$ with $\bigcup_{j=1}^\infty A_j \in \mathcal{A}$, then $\mu(\bigcup_{j=1}^\infty A_j) \le \sum_{j=1}^\infty \mu(A_j)$ if μ is σ-additive on $\mathcal{A}$.

Theorem 2.1.2 *Suppose that Ω is a compact metric space and $\mathcal{A}$ is an algebra of compact subsets of Ω. If μ is an additive set function on $\mathcal{A}$, then μ is σ-additive.*

Proof To show that μ is σ-additive is to show that if $A_1 \subset A_2 \subset \cdots$ is an increasing sequence in $\mathcal{A}$ such that $\bigcup_n A_n \in \mathcal{A}$, then $\mu(\bigcup_n A_n) = \lim_{n\to\infty} \mu(A_n)$ (cf. Exercise 2.1.1 (iii)). Let $A = \bigcup_n A_n$ and put $C_n = A \backslash A_n$ for each n, then $\bigcap_n C_n = \emptyset$. We claim that $\lim_{n\to\infty} \mu(C_n) = 0$. If not, then $\mu(C_n) \ge \lim_{n\to\infty} \mu(C_n) > 0$ implies that $C_n \neq \emptyset$ for all n. Then $\bigcap_n C_n \neq \emptyset$, by Exercise 1.7.2, contradicting the fact that $\bigcap_n C_n = \emptyset$. Now, $\lim_{n\to\infty} \mu(A_n) = \lim_{n\to\infty}\{\mu(A_n) + \mu(C_n)\} = \mu(A)$. ■

Example 2.1.1 Consider the sequence space $\Omega = \{0, 1\} \times \{0, 1\} \times \cdots$ and the additive set function P defined on the algebra $\mathcal{Q}$ of all cylinders in Ω (cf. Section 1.3). We have seen in Example 1.7.1 that Ω is compact with a suitable metric and that sets in $\mathcal{Q}$ are compact, hence P is a σ-additive set function on $\mathcal{Q}$, by Theorem 2.1.2.

A σ-additive set function μ defined on a σ-algebra Σ on Ω is called a **measure on** $\boldsymbol{\Sigma}$.

Exercise 2.1.2 Let μ be a σ-additive set function defined on an algebra $\mathcal{A}$ on Ω with $\mu(\Omega) < \infty$. Suppose that μ_1 and μ_2 are measures defined on a σ-algebra $\Sigma \supset \mathcal{A}$, with the property that $\mu_1(A) = \mu_2(A) = \mu(A)$ for $A \in \mathcal{A}$. Show that $\mu_1(B) = \mu_2(B)$ for $B \in \sigma(\mathcal{A})$. (Hint: show that $\mathcal{L} = \{B \in \Sigma : \mu_1(B) = \mu_2(B)\}$ is a λ-system.)

2.2 Measurable spaces and measurable functions

A function f defined on a set Ω and taking values in $[-\infty, \infty] := \{-\infty\} \cup \mathbb{R} \cup \{\infty\}$ is said to be **extended real-valued**. The sets $[-\infty, \infty]$ and $[0, \infty] := [0, \infty) \cup \{\infty\}$ will often also be denoted as $\overline{\mathbb{R}}$ and $\overline{\mathbb{R}}^{+}$ respectively, while $[0, \infty)$ will also be denoted as $\mathbb{R}^{+}$. Since, except where explicitly specified otherwise, functions considered are extended real-valued; we shall often call an extended real-valued function defined on Ω simply a function on Ω; while if f takes values in $\mathbb{R}$, f is said to be **real-valued** or **finite-valued**. We recall some usual conventions concerning algebraic operations involving infinity symbols ∞ and $-\infty$: $\infty + \infty = \infty$, $-\infty + (-\infty) = -\infty$, $a + \infty = -(a - \infty) = \infty$ if a is a finite number, while for an extended real number a, $a \cdot \infty = (-a) \cdot (-\infty) = \infty$, or $-\infty$, depending on whether $a > 0$ or $a < 0$, and $0 \cdot \infty = 0 \cdot (-\infty) = 0$. The symbol ∞ is sometimes written $+\infty$ for emphasis. We shall also adopt the convention that $(-\infty)^{-1} = (\infty)^{-1} = 0$, but then $\frac{\infty}{\infty}$, $\frac{-\infty}{-\infty}$, $\frac{\infty}{-\infty}$ and $\frac{-\infty}{\infty}$ are considered not to be defined. We also observe that $\infty - \infty$ and $\infty + (-\infty)$ are not defined.

An ordered pair (Ω, Σ) is called a **measurable space** if Ω is a nonempty set and Σ is a σ-algebra on Ω.

Given a measurable space (Ω, Σ), a function f on Ω is called **Σ-measurable** if $\{x \in \Omega : f(x) > \alpha\} \in \Sigma$ for every $\alpha \in \mathbb{R}$. A Σ-measurable function will simply be called **measurable** if the measurable space (Ω, Σ) is clearly implied. More generally, a function is said to be measurable on $A \in \Sigma$ if its domain of definition contains A and if $\{x \in A : f(x) > \alpha\} \in \Sigma$ for every $\alpha \in \mathbb{R}$. Observe that a function is Σ-measurable if and only if $\{x \in \Omega : f(x) > \alpha\} \in \Sigma$ for all $\alpha \in \overline{\mathbb{R}}$. This is clear, because $\{x \in \Omega : f(x) > \infty\} = \phi$ and $\{x \in \Omega : f(x) > -\infty\} = \bigcup_{n \in \mathbb{N}} \{x \in \Omega : f(x) > -n\}$. For notational simplicity, we shall presently introduce simplified notations for sets like $\{x \in \Omega : f(x) > \alpha\}$. For a set $C \subset \overline{\mathbb{R}}$ and a function f on Ω, the set $\{x \in \Omega : f(x) \in C\}$ will be denoted simply as $\{f \in C\}$. With this notation, f is Σ-measurable if $\{f \in (\alpha, \infty]\} \in \Sigma$ for all $\alpha \in \mathbb{R}$. $\{f \in (\alpha, \infty]\}$ will also be denoted as $\{f > \alpha\}$. Similarly, for $\alpha \leq \beta$ in $\overline{\mathbb{R}}$, the sets $\{f \in (\alpha, \beta)\}$, $\{f \in (\alpha, \beta]\}$, $\{f \in [\alpha, \beta)\}$ and $\{f \in [\alpha, \beta]\}$ in this order will be denoted as $\{\alpha < f < \beta\}$, $\{\alpha < f \leq \beta\}$, $\{\alpha \leq f < \beta\}$, and $\{\alpha \leq f \leq \beta\}$ respectively.

Constant functions are certainly measurable functions; after constant functions, measurable functions of the simplest structure are the simple functions that will now be introduced. For $A \subset \Omega$, we denote by I_A the function defined by $I_A(x) = 1$ or 0, according to whether $x \in A$ or not. The function I_A is called the **indicator** function of the set A; clearly, I_A is measurable if and only if $A \in \Sigma$. A function of the form $\sum_{j=1}^{k} \alpha_j I_{A_j}$, $k \in \mathbb{N}$, $\alpha_j \in \mathbb{R}$, $A_j \in \Sigma$, is called a **simple** function. One can verify directly that simple functions are measurable and form a real vector space of functions.

For a metric space M we shall denote by $\mathcal{B}(M)$ the smallest σ-algebra on M containing all open subsets of M and call a $\mathcal{B}(M)$-measurable function defined on M a **Borel** measurable function (or simply a **Borel** function). It is easily seen that a monotone increasing (decreasing) function defined on an interval of $\mathbb{R}$ is Borel measurable. One also verifies readily that lower semi-continuous functions and upper semi-continuous functions are Borel measurable. Sets in $\mathcal{B}(M)$ are called **Borel** sets in M and $\mathcal{B}(M)$ is usually referred to as the **Borel** field on M. $\mathcal{B}(\mathbb{R})$ will be simply denoted by $\mathcal{B}$. The smallest

σ-algebra on $\overline{\mathbb{R}}$ containing all open subsets of $\mathbb{R}$ as well as sets of the form $(\alpha, \infty]$ for all $\alpha \in \mathbb{R}$ is denoted by $\overline{\mathcal{B}}$. Sets in $\overline{\mathcal{B}}$ are called Borel sets in $\overline{\mathbb{R}}$. For $n \geq 2$, $\mathcal{B}(\mathbb{R}^n)$ is simply denoted by $\mathcal{B}^n$.

Example 2.2.1 Let $\{f_n\}$ be a sequence of real-valued continuous functions defined on a metric space M, and let $C = \{x \in M : \lim_{n\to\infty} f_n(x) \text{ exists}\}$. Then $C = \bigcap_{k\in\mathbb{N}} \bigcup_{l\in\mathbb{N}} \bigcap_{n,m\geq l} A_{nm}^{(k)}$, where for k, m, n in $\mathbb{N}$, $A_{nm}^{(k)} = \{x \in M : |f_n(x) - f_m(x)| < \frac{1}{k}\}$. Since each $A_{nm}^{(k)}$ is open, C is a Borel set in M.

Given a measurable space (Ω, Σ), a function defined on Ω is often referred to as a function on (Ω, Σ), by abuse of language, if the role of Σ is to be emphasized; in particular, a measurable function on (Ω, Σ) means a Σ-measurable function defined on Ω.

Remark If f is a measurable function, then $\{f \geq \alpha\} = \bigcap_{m\in\mathbb{N}}\{f > \alpha - \frac{1}{m}\}$ is in Σ; similarly, $\{f < \alpha\} = \bigcup_{m\in\mathbb{N}}\{f \leq \alpha - \frac{1}{m}\}$ is in Σ, because each $\{f \leq \alpha - \frac{1}{m}\} = \{f > \alpha - \frac{1}{m}\}^c$ is in Σ.

Exercise 2.2.1

(i) Show that $\overline{\mathcal{B}}$ is the smallest σ-algebra on $\overline{\mathbb{R}}$ containing $\{(\alpha, \infty] : \alpha \in \mathbb{R}\}$.

(ii) Let (Ω, Σ) be a measurable space. Show that a function f on Ω is Σ-measurable if and only if $\{f \in B\} \in \Sigma$ for all $B \in \overline{\mathcal{B}}$.

(iii) Let (Ω, Σ) be a measurable space. Show that if f is a finite-valued function on Ω, then f is Σ-measurable if and only if $\{f \in B\} \in \Sigma$ for all $B \in \mathcal{B}$.

For a family $\{f_\alpha\}$ of functions defined on a set Ω, define functions $\inf_\alpha f_\alpha$ and $\sup_\alpha f_\alpha$ by

$$\left(\inf_\alpha f_\alpha\right)(x) = \inf_\alpha f_\alpha(x); \quad \left(\sup_\alpha f_\alpha\right)(x) = \sup_\alpha f_\alpha(x)$$

for $x \in \Omega$. $\mathrm{Inf}_\alpha f_\alpha$ and $\sup_\alpha f_\alpha$ are sometimes expressed respectively by $\bigwedge_\alpha f_\alpha$ and $\bigvee_\alpha f_\alpha$. If $\{f_n\}$ is a sequence of functions defined on Ω, define functions $\liminf_{n\to\infty} f_n$ and $\limsup_{n\to\infty} f_n$ by

$$\left(\liminf_{n\to\infty} f_n\right)(x) = \lim_{n\to\infty}\left(\inf_{m\geq n} f_m(x)\right); \quad \left(\limsup_{n\to\infty} f_n\right)(x) = \lim_{n\to\infty}\left(\sup_{m\geq n} f_m(x)\right)$$

for $x \in \Omega$. Since uncertainty is not likely, $(\liminf_{n\to\infty} f_n)(x)$ and $(\limsup_{n\to\infty} f_n)(x)$ will be simply written as $\liminf_{n\to\infty} f_n(x)$ and $\limsup_{n\to\infty} f_n(x)$ respectively.

Naturally, if $\liminf_{n\to\infty} f_n(x) = \limsup_{n\to\infty} f_n(x)$, the common value is denoted by $\lim_{n\to\infty} f_n(x)$ and we say that the sequence $\{f_n\}$ converges at x. If $\{f_n\}$ converges at all $x \in A \subset \Omega$, and if we define a function f on A by $f(x) = \lim_{n\to\infty} f_n(x)$, then we say that the sequence $\{f_n\}$ converges **pointwise** on A to f (notationally, $f_n \to f$ on A).

Exercise 2.2.2 Let (Ω, Σ) be a measurable space and $\{f_n\}$ a sequence of measurable functions on Ω.

(i) Show that both $\inf_n f_n$ and $\sup_n f_n$ are measurable functions on Ω. (Hint: $\{\inf_n f_n > \alpha\} = \bigcup_m \bigcap_n \{f_n > \alpha + \frac{1}{m}\}$.)

(ii) Show that both $\liminf_{n\to\infty} f_n$ and $\limsup_{n\to\infty} f_n$ are measurable functions on Ω.

(iii) Show that $\{x \in \Omega : \liminf_{n\to\infty} f_n(x) = \limsup_{n\to\infty} f_n(x)\} \in \Sigma$.

(iv) Show that if $\lim_{n\to\infty} f_n(x)$ exists for all $x \in \Omega$, then $f = \lim_{n\to\infty} f_n$ is measurable.

Exercise 2.2.3 Let f be measurable. For each positive integer n, let $A^{(n)} = \{f < -n\}$, $C^{(n)} = \{f \geq n\}$, $B_i^{(n)} = \{-n + \frac{i}{n} \leq f < -n + \frac{i+1}{n}\}$, $i = 0, 1, 2, \ldots, 2n^2 - 1$, and let

$$g_n = -nI_{A^{(n)}} + \sum_{i=0}^{2n^2-1} \left(-n + \frac{i}{n}\right) I_{B_i^{(n)}} + nI_{C^{(n)}}.$$

Show that $g_n \to f$ pointwise and show that if f, g are measurable, then fg is measurable; furthermore if $g \neq 0$ everywhere on X, then f/g is also measurable.

Exercise 2.2.4 Let (Ω, Σ) be a measurable space and f, g measurable functions on Ω. Then $f + g$ is defined on Ω if and only if $\{f(x), g(x)\} \neq \{-\infty, \infty\}$ for all $x \in \Omega$. Show that if $f + g$ is defined on Ω, then $f + g$ is measurable. (Hint: $\{f + g > \alpha\} = \bigcup_{q\in\mathbb{Q}} \{f > q\} \cap \{g > \alpha - q\}$ for $\alpha \in \mathbb{R}$, where $\mathbb{Q}$ is the set of all rational numbers.)

Since for a measurable function f on Ω and $\lambda \in \mathbb{R}$, λf is clearly measurable, we infer from Exercise 2.2.4 that the space of all finite-valued measurable functions is a real vector space which contains the space of all simple functions as a vector subspace.

To conclude this section, we present a useful representation for nonnegative measurable functions.

Theorem 2.2.1 *Suppose that (Ω, Σ) is a measurable space and f is a nonnegative measurable function defined on Ω, then there is a sequence $\{A_j\}_{j=1}^{\infty} \subset \Sigma$ such that*

$$f(\omega) = \sum_{j=1}^{\infty} \frac{1}{j} I_{A_j}(\omega) \tag{2.1}$$

for all $\omega \in \Omega$.

Proof Define sets $A_1, \ldots, A_j, \ldots$ recursively as follows: $A_1 = \{f \geq 1\}$, $A_2 = \{f \geq \frac{1}{2} + I_{A_1}\}, \ldots, A_j = \{f \geq \frac{1}{j} + \sum_{k<j} \frac{1}{k} I_{A_k}\}, \ldots$. Clearly each A_j is in Σ. We now show that (2.1) holds for $\omega \in \Omega$.

Observe first, that $\omega \in \Omega \setminus \bigcup_{j=1}^{\infty} A_j$ if and only if $f(\omega) = 0$ and that when $\omega \in \Omega \setminus \bigcup_{j=1}^{\infty} A_j$, both sides of (2.1) are equal to zero. It remains to show that (2.1) holds for $\omega \in \bigcup_{j=1}^{\infty} A_j$.

For $\omega \in \bigcup_{j=1}^{\infty} A_j$, we distinguish two cases:

[Case 1] $\omega \in A_j$ for only finitely many j.
Let j_0 be the largest j such that $\omega \in A_j$. Then,

$$f(\omega) \geq \frac{1}{j_0} + \sum_{k<j_0} \frac{1}{k} I_{A_k}(\omega) = \sum_{k=1}^{j_0} \frac{1}{k} I_{A_k}(\omega) = \sum_{k=1}^{\infty} \frac{1}{k} I_{A_k}(\omega);$$

on the other hand, for $j > j_0$,

$$f(\omega) < \frac{1}{j} + \sum_{k<j} \frac{1}{k} I_{A_k}(\omega) = \frac{1}{j} + \sum_{k=1}^{\infty} \frac{1}{k} I_{A_k}(\omega);$$

hence, by letting $j \to \infty$, we have

$$f(\omega) \leq \sum_{k=1}^{\infty} \frac{1}{k} I_{A_k}(\omega).$$

Thus (2.1) holds in this case.
[Case 2] $\omega \in A_j$ for infinitely many j.
For infinitely many j, we have

$$f(\omega) \geq \frac{1}{j} + \sum_{k<j} \frac{1}{k} I_{A_k}(\omega) = \sum_{k=1}^{j} \frac{1}{k} I_{A_k}(\omega);$$

let $j \to \infty$ through such j's, it follows that

$$f(\omega) \geq \sum_{k=1}^{\infty} \frac{1}{k} I_{A_k}(\omega). \tag{2.2}$$

Now either $\omega \in A_j$ for $j \geq N$ for some $N \in \mathbb{N}$ or $\omega \notin A_j$ for infinitely many j. In the former case,

$$f(\omega) \geq \sum_{k=1}^{\infty} \frac{1}{k} I_{A_k}(\omega) \geq \sum_{k=N}^{\infty} \frac{1}{k} = \infty,$$

thus $f(\omega) = \infty = \sum_{k=1}^{\infty} \frac{1}{k} I_{A_k}(\omega)$; in the latter case,

$$f(\omega) < \frac{1}{j} + \sum_{k<j} \frac{1}{k} I_{A_k}(\omega)$$

for infinitely many j and hence when $j \to \infty$ through such j's, it follows that

$$f(\omega) \leq \sum_{k=1}^{\infty} \frac{1}{k} I_{A_k}(\omega),$$

which together with (2.2) shows that (2.1) holds. ■

Corollary 2.2.1 *If f is a nonnegative measurable function, then there is a nondecreasing sequence $\{s_n\}$ of nonnegative simple functions which converges to f pointwise.*

Proof Let $\{A_j\}$ be the sequence of measurable sets in Theorem 2.2.1. Choose the sequence $\{s_n\}$ defined by

$$s_n = \sum_{j=1}^{n} \frac{1}{j} I_{A_j}.$$ ■

Exercise 2.2.5 Let f be a measurable function; show that there is a sequence $\{f_n\}$ of simple functions such that $|f_n| \leq |f|$ and $f_n(\omega) \to f(\omega)$ for all $\omega \in \Omega$.

2.3 Measure space and integration

A triple (Ω, Σ, μ) is called a **measure space** if (Ω, Σ) is a measurable space and μ is a measure on Σ. When $\mu(\Omega) = 1$, (Ω, Σ, μ) is called a **probability space**, and in this case μ is usually denoted by P.

Example 2.3.1 Let Ω be an arbitrary nonempty set and for $A \subset \Omega$ let $\mu(A)$ be the cardinality of A if A is finite; otherwise let $\mu(A) = \infty$. Obviously μ is a measure on 2^{Ω}, the σ-algebra of all subsets of Ω, and is called the **counting measure** on Ω. The measure space $(\Omega, 2^{\Omega}, \mu)$ will be called the measure space with counting measure on Ω.

Example 2.3.2 Let Ω be a countable set, say $\Omega = \{\omega_1, \omega_2, \dots, \omega_n, \dots\}$, and $\{p_n\}$ a sequence of nonnegative real numbers with $\sum_{n=1}^{\infty} p_n = 1$. For $A \subset \Omega$, let $\mathbb{N}(A) = \{n \in \mathbb{N} : \omega_n \in A\}$ and $\mu(A) = \sum_{n \in \mathbb{N}(A)} p_n$; then the measure space $(\Omega, 2^{\Omega}, \mu)$ is called a **discrete** probability space.

Given a measure space (Ω, Σ, μ), measurable functions are extended real-valued functions measurable in reference to the measurable space (Ω, Σ).

We now fix a measure space (Ω, Σ, μ) and define the integral for certain measurable functions. Recall that a simple function is a finite linear combination of indicator functions of sets in Σ. Clearly if f is a simple function, then $f = \sum_{i=1}^{k} \alpha_i I_{A_i}$, where $\alpha_1, \dots, \alpha_k$ are the different values assumed by f and $A_i = \{f = \alpha_i\}$; we define then

$$\int_{\Omega} f d\mu = \sum_{i=1}^{k} \alpha_i \mu(A_i), \tag{2.3}$$

if the right-hand side of (2.3) has a meaning. It is easy to see that if $\int_\Omega f d\mu$ is defined and f is expressed as $f = \sum_{i=1}^{l} \beta_i I_{B_i}$, where $B_1, \ldots, B_l$ are in Σ and are disjoint, then

$$\int_\Omega f d\mu = \sum_{i=1}^{l} \beta_i \mu(B_i).$$

In particular, $\int_\Omega f d\mu$ has a meaning if f is simple and nonnegative, although it is possible that $\int_\Omega f d\mu = +\infty$.

If f is measurable and nonnegative, define

$$\int_\Omega f d\mu = \sup \int_\Omega g d\mu,$$

where the supremum is taken over all simple functions g with $0 \leq g \leq f$. Obviously, if f is nonnegative and simple, this definition coincides with the previously defined $\int_\Omega f d\mu$ for simple functions.

For a function f defined on a set Ω, define nonnegative functions f^+ and f^- by

$$\begin{aligned} f^+(x) &= f(x) \ \text{ if } f(x) \geq 0, \\ &= 0 \text{ otherwise;} \\ f^-(x) &= -f(x) \ \text{ if } f(x) \leq 0, \\ &= 0 \text{ otherwise.} \end{aligned}$$

Then $f = f^+ - f^-$ and $|f| = f^+ + f^-$; furthermore, if f is measurable on a measurable space (Ω, Σ), then both f^+ and f^- are measurable.

Return now to the discourse interrupted by the last paragraph and let f be a measurable function. Define

$$\int_\Omega f d\mu = \int_\Omega f^+ d\mu - \int_\Omega f^- d\mu$$

if the right-hand side has a meaning. In this case, $\boldsymbol{\int_\Omega f d\mu}$ is said to **exist** and is called the **integral** of f. One notes that if f is a simple function this definition of $\int_\Omega f d\mu$ coincides with that given by (2.3). If $\int_\Omega f d\mu$ is finite, then f is said to be **integrable**. Integrability and the integral of a measurable function so defined will be referred to more precisely as μ-**integrability** and the μ-**integral** respectively, if the measure μ is to be emphasized. It will be shown later that a measurable function f is integrable if and only if $|f|$ is integrable (see Theorem 2.5.3).

Suppose that f is a measurable function and $A \in \Sigma$; if $\int_\Omega f I_A d\mu$ exists, it is denoted by $\int_A f d\mu$ and is called the integral of f **over** A. Obviously, if $\int_\Omega f d\mu$ exists, then $\int_A f d\mu$ exists for all $A \in \Sigma$.

Example 2.3.3 Let Ω be an arbitrary set and consider the counting measure μ on Ω; then every function f on Ω is measurable and f is integrable if and only if $f(x)$ is finite for $x \in \Omega$ and $\{f(x)\}_{x\in\Omega}$ is summable.

Example 2.3.4 Consider the discrete probability space of Example 2.3.2. Let f be a function on Ω. Since every subset of Ω is measurable, f is measurable and is called a random variable. If $\int_\Omega f d\mu$ exists, it is called the expectation of f. It is easily verified that f is integrable if and only if $\{f(\omega_n)p_n\}_{n\in\mathbb{N}}$ is summable.

Exercise 2.3.1 If f and g are nonnegative simple functions and $\alpha,\ \beta \geq 0$, show that

$$\int_\Omega (\alpha f + \beta g) d\mu = \alpha \int_\Omega f d\mu + \beta \int_\Omega g d\mu.$$

Exercise 2.3.2 If $f \leq g$ are two nonnegative measurable functions, show that $\int_\Omega f d\mu \leq \int_\Omega g d\mu$.

Exercise 2.3.3 Suppose that f and g are measurable functions such that $f \leq g$, and suppose that $\int_\Omega g^+ d\mu < \infty$. Show that $\int_\Omega f d\mu$ exists and $\int_\Omega f d\mu \leq \int_\Omega g d\mu$.

Exercise 2.3.4 Let f be a measurable function on a measure space (Ω, Σ, μ) and for each $k \in \mathbb{N}$ let $A_k = \{2^{k-1} \leq |f| < 2^k\}$. Show that f is integrable if and only if $\sum_{k=1}^{\infty} 2^{k-1}\mu(A_k) < \infty$ and $\mu(\{|f| = \infty\}) = 0$.

Example 2.3.5 Suppose that f is a nonnegative measurable function and $0 \leq p < r < q < \infty$. Then $\int_\Omega f^r d\mu \leq \int_\Omega f^p d\mu + \int_\Omega f^q d\mu$. Actually, if we let $A = \{f \leq 1\}$ and $B = \{f > 1\}$, then $\int_\Omega f^r d\mu = \int_\Omega I_A f^r d\mu + \int_\Omega I_B f^r d\mu \leq \int_\Omega I_A f^p d\mu + \int_\Omega I_B f^q d\mu \leq \int_\Omega f^p d\mu + \int_\Omega f^q d\mu$.

2.4 Egoroff theorem and monotone convergence theorem

Suppose that Ω is a set and $\{A_n\}_{n=1}^{\infty}$ is a sequence of subsets of Ω, define

$$\limsup_{n\to\infty} A_n = \bigcap_{n\in\mathbb{N}} \bigcup_{k\geq n} A_k;$$
$$\liminf_{n\to\infty} A_n = \bigcup_{n\in\mathbb{N}} \bigcap_{k\geq n} A_k.$$

If $\limsup_{n\to\infty} A_n = \liminf_{n\to\infty} A_n$, then we say that the limit of the sequence $\{A_n\}$ exists and has the common set as its limit, which is denoted by $\lim_{n\to\infty} A_n$. In particular, if $A_1 \subset A_2 \subset \cdots \subset A_n \subset A_{n+1} \subset \cdots$ i.e. $\{A_n\}$ is monotone increasing, or $A_1 \supset A_2 \supset \cdots \supset A_n \supset A_{n+1} \supset \cdots$ i.e. $\{A_n\}$ is monotone decreasing, then $\lim_{n\to\infty} A_n$ exists and equals $\bigcup_{n\in\mathbb{N}} A_n$ or $\bigcap_{n\in\mathbb{N}} A_n$ according to whether $\{A_n\}$ is monotone increasing or monotone decreasing. Hence $\limsup_{n\to\infty} A_n = \lim_{n\to\infty} \bigcup_{k\geq n} A_k$ and $\liminf_{n\to\infty} A_n = \lim_{n\to\infty} \bigcap_{k\geq n} A_k$.

Exercise 2.4.1 Let $\{A_n\}_{n=1}^{\infty} \subset 2^{\Omega}$, where Ω is an arbitrary set, and let $B = \liminf_{n\to\infty} A_n$, $C = \limsup_{n\to\infty} A_n$. Show that for each $x \in \Omega$ we have

$$I_B(x) = \liminf_{n\to\infty} I_{A_n}(x) \text{ and } I_C(x) = \limsup_{n\to\infty} I_{A_n}(x).$$

In the following, a measure space (Ω, Σ, μ) is considered and fixed throughout.

Lemma 2.4.1 (Monotone limit lemma) *Let $\{A_n\}_{n=1}^{\infty} \subset \Sigma$ be monotone increasing, then*

$$\mu\left(\lim_{n\to\infty} A_n\right) = \mu\left(\bigcup_n A_n\right) = \lim_{n\to\infty} \mu(A_n).$$

Proof For each positive integer n let $B_n = A_n \setminus A_{n-1}$, where we put $A_0 = \emptyset$, and for convenience let $A = \bigcup_n A_n$. Then $A_n = \bigcup_{k=1}^{n} B_k$ and $A = \bigcup_k B_k$. Since $\{B_k\}$ is disjoint, we have

$$\mu(A) = \sum_{k=1}^{\infty} \mu(B_k) = \lim_{n\to\infty} \sum_{k=1}^{n} \mu(B_k) = \lim_{n\to\infty} \mu(A_n). \qquad \blacksquare$$

Corollary 2.4.1 *Let $\{A_n\}_{n=1}^{\infty} \subset \Sigma$ be monotone decreasing and $\mu(A_1) < \infty$, then*

$$\mu\left(\lim_{n\to\infty} A_n\right) = \mu\left(\bigcap_n A_n\right) = \lim_{n\to\infty} \mu(A_n).$$

Proof For each positive integer n let $B_n = A_1 \setminus A_n$, and for convenience let $A = \bigcap_n A_n$. Then $\{B_k\}$ is monotone increasing and $A_1 \setminus A = \bigcup_k B_k$. From Lemma 2.4.1, we have

$$\mu(A_1 \setminus A) = \mu\left(\bigcup_k B_k\right) = \lim_{n\to\infty} \mu(B_n).$$

But $\mu(A_1 \setminus A) = \mu(A_1) - \mu(A)$ and $\mu(B_n) = \mu(A_1) - \mu(A_n)$; this completes the proof of the corollary. $\blacksquare$

Remark In Corollary 2.4.1 one may assume that $\mu(A_n) < \infty$ for some n, instead of $\mu(A_1) < \infty$.

Exercise 2.4.2 Let (Ω, Σ, μ) be a measure space. Suppose $\{A_n\}_{n\in\mathbb{N}} \subset \Sigma$.

(i) Show that $\mu\left(\liminf_{n\to\infty} A_n\right) \le \liminf_{n\to\infty} \mu(A_n)$.

(ii) If $\mu\left(\bigcup_{j\ge n} A_j\right) < +\infty$ for some n, then show that $\mu\left(\limsup_{n\to\infty} A_n\right) \ge \limsup_{n\to\infty} \mu(A_n)$.

(iii) If the limit of $\{A_n\}$ exists and $\mu\left(\bigcup_{j\ge n} A_j\right) < \infty$ for some n, show that $\lim_{n\to\infty} \mu(A_n)$ exists and

$$\mu\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} \mu(A_n).$$

Theorem 2.4.1 (Egoroff theorem) *If $\{f_n\}$ is a sequence of measurable functions and $f_n \to f$ with finite limit on $A \in \Sigma$, where $\mu(A) < +\infty$, then for any given $\varepsilon > 0$, there is $B \in \Sigma$ with $B \subset A$, such that $\mu(A \setminus B) < \varepsilon$ and $f_n \to f$ uniformly on B.*

Proof

[Step 1] We claim that for $\varepsilon > 0$, $\eta > 0$, there is integer $N > 0$ and $C \in \Sigma$ such that $C \subset A$, $\mu(A \setminus C) < \varepsilon$, and $\sup_{x\in C}|f(x) - f_n(x)| \le \eta$ whenever $n \ge N$.

To show this, for each n let $C_n = \bigcap_{m\ge n}\{x \in A : |f(x) - f_m(x)| \le \eta\}$. Then $C_n \nearrow A$. Since $\mu(A) < \infty$, there is N such that $\mu(A \backslash C_N) = \mu(A) - \mu(C_N) < \varepsilon$ by Lemma 2.4.1. Take $C = C_N$.

[Step 2] Now given $\varepsilon > 0$. By [Step 1] for each positive integer m there is integer N_m and $C_m \subset A$ with $C_m \in \Sigma$ such that

$$\mu(A \setminus C_m) < \varepsilon/2^m$$

and

$$\sup_{x\in C_m} |f(x) - f_n(x)| \le \frac{1}{m}$$

whenever $n \ge N_m$.

Take $B = \bigcap_{m=1}^{\infty} C_m$, then $\mu(A\backslash B) = \mu(\bigcup_{m=1}^{\infty}(A\backslash C_m)) < \varepsilon$. Given $\sigma > 0$, choose $m_0 \in \mathbb{N}$ such that $\frac{1}{m_0} < \sigma$. Then for $n \ge N_{m_0}$, we have $|f(x) - f_n(x)| \le \frac{1}{m_0} < \sigma$ for all $x \in B$, because $B \subset C_{m_0}$. This shows that $f_n \to f$ uniformly on B. ■

In plain language, Theorem 2.4.1 says that convergence of a sequence of measurable functions on a set of finite measure implies approximate uniform convergence. From its proof, one sees clearly that σ-additivity of μ plays a salient role through Lemma 2.4.1. The following theorem which is called the **monotone convergence theorem** reveals the distinguished feature of σ-additivity of measure μ through integrals.

Example 2.4.1 Suppose $\mu(\Omega) < \infty$ and $\{f_n\}$ is a sequence of real-valued measurable functions such that $\lim_{n\to\infty} f_n(x) = f(x)$ exists and is finite for μ-a.e. x in Ω. For each $k \in \mathbb{N}$, by the Egoroff theorem there is $B_k \in \Sigma$ such that $\mu(\Omega\backslash B_k) < \frac{1}{k}$ and $f_n(x) \to f(x)$ uniformly for $x \in B_k$. Put $\mathbb{Z} = \Omega\backslash \bigcup_k B_k$, then $\mu(\mathbb{Z}) \le \mu(\Omega\backslash B_k) < \frac{1}{k}$ for all k and hence $\mu(\mathbb{Z}) = 0$. Therefore we have shown that there are $B_1, B_2, \ldots, \mathbb{Z}$ in Σ with $\mu(\mathbb{Z}) = 0$ such that $\Omega = \bigcup_k B_k \cup \mathbb{Z}$ and $\lim_{n\to\infty} f_n(x) = f(x)$ uniformly on each B_k.

Exercise 2.4.3 Show that the conclusion in Example 2.4.1 still holds if (Ω, Σ, μ) is σ-finite.

Theorem 2.4.2 (Monotone convergence theorem) *Let $\{f_n\}$ be a monotone nondecreasing sequence of nonnegative measurable functions. Then*

$$\int_\Omega \lim_{n\to\infty} f_n d\mu = \lim_{n\to\infty} \int_\Omega f_n d\mu.$$

Proof Put $f = \lim_{n\to\infty} f_n$; then $f_n \leq f$ for all n. Since $\int_\Omega f_1 d\mu \leq \int_\Omega f_2 d\mu \leq \cdots \leq \int_\Omega f_n d\mu \leq \cdots \leq \int_\Omega f d\mu$, we have

$$\lim_{n\to\infty} \int_\Omega f_n d\mu \leq \int_\Omega f d\mu.$$

It remains to show that $\int_\Omega f d\mu \leq \lim_{n\to\infty} \int_\Omega f_n d\mu$. For this, it suffices to show that $\lim_{n\to\infty} \int_\Omega f_n d\mu \geq \lambda$ for each finite real number $\lambda < \int_\Omega f d\mu$. For such a λ, there is a simple function $g = \sum_{j=1}^{l} \alpha_i I_{A_j}$ such that $0 \leq g \leq f$ and $\int_\Omega g d\mu = \sum_{j=1}^{n} \alpha_j \, \mu\,(A_j) > \lambda$. In the above expression for g, we may assume that $\alpha_1, \ldots, \alpha_l$ are the different positive values taken by g, and hence $A_1, \ldots, A_l$ are disjoint sets in Σ. Then $\alpha_j \leq f$ on each A_j. Choose $\varepsilon > 0$ small enough so that $\alpha_j - \varepsilon > 0$, $j = 1, \ldots, l$. For each $j = 1, \ldots, l$ and positive integer n, let $A_j^{(n)} = \{x \in A_j : f_n(x) > \alpha_j - \varepsilon\}$ and define $g_n = \sum_{j=1}^{l} (\alpha_j - \varepsilon) I_{A_j^{(n)}}$; then $0 \leq g_n \leq f_n$ and hence

$$\lim_{n\to\infty} \int_\Omega f_n d\mu \geq \lim_{n\to\infty} \sum_{j=1}^{l} (\alpha_j - \varepsilon)\mu(A_j^{(n)}) = \sum_{j=1}^{l} (\alpha_j - \varepsilon)\mu(A_j),$$

because for each j, $A_j^{(n)}$ is a nondecreasing sequence with A_j as its limit. It follows then that $\lim_{n\to\infty} \int_\Omega f_n d\mu \geq \sum_{j=1}^{l} \alpha_j \mu(A_j)$ by letting $\varepsilon \searrow 0$. The proof is complete. ■

Exercise 2.4.4

(i) If f and g are nonnegative measurable functions and $\alpha, \beta \geq 0$, show that

$$\int_\Omega (\alpha f + \beta g) d\mu = \alpha \int_\Omega f d\mu + \beta \int_\Omega g d\mu.$$

(ii) Suppose that f is integrable and $\alpha \in \mathbb{R}$. Show that $\int_\Omega \alpha f d\mu = \alpha \int_\Omega f d\mu$.

2.5 Concepts related to sets of measure zero

We now make some remarks on concepts connected with measure zero sets (as previously, a measure space (Ω, Σ, μ) is considered and fixed). For this purpose, a subset A of Ω is called a **null** set (or more precisely $\boldsymbol{\mu}$**-null** set), if $A \subset B \in \Sigma$ and $\mu(B) = 0$.

Note that countable unions of null sets are null sets. Let $A = \{x \in \Omega : x$ does not have a property $P\}$, if A is a null set, we say that the property P holds **almost everywhere** on Ω (or simply P holds almost everywhere). For example, if outside a null set, f is finite, then we say that f is finite almost everywhere; also if $\lim_{n\to\infty} f_n(x) = f(x)$ exists for each x outside a null set, then we say that f_n converges almost everywhere. If a property P holds almost everywhere, we simply say that P holds a.e. (more precisely, μ-almost everywhere or μ-a.e. if other measures might also be in question). Two measurable functions f and g are said to be **equivalent** if $f = g$ a.e. Clearly, if f and g are equivalent and if the integral of one of them exists, then both of their integrals exist and are equal. If g is equivalent to f, g is sometimes referred to as a **version** of f.

As we shall see, functions which appear naturally are often not defined at every point of Ω. The most important case is when they are defined outside null sets. A function f is said to be defined a.e. on Ω if f is defined on $\Omega\backslash A$, with A being a null set; and f is measurable if f is measurable on $\Omega\backslash N$ for some measurable null set $N \supset A$, or, equivalently, if a new function $\hat{f}$ is defined by $\hat{f}(x) = f(x)$ for $x \in \Omega\backslash N$ and $\hat{f}(x) = 0$ for $x \in N$, then $\hat{f}$ is measurable. Hence, a measurable function f which is defined a.e. on Ω can be considered as defined on Ω if it is replaced by one of $\hat{f}$ defined above; this is legitimate because any pair of such functions $\hat{f}$ are equivalent measurable functions.

Exercise 2.5.1 Show that if f is measurable, then $\{f = +\infty\}$ and $\{f = -\infty\}$ are in Σ. Show also that if f is integrable, then f is finite a.e.

All the results we have established so far remain true if the pointwise conditions are replaced by conditions held almost everywhere. For example:

Theorem 2.5.1 (Egoroff theorem) *If a sequence $\{f_n\}$ of almost everywhere finite measurable functions converges a.e. to a finite function f on A, where $A \in \Sigma$, and $\mu(A) < +\infty$, then for every $\varepsilon > 0$, there is $B \in \Sigma$, $B \subset A$ such that $\mu(A \setminus B) < \varepsilon$ and $f_n \to f$ uniformly on B.*

Theorem 2.5.2 (Monotone convergence theorem) *Let $\{f_n\}$ be a sequence of measurable functions which are nonnegative and nondecreasing a.e., then*

$$\int_\Omega \lim_{n\to\infty} f_n d\mu = \lim_{n\to\infty} \int_\Omega f_n d\mu.$$

From Theorem 2.5.2 and Exercise 2.4.4 (i) there follows the following corollary.

Corollary 2.5.1 *If $\{f_n\}$ is a sequence of a.e. nonnegative measurable functions, then*

$$\int_\Omega \sum_{n=1}^{\infty} f_n d\mu = \sum_{n=1}^{\infty} \int_\Omega f_n d\mu.$$

Exercise 2.5.2 Let f be a measurable function. Prove the following statements:

(i) Suppose that $\int_\Omega f d\mu$ exists, i.e. $\int_\Omega f d\mu = \int_\Omega f^+ d\mu - \int_\Omega f^- d\mu$, where the right-hand side has a meaning. If $f = f_1 - f_2$ where f_1 and f_2 are nonnegative and measurable, then

$$\int_\Omega f d\mu = \int_\Omega f_1 d\mu - \int_\Omega f_2 d\mu,$$

if the right-hand side has a meaning.

(ii) $\int_\Omega f d\mu$ exists if and only if $f = f_1 - f_2$ for some nonnegative measurable functions f_1 and f_2, such that $\int_\Omega f_1 d\mu - \int_\Omega f_2 d\mu$ is meaningful. (Hint: for f_1 and f_2 as above, observe that $f^+ \leq f_1$ and $f^- \leq f_2$.)

(iii) If f, g are measurable functions such that $\int_\Omega f d\mu$, $\int_\Omega g d\mu$, $\int_\Omega f d\mu + \int_\Omega g d\mu$ are meaningful, then $f + g$ is defined a.e. and

$$\int_\Omega (f+g) d\mu = \int_\Omega f d\mu + \int_\Omega g d\mu.$$

In particular, this holds true if f and g are integrable.

Exercise 2.5.3 Show that Theorem 2.5.2 still holds if $\{f_n\}$ is a sequence of measurable functions bounded from below by an integrable function a.e. and is nondecreasing a.e. (Hint: show first that f_n^- is integrable and hence $\int_\Omega f_n d\mu$ is meaningful for each n.)

Exercise 2.5.4 If $f \geq 0$ a.e. and is measurable, then show that $\int_\Omega f d\mu = 0$ if and only if $f = 0$ a.e.

Exercise 2.5.5 (Beppo–Levi) Let $\{f_n\}$ be a monotone increasing sequence of integrable functions such that $\sup_n \int f_n d\mu < +\infty$. Let $f = \lim_{n\to\infty} f_n$. Show that $-\infty < f < +\infty$ a.e., f is integrable, and $\lim_{n\to\infty} \int |f_n - f| d\mu = 0$.

The following theorems follow from Exercise 2.5.2 (iii) and Exercise 2.4.4 (ii):

Theorem 2.5.3 *A measurable function f is integrable if and only if $|f|$ is integrable.*

Theorem 2.5.4 *Suppose that f and g are integrable and α, β are finite real numbers, then*

$$\int_\Omega (\alpha f + \beta g) d\mu = \alpha \int_\Omega f d\mu + \beta \int_\Omega g d\mu.$$

In particular, if $f \leq g$ a.e. then $\int_\Omega f d\mu \leq \int_\Omega g d\mu$.

Exercise 2.5.6 Suppose that (Ω, Σ, μ) is a finite measure space and f a measurable function on Ω. For $k \in \mathbb{N}$ let $\omega_k := \mu(\{|f| > k\})$. Show that f is integrable if and only if $\sum_{k=1}^\infty \omega_k < \infty$. (Hint: show that $\sum_{k=1}^\infty \omega_k \leq \int_\Omega |f| d\mu \leq \sum_{k=1}^\infty \omega_k + \mu(\Omega)$.)

Exercise 2.5.7 Suppose that f is a nonnegative measurable function. Let $\nu : \Sigma \to [0, +\infty]$ be defined by $\nu(A) = \int_A f d\mu := \int_\Omega f I_A d\mu$; show that (Ω, Σ, ν) is a measure space and if $g \geq 0$ is Σ-measurable, then $\int_\Omega g d\nu = \int_\Omega g f d\mu$ (this fact is usually expressed by $d\nu = fd\mu$). Show also that a measurable function g is ν-integrable if and only if gf is μ-integrable.

Exercise 2.5.8 Suppose that f is a nonnegative integrable function. Show that for every $\varepsilon > 0$, there is $A \in \Sigma$ with $\mu(A) < +\infty$, such that

$$\int_A fd\mu > \int_\Omega fd\mu - \varepsilon.$$

Exercise 2.5.9 Let (Ω, Σ, μ) be a measure space, and $\{A_k\}_{k=1}^\infty \subset \Sigma$.

(i) Show that if $\sum_{k=1}^\infty \mu(A_k) < \infty$, then $\mu\left(\limsup_{k\to\infty} A_k\right) = 0$.

(ii) Show that if f is integrable, then

$$\int_{\limsup\limits_{k\to\infty} A_k} fd\mu = \lim_{k\to\infty} \int_{\bigcup\limits_{j=k}^\infty A_j} fd\mu.$$

(iii) Let f be integrable and $\varepsilon > 0$. Show that there is $\delta > 0$ such that if $A \in \Sigma$ and $\mu(A) < \delta$, then $\int_A |f|d\mu < \varepsilon$.
(Hint: suppose the contrary. Then for each k, there is $A_k \in \Sigma$ such that $\mu(A_k) < \frac{1}{k^2}$ and $\int_{A_k} |f|d\mu \geq \varepsilon$. Then apply (i) and (ii).)

Exercise 2.5.10 Let (Ω, Σ, μ) be a measure space and f a measurable function on Ω. Define a σ-algebra $\sigma(f)$ on Ω by

$$\sigma(f) = \{f^{-1}B : B \in \overline{\mathcal{B}}\}.$$

(i) Suppose that $\int_\Omega fd\mu$ exists and $\int_A fd\mu = 0$ for all $A \in \sigma(f)$. Show that $f = 0$ a.e.

(ii) Suppose now that f is integrable and g is $\sigma(f)$-measurable on Ω such that

$$\int_A gd\mu = \int_A fd\mu$$

for all $A \in \sigma(f)$. Show that there is a null set N in $\sigma(f)$ such that $g = f$ on $\Omega\backslash N$.

2.6 Fatou lemma and Lebesgue dominated convergence theorem

It is indicated in Section 2.4 that the monotone convergence theorem reveals the distinguished feature of σ-additivity of measure through integrals. We now present two consequences of the monotone convergence theorem which manifest behaviors of integral under limit processes. These are the **Fatou lemma** and **Lebesgue dominated convergence theorem** (hereafter abbreviated as **LDCT**).

Theorem 2.6.1 (Fatou lemma) *Let $\{f_n\}$ be a sequence of extended real-valued measurable functions which is bounded from below by an integrable function. Then*

$$\int_\Omega \liminf_{n\to\infty} f_n d\mu \leq \liminf_{n\to\infty} \int_\Omega f_n d\mu.$$

Proof Let $g_n = \inf_{k\geq n} f_k$, then g_n is nondecreasing and is bounded from below by an integrable function. By the monotone convergence theorem (see Exercise 2.5.3),

$$\begin{aligned}\int_\Omega \liminf_{n\to\infty} f_n d\mu &= \int_\Omega \lim_{n\to\infty} g_n d\mu \\ &= \lim_{n\to\infty} \int_\Omega g_n d\mu \leq \liminf_{n\to\infty} \int_\Omega f_n d\mu.\end{aligned}$$

■

Exercise 2.6.1 Show that if $\{f_n\}$ is bounded from above by an integrable function, then

$$\int_\Omega \limsup_{n\to\infty} f_n d\mu \geq \limsup_{n\to\infty} \int_\Omega f_n d\mu.$$

Later, both Theorem 2.6.1 and the statement shown in Exercise 2.6.1 will be referred to as the Fatou lemma. One notes that Theorem 2.6.1 is equivalent to a particular case of itself, with $\{f_n\}$ being a sequence of nonnegative measurable functions. This particular case is the original form of the Fatou lemma.

Theorem 2.6.2 (Lebesgue dominated convergence theorem (LDCT)) *If f_n, $n = 1, 2, \ldots$ and f are measurable functions and $f_n \to f$ a.e. Suppose further that $|f_n| \leq g$ a.e. for all n with g being an integrable function. Then*

$$\int_\Omega f d\mu = \lim_{n\to\infty} \int_\Omega f_n d\mu.$$

Proof $\{f_n\}$ is bounded from below and from above by integrable functions. Hence, by the Fatou lemma,

$$\limsup_{n\to\infty} \int_\Omega f_n d\mu \leq \int_\Omega \lim_{n\to\infty} f_n d\mu \leq \liminf_{n\to\infty} \int_\Omega f_n d\mu,$$

and consequently

$$\int_\Omega \lim_{n\to\infty} f_n d\mu = \lim_{n\to\infty} \int_\Omega f_n d\mu.$$

■

The Lebesgue dominated convergence theorem will henceforth be abbreviated as LDCT.

Exercise 2.6.2 Show that under the same conditions as in LDCT we have

$$\lim_{n\to\infty} \int_\Omega |f_n - f| d\mu = 0.$$

Example 2.6.1 Let $\{f_n\}$ be a sequence of nonnegative integrable functions such that $f_1(x) \geq \cdots \geq f_n(x) \geq f_{n+1}(x) \geq \cdots$ and $\lim_{n\to\infty} f_n(x) = 0$ for μ-a.e. x in Ω; then $\sum_{n=1}^\infty (-1)^{n+1} f_n$ is integrable and $\int_\Omega \sum_{n=1}^\infty (-1)^{n+1} f_n d\mu = \sum_{n=1}^\infty (-1)^{n+1} \int_\Omega f_n d\mu$. Note first, from the well-known alternating series's estimate $|\sum_{n=l}^{l+p} (-1)^{n+1} f_n(x)| \leq f_l(x)$ for μ-a.e. x and any l, p in $\mathbb{N}$, that $\sum_{n=1}^\infty (-1)^{n+1} f(x)$ converges for μ-a.e. x. Since $|\sum_{n=1}^k (-1)^{n+1} f_n(x)| \leq f_1(x)$ for μ-a.e. x and $k \in \mathbb{N}$, our assertion follows from LDCT.

Exercise 2.6.3 Let $\{f_k\}$ and $\{g_k\}$ be sequences of integrable functions such that $|f_k| \leq g_k$ a.e. on Ω for each $k \in \mathbb{N}$. Suppose that $\{f_k\}$ and $\{g_k\}$ converge a.e. to f and g respectively, and that g is integrable and $\int_\Omega g d\mu = \lim_{k\to\infty} \int_\Omega g_k d\mu$. Show that f is integrable and $\int_\Omega f d\mu = \lim_{k\to\infty} \int_\Omega f_k d\mu$. (Hint: apply the Fatou lemma to the sequences $\{g_k + f_k\}$ and $\{g_k - f_k\}$.)

Exercise 2.6.4 Suppose that $\{f_n\}$ is a sequence of measurable functions on (Ω, Σ, μ). Show that if $\int_\Omega \sum_{n=1}^\infty |f_n| d\mu < \infty$, then $\sum_{n=1}^\infty f_n(x)$ converges and is finite for a.e. x, $\sum_{n=1}^\infty f_n$ is integrable, and

$$\int_\Omega \sum_{n=1}^\infty f_n d\mu = \sum_{n=1}^\infty \int_\Omega f_n d\mu.$$

Exercise 2.6.5 A family $\{f_\alpha\}$ of integrable functions on a finite measure space (Ω, Σ, μ) is called uniformly integrable if for any $\varepsilon > 0$, there is $\delta > 0$ such that if $A \subset \Sigma$ with $\mu(A) \leq \delta$, then $\int_A |f_\alpha| d\mu \leq \varepsilon$ for all α. Show that if $\{f_n\}$ is a uniformly integrable sequence of functions on Ω which converges a.e. to an integrable function f on Ω, then

$$\lim_{n\to\infty} \int_\Omega |f_n - f| d\mu = 0.$$

2.7 The space $L^p(\Omega, \Sigma, \mu)$

Associated with a measure space (Ω, Σ, μ) is a family $\{L^p(\Omega, \Sigma, \mu)\}_{p\geq 1}$ of Banach spaces which plays an important role in many fields of mathematics. The introduction and first properties of spaces $L^p(\Omega, \Sigma, \mu)$, $p \geq 1$, are our concern in this section. A more advanced account of these spaces will be given in Chapter 6, when Ω is an open set in $\mathbb{R}^n$.

For a measurable function f, let

$$\|f\|_p = \left(\int_\Omega |f|^p d\mu \right)^{1/p} \text{ if } 0 < p < +\infty;$$
$$\|f\|_\infty = \inf\{M \geq 0 : |f| \leq M \ a.e.\}.$$

$\|f\|_p$ is called the L^p-**norm** of f; $\|f\|_\infty$ is also called the **essential sup-norm** of f.

Exercise 2.7.1 Show that $|f| \le \|f\|_\infty$ a.e.

Recall that if $p, q \ge 1$ are such that $\frac{1}{p} + \frac{1}{q} = 1$, then they are called conjugate exponents.

Theorem 2.7.1 (Hölder's inequality) *If $p, q \ge 1$ are conjugate exponents, then*

$$\int_\Omega |fg| d\mu = \|fg\|_1 \le \|f\|_p \|g\|_q$$

for any measurable functions f and g.

Proof We may assume that $0 < \|f\|_p, \|g\|_q < +\infty$, hence $|f|, |g| < \infty$ a.e. We may further assume that $1 < p, q < \infty$. Now let $\zeta = \left(\frac{|f|}{\|f\|_p}\right)^p, \eta = \left(\frac{|g|}{\|g\|_q}\right)^q, \alpha = \frac{1}{p}$, and $\beta = \frac{1}{q}$ in Lemma 1.6.1; we have

$$\frac{|f|\,|g|}{\|f\|_p \|g\|_q} \le \frac{1}{p} \frac{|f|^p}{\|f\|_p^p} + \frac{1}{q} \frac{|g|^q}{\|g\|_q^q}$$

a.e. on Ω, from which on integrating both sides we complete the proof. ■

Exercise 2.7.2 Suppose that $1 < p, q < \infty$ are conjugate exponents and $\|f\|_p, \|g\|_q$ are both finite. Show that $\|fg\|_1 = \|f\|_p \|g\|_q$ if and only if either $\|f\|_p \|q\|_q = 0$ or $|g|^q = \lambda |f|^p$ a.e. for some $\lambda > 0$. (Hint: use Exercise 1.6.1.)

The following example is a variation of Hölder's inequality.

Example 2.7.1 Let p, q, and r be positive numbers satisfying $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$, and suppose that f and g are measurable functions. Since $1 = \frac{r}{p} + \frac{r}{q}$, $\frac{p}{r}$ and $\frac{q}{r}$ are conjugate exponents; then, $\int_\Omega |fg|^r d\mu = \int_\Omega (|f|^p)^{\frac{r}{p}} (|g|^q)^{\frac{r}{q}} d\mu \le (\int_\Omega |f|^p d\mu)^{\frac{r}{p}} (\int_\Omega |g|^q d\mu)^{\frac{r}{q}}$, by Hölder's inequality. Hence, $\|fg\|_r \le \|f\|_p \|g\|_q$. When $r = 1$, this is Hölder's inequality.

Theorem 2.7.2 (Minkowski's inequality) *Let f, g be measurable, $1 \le p \le +\infty$, then*

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

whenever $f + g$ is meaningful a.e. on Ω.

Proof This is obvious when $p = 1$ or $+\infty$. We now consider the case $1 < p < +\infty$, then

$$\begin{aligned}\|f+g\|_p^p &= \int_\Omega |f+g|^p d\mu = \int_\Omega |f+g|^{p-1}|f+g| d\mu \\ &\leq \int_\Omega |f+g|^{p-1}|f| d\mu + \int_\Omega |f+g|^{p-1}|g| d\mu \\ &\leq \left[\int_\Omega |f+g|^{(p-1)q} d\mu\right]^{1/q} \{\|f\|_p + \|g\|_p\} \\ &= \|f+g\|_p^{p/q}\{\|f\|_p + \|g\|_p\},\end{aligned}$$

by Hölder's inequality, where $\frac{1}{p} + \frac{1}{q} = 1$. The theorem follows by dividing extreme ends of the above sequence of inequalities by $\|f+g\|_p^{p-1}$, because we may assume that $0 < \|f+g\|_p < \infty$. ■

Exercise 2.7.3 Verify the last statement of the proof of Theorem 2.7.2. (Hint: show that if $\|f\|_p + \|g\|_p < +\infty$, then $\|f+g\|_p < +\infty$ by using Exercise 1.6.3.)

Exercise 2.7.4 Suppose $1 < p < \infty$ and both $\|f\|_p$ and $\|g\|_p$ are finite. Show that

$$\|f+g\|_p = \|f\|_p + \|g\|_p$$

if and only if either $\|f\|_p\|g\|_p = 0$ or $g = \lambda f$ a.e. for some $\lambda > 0$.

Let now $\mathcal{L}^p(\Omega, \Sigma, \mu)$ be the family of all measurable functions f with $\|f\|_p < +\infty$. From the Minkowski inequality, it is readily seen that $\mathcal{L}^p(\Omega, \Sigma, \mu)$ is a real vector space. If we let

$$\mathcal{N} = \{f \in \mathcal{L}^p(\Omega, \Sigma, \mu) : \|f\|_p = 0\},$$

then $f \in \mathcal{N}$ if and only if $f = 0$ a.e. on Ω. Now consider the space $L^p(\Omega, \Sigma, \mu) = \mathcal{L}^p(\Omega, \Sigma, \mu)/\mathcal{N}$; then $L^p(\Omega, \Sigma, \mu)$ is a vector space which consists of equivalence classes of $\mathcal{L}^p(\Omega, \Sigma, \mu)$ w.r.t. the equivalence relation $\sim$, defined by $f \sim g$ if and only if $f = g$ a.e. on Ω.

We shall allow ourselves the liberty of not distinguishing between a class of functions in $L^p(\Omega, \Sigma, \mu)$ and a function representing the class; hence, by $f \in L^p(\Omega, \Sigma, \mu)$ we shall mean that f is to be considered as a class of equivalent functions in $L^p(\Omega, \Sigma, \mu)$ as well as any function from that class.

For $f \in L^p(\Omega, \Sigma, \mu)$, let

$$\|f\|_p = \left(\int_\Omega |f|^p d\mu\right)^{1/p} \text{ if } 1 \leq p < +\infty,$$

and

$$\|f\|_\infty = \text{ essential sup-norm of } f.$$

Remember that in the definition above, f on the left-hand side is a class of function and f on the right-hand side is a function representing that class. We note that the above definition is well defined. $\|f\|_p$ is called the L^p-norm of f in $L^p(\Omega, \Sigma, \mu)$.

$L^p(\Omega, \Sigma, \mu)$ is called the L^p space of the measure space (Ω, Σ, μ), and is often more compactly denoted by $L^p(\Omega)$ or $L^p(\mu)$ when Σ and μ are assumed to be known, or when Ω and Σ are assumed to be known.

Example 2.7.2 One notes readily that the space $\ell^p(S)$ introduced in the remark at the end of Section 1.6 is the L^p space of the measure space with counting measure on S. It is easily verified that if S is infinite, then $\ell^p(S) \subsetneq \ell^q(S)$ if $1 \leq p < q$.

Exercise 2.7.5 Suppose that the measure space (Ω, Σ, μ) is finite and $f \in L^\infty(\Omega, \Sigma, \mu)$.

(i) Show that $(\frac{1}{\mu(\Omega)} \int_\Omega |f|^p d\mu)^{1/p} \leq (\frac{1}{\mu(\Omega)} \int_\Omega |f|^{p'} d\mu)^{1/p'}$, if $1 \leq p \leq p' < \infty$.

(ii) Show that $\lim_{p\to\infty} \left(\frac{1}{\mu(\Omega)} \int_\Omega |f|^p d\mu\right)^{1/p} = \|f\|_\infty$.

Exercise 2.7.6 Suppose that $\{f_k\}$ is a sequence in $L^p(\Omega, \Sigma, \mu)$ and that $\{f_k\}$ converges a.e. to $f \in L^p(\Omega, \Sigma, \mu)$ with $\|f\|_p = \lim_{k\to\infty} \|f_k\|_p$ $(1 \leq p < \infty)$. Show that $\{f_k\}$ converges in $L^p(\Omega, \Sigma, \mu)$ to f. (Hint: cf. Exercise 2.6.3 or observe that $2^{p-1}(|f|^p + |f_k|^p) - |f - f_k|^p \geq 0$.)

Theorem 2.7.3 *$L^p(\Omega, \Sigma, \mu)$ with norm $\|\cdot\|_p$ is a Banach space.*

Proof This is obvious when $p = +\infty$, if one notes that when $\{f_n\}$ is a Cauchy sequence in $L^\infty(\Omega, \Sigma, \mu)$, there is a measurable null set N such that $\sup_{\omega\in\Omega\setminus N} |f_n(\omega) - f_m(\omega)| \leq \|f_n - f_m\|_\infty$ for all n, m in $\mathbb{N}$.

Assume now that $1 \leq p < +\infty$ and let $\{f_n\}$ be a Cauchy sequence in $L^p(\Omega, \Sigma, \mu)$. There is an increasing sequence $\{n_k\}_{k=1}^\infty$ of positive integers such that $\|f_{n_{k+1}} - f_{n_k}\|_p \leq 2^{-k}$, $k = 1, 2, \ldots$. Put $g = \sum_{k=1}^\infty |f_{n_{k+1}} - f_{n_k}|$; monotone convergence theorem and Minkowski inequality imply

$$\begin{aligned}
\|g\|_p^p &= \int_\Omega \left(\sum_{k=1}^\infty |f_{n_{k+1}} - f_{n_k}|\right)^p d\mu = \int_\Omega \lim_{l\to\infty} \left(\sum_{k=1}^l |f_{n_{k+1}} - f_{n_k}|\right)^p d\mu \\
&= \lim_{l\to\infty} \int_\Omega \left(\sum_{k=1}^l |f_{n_{k+1}} - f_{n_k}|\right)^p d\mu = \lim_{l\to\infty} \left\|\sum_{k=1}^l |f_{n_{k+1}} - f_{n_k}|\right\|_p^p \\
&\leq \lim_{l\to\infty} \left(\sum_{k=1}^l \|f_{n_{k+1}} - f_{n_k}\|_p\right)^p = \left(\sum_{k=1}^\infty \|f_{n_{k+1}} - f_{n_k}\|_p\right)^p \leq 1,
\end{aligned}$$

hence, $g \in L^p(\Omega, \Sigma, \mu)$. Observe that if $g(x) < \infty$, then $\sum_{j=1}^{\infty} |f_{n_{j+1}}(x) - f_{n_j}(x)| < \infty$ and for $k > l$, we have

$$|f_{n_k}(x) - f_{n_l}(x)| = \left|\sum_{j=l}^{k-1}(f_{n_{j+1}}(x) - f_{n_j}(x))\right| \le \sum_{j=l}^{k-1} |f_{n_{j+1}}(x) - f_{n_j}(x)| \to 0$$

as $l \to \infty$. This means that $\{f_{n_k}(x)\}$ is a Cauchy sequence in $\mathbb{R}$. Hence, $f_{n_k} \to f$ a.e. with f being finite a.e. But $|f_{n_k}| \le |f_{n_1}| + g$, $k = 1, 2, \ldots$, implies that $f \in L^p(\Omega, \Sigma, \mu)$. Now $|f_{n_k} - f|^p \le (|f| + |f_{n_1}| + g)^p$ a.e.; thus by LDCT we know that $\|f_{n_k} - f\|_p \to 0$ as $k \to \infty$; this fact, together with $\{f_n\}$ being a Cauchy sequence, implies that $\| f_n - f\|_p \to 0$ as $n \to \infty$. Hence $L^p(\Omega, \Sigma, \mu)$ is complete. ■

Exercise 2.7.7 Suppose that $\{f_k\}$ is a sequence in $L^p(\Omega, \Sigma, \mu)$, $1 \le p < \infty$ such that $|f_k| \le g$ a.e. for each k for some $g \in L^p(\Omega, \Sigma, \mu)$. Assume that $\lim_{k\to\infty} f_k = f$ a.e. Show that $f \in L^p(\Omega, \Sigma, \mu)$ and $\lim_{k\to\infty} \|f_k - f\|_p = 0$.

Exercise 2.7.8 Let $f \in L^p(\Omega, \Sigma, \mu)$, $1 \le p < \infty$. Show that for any $\varepsilon > 0$, there is a bounded function g in $L^P(\Omega, \Sigma, \mu)$ such that $\|f - g\|_p < \varepsilon$. (Hint: choose g as a truncated function of f, i.e., for some $M > 0$, $g(x) = f(x)$ if $|f(x)| \le M$, and $g(x) = 0$ otherwise.)

Exercise 2.7.9 Suppose that $\{f_k\}$ is a sequence in $L^p(\Omega, \Sigma, \mu)$ with $\sum_{k=1}^{\infty} \|f_k\|_p < \infty$. Show that $\sum_{k=1}^{\infty} f_k$ converges and is finite a.e. on Ω and is in $L^p(\Omega, \Sigma, \mu)$ with $\| \sum_{k=1}^{\infty} f_k\|_p \le \sum_{k=1}^{\infty} \|f_k\|_p$.

Exercise 2.7.10 Suppose that $\{f_n\}$ is a convergent sequence in $L^p(\Omega, \Sigma, \mu)$, $p \ge 1$. Show that $\{f_n\}$ has a subsequence which converges a.e. on Ω. (Hint: there is a subsequence $\{f_{n_k}\}$ of $\{f_n\}$ such that $\sum_{k=2}^{\infty} \|f_{n_k} - f_{n_{k-1}}\|_p < \infty$.)

Exercise 2.7.11 If $\mu(\Omega) < \infty$, show that $L^q(\Omega, \Sigma, \mu) \subset L^p(\Omega, \Sigma, \mu)$ for $1 \le p < q$. Show also that for $f \in L^p(\Omega, \Sigma, \mu)$, $\|f\|_p \le \|f\|_q \mu(\Omega)^{\frac{1}{p} - \frac{1}{q}}$ for $q \ge p$.

Exercise 2.7.12 Suppose that $1 \le p < r$. Show that for any q strictly between p and r, $L^q(\Omega, \Sigma, \mu) \subset L^p(\Omega, \Sigma, \mu) + L^r(\Omega, \Sigma, \mu)$.

2.8 Miscellaneous remarks

Some remarks complementing discussions presented so far in this chapter are now in order.

2.8.1 Restriction of measure spaces

If Σ is a σ-algebra on Ω and $A \in \Sigma$, then the family $\Sigma|A := \{B \cap A : B \in \Sigma\}$ is a σ-algebra on A, called the **restriction** of Σ to A. If, further, (Ω, Σ, μ) is a measure space, the measure space $(A, \Sigma|A, \mu)$ is called the restriction to A of the original one. Since

$\Sigma|A \subset \Sigma$, μ is defined on $\Sigma|A$, and hence $(A, \Sigma|A, \mu)$ is indeed a measure space with μ being understood to be restricted to $\Sigma|A$. Suppose now f is a Σ-measurable function on Ω, $f|_A$ is then clearly a $\Sigma|A$-measurable function on A, and if $\int_\Omega f d\mu$ exists, so does $\int_A f|_A d\mu$, and $\int_A f|_A d\mu$ is obviously the same as $\int_A f d\mu := \int_\Omega f I_A d\mu$ (cf. Exercise 2.5.7). But it might happen that $\int_A f|_A d\mu$ exists without $\int_\Omega f d\mu$ being defined, suggesting that it is convenient sometimes to consider $(A, \Sigma|A, \mu)$ instead of (Ω, Σ, μ); when this happens, it will be clear from the context and one does not revert to the formal procedure described previously.

2.8.2 Measurable maps

Suppose (Ω, Σ) and $(\widehat{\Omega}, \widehat{\Sigma})$ are measurable spaces. We say that a map T from Ω into $\widehat{\Omega}$ is **measurable** (more precisely, **$\Sigma|\widehat{\Sigma}$-measurable**) if $T^{-1}A \in \Sigma$ for every $A \in \widehat{\Sigma}$. In particular, if $\widehat{\Omega} = \overline{\mathbb{R}}$ and $\widehat{\Sigma} = \overline{\mathcal{B}}$, then T is what we call a measurable function on Ω. If (Ω, Σ, μ) and $(\widehat{\Omega}, \widehat{\Sigma}, \hat{\mu})$ are measure spaces, then a measurable map T from Ω into $\widehat{\Omega}$ is **measure preserving** if $\mu(T^{-1}A) = \hat{\mu}(A)$ for every $A \in \widehat{\Sigma}$. Now, if f is a measurable function on $\widehat{\Omega}$ and T is a measure-preserving map from Ω into $\widehat{\Omega}$, then $f \circ T$ is measurable on Ω; furthermore, $\int_{\widehat{\Omega}} f d\hat{\mu}$ exists if and only if $\int_\Omega f \circ T d\mu$ exists, and

$$\int_{\widehat{\Omega}} f d\hat{\mu} = \int_\Omega f \circ T d\mu$$

if either side exists. This is easily verified, if f is nonnegative; in the general case, one needs only to note that $f \circ T = f^+ \circ T - f^- \circ T$.

We note at this point that if the measurable space structure of (Ω, Σ) and $(\widehat{\Omega}, \widehat{\Sigma})$ is to be emphasized, a map $T : \Omega \to \widehat{\Omega}$ will also be called, by abuse of language, a map from (Ω, Σ) to $(\widehat{\Omega}, \widehat{\Sigma})$, and a measurable map from (Ω, Σ) to $(\widehat{\Omega}, \widehat{\Sigma})$ means a $\Sigma|\widehat{\Sigma}$-measurable map from Ω to $\widehat{\Omega}$.

It is readily verified that if (Ω_i, Σ_i) is a measurable space for $i = 1, 2, 3$ and T_i is a measurable map from (Ω_i, Σ_i) to $(\Omega_{i+1}, \Sigma_{i+1})$ for $i = 1, 2$, then $T_2 \circ T_1$ is a measurable map from (Ω_1, Σ_1) to (Ω_3, Σ_3); in particular, if f is a measurable function on (Ω, Σ) and g a Borel function on $\overline{\mathbb{R}}$, then $g \circ f$ is a measurable function on (Ω, Σ). In words, this means that a Borel function of a measurable function is measurable; however, we shall see in Example 4.7.2 that a measurable function of a continuous function may not be measurable.

2.8.3 Complete measure spaces

A measure space (Ω, Σ, μ) is **complete** if every null set is in Σ. One can construct a complete measure space $(\Omega, \overline{\Sigma}, \bar{\mu})$ from a measure space (Ω, Σ, μ) in the following way. Let $\overline{\Sigma} = \{B \subset \Omega : \exists C, D \text{ in } \Sigma \text{ such that } C \subset B \subset D \text{ and } \mu(D \backslash C) = 0\}$. It is clear that $\overline{\Sigma}$ is a σ-algebra on Ω. Now define a set function $\bar{\mu}$ on $\overline{\Sigma}$ by

$$\bar{\mu}(B) = \mu(C), \tag{2.4}$$

if $C \subset B \subset D$, where C and D are in Σ with $\mu(D\backslash C) = 0$. We claim that (2.4) is well defined; this amounts to showing that if $\widehat{C}$, $\widehat{D}$ are in Σ such that $\widehat{C} \subset B \subset \widehat{D}$ and $\mu(\widehat{D}\backslash\widehat{C}) = 0$, then $\mu(\widehat{C}) = \mu(C)$. Now from $C \cup \widehat{C} \subset B \subset D$ and $\mu(D\backslash[C \cup \widehat{C}]) \leq \mu(D\backslash C) = 0$, we infer that $\mu(C \cup \widehat{C}) = \mu(D) = \mu(C)$. Similarly, $\mu(C \cup \widehat{C}) = \mu(\widehat{C})$; hence $\mu(\widehat{C}) = \mu(C)$ as claimed. $\bar{\mu}$ is obviously a measure on $\overline{\Sigma}$. Suppose $B \in \overline{\Sigma}$ with $\bar{\mu}(B) = 0$ and consider $S \subset B$. There are C and D in Σ such that $C \subset B \subset D$, $\mu(D\backslash C) = 0$, and $\mu(C) = 0$. Observe that $\mu(D) = 0$. Since $\emptyset \subset S \subset D$ and $\mu(D\backslash\emptyset) = \mu(D) = 0$, $S \in \overline{\Sigma}$. This means that $(\Omega, \overline{\Sigma}, \bar{\mu})$ is complete. When (Ω, Σ, μ) is complete, one sees readily that $(\Omega, \overline{\Sigma}, \bar{\mu})$ is the same as (Ω, Σ, μ). The measure space $(\Omega, \overline{\Sigma}, \bar{\mu})$ is called the **completion** of (Ω, Σ, μ). Clearly, $\Sigma \subset \overline{\Sigma}$ and $\bar{\mu}$ is an extension of μ.

Exercise 2.8.1 Show that if f is $\overline{\Sigma}$-measurable, then there is a Σ-measurable function $\hat{f}$ such that $f = \hat{f}$ $\bar{\mu}$-a.e. and that f is $\bar{\mu}$-integrable if and only if $\hat{f}$ is μ-integrable.

2.8.4 Integral of complex-valued functions

So far only real-valued functions are considered in regard to measurability and integration; now a brief account will be given for complex-valued functions.

A complex-valued function f defined on a set Ω can be expressed as

$$f = f_1 + if_2,$$

where f_1 and f_2 are finite real-valued functions defined by

$$\begin{aligned} f_1(\omega) &= \text{real part of } f(\omega); \\ f_2(\omega) &= \text{imaginary part of } f(\omega), \end{aligned}$$

for $\omega \in \Omega$. Usually f_1 and f_2 are denoted respectively by $\operatorname{Re} f$ and $\operatorname{Im} f$. If now (Ω, Σ, μ) is a measure space, f is said to be measurable (more precisely, Σ-measurable), if both $\operatorname{Re} f$ and $\operatorname{Im} f$ are measurable.

Exercise 2.8.2 Show that a complex-valued function f defined on Ω is measurable if and only if it is $\Sigma|\mathcal{B}(\mathbb{C})$-measurable; where $\mathcal{B}(\mathbb{C})$ is the σ-algebra generated by the family of all open subsets of the complex field $\mathbb{C}$.

If both $\operatorname{Re} f$ and $\operatorname{Im} f$ are integrable, f is said to be integrable and the integral $\int_\Omega f d\mu$ of f is defined as $\int_\Omega \operatorname{Re} f d\mu + i \int_\Omega \operatorname{Im} f d\mu$. Obviously, f is integrable if and only if $|f|$ is integrable, where $|f|$ is the function defined by $|f|(\omega) = |f(\omega)| = \{\operatorname{Re} f(\omega)^2 + \operatorname{Im} f(\omega)^2\}^{\frac{1}{2}}$ for $\omega \in \Omega$. One verifies easily that $|\int_\Omega f d\mu| \leq \int_\Omega |f| d\mu$, if f is integrable, and that if f and g are integrable, then $\alpha f + \beta g$ are integrable and $\int_\Omega (\alpha f + \beta g) d\mu = \alpha \int_\Omega f d\mu + \beta \int_\Omega g d\mu$ for any complex numbers α and β. For a complex-valued measurable function f, its L^p-norm $\|f\|_p$, $p \geq 1$, is defined by

$$\|f\|_p = \| |f| \|_p.$$

Then Hölder inequality holds for complex-valued measurable functions, i.e.

$$\int_{\Omega} |fg| d\mu \leq \|f\|_p \cdot \|g\|_q,$$

where p, q are conjugate exponents; in particular,

$$\left| \int_{\Omega} fg d\mu \right| \leq \|f\|_p \cdot \|g\|_q$$

if fg is integrable. What also holds true is the Minkowski inequality,

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p,$$

as can easily be verified. It is to be noted that since f and g are complex-valued, $f + g$ is defined on Ω.

Now consider the space $\mathcal{L}^p(\Omega, \Sigma, \mu)$ of all complex-valued measurable functions f such that $\|f\|_p < \infty$. It follows from the Minkowski inequality that $\mathcal{L}^p(\Omega, \Sigma, \mu)$ is a complex vector space. As in Section 2.7, if we let $L^p(\Omega, \Sigma, \mu)$ be the quotient space $\mathcal{L}^p(\Omega, \Sigma, \mu)/\mathcal{N}$, where $\mathcal{N}$ is the vector subspace of $\mathcal{L}^p(\Omega, \Sigma, \mu)$ consisting of all those functions which are zero-valued almost everywhere. For $[f] = f + \mathcal{N}, f \in \mathcal{L}^p(\Omega, \Sigma, \mu)$, let $\|[f]\|_p = \|f\|_p$, then $\|[f]\|_p$ is well defined and $L^p(\Omega, \Sigma, \mu)$ is a complex Banach space with this norm. As before, for $f \in \mathcal{L}^p(\Omega, \Sigma, \mu)$, $[f]$ will also be denoted by f, and $\|[f]\|_p$ by $\|f\|_p$; thus f may denote an element either of $\mathcal{L}^p(\Omega, \Sigma, \mu)$ or of $L^p(\Omega, \Sigma, \mu)$ as occasion prompts, and no confusion is possible.

Henceforth, $L^p(\Omega, \Sigma, \mu)$ will denote a real or complex Banach space as the situation suggests.

3 Construction of Measures

Measure spaces provide a framework for classifying functions and for construction of certain spaces of functions which prove to be useful in various disciplines of mathematics; but appropriate measure spaces have to be available beforehand.

We therefore devote this early chapter to construction of measure spaces. A general method, the inception of which began with the introduction of the Lebesgue measure on $\mathbb{R}$ and Lebesgue measurable sets in $\mathbb{R}$ by **H. Lebesgue**, will be treated firstly. This is the method of outer measure. We shall follow the approach of **C. Carathéodory**, which defines measurable sets without introducing the concept of inner measure of Lebesgue. Construction of measure spaces from given ones by various operations will be considered in Chapter 4.

3.1 Outer measures

A nonnegative set function μ defined for all subsets A of a given set Ω is called an **outer measure on** Ω if it is monotone and σ-subadditive, i.e. (i) $\mu(\emptyset) = 0$; (ii) $0 \leq \mu(A) \leq \mu(B)$ if $A \subset B$; and (iii) $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu(A_n)$, where $\{A_n\}_{n=1}^{\infty}$ is any sequence of subsets of Ω. Recall that a set function is required to take zero as its value at $\emptyset$ if $\emptyset$ is in its domain of definition; (ii) is the condition of monotony; and condition (iii) is σ-subadditivity. A nonnegative set function τ is said to be $\boldsymbol{\sigma}$**-subadditive** if $\tau\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \tau(A_n)$ whenever $A_1, A_2, \ldots$ and $\bigcup_{n=1}^{\infty} A_n$ are in its domain of definition.

An outer measure μ on Ω is usually simply called a **measure on** $\boldsymbol{\Omega}$. Sometimes we also say that μ measures Ω. We emphasize that a measure on a set Ω and a measure on a σ-algebra on Ω are different objects; the former is an outer measure which is in general not σ-additive on 2^{Ω}.

Let μ be an outer measure on Ω. Following Carathéodory, we say that a subset A of Ω is μ-**measurable** if

$$\mu(B) = \mu(B \cap A) + \mu(B \cap A^c) \tag{3.1}$$

for all $B \subset \Omega$ i.e., if for any $C \subset A$ and $D \subset A^c$ we have

$$\mu(C \cup D) = \mu(C) + \mu(D).$$

Remark Since $\mu(B) \leq \mu(B \cap A) + \mu(B \cap A^c)$, (3.1) is equivalent to

$$\mu(B) \geq \mu(B \cap A) + \mu(B \cap A^c). \tag{3.2}$$

It is easily verified that Ω is μ-measurable and that if $\mu(A) = 0$, then A is μ-measurable.

Example 3.1.1 Let $\mu : 2^\Omega \mapsto [0, +\infty]$ be defined by

$$\begin{aligned} \mu(A) &= \text{cardinality of } A \text{ if } A \text{ is a finite set;} \\ &= \infty \text{ otherwise.} \end{aligned}$$

Obviously, μ is an outer measure on Ω (recall that μ is called the counting measure on Ω), and that every subset of Ω is μ-measurable. It happens that μ is a measure on 2^Ω.

Exercise 3.1.1 Let $S \subset 2^\Omega$ have the following properties:
(i) $\emptyset \in S$, (ii) if $A \in S$ and $B \subset A$, then $B \in S$, and (iii) if $\{A_n\}_{n=1}^{\infty} \subset S$, then $\bigcup_n A_n \in S$.
Define $\mu : 2^\Omega \mapsto [0, \infty]$ by

$$\mu(A) = \begin{cases} 0 & \text{if } A \in S \\ +\infty & \text{otherwise.} \end{cases}$$

Show that μ is an outer measure on Ω. What are the μ-measurable subsets of Ω? If now $\nu : 2^\Omega \to [0, 1]$ is defined by

$$\begin{aligned} \nu(A) &= 0 \text{ if } A \in S, \\ &= 1 \text{ otherwise,} \end{aligned}$$

then ν is an outer measure on Ω. What are the ν-measurable subsets of Ω?

Exercise 3.1.2 Let (Ω, Σ, μ) be a measure space and w a nonnegative measurable function. For $A \subset \Omega$, define $\mu_w(A) = \inf\{\int_B w d\mu : B \in \Sigma, A \subset B\}$. Show that μ_w measures Ω and every set in Σ is μ_w-measurable.

Suppose that μ is an outer measure on Ω and $A \subset \Omega$, then the **restriction** of μ to A denoted by $\mu \lfloor A$ is defined by

$$\mu \lfloor A(B) = \mu(A \cap B)$$

for $B \subset \Omega$.

Exercise 3.1.3 Let μ measure Ω.

(i) Show that $A \subset \Omega$ is μ-measurable if and only if A is $\mu\lfloor B$-measurable for every subset B of Ω.

(ii) Show that A is $\mu\lfloor A$-measurable as well as every μ-measurable set.

Exercise 3.1.4 Suppose that μ measures Ω and that A is a μ-measurable subset of Ω. Show that for any $B \subset \Omega$, $\mu(A) + \mu(B) = \mu(A \cup B) + \mu(A \cap B)$. (Hint: evaluate $\mu(B)$ and $\mu(A \cup B)$ by using the definition of μ-measurability for A.)

Exercise 3.1.5 Let μ be an outer measure on Ω. For $A \subset \Omega$, define

$$\mu_e(A) := \inf\{\mu(B) : A \subset B,\ B \text{ is } \mu\text{-measurable}\};$$
$$\mu_i(A) := \sup\{\mu(B) : B \subset A,\ B \text{ is } \mu\text{-measurable}\}.$$

Show that if $\mu(A) < \infty$, then A is μ-measurable if and only if $\mu_e(A) = \mu_i(A)$.

3.2 Lebesgue outer measure on $\mathbb{R}$

We construct in this section the Lebesgue outer measure on $\mathbb{R}$. This measure opens the way for the development of modern theory of measure and integration.

For an open finite interval $I = (a, b)$, let $|I| = b - a$ be the length of I. If A is a subset of $\mathbb{R}$, we denote by $\Lambda(A)$ the set of all numbers of the form $\sum_{n=1}^{\infty} |I_n|$, where $\{I_n\}$ is a sequence of open finite intervals such that $\bigcup_n I_n \supset A$, and let

$$\lambda(A) = \inf \Lambda(A).$$

Theorem 3.2.1 *The set function λ is an outer measure on $\mathbb{R}$.*

Proof Let $\varepsilon > 0$, and for each n let I_n be an open interval of length $\varepsilon/2^n$; then since $\bigcup I_n \supset \phi$, we have

$$\lambda(\phi) \le \sum_{n=1}^{\infty} |I_n| = \varepsilon \sum_{n=1}^{\infty} 2^{-n} = \varepsilon;$$

thus $\lambda(\phi) = 0$. If $A \subset B$, then $\Lambda(A) \supset \Lambda(B)$, and hence $\lambda(A) \le \lambda(B)$. It remains to show that if $\{A_k\}$ is a sequence of subsets of $\mathbb{R}$, then

$$\lambda\left(\bigcup_k A_k\right) \le \sum_{k=1}^{\infty} \lambda(A_k).$$

For this purpose, we may obviously assume that $\lambda(A_k) < \infty$ for all k. Now let $\varepsilon > 0$ be given; for each k there is $\lambda_k \in \Lambda(A_k)$ such that

$$\lambda(A_k) \le \lambda_k < \lambda(A_k) + \frac{\varepsilon}{2^k}.$$

Let $\lambda_k = \sum_{n=1}^{\infty} |I_n^{(k)}|$, where $\{I_n^{(k)}\}_n$ is a sequence of open intervals such that $\bigcup_{n=1}^{\infty} I_n^{(k)} \supset A_k$. Then, $\bigcup_{n,k=1}^{\infty} I_n^{(k)} \supset \bigcup_{k=1}^{\infty} A_k$, hence (cf. Section 1.2),

$$\begin{aligned}\sum_{n,k}\left|I_n^{(k)}\right| &= \sum_{k=1}^{\infty}\sum_{n=1}^{\infty}\left|I_n^{(k)}\right| = \sum_{k=1}^{\infty}\lambda_k < \sum_{k=1}^{\infty}\left\{\lambda(A_k) + \frac{\varepsilon}{2^k}\right\} \\ &= \sum_{k=1}^{\infty}\lambda(A_k) + \varepsilon;\end{aligned}$$

but since $\sum_{n,k} |I_n^{(k)}| \in \Lambda(\bigcup_{k=1}^{\infty} A_k)$, $\lambda(\bigcup_{k=1}^{\infty} A_k) \leq \sum_{n,k}\left|I_n^{(k)}\right| < \sum_{k=1}^{\infty}\lambda(A_k) + \varepsilon$. Now let ε decrease to zero; we obtain

$$\lambda\left(\bigcup_{k=1}^{\infty} A_k\right) \leq \sum_{k=1}^{\infty}\lambda(A_k).$$

This proves that λ is an outer measure on $\mathbb{R}$. ■

The measure λ is called the **Lebesgue** measure on $\mathbb{R}$. We shall show later that λ admits a fairly large class of λ-measurable sets, but, for the moment, we content ourselves by showing that every finite open interval I is λ-measurable and $\lambda(I) = |I|$. For this purpose, we prove first a lemma which foresees the method of Carathéodory outer measures, to be introduced in Section 3.5.

Lemma 3.2.1 *For each $\varepsilon > 0$, and $A \subset \mathbb{R}$, let $\Lambda_\varepsilon(A)$ be the set of all numbers of the form $\sum_{n=1}^{\infty} |I_n|$, where $\{I_n\}$ is a sequence of open intervals such that $A \subset \bigcup_n I_n$ and $|I_n| < \varepsilon$ for each n. Then $\lambda(A) = \inf \Lambda_\varepsilon(A)$.*

Proof Since $\Lambda_\varepsilon(A) \subset \Lambda(A)$, $\lambda(A) \leq \inf \Lambda_\varepsilon(A)$. Observe now for any finite interval I and $\delta > 0$, there are a finite number $I^{(1)}, \ldots, I^{(k)}$ of open intervals such that $I \subset \bigcup_{j=1}^{k} I^{(j)}$, $|I^{(j)}| < \varepsilon$, $j = 1, \ldots, k$, and $\sum_{j=1}^{k} |I^{(j)}| < |I| + \delta$. Suppose that $\{I_n\}$ is a sequence of open intervals such that $\bigcup I_n \supset A$; then for any $\delta > 0$ and each n, let $I_n^{(1)}, \ldots, I_n^{(k_n)}$ be open intervals such that $|I_n^{(j)}| < \varepsilon$, $j = 1, \ldots, k_n$, $I_n \subset \bigcup_{j=1}^{k_n} I_n^{(j)}$, and $\sum_{j=1}^{k_n} |I_n^{(j)}| < |I_n| + \delta/2^n$. Obviously, $\alpha = \sum_{n=1}^{\infty}\sum_{j=1}^{k_n} |I_n^{(j)}|$ is in $\Lambda_\varepsilon(A)$ and $\alpha < \sum_{n=1}^{\infty} |I_n| + \delta$. We have shown that given $\delta > 0$, for $\beta \in \Lambda(A)$ there is $\alpha \in \Lambda_\varepsilon(A)$ such that $\alpha < \beta + \delta$. This, means that $\inf \Lambda_\varepsilon(A) \leq \lambda(A) + \delta$; let $\delta \searrow 0$, we have $\inf \Lambda_\varepsilon(A) \leq \lambda(A)$. Hence, $\lambda(A) = \inf \Lambda_\varepsilon(A)$. ■

Proposition 3.2.1 *Every finite open interval I is λ-measurable and $\lambda(I) = |I|$.*

Proof Let $I = (a, b)$ and, for $0 < \varepsilon < \frac{1}{2}(b - a)$, let $J = (a + \varepsilon, b - \varepsilon)$. For a subset A of $\mathbb{R}$, consider any sequence $\{I_n\}$ of open intervals with $|I_n| < \varepsilon$ for all n and $A \subset \bigcup_{n=1}^{\infty} I_n$. Let $\vartheta_1 = \{n : I_n \cap J \neq \phi\}$ and $\vartheta_2 = \{n : I_n \cap (A \cap I^c) \neq \phi\}$, then $\vartheta_1 \cap \vartheta_2 = \phi$ and

$$\sum_{n=1}^{\infty}|I_n| \geq \sum_{n\in\vartheta_1}|I_n| + \sum_{n\in\vartheta_2}|I_n| \geq \lambda(A \cap J) + \lambda(A \cap I^c),$$

from which it follows, by Lemma 3.2.1, that

$$\lambda(A) \geq \lambda(A \cap J) + \lambda(A \cap I^c).$$

But it is clear that

$$\lambda(A \cap I) \leq \lambda(A \cap J) + 2\varepsilon,$$

hence,

$$\lambda(A) \geq \lambda(A \cap I) + \lambda(A \cap I^c) - 2\varepsilon.$$

Let $\varepsilon \searrow 0$; we have

$$\lambda(A) \geq \lambda(A \cap I) + \lambda(A \cap I^c).$$

Therefore I is λ-measurable.

To show that $\lambda(I) = |I|$, we observe first that $\lambda(I) \leq |I|$. It remains to show that $\lambda(I) \geq |I|$. For this purpose, we claim first that if $I_1, \ldots, I_k$ are finite open intervals such that $\bigcup_{j=1}^k I_j \supset J$, where J is a closed interval, then $\sum_{j=1}^k |I_j| \geq |J|$. This claim follows by induction on k: if $k = 1$, this claim obviously holds; suppose that the claim holds for $k - 1$ and assume as we may that I_k contains the right endpoint of J, then $\bigcup_{j=1}^{k-1} I_j \supset J \backslash I_k$ and hence by our induction hypotheses,

$$\sum_{j=1}^{k-1} |I_j| \geq |J \backslash I_k|,$$

thus,

$$|J| \leq |J \backslash I_k| + |I_k| \leq \sum_{j=1}^{k} |I_j|.$$

Let now $\{I_n\}$ be any sequence of finite open intervals with $I \subset \bigcup_{n=1}^\infty I_n$. Consider any closed interval J in I. Since J is compact, there is $k \in \mathbb{N}$ such that $\bigcup_{j=1}^k I_j \supset J$. From the claim just established, we have

$$|J| \leq \sum_{j=1}^{k} |I_j| \leq \sum_{j=1}^{\infty} |I_j|,$$

hence, $|J| \leq \inf \Lambda(I) = \lambda(I)$. Since $|J|$ can be chosen as close to $|I|$ as one wishes, $|I| \leq \lambda(I)$. This proves the proposition. ∎

Exercise 3.2.1 Show that any finite closed interval J is λ-measurable and $\lambda(J) = |J|$. (Hint: $\lambda(\{x\}) = 0$ for $x \in \mathbb{R}$.)

Exercise 3.2.2 Show that sets of the form (a, ∞) or $(-\infty, a)$ are λ-measurable.

Exercise 3.2.3 Let $A \subset \mathbb{R}$. Show that there is a sequence $\{G_n\}$ of open sets containing A such that $\lambda(A) = \lambda(\bigcap_{n=1}^{\infty} G_n)$.

3.3 Σ-algebra of measurable sets

Suppose that μ is an outer measure on Ω in this section. We reiterate that an outer measure on a set is also simply called a measure on the set.

Proposition 3.3.1 *If A is μ-measurable, then so is $\Omega \setminus A = A^c$.*

Proof Obvious. ■

Proposition 3.3.2 *If A_1, A_2 are μ-measurable, then so is $A_1 \cup A_2$.*

Proof Let $B \subset \Omega$, then

$$\begin{aligned}\mu(B) &= \mu(B \cap A_1) + \mu(B \cap A_1^c)\\ &= \mu(B \cap A_1) + \mu((B \cap A_1^c) \cap A_2) + \mu((B \cap A_1^c) \cap A_2^c)\\ &\geq \mu(B \cap (A_1 \cup A_2)) + \mu(B \cap (A_1 \cup A_2)^c),\end{aligned}$$

because $B \cap (A_1 \cup A_2) = (B \cap A_1) \cup (B \cap A_2) = (B \cap A_1) \cup (B \cap A_1^c \cap A_2)$. ■

Remark By induction, the union of finitely many μ-measurable sets is μ-measurable. This fact, together with Proposition 3.3.1, implies that the intersection of finitely many μ-measurable sets is μ-measurable.

Proposition 3.3.3 *If $\{A_j\}_{j=1}^{\infty}$ is a disjoint sequence of μ-measurable sets in Ω and $B \subset \Omega$, then*

$$\mu\left(B \cap \left\{\bigcup_{j=1}^{\infty} A_j\right\}\right) = \sum_{j=1}^{\infty} \mu(B \cap A_j).$$

Proof Let n be a positive integer, then, since $\bigcup_{j=1}^{n-1} A_j$ is μ-measurable, we have

$$\begin{aligned}\mu\left(B \cap \left\{\bigcup_{j=1}^{n} A_j\right\}\right) &= \mu\left(B \cap \left\{\bigcup_{j=1}^{n} A_j\right\} \cap \left\{\bigcup_{j=1}^{n-1} A_j\right\}\right) + \mu\left(B \cap \left\{\bigcup_{j=1}^{n} A_j\right\} \cap \left\{\bigcup_{j=1}^{n-1} A_j\right\}^c\right)\\ &= \mu\left(B \cap \left\{\bigcup_{j=1}^{n-1} A_j\right\}\right) + \mu(B \cap A_n) = \cdots = \sum_{j=1}^{n} \mu(B \cap A_j);\end{aligned}$$

then,

$$\mu\left(B \cap \left\{\bigcup_{j=1}^{\infty} A_j\right\}\right) \geq \mu\left(B \cap \left\{\bigcup_{j=1}^{n} A_j\right\}\right) = \sum_{j=1}^{n} \mu(B \cap A_j)$$

for all n, hence

$$\mu\left(B \cap \left\{\bigcup_{j=1}^{\infty} A_j\right\}\right) \geq \sum_{j=1}^{\infty} \mu(B \cap A_j).$$

But $\mu(B \cap \{\bigcup_{j=1}^{\infty} A_j\}) = \mu(\bigcup_{j=1}^{\infty} B \cap A_j) \leq \sum_{j=1}^{\infty} \mu(B \cap A_j)$, by σ-subadditivity of outer measures. ■

Proposition 3.3.4 *If $\{A_j\}_{j=1}^{\infty}$ is a disjoint sequence of μ-measurable sets, then $\bigcup_{j=1}^{\infty} A_j$ is μ-measurable.*

Proof Let $B \subset \Omega$, then

$$\begin{aligned}
&\mu\left(B \cap \bigcup_{j=1}^{\infty} A_j\right) + \mu\left(B \cap \left\{\bigcup_{j=1}^{\infty} A_j\right\}^c\right) \\
&\leq \sum_{j=1}^{n} \mu(B \cap A_j) + \mu\left(B \cap \left\{\bigcup_{j=1}^{n} A_j\right\}^c\right) + \sum_{j=n+1}^{\infty} \mu(B \cap A_j) \\
&= \mu\left(B \cap \left\{\bigcup_{j=1}^{n} A_j\right\}\right) + \mu\left(B \cap \left\{\bigcup_{j=1}^{n} A_j\right\}^c\right) + \sum_{j=n+1}^{\infty} \mu(B \cap A_j) \\
&= \mu(B) + \sum_{j=n+1}^{\infty} \mu(B \cap A_j).
\end{aligned}$$

If $\sum_{j=1}^{\infty} \mu(B \cap A_j) < \infty$, by letting $n \to \infty$ in the above inequality, we have

$$\mu\left(B \cap \bigcup_{j=1}^{\infty} A_j\right) + \mu\left(B \cap \left\{\bigcup_{j=1}^{\infty} A_i\right\}^c\right) \leq \mu(B); \tag{3.3}$$

while if $\sum_{j=1}^{\infty} \mu(B \cap A_j) = \infty$, then $\mu(B) \geq \mu\left(B \cap \bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mu(B \cap A_j) = \infty$; hence (3.3) also holds. ■

If we denote by Σ^{μ} the family of all μ-measurable sets, it follows from Propositions 3.3.1, 3.3.2, and 3.3.4 that Σ^{μ} is both a π-system and a λ-system and is therefore a σ-algebra; while μ is σ-additive on Σ^{μ} by Proposition 3.3.3. Since Σ^{μ} contains all those subsets A of Ω such that $\mu(A) = 0$, we have shown the following theorem.

Theorem 3.3.1 *Σ^{μ} is a σ-algebra and $(\Omega, \Sigma^{\mu}, \mu)$ is a complete measure space.*

For later reference, $(\Omega, \Sigma^{\mu}, \mu)$ is called the **measure space** for μ; and Σ^{μ}-measurable functions are sometimes said to be μ-measurable.

We have pointed out in Section 2.4 that the monotone limit property for increasing measurable sets, as stated in Lemma 2.4.1, reveals in a simple way the salient role played by σ-additivity of measures in the theory of measure and integration. Some outer measures possess the monotone limit property for increasing sets without requiring them to be measurable; regular measures are among them. A measure μ on Ω is said to be ***regular*** if for each $B \subset \Omega$, there is a μ-measurable set $A \supset B$ such that $\mu(A) = \mu(B)$; more generally, if Σ is a subσ-algebra of Σ^μ, we say that μ is Σ-**regular** if for each $B \subset \Omega$, there is $A \in \Sigma$ such that $A \supset B$ and $\mu(A) = \mu(B)$.

Theorem 3.3.2 *If $A_1 \subset A_2 \subset \cdots \subset \cdots$ is a sequence of sets in Ω and μ is a regular measure on Ω, then*

$$\mu\left(\bigcup_j A_j\right) = \lim_{n\to\infty} \mu(A_n).$$

Proof We always have

$$\mu\left(\bigcup_j A_j\right) \geq \lim_{n\to\infty} \mu(A_n). \tag{3.4}$$

For each j, let B_j be a μ-measurable set such that $A_j \subset B_j$ and $\mu(A_j) = \mu(B_j)$. Now let $C_j = \bigcap_{n\geq j} B_n$, then $C_j \supset A_j$ and $\mu(C_j) = \mu(A_j)$ for each j and $C_n \nearrow \bigcup_j C_j$. Therefore,

$$\mu\left(\bigcup_j A_j\right) \leq \mu\left(\bigcup_j C_j\right) = \lim_{n\to\infty} \mu(C_n) = \lim_{n\to\infty} \mu(A_n),$$

or

$$\mu\left(\bigcup_j A_j\right) \leq \lim_{n\to\infty} \mu(A_n).$$

This last inequality, together with (3.4), proves the theorem. ∎

Example 3.3.1 The Lebesgue measure λ on $\mathbb{R}$ is a regular measure. This follows from Exercise 3.2.3.

Exercise 3.3.1 Suppose that μ is a regular measure on Ω and that $B \subset \Omega$ with $\mu(B) < \infty$. Show that there is $A \in \Sigma^\mu$ such that $A \supset B$ and $\mu(C \cap A) = \mu(C \cap B)$, for every $C \in \Sigma^\mu$. (Hint: show that any $A \in \Sigma^\mu$ satisfying $A \supset B$ and $\mu(A) = \mu(B)$ will do.)

3.4 Premeasures and outer measures

Let Ω be a nonempty set, $\mathcal{G}$ a class of subsets of Ω containing $\emptyset$, and $\tau \colon \mathcal{G} \to [0, +\infty]$ satisfy $\tau(\emptyset) = 0$. Recall that such a set function τ is called a premeasure.

For a premeasure τ, define $\tau^* : 2^\Omega \to [0, +\infty]$ by

$$\tau^*(A) = \inf_{\substack{\{C_i\}_{i=1}^\infty \subset \mathcal{G} \\ \bigcup C_i \supset A}} \sum_i \tau(C_i), \; A \subset \Omega.$$

Then τ^* measures Ω and is called the (outer) measure on Ω **constructed** from τ by **Method I**. That τ^* is an outer measure on Ω follows from the same arguments as in the proof of Theorem 3.2.1 to show that λ is an outer measure on $\mathbb{R}$.

Example 3.4.1 The Lebesgue measure on $\mathbb{R}^n$.

A set of the form $I_1 \times \cdots \times I_n$ in $\mathbb{R}^n$, where $I_1, \ldots, I_n$ are finite intervals in $\mathbb{R}$, is called an **oriented rectangle** or an **oriented interval**, and $\prod_{j=1}^n |I_j|$ is called the volume of the rectangle. Let $\mathcal{G}$ be the class of all open oriented rectangles in $\mathbb{R}^n$ and let

$$\tau(I) = \text{volume of } I \text{ if } I \text{ is an open oriented rectangle.}$$

For convenience, the empty set is considered as a degenerate open oriented rectangle and hence $\mathcal{G}$ contains the empty set $\emptyset$ and $\tau(\emptyset) = 0$. The measure τ^* on $\mathbb{R}^n$ is called the **Lebesgue measure** on $\mathbb{R}^n$. The Lebesgue measure on $\mathbb{R}^n$ will be denoted by λ^n and λ^n-measurable sets are called **Lebesgue measurable** sets. In conformity with the notation for Lebesgue measure on $\mathbb{R}$, introduced in Section 3.2, λ^1 will be replaced by λ. We shall denote by $\mathcal{L}^n$ the σ-algebra of all λ^n-measurable sets in $\mathbb{R}^n$ and call $\mathcal{L}^n$-measurable functions **Lebesgue measurable** functions. Naturally, $\mathcal{L}^1$ is to be replaced by $\mathcal{L}$. But, habitually, Lebesgue measurable sets and Lebesgue measurable functions are usually called measurable sets and measurable functions, in this order. Accordingly, λ^n-integrable functions are **Lebesgue integrable** and usually simply called integrable functions. It is easily verified that if one considers closed oriented rectangles instead of open ones in the above construction, one arrives also at λ^n.

Exercise 3.4.1 For $\varepsilon > 0$, let $\mathcal{G}_\varepsilon$ be the class of all open oriented rectangles in $\mathbb{R}^n$ with diameter $< \varepsilon$, and $\tau_\varepsilon(I) = \text{volume of } I$ for $I \in \mathcal{G}_\varepsilon$. Show that the measure τ_ε^* on $\mathbb{R}^n$ is the Lebesgue measure.

Exercise 3.4.2 Let λ^n be the Lebesgue measure on $\mathbb{R}^n$.

(i) If $A, B \subset \mathbb{R}^n$ and $\operatorname{dist}(A, B) := \inf_{\substack{x \in A \\ y \in B}} |x - y| > 0$, then $\lambda^n(A \cup B) = \lambda^n(A) + \lambda^n(B)$.

(ii) Show that $\lambda^n(I) = \text{volume of I}$ if I is an open oriented rectangle. (Hint: use Lemma 1.7.2 to show $\lambda^n(I) \geq \text{volume of I}$.)

(iii) Show that every open oriented rectangle is λ^n-measurable and hence so are open sets and closed sets in $\mathbb{R}^n$. (Hint: pattern the first part of the proof of Proposition 3.2.1.)

(iv) Show that any hyperplane in $\mathbb{R}^n$ has Lebesgue measure zero.

(v) Show that $\{x \in \mathbb{R}^n : |x| = r\}$ has Lebesgue measure zero.

(vi) Show that for any $A \subset \mathbb{R}^n$, $\lambda^n(A + x) = \lambda^n(A)$ for $x \in \mathbb{R}^n$, and $\lambda^n(\alpha A) = |\alpha|^n \lambda^n(A)$ for $\alpha \in \mathbb{R}$.

Example 3.4.2 Let $I = [a_1, b_1] \times \cdots \times [a_n, b_n]$ be a finite closed oriented interval in $\mathbb{R}^n$. We assume that I is nondegenerate, i.e., $a_k < b_k$ for all $k = 1, \ldots, n$. By

Exercise 3.4.2 (iii), continuous functions on I are Lebesgue measurable. Since continuous functions on I are bounded, they are Lebesgue integrable due to the fact that $\lambda^n(I) < \infty$. We claim that for a continuous function f on I, $\int_I f d\lambda^n$ is the same as $\int_{a_n}^{b_n} \cdots \int_{a_1}^{b_1} f(x_1, \ldots, x_n) dx_1 \cdots dx_n$, the Riemann integral of f over I. To see this, recall first that a step function g on I is a function which takes constant value on each of a finite number of disjoint oriented intervals in I; the union of which is I. Since f is continuous, there is a sequence $\{g_k\}$ of step functions converging uniformly to f on I; then $\lim_{k\to\infty} \int_I g_k d\lambda^n = \int_I f d\lambda^n$. But $\{\int_I g_k d\lambda^n\}$ is a sequence of Riemann sums of f which tends to $\int_{a_n}^{b_n} \cdots \int_{a_1}^{b_1} f(x_1, \ldots, x_n) dx_1 \cdots dx_n$, hence our claim holds. We shall show in Section 4.2 that a Riemann integrable function on I is Lebesgue integrable, and its Lebesgue integral and Riemann integral are the same.

Example 3.4.3 A continuous function f on $\mathbb{R}$ is clearly Lebesgue measurable. We claim that f is Lebesgue integrable if and only if the improper integral $\int_{-\infty}^{\infty} f(x)dx$ converges absolutely. Suppose first that f is Lebesgue integrable. Then $|f|$ is Lebesgue integrable, hence $\int_{-\infty}^{\infty} |f| d\lambda = \lim_{n\to\infty} \int_{\mathbb{R}} |f| I_{[-n,n]} d\lambda = \lim_{n\to\infty} \int_{[-n,n]} |f| d\lambda$, by the monotone convergence theorem as well as by LCDT. But $\int_{[-n,n]} |f| d\lambda = \int_{-n}^{n} |f(x)| dx$, as we have shown in the previous example, thus, $\int_{-\infty}^{\infty} |f(x)| dx = \lim_{n\to\infty} \int_{-n}^{n} |f(x)| dx = \lim_{n\to\infty} \int_{[-n,n]} |f| d\lambda = \int_{-\infty}^{\infty} |f| d\lambda < \infty$, or $\int_{-\infty}^{\infty} f(x)dx$ converges absolutely. Conversely, if $\int_{-\infty}^{\infty} f(x)dx$ converges absolutely, then $\int_{-\infty}^{\infty} |f| d\lambda = \lim_{n\to\infty} \int_{[-n,n]} |f| d\lambda = \lim_{n\to\infty} \int_{-n}^{n} |f(x)| dx = \int_{-\infty}^{\infty} |f(x)| dx < \infty$. Hence, $|f|$ is Lebesgue integrable, and so is f. One sees easily that if either f is Lebesgue integrable or $\int_{-\infty}^{\infty} f(x)dx$ converges absolutely, then $\int_{\mathbb{R}} f d\lambda = \int_{-\infty}^{\infty} f(x)dx$.

Exercise 3.4.3 Let f be a real-valued continuous function on $\mathbb{R}$. Show that f is Lebesgue integrable on $\mathbb{R}$ if and only if for every sequence $\{I_n\}$ of finite disjoint open intervals, the system $\{\int_{I_n} f(x)dx\}_n$ is summable.

Exercise 3.4.4 Show that

$$\int_0^t \frac{2x}{1+x^2} dx = 2 \sum_{j=0}^{\infty} (-1)^j \int_0^t x^{2j+1} dx$$

for $0 < t < 1$; then show that

$$\int_0^1 \frac{2x}{1+x^2} dx = \sum_{j=0}^{\infty} (-1)^j \frac{1}{j+1},$$

and evaluate $\sum_{j=0}^{\infty} (-1)^j \frac{1}{j+1}$.

Exercise 3.4.5 Suppose that f is Lebesgue integrable on $\mathbb{R}$. Define a function g on $\mathbb{R}$ by

$$g(x) := \int_{(-\infty,x)} f d\lambda, \quad x \in \mathbb{R}.$$

Show that g is a bounded and uniformly continuous on $\mathbb{R}$.

Exercise 3.4.6 Find continuous functions f and g on $(0,\infty)$ such that f and g^2 are Lebesgue integrable on $(0,\infty)$, while f^2 and g are not Lebesgue integrable on $(0,\infty)$. Compare this exercise with Example 2.7.2 and Exercise 2.7.11.

Exercise 3.4.7 Let f be a continuous function on $\mathbb{R}^2$ and suppose that its improper integral on $\mathbb{R}^2$ is absolutely convergent. For integers m and n, let

$$\alpha_{mn} = \int_n^{n+1} \int_m^{m+1} f(x,y)dxdy.$$

(i) Show that $\{\alpha_{mn}\}_{(m,n)\in\mathbb{Z}\times\mathbb{Z}}$ is summable.
(ii) Show that $\int_{\mathbb{R}} f(x,y)d\lambda(x)$ is a Borel measurable function of y.
(iii) Show that $\iint_{\mathbb{R}^2} f(x,y)dxdy = \int_{\mathbb{R}^2} fd\lambda^2 = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x,y)d\lambda(x)\right) d\lambda(y)$.

(Hint: assume first that $f(x,y) \geq 0$. For positive integer n, $F_n(y) = \int_{[-n,n]} f(x,y)d\lambda(x)$ is a continuous function of y.)

Exercise 3.4.8

(i) Show that $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)}dxdy = (\int_{-\infty}^{\infty} e^{-t^2}dt)^2$.
(ii) Evaluate $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)}dxdy$, by using polar coordinates, and then find $\int_{-\infty}^{\infty} e^{-t^2}dt$.

Exercise 3.4.9 Find the following limits:

(i) $\lim_{n\to\infty}\int_0^{\infty}(1+\frac{x}{n})^{-n}\sin(\frac{x}{n})dx$.
(ii) $\lim_{n\to\infty}\int_0^1(1+nx^2)(1+x^2)^{-n}dx$.
(iii) $\lim_{n\to\infty}\int_0^{\infty} n\sin(\frac{x}{n})[x(1+x^2)]^{-1}dx$.
(iv) $\lim_{n\to\infty}\int_0^n(1+\frac{x}{n})^n e^{-2x}dx$.

Exercise 3.4.10 Let $\alpha = \int_{-\infty}^{\infty} e^{-x^2}dx$; show that

$$\int_{-\infty}^{\infty} x^{2n}e^{-x^2}dx = (2n)!(4^n n!)^{-1}\alpha.$$

Exercise 3.4.11 Show that $\lim_{k\to\infty}\int_0^k x^n(1-x/k)^k dx = n!$.

Exercise 3.4.12 Show that the improper integral $\int_0^1 \frac{x^p}{1-x}\ln\frac{1}{x}dx$ exists and equals $\sum_{j=1}^{\infty}\frac{1}{(p+j)^2}$ $(p > 0)$. (Hint: expand $\frac{1}{1-x}$ as a geometric series over $[0, 1-\varepsilon]$ for $0 < \varepsilon < 1$.)

Exercise 3.4.13 Suppose that f is a Lebesgue integrable function and φ is a bounded continuous function on $\mathbb{R}$. Show that $F(x) = \int_{\mathbb{R}} f(y)\varphi(x-y)d\lambda(y)$ is a continuous function of x in $\mathbb{R}$.

Example 3.4.4 Suppose that f is a function defined on $\mathbb{R}^2$ such that (i) $x \mapsto f(x,y)$ is Lebesgue measurable for each y, (ii) for λ-a.e. $x \in \mathbb{R}$, $f(x,y)$ is a continuous function of y, and (iii) there is a Lebesgue integrable function g on $\mathbb{R}$ such that $|f(x,y)| \leq g(x)$ for λ-a.e. x and for all y. Show that the function defined by

$$F(y) = \int_{\mathbb{R}} f(x,y)dx, \quad y \in \mathbb{R}$$

is a continuous function on $\mathbb{R}$. Let $y \in \mathbb{R}$ and $\{y_n\}$ a sequence in $\mathbb{R}$ converging to y. Put $f_n(x) = f(x,y_n)$, then $f_n(x) \to f(x,y)$ and $|f_n(x)| \leq g(x)$ for λ-a.e. x in $\mathbb{R}$. It follows then from LDCT that $\lim_{n\to\infty} F(y_n) = F(y)$. Hence F is continuous on $\mathbb{R}$.

Exercise 3.4.14 Let f and g be as in Example 3.4.4. Assume further that $y \mapsto f(x,y)$ is continuously differentiable for λ-a.e. x and there is an integrable function h on $\mathbb{R}$ such that $|\frac{\partial}{\partial y} f(x,y)| \leq h(x)$ for λ-a.e. x and for all y. Let F be defined as in Example 3.4.4 show that F is continuously differentiable on $\mathbb{R}$ and

$$F'(y) = \int_{\mathbb{R}} \frac{\partial}{\partial y} f(x,y)dx, \quad y \in \mathbb{R}.$$

Exercise 3.4.15 Define a function f on $(0,\infty)$ by

$$f(x) = \int_0^\infty \frac{e^{-t^2 x}}{1+t^2} dt, \quad x \in (0,\infty).$$

Show that f is continuously differentiable on $(0,\infty)$ and is a solution of the equation $y' - y + \frac{\sqrt{\pi}}{2}\frac{1}{\sqrt{x}} = 0$.

Exercise 3.4.16 Suppose that f is a continuous integrable function on $\mathbb{R}$. Show that the function $F : \mathbb{R} \to \mathbb{R}$, defined by

$$F(x) = \frac{1}{2}\int_{\mathbb{R}} e^{-|x-y|} f(y) d\lambda(y),$$

solves $F'' - F = f$ on $\mathbb{R}$.

Measurability of a given function is sometimes an issue, and is usually decided by whether it is the limit a.e. of a sequence of measurable functions. We illustrate this using an example.

Example 3.4.5 Suppose that $f(\cdot, y)$ is continuous on $[0,1]$ for each $y \in [0,1]$ and $f(x,\cdot)$ is continuous on $[0,1]$ for each $x \in [0,1]$. Then f is Lebesgue measurable on $[0,1] \times [0,1]$.

Proof For each $n \in \mathbb{R}$, define $f_n : [0,1] \times [0,1]$ by

$$f_n(x,y) = f\left(x, \frac{k}{n}\right),$$

if $\frac{k}{n} \leq y < \frac{k+1}{n}$, $k = 0, 1, \ldots, n-1$. Since the restriction of f_n to $[0,1] \times [\frac{k}{n}, \frac{k+1}{n})$ is continuous for $k = 0, \ldots, n-1$, f_n is Lebesgue measurable on $[0,1] \times [0,1)$ for each $n \in \mathbb{N}$. To show the measurability of f, it suffices to show that f_n converges to f pointwise as $n \to \infty$. Fix $(x_0, y_0) \in [0,1] \times [0,1)$. For each $\varepsilon > 0$ given, there is $\delta = \delta(x_0, y_0) > 0$ such that $|f(x_0, y) - f(x_0, y_0)| < \varepsilon$ if $|y - y_0| < \delta$ by the continuity of $f(x_0, \cdot)$. Thus for each $n > \frac{1}{\delta}$,

$$|f(x_0, y_0) - f_n(x_0, y_0)| = \left| f(x_0, y_0) - f\left(x_0, \frac{k}{n}\right) \right| < \varepsilon,$$

where $k = k(y_0, n)$ with $\frac{k}{n} \leq y_0 < \frac{k+1}{n}$. Therefore, $\lim_{n\to\infty} f_n(x_0, y_0) = f(x_0, y_0)$, and hence f is Lebesgue measurable. ■

For a nonempty class $\mathcal{G}$ of subsets of a set Ω, denote by $\mathcal{G}_\sigma$ the family of all those countable unions of sets from $\mathcal{G}$, and by $\mathcal{G}_{\sigma\delta}$ the family of all those countable intersections of sets from $\mathcal{G}_\sigma$; in parallel, the families $\mathcal{G}_\delta$ and $\mathcal{G}_{\delta\sigma}$ are defined by interchanging countable unions and countable intersections. In a metric space, a countable intersection of open sets is called a $\mathbf{G}_\delta$**-set** and a countable union of closed sets is called a $\mathbf{F}_\sigma$**-set**.

Proposition 3.4.1 *Let τ be a premeasure with domain $\mathcal{G}$ and suppose that there is $\{G_n\}_{n=1}^\infty \subset \mathcal{G}$ such that $\bigcup_n G_n = \Omega$. Then for every $B \subset \Omega$, there is $A \in \mathcal{G}_{\sigma\delta}$ such that $A \supset B$ and $\tau^*(A) = \tau^*(B)$.*

Proof From the definition of τ^* and the assumption that there is $\{G_n\} \subset \mathcal{G}$ such that $\bigcup_n G_n = \Omega \supset B$, one infers that there are $\{G_n^{(k)}\}_n \subset \mathcal{G}$, $k = 1, 2, 3, \ldots$, with the property $\bigcup_n G_n^{(k)} \supset B$ for each k and $\lim_{k\to\infty} \sum_n \tau(G_n^{(k)}) = \tau^*(B)$. Put $A = \bigcap_k \bigcup_n G_n^{(k)}$, then $A \in \mathcal{G}_{\sigma\delta}$ and $A \supset B$. It is clear from the definition of τ^* that $\tau^*(\bigcup_n G_n^{(k)}) \leq \sum_n \tau(G_n^{(k)})$, and consequently that $\tau^*(A) \leq \inf_k \tau^*(\bigcup_n G_n^{(k)}) \leq \liminf_{k\to\infty} \sum_n \tau(G_n^{(k)}) = \tau^*(B)$. But $B \subset A$ implies $\tau^*(B) \leq \tau^*(A)$, hence $\tau^*(A) = \tau^*(B)$. ■

Exercise 3.4.17

(i) Show that for any $B \subset \mathbb{R}^n$ and $\varepsilon > 0$, there is an open set $G \supset B$ such that $\lambda^n(G) \leq \lambda^n(B) + \varepsilon$.

(ii) Show that for any $B \subset \mathbb{R}^n$, there is a G_δ-set A in $\mathbb{R}^n$ such that $A \supset B$ and $\lambda^n(A) = \lambda^n(B)$.

Some applications of the method of constructing measures presented in this section will now be considered. Firstly, an extension theorem of Carathéodory–Hahn is to be established.

Theorem 3.4.1 (Carathéodory–Hahn) *Suppose that τ is a σ-additive set function on an algebra $\mathcal{A}$ on Ω, and let τ^* be the measure on Ω constructed from τ by Method I. Then $\sigma(\mathcal{A}) \subset \Sigma^{\tau^*}$ and $\tau(A) = \tau^*(A)$ for $A \in \mathcal{A}$. Furthermore, if τ is σ-finite, then the restriction of τ^* to $\sigma(\mathcal{A})$ is the unique measure on $\sigma(\mathcal{A})$ extending τ.*

Proof If we show that $\mathcal{A} \subset \Sigma^{\tau^*}$ and $\tau^*(A) = \tau(A)$ for $A \in \mathcal{A}$, then the first part of the theorem is proved. For $A \in \mathcal{A}$ and $B \subset \Omega$, consider an arbitrary sequence $\{A_n\}$ in $\mathcal{A}$ satisfying $\bigcup_n A_n \supset B$, then

$$\{A_n \cap A\} \subset \mathcal{A}, \quad \{A_n \cap A^c\} \subset \mathcal{A};$$
$$\bigcup_n (A_n \cap A) \supset B \cap A, \quad \bigcup_n (A_n \cap A^c) \supset B \cap A^c.$$

Hence,

$$\sum_n \tau(A_n) = \sum_n \tau(A_n \cap A) + \sum_n \tau(A_n \cap A^c) \geq \tau^*(B \cap A) + \tau^*(B \cap A^c),$$

from which follows that

$$\tau^*(B) \geq \tau^*(B \cap A) + \tau^*(B \cap A^c),$$

and thus $A \in \Sigma^{\tau^*}$.

To see that $\tau(A) = \tau^*(A)$, observe first that $\tau(A) \geq \tau^*(A)$; to show $\tau(A) \leq \tau^*(A)$, pick any sequence $\{A_n\}$ in $\mathcal{A}$ with $\bigcup_n A_n \supset A$ and verify that

$$\sum_n \tau(A_n) \geq \sum_n \tau(A_n \cap A) \geq \tau\left(\bigcup_n [A_n \cap A]\right) = \tau(A)$$

from σ-subadditivity of τ (cf. Exercise 2.1.1. (iv)), concluding that $\tau^*(A) \geq \tau(A)$.

Suppose now that ν is a measure on $\sigma(\mathcal{A})$ such that $\nu(A) = \tau(A)$ for $A \in \mathcal{A}$. We claim that $\nu(A) \leq \tau^*(A)$ for $A \in \sigma(\mathcal{A})$. Let $A \in \sigma(\mathcal{A})$, and consider an arbitrary sequence $\{A_n\}$ in $\mathcal{A}$ with $\bigcup_n A_n \supset A$. Then,

$$\nu(A) \leq \sum_n \nu(A_n) = \sum_n \tau(A_n),$$

concluding $\nu(A) \leq \tau^*(A)$.

If τ is σ-finite, there is an increasing sequence $\Omega_1 \subset \Omega_2 \subset \cdots \subset \Omega_n \subset \cdots$ in $\mathcal{A}$ such that $\tau(\Omega_n) < \infty$ for all n and $\bigcup_n \Omega_n = \Omega$. For each n, from what we have just claimed, we have for $A \in \sigma(\mathcal{A})$,

$$\nu(\Omega_n \backslash [\Omega_n \cap A]) \leq \tau^*(\Omega_n \backslash [\Omega_n \cap A]),$$

or

$$\nu(\Omega_n) - \nu(\Omega_n \cap A) \leq \tau^*(\Omega_n) - \tau^*(\Omega_n \cap A),$$

from which, using the fact that $\nu(\Omega_n) = \tau^*(\Omega_n) = \tau(\Omega_n) < \infty$, we have

$$\nu(\Omega_n \cap A) \geq \tau^*(\Omega_n \cap A).$$

Let $n \to \infty$ in the last inequality; it follows that $\nu(A) \geq \tau^*(A)$. This shows that $\nu(A) = \tau^*(A)$ for $A \in \sigma(\mathcal{A})$, completing the proof of the second part of the theorem. ∎

Example 3.4.6 (Continuation of Example 2.1.1) Consider the sequence space Ω, the algebra $\mathcal{Q}$ of all cylinders in Ω, and the set function P, defined in Section 1.3. We know from Example 2.1.1 that P is σ-additive on $\mathcal{Q}$. Note that $P(\Omega) = 1$. Now by Theorem 3.4.1, P can be extended uniquely to be a measure on $\sigma(\mathcal{Q})$; then the probability space $(\Omega, \sigma(\mathcal{Q}), P)$ is referred to as the Bernoulli sequence space. One can verify easily that the set E defined in the last paragraph of Section 1.3 is actually in $\sigma(\mathcal{Q})$ by observing that $E_{nk} := \{w \in \Omega : \frac{1}{2} - \frac{1}{k} < \frac{S_n(w)}{n} < \frac{1}{2} + \frac{1}{k}\} \in \mathcal{Q}$ for n, k in $\mathbb{N}$; $P(E)$ therefore has a meaning. Note that if $w = (w_k) \in \Omega$, then $\{w\} = E(w_1) \cap E(w_1, w_2) \cap \cdots \cap E(w_1, \ldots, w)n) \cap \cdots$; hence any singleton set in Ω is in $\sigma(\mathcal{Q})$, and clearly the probability of any singleton set is zero.

Theorem 3.4.1 contains the fact that the method of outer measure is universal in constructing measure spaces.

Corollary 3.4.1 *Given a measure space (Ω, Σ, μ), the measure μ^* on Ω constructed from μ (considered as defined on Σ) by Method I is the unique Σ-regular measure on Ω such that $\mu^*(A) = \mu(A)$ for $A \in \Sigma$.*

Proof By Theorem 3.4.1, $\Sigma \subset \Sigma^{\mu^*}$ and $\mu^*(A) = \mu(A)$ for $A \in \Sigma$. Since $\Sigma_{\sigma\delta} = \Sigma$, it follows from Proposition 3.4.1 that μ^* is Σ-regular.

To prove uniqueness, let ν be a Σ-regular measure on Ω such that $\nu(A) = \mu(A)$ for $A \in \Sigma$. We claim that $\nu = \mu^*$. Actually, for any set $B \subset \Omega$, there are A_1 and A_2 in Σ such that $A_1 \supset B, A_2 \supset B, \mu^*(A_1) = \mu^*(B)$, and $\nu(A_2) = \nu(B)$. Put $A = A_1 \cap A_2$, then

$$\mu^*(A_1) \geq \mu^*(A) \geq \mu^*(B) = \mu^*(A_1);$$
$$\nu(A_2) \geq \nu(A) \geq \nu(B) = \nu(A_2),$$

hence, $\mu^*(B) = \mu^*(A)$ and $\nu(B) = \nu(A)$. But $A \in \Sigma$ implies that $\nu(A) = \mu(A) = \mu^*(A)$. Thus $\mu^*(B) = \nu(B)$. ∎

Exercise 3.4.18

(i) If (Ω, Σ, μ) is σ-finite, show that for $A \in \Sigma^{\mu^*}$ there is $B \in \Sigma$ such that $B \supset A$ and $\mu^*(B \backslash A) = 0$.

(ii) If (Ω, Σ, μ) is σ-finite, show that $(\Omega, \Sigma^{\mu^*}, \mu^*)$ is the completion of (Ω, Σ, μ) (cf. Section 2.8.3).

(iii) If μ measures Ω and $\Sigma = \Sigma^{\mu}$, show that $\mu^* = \mu$ if and only if μ is regular.

Remark Because of Corollary 3.4.1, we may consider any measure space (Ω, Σ, μ) as obtained by restricting to Σ the Σ-regular measure μ^* on Ω. Note that if μ is a measure on Ω, the measure μ^* on Ω constructed from μ as a measure on Σ^μ by Method I is in general different from the original measure μ on Ω (cf. Exercise 3.4.18 (iii)).

Theorem 3.4.2 *Let $\mathcal{A}$, τ be as in Theorem 3.4.1. Then Σ^{τ^*} is the largest σ-algebra containing $\mathcal{A}$ on which τ^* is σ-additive.*

Proof Let Σ' be a σ-algebra containing $\mathcal{A}$ on which τ^* is σ-additive. We shall show that $\Sigma' \subset \Sigma^{\tau^*}$. Let $A \in \Sigma'$ and $B \subset \Omega$. For $\varepsilon > 0$, there is a sequence $\{A_n\}$ in $\mathcal{A}$ such that $B \subset \bigcup_n A_n$ and $\sum_n \tau(A_n) \leq \tau^*(B) + \varepsilon$. Put $H = \bigcup_n A_n$, then $H, H \cap A, H \cap A^c$ are in Σ', and

$$\begin{aligned}\tau^*(B) + \varepsilon \geq \sum_n \tau(A_n) &= \sum_n \tau^*(A_n) \geq \tau^*(H)\\ &= \tau^*(H \cap A) + \tau^*(H \cap A^c) \geq \tau^*(B \cap A) + \tau^*(B \cap A^c)\\ &\geq \tau^*(B).\end{aligned}$$

Let $\varepsilon \searrow 0$ in the last sequence of inequalities; we obtain $\tau^*(B) = \tau^*(B \cap A) + \tau^*(B \cap A^c)$, concluding that $A \in \Sigma^{\tau^*}$. ■

Exercise 3.4.19 Use the (π-λ) Theorem to prove the second part of Theorem 3.4.1.

Exercise 3.4.20

(i) Show that the measure on $\mathbb{R}^n$ constructed from the restriction of λ^n to $\mathcal{B}^n$ by Method I is λ^n.

(ii) Show that λ^n is not σ-additive on any σ-algebra on $\mathbb{R}^n$ which contains $\mathcal{L}^n$ strictly.

3.5 Carathéodory measures

We shall consider in this section a class of measures on metric spaces which plays an important role in analysis. For this purpose, we first introduce some useful notations. For a metric space X with metric ρ and for nonempty subsets A, B of X, let

$$\rho(A, B) = \inf_{x \in A, y \in B} \rho(x, y).$$

When $A = \{x\}$, $\rho(\{x\}, A)$ is written simply as $\rho(x, A)$. In the case of $\mathbb{R}^n$ with the Euclidean metric ρ, $\rho(A, B)$ is usually denoted by $\text{dist}(A, B)$ and is called the **distance** between A and B. Recall that for a metric space X, we use $\mathcal{B}(X)$ to denote the σ-algebra generated by the family of all open sets of X and that sets in $\mathcal{B}(X)$ are called Borel sets.

Let μ be a measure on X, with X being a metric space, μ is called a **Carathéodory** measure on X if $\mu(A \cup B) = \mu(A) + \mu(B)$ whenever $\rho(A, B) > 0$.

Example 3.5.1 The Lebesgue measure on $\mathbb{R}^n$ is a Carathéodory measure (cf. Exercise 3.4.2 (i)).

Theorem 3.5.1 *If μ is a Carathéodory measure on a metric space X, then every closed subset of X is μ-measurable.*

A lemma precedes the proof of the theorem.

Lemma 3.5.1 *Let $A_1 \subset A_2 \subset \cdots \subset A_n \subset A_{n+1} \subset \cdots$ be an increasing sequence of subsets of X such that for each n, $\rho(A_n, A_{n+1}^c) > 0$. Then,*

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sup_n \mu(A_n).$$

Proof Obviously, $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) \geq \sup_n \mu(A_n)$.

To show that $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sup_n \mu(A_n)$, we may assume that $\sup_n \mu(A_n) < +\infty$.

Let $D_1 = A_1, D_2 = A_2 \setminus A_1, \ldots, D_n = A_n \setminus A_{n-1}, \ldots$. By our assumption, for any n and $m \geq n+2$, we have $\operatorname{dist}(D_n, D_m) > 0$. Then,

$$\mu(D_1 \cup D_3 \cup \cdots \cup D_{2k-1}) = \mu(D_1) + \mu(D_3) + \cdots + \mu(D_{2k-1})\,;$$
$$\mu(D_2 \cup D_4 \cup \cdots \cup D_{2k}) = \mu(D_2) + \mu(D_4) + \cdots + \mu(D_{2k})$$

for each k. Now,

$$\sum_{j=1}^{k} \mu(D_{2j-1}) = \mu(D_1 \cup D_3 \cup \cdots \cup D_{2k-1}) \leq \mu(A_{2k-1}) \leq \sup_n \mu(A_n) < +\infty,$$

implying that $\sum_{j=1}^{\infty} \mu(D_{2j-1}) < \infty$. Similarly, $\sum_{j=1}^{\infty} \mu(D_{2j}) < +\infty$. Then,

$$\mu\left(\bigcup_{j=1}^{\infty} A_j\right) = \mu\left(A_n \cup \bigcup_{j=n+1}^{\infty} A_j\right) = \mu\left(A_n \cup \bigcup_{j=n+1}^{\infty} D_j\right)$$
$$\leq \mu(A_n) + \sum_{j=n+1}^{\infty} \mu(D_j),$$

from which by letting $n \to \infty$, we have

$$\mu\left(\bigcup_{j=1}^{\infty} A_j\right) \leq \sup_n \mu(A_n).$$

■

Proof of Theorem 3.5.1 Let $F \subset X$ be a closed set, and let $A \subset F$, $B \subset F^c$. For each $n \in \mathbb{N}$, let

$$B_n = \left\{x \in B : \rho(x, F) > \frac{1}{n}\right\}.$$

Then, since F is closed, we have $\bigcup_{n=1}^{\infty} B_n = B$. Obviously, $B_1 \subset B_2 \subset \cdots \subset B_n \subset B_{n+1} \subset \cdots$. Now,

$$\rho(B_n, B \setminus B_{n+1}) \geq \frac{1}{n(n+1)} > 0,$$

hence, by Lemma 3.5.1 (applied to the metric space (B, ρ)),

$$\sup_n \mu(B_n) = \mu\left(\bigcup_{n=1}^{\infty} B_n\right) = \mu(B),$$

and since $\rho(A, B_n) \geq \rho(F, B_n) \geq \frac{1}{n} > 0$,

$$\mu(A \cup B) \geq \mu(A \cup B_n) = \mu(A) + \mu(B_n)$$

for each n; thus,

$$\mu(A \cup B) \geq \mu(A) + \sup_n \mu(B_n) = \mu(A) + \mu(B). \qquad \blacksquare$$

Corollary 3.5.1 *If μ is a Carathéodory measure on a metric space X, then all Borel subsets of X are μ-measurable.*

3.6 Construction of Carathéodory measures

Let X be a metric space and $\tau : \mathcal{G} \to [0, +\infty]$ a premeasure on X. For $\varepsilon > 0$, define a measure τ_ε on X as follows. For $A \subset X$, let

$$\tau_\varepsilon(A) = \inf \sum_i \tau(C_i),$$

where the infimum is taken over all sequences $\{C_i\} \subset \mathcal{G}$ such that $\bigcup_i C_i \supset A$ and $\operatorname{diam} C_i \leq \varepsilon$ for each i; τ_ε is the measure constructed from the restriction of τ to $\mathcal{G}_\varepsilon = \{C \in \mathcal{G} : \operatorname{diam} C \leq \varepsilon\}$ by Method I. Since $\tau_\varepsilon(A)$ increases as ε decreases for $A \subset X$, $\lim_{\varepsilon \to 0} \tau_\varepsilon(A)$ exists and we define

$$\tau^d(A) = \lim_{\varepsilon \to 0} \tau_\varepsilon(A), \;\; A \subset X.$$

Exercise 3.6.1

(i) Show that τ^d is a Carathéodory measure on X.

(ii) Show that if $\mathcal{G}$ consists of open sets, then for any $A \subset X$ there is a G_δ-set $B \supset A$ such that $\tau^d(A) = \tau^d(B)$.

We shall call τ^d the measure **constructed** from premeasure τ by **Method II**.

Exercise 3.6.2 Let $\mathcal{G}$ be the family of all bounded open intervals in $\mathbb{R}$ and suppose that f is a nonnegative integrable function on $\mathbb{R}$. Define $\tau(I) = \int_I f d\lambda$ for $I \in \mathcal{G}$ and let τ^d be the measure on $\mathbb{R}$ constructed from τ by Method II. Show that every measurable set in $\mathbb{R}$ is τ^d-measurable and $\tau^d(A) = \int_A f d\lambda$ for every measurable set A. (Hint: show first that $\tau^d(I) = \tau(I)$ for bounded open interval I.)

Example 3.6.1 Let X be a metric space and $0 \leq s < +\infty$. Take $\mathcal{G} = 2^X$ and let τ^s be the premeasure defined by $\tau^s(\emptyset) = 0$ and $\tau^s(A) = (\operatorname{diam} A)^s$ if $A \neq \emptyset$. The measure

H^s constructed from τ^s by Method II is called the **s-dimensional Hausdorff** measure on X. Note that if we take $\mathcal{G}$ to be the family of all open subsets of X or the family of all closed subsets of X, we shall arrive at the same measure H^s.

Exercise 3.6.3

(i) Show that H^0 is the counting measure on X.

(ii) If $H^s(A) < +\infty$, show that $H^{s+\delta}(A) = 0$ if $\delta > 0$.

(iii) If $H^s(A) > 0$, show that $H^t(A) = +\infty$ if $0 \le t < s$.

Exercise 3.6.4 Show that H^1 on $\mathbb{R}$ is the Lebesgue measure on $\mathbb{R}$.

Since Hausdorff dimensional measures will not be our main concern, we shall content ourselves by showing that the arclength of a rectifiable arc in $\mathbb{R}^2$ is its one-dimensional Hausdorff measure. By an **arc** C in $\mathbb{R}^2$ we shall mean the image of a continuous injective map from a finite closed interval $[a,b]$ into $\mathbb{R}^2$. Any continuous injective map with C as its image is called a parametric representation of $\mathcal{C}$. Let $t : [a,b] \to \mathbb{R}^2$ be a parametric representation of $\mathcal{C}$ and consider a partition $\mathcal{P} := a = x_0 < x_1 < \cdots < x_k = b$ of $[a,b]$. Define

$$l = \sup_{\mathcal{P}} \sum_{j=1}^{k} |t(x_j) - t(x_{j-1})|,$$

where $|\cdot|$ is the Euclidean norm in $\mathbb{R}^2$. If $l < \infty$, $\mathcal{C}$ is called a **rectifiable** arc, and l is called the arclength of $\mathcal{C}$. Since l is the supremum of the length of all possible inscribed polygonal arcs, it is independent of parametric representations of $\mathcal{C}$.

Proposition 3.6.1 *Let $\mathcal{C}$ be a rectifiable arc in $\mathbb{R}^2$, then $H^1(\mathcal{C})$ is the arclength of $\mathcal{C}$.*

Proof Let l be the arclength of $\mathcal{C}$ and let $t : [0,l] \to \mathbb{R}^2$ be the parametric representation of $\mathcal{C}$ by arclength, with $t(0)$ and $t(l)$ the endpoints of $\mathcal{C}$, i.e. the arclength from $t(0)$ to $t(s)$ is s for $0 \le s \le l$. Then for s_1, s_2 in $[0,l]$,

$$\operatorname{diam} t[s_1, s_2] \le |s_1 - s_2|.$$

Given $\varepsilon > 0$, let $0 = s_0 < s_1 < \cdots < s_k = l$ be a partition of $[0,l]$ such that $|s_j - s_{j-1}| < \varepsilon$ for $j = 1, \ldots, k$, then,

$$l = \sum_{j=1}^{k} |s_j - s_{j-1}| \ge \sum_{j=1}^{k} \operatorname{diam} t[s_{j-1}, s_j] \ge \tau_\varepsilon^1(\mathcal{C}),$$

hence $l \ge H^1(\mathcal{C})$.

To show $l \le H^1(\mathcal{C})$, we observe first that if L is a line in $\mathbb{R}^2$ and P the orthogonal projection from $\mathbb{R}^2$ onto L, then for any $A \subset \mathbb{R}^2$, $H^1(PA) \le H^1(A)$. Now let $0 = s_0 < s_1 < \cdots < s_k = l$ be a partition of $[0,l]$, and for each $j = 1, \ldots, k$ consider the line L which passes through $t(s_{j-1})$ and $t(s_j)$ and the orthogonal projection P from

$\mathbb{R}^2$ onto L. From the above observation, $H^1(t([s_{j-1}, s_j])) \geq H^1([t(s_{j-1}), t(s_j)]) = |t(s_{j-1}) - t(s_j)|$, where $[t(s_{j-1}), t(s_j)]$ is the line segment connecting $t(s_{j-1})$ and $t(s_j)$; consequently,

$$H^1(\mathcal{C}) = \sum_{j=1}^{k} H^1(t([s_{j-1}, s_j])) \geq \sum_{j=1}^{k} |t(s_{j-1}) - t(s_j)|,$$

from which one infers that $H^1(\mathcal{C}) \geq l$. ∎

3.7 Lebesgue–Stieltjes measures

Given a monotone increasing function g on $\mathbb{R}$, a measure μ_g on $\mathbb{R}$ will be constructed, which is suggested by the Riemann–Stieltjes integral of functions with respect to g.

For a finite open interval, $I = (a, b)$, $a \leq b$, let $\tau(I) = g(b) - g(a)$, then τ is a pre-measure on $\mathbb{R}$. The measure τ^* on $\mathbb{R}$ constructed from τ by Method I is called the **Lebesgue–Stieltjes** measure generated by g and is denoted by μ_g; when $g(x) = x$, μ_g is the Lebesgue measure on $\mathbb{R}$.

It turns out that μ_g is also the measure τ^d on $\mathbb{R}$ constructed from τ by Method II. To see this, a preliminary result on the set of points of discontinuity of g will first be shown.

Lemma 3.7.1 *The set D of points of discontinuity of g is at most countable. Furthermore D consists only of points of jump of g.*

Proof Since g is monotone, $g(x+) = \lim_{y \to x+} g(y)$ and $g(x-) = \lim_{y \to x-} g(y)$ exist and are finite at every point x of $\mathbb{R}$. It is clear that $x \in D$ if and only if $g(x+) - g(x-) > 0$, hence D consists only of points of jump of g. To show that D is at its most countable, it is sufficient to show that $D_n := D \cap (-n, n)$ is at its most countable for all $n \in \mathbb{N}$. Fix $n \in \mathbb{N}$ and for $x \in D_n$, let I_x be the open interval $(g(x-), g(x+))$ and $c_x = g(x+) - g(x-)$. Consider any nonempty finite subset A of D_n, we have

$$\sum_{x \in A} c_x \leq g(n+) - g((-n)-),$$

because $\{I_x : x \in A\}$ is a finite disjoint family of open intervals. Hence the system $\{c_x\}$ indexed by $x \in D_n$ is summable by Theorem 1.1.2. But the fact that $c_x > 0$ for $x \in D_n$ implies, by Exercise 1.1.6, that D_n is at its most countable. ∎

We are now going to verify that $\tau^d = \mu_g$. Fix $\varepsilon > 0$. Consider a finite open interval $I = (a, b)$, $a < b$, and let $\delta > 0$ be given. By Lemma 3.7.1 we can find a partition, $a = a_0 < x_1 < \cdots < x_k = b$, such that $x_j - x_{j-1} < \varepsilon$ for $j = 1, \ldots, k$ and such that each x_j, $j = 1, \ldots, k-1$, is a point of continuity of g; then for each $j = 1, \ldots, k-1$, choose a point y_j in (x_j, x_{j+1}) such that $g(y_j) - g(x_j) < \frac{\delta}{k}$ and $y_j - x_{j-1} \leq \varepsilon$. The intervals (a, y_1),

$(x_1, y_2), \ldots, (x_{k-2}, y_{k-1})$, and (x_{k-1}, b) form a covering of $I = (a, b)$, and each of them has length $\leq \varepsilon$. Call these intervals $I_1, \ldots, I_k$ in this order, then,

$$\tau(I) = g(b) - g(a) = \sum_{j=1}^{k} \{g(x_j) - g(x_{j-1})\} > \sum_{j=1}^{k} \tau(I_j) - \delta,$$

from which one infers (cf. the method of proof of Lemma 3.2.1) that $\tau_\varepsilon(A) = \mu_g(A)$ for $A \subset \mathbb{R}$, and hence $\tau^d = \mu_g$ (see Section 3.6 for definitions of τ_ε and τ^d).

Theorem 3.7.1 *The measure μ_g is a Carathéodory measure on $\mathbb{R}$ which takes finite value on each bounded set. Furthermore, there is a sequence $\{G_k\}$ of open sets such that $A \subset \bigcap_k G_k$ and $\mu_g(A) = \inf_k \mu_g(G_k)$; in particular, for any $A \subset \mathbb{R}$, there is a G_δ-set $B \supset A$ such that $\mu_g(A) = \mu_g(B)$ (recall that the intersection of a sequence of open sets is called a G_δ-set).*

Proof Since, as we have just shown, μ_g is a measure on $\mathbb{R}$ constructed from the premeasure τ by Method II, μ_g is a Carathéodory measure. That $\mu_g(A) < \infty$ if A is bounded is obvious.

Now let $A \subset \mathbb{R}$. There is a sequence $\{I_n^{(1)}\}, \{I_n^{(2)}\}, \ldots$ of countable coverings of A consisting of finite open intervals such that

$$\mu_g(A) = \lim_{k \to \infty} \sum_n \tau(I_n^{(k)}).$$

For each k, let $G_k = \bigcup_n I_n^{(k)}$, then

$$\mu_g(A) \leq \mu_g(G_k) \leq \sum_n \tau(I_n^{(k)}),$$

from which we obtain $\mu_g(A) = \inf_k \mu_g(G_k)$ by letting $k \to \infty$. Finally, let $B = \bigcap_k G_k$, then B is a G_δ-set containing A and $\mu_g(A) \leq \mu_g(B) \leq \inf_k \mu_g(G_k) = \mu_g(A)$. Hence, $\mu_g(A) = \mu_g(B)$. ∎

Lemma 3.7.2 $\mu_g([a, b]) = g(b+) - g(a-), -\infty < a \leq b < \infty$.

Proof Since $\mu_g([a, b]) \leq g(d) - g(c)$ for $(c, d) \supset [a, b]$, $\mu_g([a, b]) \leq g(b+) - g(a-)$. It remains to show that $g(b+) - g(a-) \leq \mu_g([a, b])$.

Let $\{I_n\}$ be a sequence of finite open intervals such that $\bigcup_n I_n \supset [a, b]$, and write $I_n = (a_n, b_n), n = 1, 2, \ldots$. $\{I_n\}$ is an open covering of $J = [a', b']$ for some $a' < a$ and some $b' > b$. Let $\delta > 0$ be the Lebesgue number of J w.r.t. the open covering $\{I_n\}$ (cf. Lemma 1.7.2), and let $a' = x_0 < x_1 < \cdots < x_k = b'$ be a partition of J with $(x_j - x_{j-1}) \leq \delta, j = 1, \ldots, k$. Put $J_j = [x_{j-1}, x_j]$ for $j = 1, \ldots, k$ and proceed as follows. First pick $n_1 \in \mathbb{N}$ with $[x_0, x_1] \subset I_{n_1}$ according to Lemma 1.7.2, and let j_1 be the largest integer between 1 and k such that $[x_0, x_{j_1}] \subset I_{n_1}$. If $j_1 = k$, stop the process; otherwise, there is $n_2 \in \mathbb{N}$ with $[x_{j_1}, x_{j_1+1}] \subset I_{n_2}$ (again by Lemma 1.7.2), and let j_2 be the largest integer between $j_1 + 1$ and k such that $[x_{j_1}, x_{j_2}] \subset I_{n_2}$. Obviously, $n_1 \neq n_2$. Continue in this fashion, we obtain mutually different positive integers $n_1, \ldots, n_l$ and integers $1 \le j_1 < \cdots < j_l = k$ such that $[x_{j_m} - x_{j_{m+1}}] \subset I_{n_{m+1}}$ for $m = 0, 1, \ldots, l-1$. Now,

$$g(b+) - g(a-) \le g(b') - g(a') = \sum_{m=1}^{l} \{g(x_{j_m}) - g(x_{j_{m-1}})\}$$
$$\le \sum_{m=1}^{l} \tau(I_{n_m}) \le \sum_{n} \tau(I_n),$$

from which, since $\{I_n\}$ is any sequence of finite open intervals with $\bigcup_n I_n \supset [a, b]$, it follows that $g(b+) - g(a-) \le \mu_g([a, b])$. ■

Exercise 3.7.1 Show that for $a < b$ in $\mathbb{R}$,

$$\mu_g((a, b]) = g(b+) - g(a+);$$
$$\mu_g((a, b)) = g(b-) - g(a+);$$
$$\mu_g([a, b)) = g(b-) - g(a-).$$

Exercise 3.7.2 Let w be a nonnegative measurable function on $\mathbb{R}$ such that $\int_{(-\infty,x]} w d\lambda < \infty$ for all $x \in \mathbb{R}$. Define a monotone increasing function g on $\mathbb{R}$ by $g(x) = \int_{(-\infty,x]} w d\lambda$. Show that $\mu_g(B) = \int_B w d\lambda$ for $B \in \mathcal{B}$.

From Exercise 3.7.1, we know that if g is right-continuous, then $\mu_g((a, b]) = g(b) - g(a)$. Recall that a function is **right-continuous** if it is continuous from the right-hand side at each point of its domain of definition. We show now that for any monotone increasing function g on $\mathbb{R}$, μ_g is the same as the Lebesgue–Stieltjes measure generated by a right-continuous monotone increasing function.

Theorem 3.7.2 *For a monotone increasing function g on $\mathbb{R}$, define a function $\hat{g}$ on $\mathbb{R}$ by $\hat{g}(x) = g(x+)$. Then $\hat{g}$ is right-continuous and $\mu_{\hat{g}} = \mu_g$.*

Proof Proof of right-continuity of $\hat{g}$ is left as an exercise.

To show that $\mu_{\hat{g}} = \mu_g$, we note first that an open interval (a, b) is a union of a sequence $(a_n, b_n]$, $n = 1, 2, \ldots$, of increasing half open intervals such that $a_n \searrow a$ and $b_n \nearrow b$, hence (cf. Exercise 3.7.1),

$$\mu_g((a, b)) = \lim_{n\to\infty} \mu_g((a_n, b_n]) = \lim_{n\to\infty} \{g(b_n+) - g(a_n+)\}$$
$$= \lim_{n\to\infty} \{\hat{g}(b_n) - \hat{g}(a_n)\} = \lim_{n\to\infty} \{\hat{g}(b_n+) - \hat{g}(a_n+)\}$$
$$= \lim_{n\to\infty} \mu_{\hat{g}}((a_n, b_n]) = \mu_{\hat{g}}((a, b));$$

consequently, $\mu_{\hat{g}}(G) = \mu_g(G)$ if G is open. Now let A be any subset of $\mathbb{R}$; by Theorem 3.7.1 there are sequences $\{G_n\}$ and $\{\widehat{G}_n\}$ of open sets such that $\bigcap_n G_n \supset A$, $\bigcap_n \widehat{G}_n \supset A$, $\mu_g(A) = \inf_k \mu_g(G_k)$, and $\mu_{\hat{g}}(A) = \inf_k \mu_{\hat{g}}(\widehat{G}_k)$. Observe that $\mu_g(A) = \inf_k \mu_g(G_k \cap \widehat{G}_k)$ and $\mu_{\hat{g}}(A) = \inf_k \mu_{\hat{g}}(G_k \cap \widehat{G}_k)$; then, since $\mu_g(G_k \cap \widehat{G}_k) = \mu_{\hat{g}}(G_k \cap \widehat{G}_k)$, it follows that $\mu_g(A) = \mu_{\hat{g}}(A)$. ■

Exercise 3.7.3 Show that the function $\hat{g}$ defined in Theorem 3.7.2 is right-continuous.

Example 3.7.1 Let D be a finite or countably infinite set in $\mathbb{R}$ and ν a positive-valued function on D such that $\sum_{t\in(-\infty,x]\cap D} \nu(t) < \infty$ for all $x \in \mathbb{R}$. Define a function g on $\mathbb{R}$ by $g(x) = \sum_{t\in(-\infty,x]\cap D} \nu(t)$, $x \in \mathbb{R}$; then g is a monotone increasing function. We claim that g is right-continuous. For $x \in \mathbb{R}$, fix $y_0 > x$. Then, $g(y) - g(x) = \sum_{t\in(x,y]\cap D} \nu(t)$ if $y \in (x, y_0]$. If $(x, y_0] \cap D$ is finite, $g(y) = g(x)$, when y is sufficiently near to x, and hence $g(x+) = g(x)$. We may therefore assume that $D \cap (x, y_0]$ is infinite and denote it by $\{t_n\}_{n\in\mathbb{N}}$. Since $\sum_{n\in\mathbb{N}} \nu(t_n) < \infty$, for given $\varepsilon > 0$ there is $n_0 \in \mathbb{N}$ such that $\sum_{n>n_0} \nu(t_n) < \varepsilon$. Let $y > x$ be smaller than $t_1, \ldots, t_{n_0}$, then $g(y) - g(x) \le \sum_{n>n_0} \nu(t_n) < \varepsilon$, and consequently $g(x+) = g(x)$. Hence g is right-continuous at every $x \in \mathbb{R}$. The same argument also shows that D is the set of points of discontinuity of g and $g(t) - g(t-) = \nu(t)$ for $t \in D$. Similarly, if D and ν satisfy the condition that $\sum_{t\in[x,\infty)\cap D} \nu(t) < \infty$ for all $x \in \mathbb{R}$, and if g is defined by $g(x) = -\sum_{t\in(x,\infty)\cap D} \nu(t)$, $x \in \mathbb{R}$, then g enjoys the same properties as shown previously.

Exercise 3.7.4 Let g be a monotone increasing and right-continuous function on $\mathbb{R}$, and denote by D the set of points of discontinuity of g. Define $\nu(t) = g(t) - g(t-)$ for $t \in D$, and define a function g_d on $\mathbb{R}$ by

$$g_d(x) = \begin{cases} \sum_{t\in(0,x]\cap D} \nu(t), & x \ge 0 \\ -\sum_{t\in(x,0]\cap D} \nu(t), & x < 0. \end{cases}$$

Show that g_d is a monotone increasing and right-continuous function with D as its set of points of discontinuity. Furthermore, the function $g - g_d$ is continuous.

Exercise 3.7.5 Let g, D, g_d be as in Exercise 3.7.4, and let $\mu = \mu_{g_d}$ be the Lebesgue–Stielties measure generated by g_d. Show that for $B \in \mathcal{B}$, $\mu(B) = \sum_{t\in B\cap D} \nu(t)$, where $\nu(t) = g(t) - g(t-)$ for $t \in D$. (Hint: show first that $\mu(G) = \sum_{t\in G\cap D} \nu(t)$ if G is open, and use Theorem 2.1.1.)

Suppose now that g is a monotone increasing function on a closed finite interval $[a, b]$; extend g to a function h on $\mathbb{R}$ by defining $h(x) = g(a)$ for $x < a$ and $h(x) = g(b)$ for $x > b$. Then the Lebesgue–Stieltjes measure μ_g on $[a, b]$ generated by g is the restriction of μ_h to $[a, b]$, i.e.

$$\mu_g(A) = \mu_h(A), \; A \subset [a, b].$$

For notational convenience, the integral of a function f w.r.t. a Lebesgue–Stieltjes measure μ_g on $\mathbb{R}$ or on a finite closed interval $[a, b]$ will be denoted by $\int_{-\infty}^{\infty} f d\mu_g$ or $\int_a^b f d\mu_g$, as the situation suggests.

3.8 Borel regularity and Radon measures

Recall that a measure μ on a set Ω is called *regular* if for any $A \subset \Omega$, there is a μ-measurable set $B \supset A$ such that $\mu(B) = \mu(A)$. Such a regularity endows μ with a significant monotone limit property, stated in Theorem 3.3.2. A further regularity along this line for measures on metric spaces will now be introduced.

A measure μ on a metric space X is called a **Borel** measure if every Borel set is μ-measurable. It is said to be **Borel regular** if it is Borel and if for every $A \subset X$, there is a Borel set $B \supset A$ such that $\mu(B) = \mu(A)$; in other words, a Borel regular measure on X is what we call a $\mathcal{B}(X)$-regular measure (see the paragraph preceding Theorem 3.3.2). It is called a **Radon measure** if it is Borel regular and $\mu(K) < +\infty$ for each compact set K.

We already know that every Carathéodory measure is Borel. Obviously, λ^n is a Radon measure on $\mathbb{R}^n$, by Exercise 3.4.17. More generally, all Lebesgue–Stieltjes measures on $\mathbb{R}$ are Radon measures by Theorem 3.7.1.

Example 3.8.1 Suppose that μ is a Borel measure on a metric space X and f is a nonnegative Σ^μ-measurable function on X. Let ν be the measure on $\mathcal{B}(X)$ defined by

$$\nu(A) = \int_A f d\mu$$

for $A \in \mathcal{B}(X)$ (cf. Exercise 2.5.7). We shall call ν the indefinite integral of f with respect to μ, or simply the μ-indefinite integral of f, and denote it by $\{f\mu\}$. The measure on X constructed from $\{f\mu\}$ by Method I is denoted by $\{f\mu\}^*$; $\{f\mu\}^*$ is the unique Borel regular measure on X such that $\{f\mu\}^*(A) = \{f\mu\}(A)$ for $A \in \mathcal{B}(X)$, by Corollary 3.4.1; it is for the Borel regularity of $\{f\mu\}^*$ that our construction starts, with $\{f\mu\}$ being originally defined on $\mathcal{B}(X)$. If, further, f is μ-integrable on every compact subset of X, then $\{f\mu\}^*$ is a Radon measure. Note that if μ is σ-finite and Borel regular, then for any Σ^μ-measurable set S, $\{f\mu\}^*(S) = \int_S f d\mu$. Actually, there is a Borel set $B \supset S$ such that $\mu(B \backslash S) = 0$ and then there is a Borel set $C \supset (B \backslash S)$ such that $\mu(C) = 0$, implying that $\{f\mu\}^*(B \backslash S) \leq \{f\mu\}^*(C) = \{f\mu\}(C) = \int_C f d\mu = 0$; consequently,

$$\{f\mu\}^*(B) \leq \{f\mu\}^*(S) + \{f\mu\}^*(B \backslash S) = \{f\mu\}^*(S) \leq \{f\mu\}^*(B),$$

from which follows that $\{f\mu\}^*(S) = \{f\mu\}^*(B) = \{f\mu\}(B) = \int_B f d\mu = \int_S f d\mu$.

When μ is the Lebesgue measure on $\mathbb{R}^n$ and X is a Lebesgue measurable set in $\mathbb{R}^n$, $\{f\mu\}$ and $\{f\mu\}^*$ will be replaced by $\{f\}$ and $\{f\}^*$ respectively for compactness of expression.

The following proposition asserts that a measure constructed from a premeasure by Method II on a metric space X is Borel regular if the domain of the premeasure consists of Borel sets of X.

Proposition 3.8.1 *Suppose that X is a metric space and τ a premeasure defined on $\mathcal{G} \subset \mathcal{B}(X)$. Then the measure τ^d on X constructed from τ by Method II is Borel regular.*

Proof Let $A \subset X$. We may assume that $\tau^d(A) < \infty$. For each $k \in \mathbb{N}$, there is a sequence $\{C_n^{(k)}\}$ in the domain of τ such that $\bigcup_n C_n^{(k)} \supset A$, $\operatorname{diam} C_n^{(k)} \le \frac{1}{k}$ for each n, and $\sum_n \tau(C_n^{(k)}) \le \tau_{\frac{1}{k}}(A) + \frac{1}{k} \le \tau^d(A) + \frac{1}{k}$. Let $B = \bigcap_k \bigcup_n C_n^{(k)}$, then $B \in \mathcal{B}(X)$ because each $C_n^{(k)} \in \mathcal{B}(X)$. Since $A \subset B$, $\tau^d(A) \le \tau^d(B)$; but $\tau^d(B) = \lim_{k\to\infty} \tau_{\frac{1}{k}}(B) \le \liminf_{k\to\infty} \sum_n \tau(C_n^{(k)}) \le \liminf_{k\to\infty}\{\tau^d(A) + \frac{1}{k}\} = \tau^d(A)$, hence $\tau^d(B) = \tau^d(A)$. Recall that

$$\tau_{\frac{1}{k}}(B) = \inf \sum_n \tau(C_n),$$

where the infimum is taken over all sequences $\{C_n\} \subset \mathcal{G}$ such that $\bigcup_n C_n \supset B$ and $\operatorname{diam} C_n \le \frac{1}{k}$ for all n, hence, $\tau_{\frac{1}{k}}(B) \le \sum_n \tau(C_n^{(k)})$. ∎

Recall that if μ is a measure on Ω and $A \subset \Omega$, then the restriction to A of μ, denoted $\mu\lfloor A$, is defined by $\mu\lfloor A(B) = \mu(A \cap B)$ for $B \subset \Omega$ (cf. Exercise 3.1.3).

Proposition 3.8.2 *Let μ be a Borel regular measure on a metric space X and suppose that $A \subset X$ is μ-measurable and $\mu(A) < +\infty$. Then $\mu\lfloor A$ is a Radon measure.*

Proof Let $\nu \equiv \mu\lfloor A$. Clearly, $\nu(K) < +\infty$ for compact K; actually, $\nu(S) \le \mu(A) < \infty$ for any $S \subset X$. Since every μ-measurable set is ν-measurable, ν is a Borel measure. It remains to show that ν is Borel regular. There is a Borel set B such that $A \subset B$ and $\mu(A) = \mu(B) < +\infty$. Hence, $\mu(B \setminus A) = \mu(B) - \mu(A) = 0$. For $C \subset X$, we have

$$\begin{aligned}\nu(C) &\le (\mu\lfloor B)(C) = \mu(B \cap C) = \mu(C \cap B \cap A) + \mu((C \cap B) \cap A^C)\\ &\le \mu(C \cap A) + \mu(B \cap A^C) = \nu(C).\end{aligned}$$

Hence, $\nu(C) = (\mu\lfloor B)(C)$. We may assume then that A is Borel. Let now $C \subset X$; there is a Borel set $E \supset A \cap C$ such that $\mu(E) = \mu(A \cap C)$. Let $D = E \cup A^c$; D is a Borel set and $C \subset (A \cap C) \cup A^c \subset D$. Since $D \cap A = E \cap A$,

$$\nu(C) \le \nu(D) = \mu(D \cap A) = \mu(E \cap A) \le \mu(E) = \mu(A \cap C) = \nu(C),$$

implying, $\nu(C) = \nu(D)$. ∎

3.9 Measure-theoretical approximation of sets in $\mathbb{R}^n$

This section is devoted to considering measure-theoretical approximation of sets in $\mathbb{R}^n$ by sets of familiar structure, such as open, closed, and compact sets. We observe first two easy and useful facts about open sets in $\mathbb{R}^n$. For this purpose, we call an oriented rectangle $I_1 \times \cdots \times I_n$ in $\mathbb{R}^n$ an oriented cube, if $|I_1| = \cdots = |I_n|$, and call it **nondegenerate** if $|I_j| > 0$ for all $j = 1, \ldots, n$. Oriented rectangles I and J are said to be **nonoverlapping** if $\mathring{I} \cap \mathring{J} = \emptyset$.

Proposition 3.9.1 *Every open set G in $\mathbb{R}^n$ is the union of a countable family of nondegenerate and mutually nonoverlapping closed oriented cubes.*

Proof Let $k \in \mathbb{N}$; we call an oriented closed cube $I_1 \times \cdots \times I_n$ a dyadic cube of order k if $I_j = \left[\frac{l_j}{2^k}, \frac{l_j+1}{2^k}\right]$, where l_j is an integer for each $j = 1, \ldots, n$. Let $\mathcal{F}_1$ be the family of all those dyadic cubes of order 1 which are contained in G; then let $\mathcal{F}_2$ be the family of all those dyadic cubes of order 2 which are contained in G and are nonoverlapping with those in $\mathcal{F}_1$; proceeding in this fashion we obtain a sequence $\{\mathcal{F}_j\}$ of families of oriented cubes in G such that cubes in each $\mathcal{F}_j$ are mutually nonoverlapping, and nonoverlapping with those in the preceding families if $j \geq 2$. Note that some of the $\mathcal{F}_j$'s might be empty. Let $\mathcal{F} = \bigcup_j \mathcal{F}_j$, then $\mathcal{F}$ is a countable family of nondegenerate and mutually nonoverlapping closed cubes such that $G = \bigcup \mathcal{F}$. ■

Proposition 3.9.2 *Let G be an open set in $\mathbb{R}^n$, then there is an increasing sequence $\{K_j\}$ of compact sets such that*

$$G = \bigcup_{j=1}^{\infty} K_j. \tag{3.5}$$

Proof By Proposition 3.9.1, there is a countable family $\{C_k\}$ of nondegenerate and mutually nonoverlapping closed oriented cubes such that $G = \bigcup_k C_k$. Put $K_j = \bigcup_{k=1}^{j} C_k$, then $\{K_j\}$ is an increasing sequence of compact sets such that (3.5) holds. ■

Remark As a consequence of Proposition 3.9.2, $\mathcal{B}^n$ is the σ-algebra generated by the family of all compact sets.

Lemma 3.9.1 *Suppose that μ is a Borel measure on $\mathbb{R}^n$ and B is a Borel set with $\mu(B) < \infty$, then for each $\varepsilon > 0$ there is a compact set $K \subset B$ such that $\mu(B \backslash K) < \varepsilon$.*

Proof Replacing μ by $\mu\lfloor B$ if necessary, we may assume that μ is a finite measure.

Let $\mathcal{M}$ be the family of all those Borel sets B such that for each $\varepsilon > 0$ there are compact sets $K' \subset B$ and $K'' \subset B^c$, such that $\mu(B\backslash K') < \varepsilon$ and $\mu(B^c\backslash K'') < \varepsilon$. We claim first that $\mathcal{M}$ contains all compact sets. Actually, if K is a compact set, for each $\varepsilon > 0$ choose $K' = K$ and choose K'' as follows: since by (3.5) $K^c = \bigcup_{j=1}^{\infty} K_j$, where $\{K_j\}$ is an increasing sequence of compact sets, $\mu(K^c) = \lim_{j\to\infty} \mu(K_j)$, which implies that $\mu(K^c\backslash K_j) < \varepsilon$ if j is sufficiently large; then choose $K'' = K_j$ for such a sufficiently large j. Thus $\mathcal{M}$ contains all compact sets. In particular, $\mathbb{R}^n \in \mathcal{M}$, because $(\mathbb{R}^n)^c = \emptyset$ which is compact. By definition, a Borel set B is in $\mathcal{M}$ if and only if B^c is in $\mathcal{M}$, hence $B^c \in \mathcal{M}$ if $B \in \mathcal{M}$. Now let $\{B_j\}$ be a disjoint sequence in $\mathcal{M}$ and put $B = \bigcup B_j$, then $B^c = \bigcap_j B_j^c$. Given that $\varepsilon > 0$, there are compact sets $K_j' \subset B_j$ and $K_j'' \subset B_j^c$ such that $\mu(B_j\backslash K_j') < \varepsilon 2^{-(j+1)}$ and $\mu(B_j^c\backslash K_j'') < \varepsilon 2^{-(j+1)}$. We have

$$\mu\left(B\backslash \bigcup_{j=1}^{l} K_j'\right) = \sum_{j=1}^{l} \mu(B_j\backslash K_j') + \sum_{j=l+1}^{\infty} \mu(B_j) < \frac{\varepsilon}{2} + \sum_{j=l+1}^{\infty} \mu(B_j) < \varepsilon,$$

if l is sufficiently large, because $\lim_{l\to\infty}\sum_{j=l+1}^{\infty}\mu(B_j)=0$; choose $K'=\bigcup_{j=1}^{l}K_j'$ for such an l. On the other hand,

$$\mu\left(B^c\setminus\bigcap_j K_j''\right)=\mu\left(\bigcap_j B_j^c\setminus\bigcap_j K_j''\right)\le\mu\left(\bigcup_j(B_j^c\setminus K_j'')\right)$$
$$\le\sum_j\mu(B_j^c\setminus K_j'')<\varepsilon;$$

hence, by choosing $K''=\bigcap_j K_j''$, we have shown that $B\in\mathcal{M}$. We have shown therefore that $\mathcal{M}$ is a λ-system. Since $\mathcal{M}$ contains all compact sets, and since the family of all compact sets is a π-system, $\mathcal{M}$ contains $\mathcal{B}^n$ by the $(\pi$-$\lambda)$ theorem, because $\mathcal{B}^n$ is the σ-algebra generated by the family of all compact sets (cf. Remark after Proposition 3.9.2). But $\mathcal{M}\subset\mathcal{B}^n$ by definition, hence $\mathcal{M}=\mathcal{B}^n$. This completes the proof. ■

Lemma 3.9.2 *If μ is a Radon measure on $\mathbb{R}^n$, then for a Borel set B in $\mathbb{R}^n$ and $\varepsilon>0$, there is an open set $U\supset B$ such that $\mu(U\setminus B)<\varepsilon$.*

Proof For each positive integer m let $U_m=B_m(0)$, the open ball with center 0 and radius m. Then $U_m\setminus B$ is a Borel set with $\mu(U_m\setminus B)\le\mu(\overline{U}_m)<+\infty$, and so for $\varepsilon>0$, by Lemma 3.9.1, there is a compact set $K_m\subset U_m\setminus B$ such that

$$\mu((U_m\setminus K_m)\setminus B)=\mu((U_m\setminus B)\setminus K_m)<\varepsilon 2^{-m}.$$

Let $U=\bigcup_m(U_m\setminus K_m)$, then U is open and

$$B=\bigcup_{m=1}^{\infty}(U_m\cap B)\subset\bigcup_{m=1}^{\infty}(U_m\setminus K_m)=U.$$

Now,

$$\mu(U\setminus B)=\mu\left(\bigcup_{m=1}^{\infty}((U_m\setminus K_m)\setminus B)\right)$$
$$\le\sum_{m=1}^{\infty}\mu((U_m\setminus K_m)\setminus B)<\sum_{m=1}^{\infty}\varepsilon\frac{1}{2^m}=\varepsilon.$$ ■

Theorem 3.9.1 *Let μ be a Radon measure on $\mathbb{R}^n$. Then*
(i) *for $A\subset\mathbb{R}^n$,*

$$\mu(A)=\inf\{\mu(U):A\subset U,\ U\text{ is open}\};$$

and
(ii) *for μ-measurable set $A\subset\mathbb{R}^n$,*

$$\mu(A)=\sup\{\mu(K):K\subset A,\ K\text{ is compact}\}.$$

Proof (i) We may assume that $\mu(A)<+\infty$. Suppose first that A is a Borel set. By Lemma 3.9.2, for each $\varepsilon>0$ there is an open $U\supset A$ such that $\mu(U\setminus A)<\varepsilon$, hence

$\mu(U) = \mu(A) + \mu(U \setminus A) < \mu(A) + \varepsilon$, which shows that (i) holds. Now let A be arbitrary. There is a Borel set $B \supset A$ with $\mu(A) = \mu(B)$. Then,

$$\mu(A) = \mu(B) = \inf\{\mu(U) : U \supset B,\ U \text{ is open}\} \geq \inf\{\mu(U) : U \supset A,\ U \text{ is open}\},$$

which establishes (i), because the reverse inequality is obvious.

(ii) Let A be μ-measurable with $\mu(A) < +\infty$ and denote $\mu\lfloor A$ by ν; then by Proposition 3.8.2, ν is a Radon measure. By (i), given $\varepsilon > 0$, there is an open set $U \supset A^c$ with $\nu(U) < \varepsilon$. Let $C = U^c$, C is closed, $C \subset A$, and

$$\mu(A \setminus C) = \nu(\mathbb{R}^n \setminus C) = \nu(C^c) = \nu(U) < \varepsilon,$$

from which,

$$0 \leq \mu(A) - \mu(C) < \varepsilon.$$

But from $\mu(C) = \lim_{k\to\infty} \mu(C_k)$, where $C_k = \{x \in C : |x| \leq k\}$, it follows that there is a compact set $K \subset A$ such that

$$0 \leq \mu(A) - \mu(K) < \varepsilon,$$

and hence,

$$\mu(A) = \sup\{\mu(K) : K \subset A,\ K \text{ is compact}\}.$$

If $\mu(A) = +\infty$, let $A_j = \{x \in A : j - 1 \leq |x| < j\}$, $j = 1, 2, \ldots$ Then each A_j is μ-measurable and

$$\mu(A) = \sum_j \mu(A_j).$$

Since μ is a Radon measure, $\mu(A_j) < +\infty$. By what is proved above, there is a compact set $K_j \subset A_j$ with $\mu(K_j) \geq \mu(A_j) - 2^{-j}$. Now, $\bigcup_j K_j \subset A$ and

$$\lim_{l\to\infty} \mu\left(\bigcup_{j=1}^{l} K_j\right) = \mu\left(\bigcup_{j=1}^{\infty} K_j\right) = \sum_{j=1}^{\infty} \mu(K_j) \geq \sum_{j=1}^{\infty} [\mu(A_j) - 2^{-j}] = \infty.$$

Since $\bigcup_{j=1}^{l} K_j$ is compact for every l, we have

$$\sup\{\mu(K) : K \subset A,\ K \text{ is compact}\} \geq \sup\left\{\mu\left(\bigcup_{j=1}^{l} K_j\right) : l = 1, 2, \ldots\right\} = +\infty. \quad \blacksquare$$

Remark Because of Theorem 3.9.1 (i), a set $E \subset \mathbb{R}^n$ is μ-measurable if and only if $\mu(G) = \mu(G \cap E) + \mu(G \cap E^c)$ for all open sets G, where μ is a Radon measure on $\mathbb{R}^n$.

Corollary 3.9.1 *The Lebesgue measure λ^n is also the measure on $\mathbb{R}^n$ constructed by Method I from the premeasure τ on the family of all oriented closed cubes I, defined by $\tau(I)$ = volume of I.*

Proof Let τ^* be the measure on $\mathbb{R}^n$ constructed from τ by Method I. For $B \subset \mathbb{R}^n$ and any sequence $\{I_k\}$ of oriented closed cubes with $B \subset \bigcup_k I_k$, we have $\lambda^n(B) \leq \sum_k \lambda^n(I_k) = \sum_k \tau(I_k)$, from which follows $\lambda^n(B) \leq \tau^*(B)$. For $B \subset \mathbb{R}^n$ and $\varepsilon > 0$, there is an open set $G \supset B$ such that $\lambda^n(G) \leq \lambda^n(B) + \varepsilon$, by Theorem 3.9.1 (i) (this fact is actually the conclusion of Exercise 3.4.17 (i)). Now, there is a sequence $\{C_k\}$ of nondegenerate and mutually nonoverlapping oriented closed cubes such that $\bigcup_k C_k = G$, by Proposition 3.9.1. Since C_k's are mutually nonoverlapping, $\sum_k \tau(C_k) = \sum_k \lambda^n(C_k) = \lambda^n(G)$, and hence $\sum_k \tau(C_k) = \lambda^n(G) \leq \lambda^n(B) + \varepsilon$. Thus, $\tau^*(B) \leq \sum_k \tau(C_k) \leq \lambda^n(B) + \varepsilon$, from which follows $\tau^*(B) \leq \lambda^n(B)$, and consequently $\tau^*(B) = \lambda^n(B)$. ■

Exercise 3.9.1

(i) Let $A \subset \mathbb{R}^n$ be Lebesgue measurable; show that there is a F_σ set $M \subset A$ with $\lambda^n(A \backslash M) = 0$ (a F_σ-set is a countable union of closed sets).

(ii) Let $f : \mathbb{R}^n \to \mathbb{R}$ be Lebesgue measurable; show that f is equivalent to a Borel measurable function. (Hint: consider first f, which is an indicator function.)

Exercise 3.9.2 Show that a set A in $\mathbb{R}^n$ is measurable if and only if for every $\varepsilon > 0$ there is an open set $G \supset A$ and a closed set $C \subset A$, such that $\lambda^n(G \backslash C) < \varepsilon$.

Exercise 3.9.3 Suppose that f is a Lebesgue integrable function on $\mathbb{R}^n$.

(i) Show that for any given $\varepsilon > 0$, there is a compact set K in $\mathbb{R}^n$ such that $\int_{\mathbb{R}^n \backslash K} |f| d\lambda^n < \varepsilon$.

(ii) Show that $\lim_{|x| \to \infty} \int_{K+x} f d\lambda^n = 0$ for any compact set K in $\mathbb{R}^n$ (recall that $K + x = \{z + x : z \in K\}$).

(iii) Show that $\lim_{|y| \to \infty} \int_{\mathbb{R}^n} |f(x+y) - f(x)| d\lambda^n(x) = 2 \int_{\mathbb{R}^n} |f| d\lambda^n$.

Exercise 3.9.4 Let $w \geq 0$ be integrable on $\mathbb{R}^n$ and let μ be a premeasure defined for open sets G in $\mathbb{R}^n$ by

$$\mu(G) = \int_G w d\lambda^n.$$

Denote by μ^* the measure on $\mathbb{R}^n$ constructed from μ by Method I.

(i) Show that $\mu^*(S) = \inf \mu(G)$, where the infimum is taken over all open sets G containing S.

(ii) Show that μ^* is a Carathéodory measure and

$$\mu^*(B) = \int_B w d\lambda^n$$

for Borel sets B.

(iii) Show that $\mathcal{L}^n \subset \Sigma^{\mu^*}$ and $\mu^*(A) = \int_A w d\lambda^n$ if $A \in \mathcal{L}^n$.

Exercise 3.9.5 Suppose that μ is a measure on a metric space X with the property that compact sets are μ-measurable. Let $E \subset A$ be subsets of X of which E is not μ-measurable. Show that there exists $\varepsilon > 0$ such that, if $K_1 \subset E$ and $K_2 \subset A \backslash E$ are compact sets, we always have $\mu(A \backslash (K_1 \cup K_2)) \geq \varepsilon$.

3.10 Riesz measures

We introduce now a class of Radon measures on a locally compact metric space X, which has its origin in the work of F. Riesz on representation of bounded linear functionals on $C[a, b]$ by measures; and we therefore refer to measures in this class as Riesz measures.

Consider and fix a locally compact metric space X. We shall denote by $\mathcal{G}$ the family of all open subsets of X, and by $\mathcal{K}$ the family of all compact subsets of X. A Radon measure μ on X is called a **Riesz measure** if it satisfies the following conditions:

(i) For $A \subset X$,

$$\mu(A) = \inf\{\mu(G) : G \supset A, G \in \mathcal{G}\};$$

(ii) for $G \in \mathcal{G}$,

$$\mu(G) = \sup\{\mu(K) : K \subset G, K \in \mathcal{K}\}.$$

Henceforth, condition (i) and condition (ii) will be referred to respectively as **outer regularity** and **inner regularity** of μ. Note that all Radon measures on $\mathbb{R}^n$ are Riesz measures, according to Theorem 3.9.1. Actually, conclusion (ii) of Theorem 3.9.1 is stronger than inner regularity for Riesz measures; but the following proposition claims that finite Riesz measures satisfy the same conclusion as that of Theorem 3.9.1 (ii).

Proposition 3.10.1 *If μ is a finite Riesz measure on X, then for any μ-measurable set A, we have*

$$\mu(A) = \sup\{\mu(K) : K \in \mathcal{K}, K \subset A\}.$$

Proof Let $\varepsilon > 0$. There is $K_0 \in \mathcal{K}$ such that

$$\mu(K_0^c) = \mu(X \backslash K_0) < \frac{\varepsilon}{2},$$

by the inner regularity of μ, and there is $G \in \mathcal{G}$ such that $G \supset A^c$ and

$$\mu(G \cap A) = \mu(G \backslash A^c) < \frac{\varepsilon}{2},$$

by the outer regularity of μ. Now, $K_0 \cap G^c$ is a compact set contained in A and

$$A \backslash (K_0 \cap G^c) = A \cap (K_0 \cap G^c)^c = A \cap (K_0^c \cup G) \subset K_0^c \cup (A \cap G),$$

hence $\mu(A \backslash (K_0 \cap G^c) \leq \mu(K_0^c) + \mu(A \cap G) < \varepsilon$, i.e.

$$\mu(A) < \mu(K_0 \cap G^c) + \varepsilon \leq \sup\{\mu(K) : K \in \mathcal{K}, K \subset A\} + \varepsilon.$$

Letting $\varepsilon \searrow 0$, we have

$$\mu(A) \leq \sup\{\mu(K) : K \in \mathcal{K}, K \subset A\}.$$

That $\mu(A) \geq \sup\{\mu(K) : K \in \mathcal{K}, K \subset A\}$ is obvious. ■

Suppose now that X is locally compact, and denote as in Section 1.10 by $C_c(X)$ the space of all real continuous functions on X with compact support, and if $G \in \mathcal{G}$ by $U_c(G)$ the family of all those functions in $C_c(X)$ such that $0 \leq f \leq 1$ and $\operatorname{supp} f \subset G$. Our main purpose of this section is to construct a Riesz measure on X for each positive linear functional on $C_c(X)$. A linear functional ℓ on a vector space of functions on a set is said to be **positive** if $\ell(f) \geq 0$ whenever $f \geq 0$. Given a positive linear functional ℓ on $C_c(X)$, a related measure μ on X is constructed as follows. Define first a premeasure τ on $\mathcal{G}$ by

$$\tau(G) = \sup\{\ell(f) : f \in U_c(G)\}, \quad G \in \mathcal{G};$$

then for $A \subset X$, define

$$\mu(A) = \inf\{\tau(G) : G \supset A, G \in \mathcal{G}\}.$$

Observe that

(1) $\mu(G) = \tau(G)$ for $G \in \mathcal{G}$;
(2) $\mu(\bigcup_{j=1}^n G_j) \leq \sum_{j=1}^n \mu(G_j)$ if $G_1, \ldots, G_n$ are in $\mathcal{G}$; furthermore, if G_j's are disjoint, then $\mu(\bigcup_{j=1}^n G_j) = \sum_{j=1}^n \mu(G_j)$.

Clearly, (1) is a direct consequence of the obvious fact that $\tau(G_1) \leq \tau(G_2)$, if G_1 and G_2 are in $\mathcal{G}$ and $G_1 \subset G_2$. To verify (2), let $u \in U_c(\bigcup_{j=1}^n G_j)$ and put $K = \operatorname{supp} u$. By Theorem 1.10.1, there is a partition of unity $\{u_1, \ldots, u_n\}$ of K subordinate to $\{G_1, \ldots, G_n\}$; one sees readily that $u = \sum_{j=1}^n uu_j$. Since each uu_j is in $U_c(G_j)$, $\ell(u) = \sum_{j=1}^n \ell(uu_j) \leq \sum_{j=1}^n \tau(G_j) = \sum_{j=1}^n \mu(G_j)$, from which it follows that $\mu(\bigcup_{j=1}^n G_j) =$

$\tau(\bigcup_{j=1}^{n} G_j) \le \sum_{j=1}^{n} \mu(G_j)$. Thus the first part of (2) is verified. Now if $G_1, \ldots, G_n$ are disjoint, we need to show that $\mu(\bigcup_{j=1}^{n} G_j) \ge \sum_{j=1}^{n} \mu(G_j)$. For this purpose, since $\mu(\bigcup_{j=1}^{n} G_j) \ge \mu(G_j)$ for each j, we may assume that $\mu(G_j) < \infty$ for each j. Given $\varepsilon > 0$, there is $u_j \in U_c(G_j)$ such that $\mu(G_j) = \tau(G_j) < \ell(u_j) + \frac{\varepsilon}{n}$ for each j. Then, $u = \sum_{j=1}^{n} u_j \in U_c(\bigcup_{j=1}^{n} G_j)$, because G_j's are disjoint, and hence

$$\mu\left(\bigcup_{j=1}^{n} G_j\right) = \tau\left(\bigcup_{j=1}^{n} G_j\right) \ge \ell(u) = \sum_{j=1}^{n} \ell(u_j) \ge \sum_{j=1}^{n} \mu(G_j) - \varepsilon,$$

from which $\mu(\bigcup_{j=1}^{n} G_j) \ge \sum_{j=1}^{n} \mu(G_j)$ follows by letting $\varepsilon \to 0$. Thus (2) is verified.

We show next that μ is a Carathéodory measure on X. Let $\{A_n\}$ be a sequence of subsets of X; we claim that $\mu(\bigcup_n A_n) \le \sum_n \mu(A_n)$. For this, we may assume that $\mu(A_n) < \infty$ for all n. Given $\varepsilon > 0$ and $n \in \mathbb{N}$, there is an open set $G_n \supset A_n$ such that $\mu(G_n) < \mu(A_n) + \frac{\varepsilon}{2^n}$. Then for $u \in U_c(\bigcup_n G_n)$, since $\operatorname{supp} u$ is compact, $u \in U_c(\bigcup_{j=1}^{n_0} G_j)$ for some n_0, and we have therefore by (2),

$$\ell(u) \le \mu\left(\bigcup_{j=1}^{n_0} G_j\right) \le \sum_{j=1}^{n_0} \mu(G_j) \le \sum_{n} \mu(G_n) \le \sum_{n} \mu(A_n) + \varepsilon;$$

consequently, $\ell(u) \le \sum_n \mu(A_n) + \varepsilon$ for each $u \in U_c(\bigcup_n G_n)$ and hence $\mu(\bigcup_n G_n) \le \sum_n \mu(A_n) + \varepsilon$. Thus, $\mu(\bigcup_n A_n) \le \mu(\bigcup_n G_n) \le \sum_n \mu(A_n) + \varepsilon$. Since $\varepsilon > 0$ is arbitrary, $\mu(\bigcup_n A_n) \le \sum_n \mu(A_n)$. As $\mu(\emptyset) = 0$ and $\mu(A) \le \mu(B)$ for $A \subset B$ are direct consequences of the definition of μ, μ is a measure on X. Suppose now that A and B are subsets of X with $\rho(A, B) > 0$; if we put $H_1 = \{x \in X : \rho(x, A) < \frac{1}{2}\rho(A, B)\}$, $H_2 = \{x \in X : \rho(x, B) < \frac{1}{2}\rho(A, B)\}$, then H_1 and H_2 are open and disjoint. Now let G be any open set containing $A \cup B$ and put $G_1 = H_1 \cap G$, $G_2 = H_2 \cap G$, then

$$\mu(G) \ge \mu(G \cap (H_1 \cup H_2)) = \mu(G_1) + \mu(G_2) \ge \mu(A) + \mu(B),$$

and consequently, $\mu(A \cup B) \ge \mu(A) + \mu(B)$, or $\mu(A \cup B) = \mu(A) + \mu(B)$. Thus, μ is a Carathéodory measure on X. The measure μ so constructed will be referred to as the **measure** for the positive linear functional ℓ.

Lemma 3.10.1 *Suppose that ℓ is a positive linear functional on $C_c(X)$ and let μ be the measure for ℓ, then μ is a Radon measure on X.*

Proof Since μ is a Carathéodory measure, it is a Borel measure. From the definition of μ, for $A \subset X$ there is a sequence $\{G_n\}$ of open sets such that $\bigcap_n G_n \supset A$ and $\mu(A) = \mu(\bigcap_n G_n)$, hence μ is Borel regular. Now let K be a compact subset of X. By (i) of Section 1.10, K has a compact neighborhood V, for which we know from Corollary 1.10.1 that there is $f \in U_c(X)$ such that $f = 1$ on V. Clearly if $u \in U_c(\mathring{V})$, then $u \le f$. Thus $\mu(K) \le \mu(\mathring{V}) = \sup\{\ell(u) : u \in U_c(\mathring{V})\} \le \ell(f) < \infty$. We have shown that μ is a Radon measure on X. ■

Lemma 3.10.2 *Suppose that ℓ is a positive linear functional on $C_c(X)$ and μ is the measure for ℓ. Then,*

$$\ell(f) = \int_X f d\mu$$

for $f \in C_c(X)$.

Proof Let $f \in C_c(X)$ and put $K = \operatorname{supp} f$. Given $\varepsilon > 0$, for $j \in \mathbb{Z}$, let $E_j = \{x \in K : \varepsilon j < f(x) \leq \varepsilon(j+1)\}$. As f is necessarily bounded, $E_j = \emptyset$ if $|j| > k$ for some $k \in \mathbb{N}$. Since $\mu(E_j) \leq \mu(K) < \infty$, for each j with $|j| \leq k$, there is an open set $G_j \supset E_j$ such that $\mu(G_j \backslash E_j) < \frac{1}{(2k+1)(|j|+2)}$ and $f(x) \leq \varepsilon(j+2)$ for $x \in G_j$. There is a partition of unity $\{u_j\}_{|j|\leq k}$ of K subordinate to the finite covering $\{G_j\}_{|j|\leq k}$ of K, by Theorem 1.10.1. Then, $f = \sum_{|j|\leq k} fu_j$ and hence

$$\begin{aligned}
\ell(f) &= \sum_{|j|\leq k} \ell(fu_j) \leq \sum_{|j|\leq k} \varepsilon(j+2)\ell(u_j) \leq \sum_{|j|\leq k} \varepsilon(j+2)\mu(G_j) \\
&\leq \sum_{|j|\leq k} \varepsilon(j+2)\left\{\mu(E_j) + \frac{1}{(2k+1)(|j|+2)}\right\} \\
&\leq \int_X f d\mu + 2\varepsilon\mu(K) + \varepsilon,
\end{aligned}$$

and consequently, since $\varepsilon > 0$ is arbitrary, we have

$$\ell(f) \leq \int_X f d\mu;$$

but in the last inequality, if we replace f by $(-f)$, we also have $\ell(f) \geq \int_X f d\mu$, and thus

$$\ell(f) = \int_X f d\mu. \qquad \blacksquare$$

Corollary 3.10.1 *If G is an open set in X, then*

$$\mu(G) = \sup\{\mu(K) : K \subset G, K \in \mathcal{K}\}.$$

Proof It is sufficient to show that

$$\mu(G) \leq \sup\{\mu(K) : K \subset G, K \in \mathcal{K}\}.$$

Let $f \in U_c(G)$, then since $f \leq 1$, we have

$$\ell(f) = \int_X f d\mu = \int_{\operatorname{supp} f} f d\mu \leq \mu(\operatorname{supp} f),$$

from which we infer that

$$\sup\{\mu(K) : K \subset G, K \in \mathcal{K}\} \geq \sup\{\ell(f) : f \in U_c(G)\} = \mu(G). \qquad \blacksquare$$

From Corollary 3.10.1 and the definition of μ, the Radon measure μ is both outer regular and inner regular. Hence, the measure for any positive linear functional on $C_c(X)$ is a Riesz measure.

Theorem 3.10.1 *The measure μ for a positive linear functional ℓ on $C_c(X)$ is the unique Riesz measure on X, such that*

$$\ell(f) = \int_X f d\mu \tag{3.6}$$

for all $f \in C_c(X)$.

Proof Since the measure μ for ℓ is a Riesz measure on X for which (3.6) holds, it remains to show that if ν is a Riesz measure on X, such that $\ell(f) = \int_X f d\nu$ for all $f \in C_c(X)$, then $\nu = \mu$. To show $\nu = \mu$, it is sufficient to show that $\nu(G) = \mu(G)$ for all $G \in \mathcal{G}$, because both ν and μ are outer regular. Let now $G \in \mathcal{G}$. For $f \in U_c(G)$, $\nu(G) \geq \int_X f d\nu = \ell(f)$ implies

$$\nu(G) \geq \sup\{\ell(f) : f \in U_c(G)\} = \mu(G).$$

To see $\nu(G) \leq \mu(G)$, consider any given compact set $K \subset G$ and choose according to Corollary 1.10.1 a function f in $U_c(G)$ such that $f = 1$ on K. For such a function f, we have

$$\nu(K) \leq \int_X f d\nu = \ell(f) \leq \mu(G).$$

Thus, $\nu(G) = \sup\{\nu(K) : K \in \mathcal{K}, K \subset G\} \leq \mu(G)$. $\blacksquare$

Exercise 3.10.1 Define a norm for $f \in C_c(X)$ by $\|f\| = \sup_{x \in X} |f(x)| = \max_{x \in X} |f(x)|$. Show that if ℓ is a bounded positive linear functional on $C_c(X)$ as a n.v.s. with the norm previously defined, then the measure μ for ℓ is a finite measure and $\|\ell\| = \mu(X)$.

Exercise 3.10.2 Suppose that X is a compact metric space. Show that a positive linear functional on $C(X)$ is necessarily a bounded linear functional on $C(X)$.

Exercise 3.10.3 Let ℓ be a positive linear functional on $C[0, 1]$ and let μ be the measure for ℓ. Define a function g on $[0, 1]$ by $g(x) = \mu([0, x])$ for $x \in (0, 1]$ and $g(0) = 0$. Show that the Lebesgue–Stielties measure μ_g is μ.

3.11 Existence of nonmeasurable sets

We exhibit here a nonmeasurable set in $\mathbb{R}$. For this purpose we prove first a remarkable property of measurable sets in $\mathbb{R}$.

Proposition 3.11.1 *Let A be a measurable set in $\mathbb{R}$ with $\lambda(A) > 0$, then $D := \{x - y : x, y \in A\}$ contains a nondegenerate interval.*

Proof We may assume that $\lambda(A) < \infty$. There is an open set $U \supset A$ such that

$$\lambda(U) < \left(1 + \frac{1}{3}\right)\lambda(A). \tag{3.7}$$

Since $U = \bigcup_k I_k$, where $\{I_k\}$ is a disjoint sequence of open intervals, we have $\lambda(A) = \sum_k \lambda(A \cap I_k)$, and hence, in view of (3.7),

$$\lambda(I_{k_0}) < \left(1 + \frac{1}{3}\right)\lambda(A \cap I_{k_0}) \tag{3.8}$$

for some k_0. We now verify that $I := (-\frac{1}{2}\lambda(I_{k_0}), \frac{1}{2}\lambda(I_{k_0})) \subset D$. Let $t \in I$, $t \neq 0$, i.e. $0 < |t| < \frac{1}{2}\lambda(I_{k_0})$, then $(A \cap I_{k_0}) \cup (A \cap I_{k_0} + t)$ is contained in an interval of length $< \frac{3}{2}\lambda(I_{k_0})$. If $(A \cap I_{k_0}) \cap (A \cap I_{k_0} + t) = \emptyset$, by (3.8),

$$\lambda((A \cap I_{k_0}) \cup (A \cap I_{k_0} + t)) = 2\lambda(A \cap I_{k_0}) > 2 \cdot \frac{3}{4}\lambda(I_{k_0}) = \frac{3}{2}\lambda(I_{k_0}),$$

which contradicts the fact that $(A \cap I_{k_0}) \cup (A \cap I_{k_0} + t)$ is contained in an interval of length $< \frac{3}{2}\lambda(I_{k_0})$. Thus, $(A \cap I_{k_0}) \cap (A \cap I_{k_0} + t) \neq \emptyset$; say $x = y + t$ for some x and y in $A \cap I_{k_0}$, then $t = x - y \in D$. This shows that $I \subset D$, because $t = 0$ is certainly in D. ■

For $x \in \mathbb{R}$, let $[x]$ denote the set of all those numbers y in $\mathbb{R}$ such that $x - y$ is rational. It is clear that for x and y in $\mathbb{R}$, $[x]$ and $[y]$ are either disjoint or the same set, and $[x] = [y]$ if and only if $x - y$ is rational; in particular, $[x]$ is the set of all rational numbers if x is rational and each set $[x]$ is countable. Let S be a subset of $\mathbb{R}$ which contains exactly one point of each $[x]$. The possibility of choosing such a set follows from the **axiom of choice**, which states that from any given family of sets in a universal set, a set can be formed by choosing exactly one element from each set of the family. We note that axiom of choice is consistent with the usual logic adopted in mathematics, and we accept it as an axiom in our discourse. Returning to our set S, we observe first that $\mathbb{R} = \bigcup_\alpha (S + \alpha)$, where the union is taken over all rational numbers α. Actually, if $s_x = S \cap [x]$, then $\mathbb{R} \supset \bigcup_\alpha (S + \alpha) \supset \bigcup_{x \in \mathbb{R}} \bigcup_\alpha \{s_x + \alpha\} = \bigcup_{x \in \mathbb{R}} [x] = \mathbb{R}$. It follows then $\lambda(S) > 0$, because if $\lambda(S) = 0$, $\lambda(S + \alpha) = 0$ for all rational number α and $\infty = \lambda(\mathbb{R}) \leq \sum_\alpha \lambda(S + \alpha) = 0$; which is absurd. Next, note that if x and y are distinct elements of S, then $x - y$ is irrational (otherwise, x and y are from $[x]$, contradicting the fact that $S \cap [x]$ consists of one element). This implies that each element of the set $D_0 := \{x - y : x, y \in S\}$ other than 0 is irrational; consequently D_0 contains no nonempty interval. Now, should S be measurable, D_0 would contain a nonempty interval, by Proposition 3.11.1. Thus, S is nonmeasurable. This asserts the existence of nonmeasurable sets in $\mathbb{R}$.

Proposition 3.11.2 *If A is a measurable subset of $\mathbb{R}$ with positive measure, then A contains a nonmeasurable set.*

Proof Let S be the nonmeasurable set, previously constructed, and let D_0 be the difference set $S - S$, as defined before. Observe first that if E is a measurable set in S, then $\lambda(E) = 0$, because if $\lambda(E) > 0$; by Proposition 3.11.1 the difference set $E - E$ contains a nonempty interval, then so does D_0, contrary to the fact that D_0 contains no nonempty interval. Similarly, if E is a measurable set in $S + \alpha$, where α is a real number, then $\lambda(E) = 0$.

Suppose now that A contains no nonmeasurable subset, then $A \cap \{S + \alpha\}$ is measurable for each rational number α and hence $\lambda(A \cap \{S + \alpha\}) = 0$, from the previous observation. But we know that $\mathbb{R} = \bigcup_\alpha \{S + \alpha\}$, where the union is over all rational numbers α, thus,

$$\lambda(A) \leq \sum_\alpha \lambda(A \cap \{S + \alpha\}) = 0,$$

contrary to the assumption that $\lambda(A) > 0$. The contradiction asserts that A contains a nonmeasurable subset. ∎

3.12 The axiom of choice and maximality principles

We have mentioned and used the **axiom of choice** in Section 3.11, when constructing a nonmeasurable set in $\mathbb{R}$. A more explicit discussion on the axiom of choice will now be made together with introduction of two maximality principles which are equivalent to the axiom of choice. The alluded maximality principles are **Hausdorff's maximality principle** and **Zorn's lemma**, which are often used in construction of mathematical objects.

Suppose that X is a nonempty set; a mapping f from $2^X \backslash \{\emptyset\}$ to X is called a **choice function** for X, if $f(A) \in A$ for each nonempty subset A of X. It is clear that the axiom of choice stated in Section 3.11 can be put in the following form:

> **Axiom of choice**. For every nonempty set X, there is a choice function for X.

A binary relation $\leq$ between some pairs of elements of a nonempty set X is called a **partial order** on X if (i) $x \leq x$ for all $x \in X$; (ii) $x \leq y$ and $y \leq z$ for x, y, and z in X, then $x \leq z$; and (iii) $x \leq y$ and $y \leq x$ result in $x = y$. X is then said to be **partially ordered** by $\leq$. By a **partially ordered set** X we understand a nonempty set partially ordered by a certain partial order.

A familiar situation is when X is a family of subsets of a given set, then X is partially ordered by set inclusion, i.e. for sets A and B in X, $A \leq B$ if and only if $A \subset B$. Such X is always considered as partially ordered in this way.

An element x in a partially ordered set X is said to be **maximal** if $x \leq y$ for y in X; then $y = x$; in the case where X is a family of subsets of a given set, then a set A in X is maximal means that A is not a proper subset of any set in X. For example, if X is the family of all proper vector subspaces of a vector space V and is ordered by set inclusion; then maximal elements of X are called **hyperplanes** in V.

Let x, y be elements of a partially ordered set X; x is said to be comparable to y if either $x \leq y$ or $y \leq x$ holds; then x and y are comparable to each other. A nonempty subset C of a partially ordered set X is called a **chain** in X if any two elements of C are comparable to each other.

Hausdorff's maximality principle. In any partially ordered set X, there exists a maximal chain. In other words, there is a chain in X which is not contained in another chain properly.

If A is a nonempty subset of a partially ordered set X, then an element b of X is called an upper bound of A if $a \leq b$ holds for all $a \in X$.

Zorn's lemma. If every chain in a partially ordered set X has an upper bound, then X has a maximal element.

It is easy to see that Zorn's lemma follows from Hausdorff's maximality principle. By Hausdorff maximality principle, there is a maximal chain C in X, then C has an upper bound b in X, by the assumption of Zorn's lemma; then b is a maximal element of X, because, otherwise, there is x in X such that $b \leq x$ and $b \neq x$, implying that the chain $C \cup \{x\}$ contains C properly.

We show next that the axiom of choice is a consequence of the validity of Zorn's lemma. Given a nonempty set X, let $\mathcal{F} = 2^X \backslash \{\emptyset\}$, and consider the set Y of all those mappings f with its domain $D(f) \subset \mathcal{F}$ and range in X, such that $f(A) \in A$ for $A \in D(f)$. Y is nonempty because, for any $x \in X$, let $D(f) = \{\{x\}\}$ and $f(\{x\}) = x$, then $f \in Y$. Define a partial order $\leq$ on Y as follows. For f, g in Y, $f \leq g$ if and only if $D(f) \subset D(g)$ and $g(A) = f(A)$ for $A \in D(f)$. Y is obviously partially ordered by $\leq$. Now let C be a chain in Y; define a mapping g with $D(g) = \bigcup_{f \in C} D(f)$ and with $g(A) = f(A)$ if $f \in C$ and $A \in D(f)$. Since C is a chain in Y, g is well defined and belongs to Y. Obviously, g is an upper bound of C. By Zorn's lemma, Y has a maximal element, say f. We claim that f is a choice function for X by showing that $D(f) = \mathcal{F}$. Suppose the contrary, then there is A in $\mathcal{F}$ but not in $D(f)$; choose $x \in A$ and let g be a mapping from $D(f) \cup \{A\}$ to X defined by $g(B) = f(B)$ for $B \in D(f)$ and $g(A) = x$. Then g is in Y, $f \leq g$, and $f \neq g$, contradicting that f is a maximal element in Y. Thus $D(f) = \mathcal{F}$ and f is a choice function for X. Hence the axiom of choice is a consequence of Zorn's lemma.

The rest of this section aims to show that Hausdorff's maximality principle follows from the axiom of choice, completing the establishment of the equivalence among axiom of choice, Hausdorff's maximality principle, and Zorn's lemma.

Let X be a partially ordered set and $\mathcal{F}$ be the family of all chains in X and $\emptyset$. Then $\mathcal{F}$ satisfies the conditions:

(a) If $A \in \mathcal{F}$, then all the subsets of A are in $\mathcal{F}$;

(b) if $\mathcal{C}$ is a chain in $\mathcal{F}$, then $\bigcup \mathcal{C}$ is in $\mathcal{F}$.

In condition (b), $\bigcup \mathcal{C}$ denotes the union of all sets in the family $\mathcal{C}$. By the axiom of choice, there is a choice function f for X. This choice function is fixed throughout the rest of this section. For $A \in \mathcal{F}$, let $\widehat{A} = \{x \in X : A \cup \{x\} \in \mathcal{F}\}$; observe that $\widehat{A} \supset A$ and $\widehat{A} = A$ if and only if A is maximal in $\mathcal{F}$. Define a mapping $\tau : \mathcal{F} \mapsto \mathcal{F}$ by $\tau(A) = A$ if $\widehat{A} = A$, while $\tau(A) = A \cup \{f(\widehat{A}\backslash A)\}$ if $\widehat{A}\backslash A \neq \emptyset$. Since $f(\widehat{A}\backslash A) \in \widehat{A}$ if $\widehat{A}\backslash A \neq \emptyset$, $A \cup \{f(\widehat{A}\backslash A)\} \in \mathcal{F}$ and τ is actually a mapping from $\mathcal{F}$ into $\mathcal{F}$. Observe that $A \subset \tau(A)$ and $\tau(A)\backslash A$ consists of at most one element. Since A is maximal in $\mathcal{F}$ if and only if $\widehat{A} = A$, A is maximal in $\mathcal{F}$ if and only if $\tau(A) = A$; but if $\tau(A) = A$, A is not empty by the fact that $\tau(\emptyset) = \{f(\bigcup \mathcal{F})\} \neq \emptyset$, and thus A is a maximal chain in X. Therefore, in order to establish Hausdorff's maximality principle, it is sufficient to show that $\tau(A) = A$ for some A in $\mathcal{F}$. This is what we shall do in the following.

A subfamily $\mathcal{T}$ of $\mathcal{F}$ is called a **tower** if it satisfies the following conditions:

(i) $\emptyset \in \mathcal{T}$;

(ii) if $A \in \mathcal{T}$, then $\tau(A) \in \mathcal{T}$; and

(iii) if $\mathcal{C}$ is a chain in $\mathcal{T}$, then $\bigcup \mathcal{C} \in \mathcal{T}$.

Since $\mathcal{F}$ is a tower, and the intersection of all towers is a tower, the smallest tower $\mathcal{T}_0$ exists. We shall claim that $\mathcal{T}_0$ is a chain. For this purpose, consider the family $\widehat{\mathcal{T}_0}$ of all those $C \in \mathcal{T}_0$ such that if $A \in \mathcal{T}_0$, either $A \subset C$ or $C \subset A$ holds, i.e. $\widehat{\mathcal{T}_0}$ is the family of all those elements of $\mathcal{T}_0$ which are comparable to all elements of $\mathcal{T}_0$; then for $C \in \widehat{\mathcal{T}_0}$ let $\xi(C)$ be the family of all those $A \in \mathcal{T}_0$ such that either $A \subset C$ or $\tau(C) \subset A$.

Proposition 3.12.1 *Let $C \in \widehat{\mathcal{T}_0}$. Suppose that $A \in \mathcal{T}_0$ and A is a proper subset of C, then $\tau(A) \subset C$.*

Proof Suppose the contrary. Then, since $\tau(A) \in \mathcal{T}_0$, C is a proper subset of $\tau(A)$; but this fact, together with the assumption that A is a proper subset of C, implies that $\tau(A)\backslash A$ contains at least two elements, contradicting the fact that $\tau(A)\backslash A$ contains at most one element. ∎

Proposition 3.12.2 *If $C \in \widehat{\mathcal{T}_0}$, then $\xi(C) = \mathcal{T}_0$.*

Proof It is sufficient to show that $\xi(C)$ is a tower. The conditions (i) and (iii) hold obviously for $\xi(C)$. It remains to show that condition (ii) holds for $\xi(C)$. Let $A \in \xi(C)$, then either $A \subset C$ or $\tau(C) \subset A$. If $\tau(C) \subset A$, then $\tau(C) \subset \tau(A)$, which implies that

$\tau(A) \in \xi(C)$. Otherwise $A \subset C$, i.e. either $A = C$ or A is a proper subset of C; in the latter case, $\tau(A) \in \xi(C)$, by Proposition 3.12.1, while in the former, $\tau(A) = \tau(C)$ implies that $\tau(A) \supset \tau(C)$ and hence $\tau(A) \in \xi(C)$. Thus, condition (ii) holds for $\xi(C)$ and $\xi(C)$ is a tower. ■

We are ready to see that $\mathcal{T}_0$ is a chain. Let $C \in \widehat{\mathcal{T}_0}$. By Proposition 3.12.2, $\xi(C) = \mathcal{T}_0$, which means that if $A \in \mathcal{T}_0$, then either $A \subset C$ or $\tau(C) \subset A$, implying that either $\tau(A) \subset \tau(C)$ or $\tau(C) \subset A$ and consequently $\tau(C) \in \widehat{\mathcal{T}_0}$. Now, $\bigcup \mathcal{C} \in \widehat{\mathcal{T}_0}$ if $\mathcal{C}$ is a chain in $\widehat{\mathcal{T}_0}$ follows immediately from the definition of $\widehat{\mathcal{T}_0}$. As $\emptyset \in \widehat{\mathcal{T}_0}$, we have shown that $\widehat{\mathcal{T}_0}$ is a tower and hence $\widehat{\mathcal{T}_0} = \mathcal{T}_0$. $\widehat{\mathcal{T}_0} = \mathcal{T}_0$ means only that $\mathcal{T}_0$ is a chain.

Finally, let $A = \bigcup \mathcal{T}_0$. Since $\mathcal{T}_0$ is a tower and a chain, $A \in \mathcal{T}_0$ and $\tau(A) \in \mathcal{T}_0$. Then $A = \bigcup \mathcal{T}_0 \supset \tau(A)$, and consequently $\tau(A) = A$. Thus A is a maximal chain in X and therefore Hausdorff's maximality principle holds.

We have concluded that the axiom of choice, Hausdorff's maximality principle, and Zorn's lemma are each equivalent to one another.

4 Functions of Real Variables

This chapter starts a systematic study of properties of functions of real variables, in terms of concepts related to measures. Properties of functions considered in this light are usually referred to as metric properties.

We begin with a characterization of measurable functions due to **N.N. Lusin**. This characterization is an intuitively satisfactory description of measurable functions and has basic and important consequences, in so far as measurable functions are concerned. Riemann integrable functions are then taken up and shown to be Lebesgue integrable and their integrals in either sense are the same.

Push-forward of measures, a natural construct of measures from those given through mappings, is then interposed for the purpose of representation of general integrals as integrals on $\mathbb{R}$, as well as for a transformation formula of the Lebesgue integral of functions on $\mathbb{R}^n$ through change of variables later in the chapter. Then there follows naturally a more detailed study of functions of a real variable, in which considerable emphasis is placed on study of differentiability of functions unfolding from the Lebesgue differentiation theorem for Radon measures on $\mathbb{R}^n$.

Product measures are treated and followed by further studies of functions of several real variables in later sections of the chapter.

A detailed presentation of polar coordinates in $\mathbb{R}^n$ is given in Section 4.11, with applications to integral operators of potential type and integral representation of C^1 functions.

4.1 Lusin theorem

Let μ be a Borel regular measure on $\mathbb{R}^n$, and f a finite-valued function defined on a μ-measurable subset A of $\mathbb{R}^n$. We suppose that $\mu(A) < \infty$. We shall show that f is Σ^{μ}-measurable if and only if it is almost a continuous function; "almost" in the sense given in Theorem 4.1.1. Theorem 4.1.1, is called the Lusin theorem in this book. In the following, μ, A, and f are fixed and specified as previously.

Lemma 4.1.1 *Let h be a simple function defined on A, then for $\varepsilon > 0$, there is a compact set $K \subset A$ such that $h|_K$ is continuous and $\mu(A\backslash K) < \varepsilon$.*

Real Analysis. Fon-Che Liu. © Fon-Che Liu 2016. Published 2016 by Oxford University Press.

Proof In view of Proposition 3.8.2, we may assume that μ is a Radon measure. The simple function h can be expressed as

$$h = \sum_{j=1}^{k} \alpha_j I_{A_j},$$

where $A_1, \ldots, A_k$ are disjoint μ-measurable subsets of A with $A = \bigcup_{j=1}^{k} A_j$. For each $j = 1, \ldots, k$, there is a compact set $K_j \subset A_j$ with $\mu(A_j \backslash K_j) < \frac{\varepsilon}{k}$, by Theorem 3.9.1 (ii). Since $K_1, \ldots, K_k$ are disjoint compact sets, $\operatorname{dist}(K_i, K_j) > 0$ if $i \neq j$; this, together with the fact that h is constant on each K_j, shows that $h|_K$ is continuous if $K := \bigcup_{j=1}^{k} K_j$. Now, $\mu(A \backslash K) = \sum_{j=1}^{k} \mu(A_j \backslash K_j) < \varepsilon$. The Lemma is proved. ■

Theorem 4.1.1 (Lusin) *Suppose that f is finite-valued and Σ^μ-measurable. Then for $\varepsilon > 0$, there is a compact set $K \subset A$ and a continuous function g defined on $\mathbb{R}^n$ such that $\mu(A \backslash K) < \varepsilon$ and $g = f$ on K.*

Proof There is a sequence $\{f_m\}$ of simple functions defined on A such that $\lim_{m\to\infty} f_m(x) = f(x)$ for $x \in A$. By the Egoroff theorem and Theorem 3.9.1 (ii), there is a compact set $K' \subset A$ such that $\mu(A \backslash K') < \frac{\varepsilon}{2}$ and $f_m(x)$ converges to $f(x)$ uniformly for $x \in K'$. For each m, by Lemma 4.1.1, there is a compact set $K_m \subset A$ such that $f_m|_{K_m}$ is continuous and $\mu(A \backslash K_m) < \frac{\varepsilon}{2^{m+1}}$. Set $K'' = \bigcap_{m=1}^{\infty} K_m$, then $f_m|_{K''}$ is continuous for each m, and

$$\mu(A \backslash K'') = \mu\left(\bigcup_{m=1}^{\infty} (A \backslash K_m)\right) < \sum_{m=1}^{\infty} \frac{\varepsilon}{2^{m+1}} = \frac{\varepsilon}{2}.$$

Now let $K = K' \cap K''$, then $\mu(A \backslash K) < \varepsilon$ and

(a) each $f_m|_K$ is continuous;
(b) $f_m|_K$ converges uniformly to $f|_K$.

From (a) and (b) follows the conclusion that $f|_K$ is continuous. By the Tietze Theorem (Theorem 1.8.1) there is a continuous function g on $\mathbb{R}^n$ such that $g = f|_K$ on K, or $g = f$ on K. ■

Concerning the Lusin theorem, we note first that it still holds if f is finite-valued μ-a.e. on A; and secondly, if f is finite-valued μ-a.e. and satisfies the conclusion of the Lusin theorem, then f is Σ^μ-measurable. To see this, we proceed as follows. For each $m \in \mathbb{N}$ there is a compact set $K_m \subset A$ and a continuous function g_m on $\mathbb{R}^n$ such that $\mu(A \backslash K_m) < \frac{1}{m^2}$ and $g_m = f$ on K_m; now $\sum_m \frac{1}{m^2} < \infty$ implies $\mu(\limsup_{m\to\infty}(A \backslash K_m)) = 0$ (cf. Exercise 2.5.9 (i)), which means that μ-a.e. x in A is in K_m if m is sufficiently large (observe that $A \backslash \limsup_{m\to\infty}(A \backslash K_m) = A \backslash \bigcap_{m=1}^{\infty} \bigcup_{l\geq m}(A \backslash K_l) = \bigcup_{m=1}^{\infty} \bigcap_{l \geq m} K_l = \liminf_{m\to\infty} K_m$), or $f(x) = \lim_{m\to\infty} g_m(x)$ and consequently f is Σ^μ-measurable because each g_m is Σ^μ-measurable due to the fact that μ is a Borel measure. Thus the conclusion of the Lusin theorem is a characterization of Σ^μ-measurable functions on A. We state this explicitly as a theorem for later reference and still call it the Lusin theorem.

Theorem 4.1.2 *Suppose that f is finite-valued μ-a.e. on A. Then f is Σ^{μ}-measurable if and only if for any given $\varepsilon > 0$, there is a compact set $K \subset A$ and a continuous function g on $\mathbb{R}^n$ such that $\mu(A \backslash K) < \varepsilon$ and $f = g$ on K.*

Exercise 4.1.1 Let f be a monotone increasing function defined on a finite open interval (a, b) in $\mathbb{R}$. Show that for any $\varepsilon > 0$, there is a continuous and monotone increasing function g on $\mathbb{R}$ such that the set $\{x \in (a, b) : f(x) \neq g(x)\}$ has Lebesgue measure less than ε. Furthermore, if f is bounded on (a, b), g can also be chosen to be bounded by the same bound as that of f.

Exercise 4.1.2 Suppose that f is integrable on $[a, b]$. Show that for each $\varepsilon > 0$ there is $g \in C[a, b]$ such that $\int_a^b |f - g| d\lambda < \varepsilon$. (Hint: prove first that the conclusion holds for bounded measurable function f.)

To conclude this section, we prove that when μ is the Lebesgue measure λ^n on $\mathbb{R}^n$, a characterization of Lebesgue measurable functions defined on an arbitrary Lebesgue measurable subset A of $\mathbb{R}^n$ similar to Theorem 4.1.2 holds.

Theorem 4.1.3 *Let A be a Lebesgue measurable set in $\mathbb{R}^n$. A function f which is defined and finite almost everywhere on A is measurable if and only if for any $\varepsilon > 0$ there is a closed set $F \subset A$ and a continuous function g on $\mathbb{R}^n$ such that $\lambda^n(A \backslash F) < \varepsilon$ and $f = g$ on F.*

Proof The sufficiency part follows from the same arguments that precede the statement of Theorem 4.1.2. We need only consider the necessity part. So, let f be a measurable function which is defined and finite almost everywhere on A, and let $\varepsilon > 0$ be given. Consider the following sequence $\{A_k\}$ of subsets of A : $A_1 = \{x \in A : |x| < 1\}$ and for $k \geq 2$ let $A_k = \{x \in A : k - 1 < |x| < k\}$. Since each set $\{x \in \mathbb{R}^n : |x| = k\}$ has measure zero (see Exercise 3.4.2), $\bigcup_{k=1}^{\infty} A_k$ consists of almost all points of A. Each A_k is measurable and has finite measure. By Theorem 4.1.1, for each k there is a compact set $F_k \subset A_k$ such that $f|_{F_k}$ is continuous and $\lambda^n(A_k \backslash F_k) < \frac{\varepsilon}{2^k}$. Now let $\{g_k\}$ be a sequence of continuous functions defined as follows: g_1 is a continuous function defined on $\{x \in \mathbb{R}^n : |x| \leq 1\}$ such that $g_1 = f|_{F_1}$ on F_1; suppose $g_1, \ldots, g_k$ have been defined, let g_{k+1} be a continuous function defined on $\{x \in \mathbb{R}^n : |x| \leq k + 1\}$ such that $g_{k+1} = g_k$ on $\{|x| \leq k\}$ and $g_{k+1} = f|F_{k+1}$ on F_{k+1}. That $\{g_k\}$ can be so defined is due to Tietze's extension theorem (Theorem 1.8.1). Then define $g(x) = g_k(x)$ if $|x| \leq k$. Obviously, from our construction of the sequence $\{g_k\}$, g is well defined and is continuous on $\mathbb{R}^n$. If we put $F = \bigcup_k F_k$, F is a closed set, $F \subset A$, and $\lambda^n(A \backslash F) = \sum_k \lambda^n(A_k \backslash F_k) < \sum_k \frac{\varepsilon}{2^k} = \varepsilon$. It is clear that $g = f$ on F. ■

4.2 Riemann and Lebesgue integral

In this section an oriented rectangle in $\mathbb{R}^n$ will be called an oriented interval. We show that a Riemann integrable function defined on a closed oriented interval in $\mathbb{R}^n$ is Lebesgue integrable and its Lebesgue integral coincides with its Riemann integral. First, we recall briefly the Riemann integrability. Fix a finite closed oriented interval

$I = [a_1, b_1] \times \cdots \times [a_n, b_n]$, which is not degenerated i.e. $a_i < b_i$, $i = 1, \ldots, n$. Unless stated otherwise, henceforth in this section, an interval is always a finite, closed, nondegenerate, and oriented interval. Two intervals are said to be nonoverlapping if their interiors are disjoint. A **partition** $\mathcal{P}$ of I is a finite family $\{I_j\}_{j=1}^k$ of nonoverlapping intervals such that $I = \bigcup_{j=1}^k I_j$, where k depends on $\mathcal{P}$; in particular, when $I = [a, b]$ is a finite closed interval in $\mathbb{R}$, a partition $\mathcal{P}$ of I is determined by a sequence $a = x_0 < x_1 < \cdots < x_l = b$ of points in $[a, b]$ and we simply call such a sequence of points in $[a, b]$ a partition of $[a, b]$. For a partition $\mathcal{P} = \{I_j\}_{j=1}^k$ of I, $\|\mathcal{P}\|$ will be used to denote $\max_{1 \le j \le k} \operatorname{diam} I_j$, and is called the **mesh** of $\mathcal{P}$.

Consider now a bounded function f defined on I. For an interval $J \subset I$, let $\bar{f}_J = \sup_{x \in J} f(x)$ and $\underline{f}_J = \inf_{x \in J} f(x)$. If $\mathcal{P} = \{I_j\}_{j=1}^k$ is a partition of I, put

$$\bar{S}(f, \mathcal{P}) = \sum_{j=1}^{k} \bar{f}_{I_j} |I_j|; \quad \underline{S}(f, \mathcal{P}) = \sum_{j=1}^{k} \underline{f}_{I_j} |I_j|,$$

where $|J|$ denote the volume of the interval J. A partition $\mathcal{P}$ is said to be **finer** than a partition Q if every interval in Q is a union of intervals in $\mathcal{P}$. One verifies easily that if $\mathcal{P}$ is finer than Q, then

$$\bar{S}(f; \mathcal{P}) \le \bar{S}(f; Q); \quad \underline{S}(f; \mathcal{P}) \ge \underline{S}(f; Q).$$

For partitions $\mathcal{P}$ and Q of I, denote by $\mathcal{P} \vee Q$ the partition of I formed by all the nondegenerate intersections of intervals of $\mathcal{P}$ and those of Q. $\mathcal{P} \vee Q$ is finer than both $\mathcal{P}$ and Q, hence

$$\bar{S}(f; \mathcal{P}) \ge \bar{S}(f; \mathcal{P} \vee Q) \ge \underline{S}(f; P \vee Q) \ge \underline{S}(f; Q),$$

and consequently

$$\inf_{\mathcal{P}} \bar{S}(f; \mathcal{P}) \ge \sup_{\mathcal{P}} \underline{S}(f; \mathcal{P}).$$

$\inf_{\mathcal{P}} \bar{S}(f; \mathcal{P})$ is called the **Darboux upper integral** of f over I and is denoted by $\overline{\int}_I f$, while $\sup_{\mathcal{P}} \underline{S}(f; \mathcal{P})$ is called the **Darboux lower integral** of f over I and is denoted by $\underline{\int}_I f$. We have shown that

$$\underline{\int}_I f \le \overline{\int}_I f;$$

if $\underline{\int}_I f = \overline{\int}_I f$, then the common value, denoted $\int_I f(x)dx$, is called the **Riemann integral** of f over I, and f is then said to be Riemann integrable over I.

Exercise 4.2.1 Show that a bounded function f defined on I is Riemann integrable if and only if for any $\varepsilon > 0$ there is a partition $\mathcal{P}$ of I such that $\bar{S}(f; \mathcal{P}) - \underline{S}(f; \mathcal{P}) < \varepsilon$. In particular, infer that continuous functions defined on I are Riemann integrable.

For a bounded function f on I, we define related functions $\underline{f}$ and $\bar{f}$ as follows:

$$\underline{f}(x) = \lim_{\delta\to 0+} \inf_{|y-x|<\delta} f(y); \quad \bar{f}(x) = \lim_{\delta\to 0+} \sup_{|y-x|<\delta} f(y).$$

Lemma 4.2.1 *$\underline{f}$ is lower semi-continuous and $\bar{f}$ is upper semi-continuous on I. Hence both are Borel measurable, and therefore are Lebesgue measurable.*

Proof Since $\bar{f} = -(\underline{-f})$, we need only show that $\underline{f}$ is lower semi-continuous.

Let $\lambda \in \mathbb{R}$; we shall show that $E_\lambda := \{\underline{f} > \lambda\}$ is open in I. Let $a \in E_\lambda$, then there is $\delta > 0$, such that

$$\inf_{\substack{|y-a|<2\delta \\ y\in I}} f(y) > \lambda.$$

Now let $x \in I$ and $|x-a| < \delta$; then $|y-x| < \delta$ entails that $|y-a| < 2\delta$ and hence,

$$\inf_{\substack{|y-x|<\delta \\ y\in I}} f(y) \geq \inf_{\substack{|y-a|<2\delta \\ y\in I}} f(y) > \lambda.$$

Consequently, $x \in E_\lambda$ and E_λ is open in I. This shows that $\underline{f}$ is lower semi-continuous on I. ■

Lemma 4.2.2 $\underline{\int}_I f = \int_I \underline{f}\, d\lambda^n$, $\overline{\int}_I f = \int_I \bar{f}\, d\lambda^n$.

Proof Choose a sequence $\{\mathcal{P}_k\}$ of partitions of I such that $\lim_{k\to\infty} \underline{S}(f;\mathcal{P}_k) = \underline{\int}_I f$. Since we still have $\lim_{k\to\infty} \underline{S}(f;Q_k) = \underline{\int}_I f$, if each Q_k is finer than $\mathcal{P}_k$, we may assume that $\|\mathcal{P}_k\| \to 0$ as $k\to\infty$. Let $\mathcal{P}_k = \{I_i^{(k)}\}_{i=1}^{n_k}$ and define $f_k(x) = \underline{f}_{I_i^{(k)}}$ if $x \in [I_i^{(k)})$, $i = 1,\ldots,n_k$ and $f_k(x) = 0$ otherwise, where for an interval $J = [c_1,d_1] \times \cdots \times [c_n,d_n]$, $[J)$ denotes the half-open interval $[c_1,d_1) \times \cdots \times [c_n,d_n)$. We claim now that if $x \in I\setminus \bigcup_{k=1}^\infty \bigcup_{i=1}^{n_k} \partial I_i^{(k)}$, then $\lim_{k\to\infty} f_k(x) = \underline{f}(x)$. For each $\delta > 0$, since $\|\mathcal{P}_k\| \to 0$ as $k\to\infty$, $\inf_{|y-x|<\delta} f(y) \leq f_k(x)$, if k is sufficiently large, hence $\inf_{|y-x|<\delta} f(y) \leq \liminf_{k\to\infty} f_k(x)$ and consequently $\underline{f}(x) \leq \liminf_{k\to\infty} f_k(x)$. On the other hand, for each k, $f_k(x) \leq \inf_{|y-x|<\delta} f(y)$ if $\delta > 0$ is small enough, or $f_k(x) \leq \underline{f}(x)$ and hence $\limsup_{k\to\infty} f_k(x) \leq \underline{f}(x)$. Thus,

$$\limsup_{k\to\infty} f_k(x) \leq \underline{f}(x) \leq \liminf_{k\to\infty} f_k(x) \leq \limsup_{k\to\infty} f_k(x),$$

or $\underline{f}(x) = \lim_{k\to\infty} f_k(x)$, as we claim. Now the set $\bigcup_{k=1}^\infty \bigcup_{i=1}^{n_k} \partial I_i^{(k)}$ has Lebesgue measure zero and $|f_k(x)| \leq M := \sup_{x\in I} |f(x)|$; we may apply the Lebesgue dominated convergence theorem to obtain the equality $\underline{\int}_I f = \lim_{k\to\infty} \underline{S}(f;\mathcal{P}_k) = \lim_{k\to\infty} \int_I f_k\, d\lambda^n = \int_I \underline{f}\, d\lambda^n$. Similarly, $\overline{\int} f = \int_I \bar{f}\, d\lambda^n$. ■

Theorem 4.2.1 *A bounded function f on I is Riemann integrable if and only if f is continuous at almost all points of I.*

Proof Since $\underline{f} \leq f \leq \bar{f}$ on I and

$$\underline{\int_I} f = \int_I \underline{f} d\lambda^n \leq \int_I \bar{f} d\lambda^n = \overline{\int_I} f,$$

f is Riemann integrable if and only if $\underline{f} = \bar{f}$ almost everywhere on I. But from Lemma 4.2.1, we know that $\underline{f}$ is lower semi-continuous and $\bar{f}$ is upper semi-continuous; it follows that $\underline{f} = \bar{f}$ almost everywhere on I means, through the inequalities $\underline{f} \leq f \leq \bar{f}$ on I, that f is continuous almost everywhere on I (cf. Exercise 1.5.2 (i) and (ii)). ■

Theorem 4.2.2 *A Riemann integrable function f on I is Lebesgue integrable and $\int_I f(x)dx = \int_I f d\lambda^n$.*

Proof Since $\underline{f} \leq f \leq \bar{f}$ on I, and $\underline{f}(x) = \bar{f}(x)$ for almost all x in I, as we have shown in the proof of Theorem 4.2.1, $f = \underline{f}$ almost everywhere on I and is therefore measurable. As f is bounded and measurable, it is Lebesgue integrable. Now $\int_I f d\lambda^n = \int_I \underline{f} d\lambda^n = \underline{\int_I} f = \int_I f(x)dx$. ■

We note in passing that the function on $[0, 1]$ which takes value 1 on irrational numbers and takes value 0 on rational numbers is not Riemann integrable, but is Lebesgue integrable with Lebesgue integral being 1.

Exercise 4.2.2 Let f be a function defined on $\mathbb{R}$ whose improper integral $\int_{-\infty}^{\infty} f(x)dx$ converges absolutely. Show that f is Lebesgue integrable on $\mathbb{R}$ and $\int_{\mathbb{R}} f d\lambda = \int_{-\infty}^{\infty} f(x)dx$.

Exercise 4.2.3 Give an example to show that the conclusion in Exercise 4.2.2 does not hold if $\int_{-\infty}^{\infty} f(x)dx$ converges, but not absolutely.

We strongly suggest that readers verify that results similar to the conclusion of Exercise 4.2.2 hold for other types of improper integrals.

Notational convention Because of Theorem 4.2.2 and Exercise 4.2.2, we often write $\int_A f d\lambda^n$ as $\int_A f(x)dx$; also, we use $\int_a^b f(x)dx$, $\int_a^{\infty} f(x)dx$, $\int_{-\infty}^b f(x)dx$, and $\int_{-\infty}^{\infty} f(x)dx$ to denote $\int_I f d\lambda$ if I is $[a, b]$, $[a, \infty)$, $(-\infty, b]$, and $(-\infty, \infty)$ in this order. More generally, for a Borel measure μ on $\mathbb{R}$, $\int_a^b f d\mu$, $\int_a^{\infty} f d\mu$, $\int_{-\infty}^b f d\mu$, and $\int_{-\infty}^{\infty} f d\mu$ are similarly connoted.

4.3 Push-forward of measures and distribution of functions

Distribution of a measurable function on a measure space is now considered with its application to representation of the integral of Borel functions of the function as integral on $\mathbb{R}$. For this purpose, a natural method of constructing new measures from one given through mappings will be presented first.

Suppose that μ measures Ω and that t is a map from Ω to a set X; define a set function $t_\#\mu$ on 2^X by

$$t_\#\mu(A) = \mu(t^{-1}A), \quad A \subset X.$$

Obviously, $t_\#\mu$ is a measure on X; it is called the **push-forward** of μ through the map t. Let $A \subset X$ be such that $t^{-1}A$ is μ-measurable, then $t^{-1}A$ is $\mu\lfloor B$-measurable for any subset B of Ω by Exercise 3.1.3 (i). Thus if C is any subset of X, we have, since $(t^{-1}A)^c = t^{-1}A^c$,

$$\mu\lfloor B(t^{-1}C) = \mu\lfloor B(t^{-1}C \cap t^{-1}A) + \mu\lfloor B(t^{-1}C \cap t^{-1}A^c),$$

or,

$$t_\#(\mu\lfloor B)(C) = t_\#(\mu\lfloor B)(C \cap A) + t_\#(\mu\lfloor B)(C \cap A^c).$$

The last equality means that A is $t_\#(\mu\lfloor B)$-measurable for any subset B of Ω. Conversely, suppose that a subset A of X is $t_\#(\mu\lfloor B)$-measurable for any subset B of Ω; then if we choose $C = X$ in the last equality, it follows that

$$t_\#(\mu\lfloor B)(X) = t_\#(\mu\lfloor B)(A) + t_\#(\mu\lfloor B)(A^c),$$

or,

$$\mu\lfloor B(\Omega) = \mu\lfloor B(t^{-1}A) + \mu\lfloor B(\{t^{-1}A\}^c),$$

and hence,

$$\mu(B) = \mu(B \cap t^{-1}A) + \mu(B \cap \{t^{-1}A\}^c)$$

for any subset B of Ω, implying that $t^{-1}A$ is μ-measurable. We have shown the following proposition.

Proposition 4.3.1 *Let A be a subset of X, then, $t^{-1}A$ is μ-measurable if and only if A is $t_\#(\mu\lfloor B)$-measurable for every subset B of Ω.*

Corollary 4.3.1 *A subset A of X is $t_\#\mu$-measurable if $t^{-1}A$ is μ-measurable.*

Exercise 4.3.1 Show that if t is injective, then $A \subset X$ is $t_\#\mu$-measurable if and only if $t^{-1}A$ is μ-measurable.

Proposition 4.3.2 *If μ is a finite regular measure on Ω, then $A \subset X$ is $t_{\#}\mu$- measurable if and only if $t^{-1}A$ is μ-measurable.*

Proof Because of Corollary 4.3.1, we need only show that if A is $t_{\#}\mu$-measurable, then $t^{-1}A$ is μ-measurable.

Choose $C \in \Sigma^{\mu}$ such that $t^{-1}A \subset C$ and $\mu(t^{-1}A) = \mu(C)$. Using the conclusion of Exercise 3.1.4, we have

$$\begin{aligned}\mu(C \cap t^{-1}A^c) + \mu(C \cup t^{-1}A^c) &= \mu(C) + \mu(t^{-1}A^c)\\ &= \mu(t^{-1}A) + \mu(t^{-1}A^c)\\ &= t_{\#}\mu(A) + t_{\#}\mu(A^c) = t_{\#}\mu(X)\\ &= \mu(\Omega) = \mu(C \cup t^{-1}A^c);\end{aligned}$$

since μ is finite, we may cancel out the term $\mu(C \cup t^{-1}A^c)$ from the far left-hand side and the far right-hand side in the above sequence of equalities to obtain $\mu(C \cap (t^{-1}A)^c) = 0$. Thus $C \cap (t^{-1}A)^c$ is μ-measurable. But from $t^{-1}A \subset C$, we have $t^{-1}A = C \backslash (C \cap (t^{-1}A)^c)$ and hence $t^{-1}A$ is μ-measurable. ■

Suppose now that (Ω, Σ, μ) is a measure space and t is a map from Ω into a set X. Let μ^* be the measure on Ω constructed from μ by Method I; μ^* is the unique Σ-regular measure on Ω such that $\mu^*(A) = \mu(A)$ for $A \in \Sigma$ as asserted by Corollary 3.4.1. Define $t_{\#}\Sigma := \{A \subset X : t^{-1}A \in \Sigma\}$. Since $\Sigma \subset \Sigma^{\mu^*}$ and $\mu^*(A) = \mu(A)$ for $A \in \Sigma$, $t_{\#}\Sigma \subset \Sigma^{t_{\#}\mu^*}$ (by Corollary 4.3.1) and $t_{\#}\mu^*(A) = \mu(t^{-1}A)$ for $A \in t_{\#}\Sigma$. For notational simplicity, denote the restriction of $t_{\#}\mu^*$ to $t_{\#}\Sigma$ by $t_{\#}\mu$; then $(X, t_{\#}\Sigma, t_{\#}\mu)$ is a measure space called the **push-forward** of (Ω, Σ, μ) through the map t. Note that the map t from Ω into X is measure-preserving from (Ω, Σ, μ) to $(X, t_{\#}\Sigma, t_{\#}\mu)$ (cf. Section 2.8.2).

Exercise 4.3.2 Let (Ω, Σ, μ) be a measure space and t a map from Ω into a set X.

(i) Show that a function f on X is $t_{\#}\Sigma$-measurable if and only if $f \circ t$ is Σ-measurable.

(ii) Show that if $f \geq 0$ is $t_{\#}\Sigma$-measurable, then $\int_X f dt_{\#}\mu = \int_\Omega f \circ t d\mu$.

(iii) Show that if f is $t_{\#}\Sigma$-measurable, then $\int_X f dt_{\#}\mu = \int_\Omega f \circ t d\mu$ if one of the integrals is meaningful.

(Hint: start with f as an indicator function of a set.)

Example 4.3.1 (Cf. Exercise 3.4.2 (vi)) Suppose that $\Omega = X = \mathbb{R}^n$, and Σ is the σ-algebra $\mathcal{L}^n$ of all Lebesgue measurable sets in $\mathbb{R}^n$.

(i) For $a \in \mathbb{R}^n$ fixed, let t be the mapping $tx = x + a$, $x \in \mathbb{R}^n$. Then $t_{\#}\mathcal{L}^n = \mathcal{L}^n$, $t_{\#}\lambda^n = \lambda^n$, hence,

$$\int_{\mathbb{R}^n} f(x+a)dx = \int_{\mathbb{R}^n} f(x)dx,$$

if $\int_{\mathbb{R}^n} f(x)dx$ exists, i.e. the Lebesgue integral is translation invariant on $\mathbb{R}^n$.

(ii) For $\alpha \in \mathbb{R}$, $\alpha \neq 0$, consider the mapping $tx = \alpha x$. Then, $t_{\#}\mathcal{L}^n = \mathcal{L}^n$, $t_{\#}\lambda^n = \frac{1}{|\alpha|^n}\lambda^n$, hence,

$$\int_{\mathbb{R}^n} f(\alpha x)dx = \frac{1}{|\alpha|^n}\int_{\mathbb{R}^n} f(x)dx,$$

if $\int_{\mathbb{R}^n} f(x)dx$ exists. In particular, take $f = I_{B_1(0)}$, then $\lambda^n(B_r(0)) = r^n\lambda^n(B_1(0))$.

Exercise 4.3.3 Suppose that f is Lebesgue measurable on $\mathbb{R}$ and is periodic with period $l > 0$ i.e. $f(x) = f(x + l)$ for $x \in \mathbb{R}$. Suppose further that f is integrable on $[0, l]$. Show that f is integrable on $[a, a + l]$ and $\int_0^l f d\lambda = \int_a^{a+l} f d\lambda$ for any $a \in \mathbb{R}$.

Exercise 4.3.4 Suppose that t is a continuous and monotone increasing function defined on a finite interval $[a, b]$. Put $c = t(a)$ and $d = t(b)$. Show that for any Borel set $A \subset [c, d]$, $t_{\#}\mu_t(A) = \lambda(A)$, where μ_t is the Lebesgue–Stieltjes measure generated by t. (Hint: for any interval I open in $[c, d]$, $t_{\#}\mu_t(I) = |I|$.)

Suppose now that f is a finite-valued measurable function on a measure space (Ω, Σ, μ). Since f is Σ-measurable, $f_{\#}\Sigma$ contains all Borel subsets of $\mathbb{R}$ and $f_{\#}\mu$ is a measure on $\mathcal{B}$. Considered as a measure on $\mathcal{B}$, $f_{\#}\mu$ is called the **distribution** of f. If g is a Borel function on $\mathbb{R}$, then $g \circ f$ is Σ-measurable and

$$\int_{\mathbb{R}} g df_{\#}\mu = \int_{\Omega} g \circ f d\mu \tag{4.1}$$

if one of the integrals exists. In particular, if g is taken to be $g(t) = |t|^p$, $1 \leq p < \infty$, then

$$\int_{\Omega} |f|^p d\mu = \int_{\mathbb{R}} |t|^p df_{\#}\mu.$$

Thus, $\int_{\Omega} |f|^p d\mu$ can be expressed as an integral on $\mathbb{R}$ w.r.t. the measure $f_{\#}\mu$. When $\mu(\{f \leq t\}) < \infty$ for every $t \in \mathbb{R}$, put

$$F(t) = \mu(\{f \leq t\}) = \mu(f^{-1}(-\infty, t]),$$

then F is a monotone increasing function and we might expect $\int_{\mathbb{R}} |t|^p df_{\#}\mu$ to be the improper Riemann–Stieltjes integral $\int |t|^p dF := \lim_{\substack{b\to\infty \\ a\to-\infty}} \int_a^b |t|^p dF$. We shall see that this is actually true (cf. Exercise 4.5.6).

Exercise 4.3.5 Show that the function F, previously defined, is right-continuous i.e. $F(t) = F(t+)$. Moreover, $\lim_{t\to-\infty} F(t) = 0$, $\lim_{t\to\infty} F(t) = \mu(\Omega)$.

The function F is called the **distribution function** of f. When a function F is mentioned as the distribution function of a measurable function f, it is implicitly assumed that $\mu(\{f \leq t\}) < \infty$ for every $t \in \mathbb{R}$. One sees easily that if f is measurable and finite a.e. on Ω, its distribution $f_{\#}\mu$ and distribution function can be similarly defined.

As we have seen in Section 3.8, F generates a Lebesgue–Stieltjes measure μ_F on $\mathbb{R}$. It turns out that $\mu_F = f_{\#}\mu$ on $\mathcal{B}$, as the following theorem claims.

Theorem 4.3.1 *Suppose that F is the distribution function of a finite-valued measurable function f on a measure space (Ω, Σ, μ). Then, $(\mathbb{R}, \mathcal{B}, f_{\#}\mu) = (\mathbb{R}, \mathcal{B}, \mu_F)$, where μ_F is the Lebesgue–Stieltjes measure generated by F, and*

$$\int_{\Omega} g \circ f d\mu = \int_{\mathbb{R}} g d\mu_F, \tag{4.2}$$

for any Borel measurable function g on $\mathbb{R}$ whose μ_F-integral exists.

Proof Since F is right-continuous, $\mu_F((a, b]) = F(b) - F(a)$, from which by letting $a \to -\infty$, we have

$$\mu_F((-\infty, b]) = F(b) = \mu(f^{-1}(-\infty, b]) = f_{\#}\mu((-\infty, b])$$

for $b \in \mathbb{R}$. Now fix $a \in \mathbb{R}$ and consider the family $\mathcal{F}$ of all $B \in \mathcal{B}$ such that $\mu_F((-\infty, a] \cap B) = f_{\#}\mu((-\infty, a] \cap B)$. It is clear that $\mathcal{F}$ is a λ-system and it contains all sets of the form $(-\infty, b]$, $b \in \mathbb{R}$. Since the family of all sets of the form $(-\infty, b]$, $b \in \mathbb{R}$, is a π-system and $\mathcal{B}$ is the smallest σ-algebra containing all sets of the form $(-\infty, b]$, it follows from the $(\pi$-$\lambda)$ theorem that $\mathcal{F} = \mathcal{B}$. Thus,

$$\mu_F((-\infty, a] \cap B) = f_{\#}\mu((-\infty, a] \cap B)$$

for all $B \in \mathcal{B}$. From this, by letting $a \to \infty$, we infer that $\mu_F(B) = f_{\#}(B)$ for all $B \in \mathcal{B}$, or $(\mathbb{R}, \mathcal{B}, \mu_F) = (\mathbb{R}, \mathcal{B}, f_{\#}\mu)$. Then (4.2) follows from (4.1). ■

In the final part of this section we demonstrate using an example the fact that measure spaces, which look very different from one another in appearance, might be the same measure space in different forms.

Example 4.3.2 Let $(\Omega, \sigma(\mathcal{Q}), P)$ be the Bernoulli sequence space of Example 3.4.6. Define a map $t : \Omega \to [0, 1]$ by

$$t(\omega) = \sum_{j=1}^{\infty} \frac{\omega_j}{2^j}, \quad \omega = (\omega_j) \in \Omega.$$

Note that $0.\omega_1\omega_2\omega_3 \cdots$ is a **binary expansion** of $t(\omega)$. For $x \in [0, 1]$, $t^{-1}x$ consists of either two elements or one element, depending on whether x is a **binary rational number** or not, except that $t^{-1}0$ consists of one element; when x is a binary rational number in $(0, 1]$, say $x = \sum_{j=1}^{n} \frac{\varepsilon_j}{2^j}$ with $\varepsilon_n = 1$, then $t^{-1}x$ consists of $(\varepsilon_1, \ldots, \varepsilon_{n-1}, 1, 0, 0, 0, \ldots)$ and $(\varepsilon_1, \ldots, \varepsilon_{n-1}, 0, 1, 1, 1, \ldots)$. Therefore if we put

$$\widehat{\Omega} = \{\omega \in \Omega : \omega_j = 1 \text{ for infinitely many } j\},$$

then $\Omega \backslash \widehat{\Omega}$ is countable and hence $\widehat{\Omega} \in \sigma(\mathcal{Q})$ with $P(\widehat{\Omega}) = 1$. One sees readily that if $\hat{t}$ is the restriction of t to $\widehat{\Omega}$, $\hat{t}$ is bijective from $\widehat{\Omega}$ to $(0, 1]$. As in Section 1.3, for a finite sequence $\varepsilon_1, \ldots, \varepsilon_n$ of 0 and 1, the elementary cylinder $\{\omega \in \Omega : \omega_j = \varepsilon_j, j = 1, \ldots, n\}$ in Ω of rank n is denoted by $E(\varepsilon_1, \ldots, \varepsilon_n)$; and we let $\mathcal{E}$ be the family of empty set $\emptyset$ and all elementary cylinders of all ranks in Ω. $\mathcal{E}$ is a π-system on Ω, and if we let $\widehat{\mathcal{E}} = \{E \cap \widehat{\Omega} : E \in \mathcal{E}\}$, then $\widehat{\mathcal{E}}$ is a π-system on $\widehat{\Omega}$.

(i) Observe first that $\sigma(\widehat{\mathcal{E}}) = \sigma(\mathcal{Q})|\widehat{\Omega}$. Actually, $\Sigma := \{A \in \sigma(\mathcal{Q}) : A \cap \widehat{\Omega} \in \sigma(\widehat{\mathcal{E}})\}$ is a σ-algebra on Ω containing $\mathcal{E}$, implying $\Sigma \supset \sigma(\mathcal{E}) = \sigma(\mathcal{Q}) \supset \Sigma$, or $\Sigma = \sigma(\mathcal{Q}) = \sigma(\mathcal{E})$, and hence $\sigma(\mathcal{Q})|\widehat{\Omega} \subset \sigma(\widehat{\mathcal{E}})$; that $\sigma(\widehat{\mathcal{E}}) \subset \sigma(\mathcal{Q})|\widehat{\Omega}$ follows from the fact that $\sigma(\mathcal{Q})|\widehat{\Omega}$ is a σ-algebra on $\widehat{\Omega}$ containing $\widehat{\mathcal{E}}$.

(ii) For any elementary cylinder $E(\varepsilon_1, \ldots, \varepsilon_n)$ of positive rank n in Ω, put $\widehat{E}(\varepsilon_1, \ldots, \varepsilon_n) = E \cap \widehat{\Omega}$. Observe that $\hat{t}\widehat{E}(\varepsilon_1, \ldots, \varepsilon_n) = (\alpha, \alpha + \frac{1}{2^n}]$, where $\alpha = \sum_{j=1}^{n} \frac{\varepsilon_j}{2^j}$, and since $\hat{t}$ is bijective on $\widehat{\Omega}$ to $(0, 1]$, $\hat{t}^{-1}(\alpha, \alpha + \frac{1}{2^n}] = \widehat{E}(\varepsilon_1, \ldots, \varepsilon_n)$, implying that $\hat{t}_{\#}P((\alpha, \alpha + \frac{1}{2^n}]) = P(\widehat{E}(\varepsilon_1, \ldots, \varepsilon_n)) = \frac{1}{2^n} = \lambda((\alpha, \alpha + \frac{1}{2^n}])$. Now, if we let $\widehat{\mathcal{I}} = \{\hat{t}A : A \in \widehat{\mathcal{E}}\}$, then $\widehat{\mathcal{I}}$ is a π-system on $(0, 1]$. Denote temporarily, in this example, by $\mathcal{B}$ and $\widehat{\mathcal{B}}$ the Borel fields on $[0, 1]$ and on $(0, 1]$ respectively, and let

$$\mathcal{M} = \{B \in \widehat{\mathcal{B}} : \hat{t}^{-1}B \in \sigma(\widehat{\mathcal{E}}) \text{ and } P(\hat{t}^{-1}B) = \lambda(B)\}.$$

As $\hat{t}$ is bijective from $\widehat{\Omega}$ to $(0, 1]$, $\mathcal{M}$ is easily seen to be a λ-system on $(0, 1]$ containing $\widehat{\mathcal{I}}$; and as $\sigma(\widehat{\mathcal{I}}) = \widehat{\mathcal{B}}$, we conclude by the $(\pi$-$\lambda)$ theorem that $\mathcal{M} = \widehat{\mathcal{B}}$, i.e. $\widehat{\mathcal{B}} \subset \hat{t}_{\#}\sigma(\widehat{\mathcal{E}})$ and $\hat{t}_{\#}P(B) = \lambda(B)$ for all $B \in \widehat{\mathcal{B}}$.

(iii) We have shown in (ii) that $\widehat{\mathcal{B}} \subset \hat{t}_{\#}\sigma(\widehat{\mathcal{E}})$ and $\hat{t}_{\#}P(B) = \lambda(B)$ for all $B \in \widehat{\mathcal{B}}$; now it will be shown that $\widehat{\mathcal{B}} = \hat{t}_{\#}\sigma(\widehat{\mathcal{E}})$ and thus $((0, 1], \widehat{\mathcal{B}}, \lambda)$ is the push-forward of $(\widehat{\Omega}, \sigma(\widehat{\mathcal{E}}), P)$ through the map $\hat{t}$. For this purpose, it is sufficient to claim that $\hat{t}A \in \widehat{\mathcal{B}}$ if $A \in \sigma(\widehat{\mathcal{E}})$. Consider $\mathcal{M} = \{A \in \sigma(\widehat{\mathcal{E}}) : \hat{t}A \in \widehat{\mathcal{B}}\}$. Clearly, $\mathcal{M}$ is a σ-algebra on $\widehat{\Omega}$ containing $\widehat{\mathcal{E}}$ and hence $\mathcal{M} = \sigma(\widehat{\mathcal{E}})$.

From the conclusions in (ii) and (iii) and the fact that $\hat{t}$ is bijective from $\widehat{\Omega}$ to $(0, 1]$, we conclude that $B \in \widehat{\mathcal{B}}$ if and only if $\hat{t}^{-1}B \in \sigma(\widehat{\mathcal{E}})$ (equivalently, $A \in \sigma(\widehat{\mathcal{E}})$ if and only if $\hat{t}A \in \widehat{\mathcal{B}}$) and that $\hat{t}_{\#}P = \lambda$ on $\widehat{\mathcal{B}}$ and $\hat{t}_{\#}^{-1}\lambda = P$ on $\sigma(\widehat{\mathcal{E}})$. Therefore, $(\widehat{\Omega}, \sigma(\widehat{\mathcal{E}}), P)$ and $((0, 1], \widehat{\mathcal{B}}, \lambda)$ are the same measure space labeled differently. Since $\sigma(\widehat{\mathcal{E}}) = \sigma(\mathcal{Q})|\widehat{\Omega}$ and $\Omega \backslash \widehat{\Omega}$ is countable, $([0, 1], \mathcal{B}, \lambda)$ is the push-forward of $(\Omega, \sigma(\mathcal{Q}), P)$ through t and $B \in \mathcal{B}$ if and only if $t^{-1}B \in \sigma(\mathcal{Q})$.

Exercise 4.3.6 Let $(\Omega, \sigma(\mathcal{Q}), P)$ and t be as in Example 4.3.2 and P^* be the measure on Ω constructed from P by Method I. Show that $t_{\#}P^* = \lambda$ on $[0, 1]$.

4.4 Functions of bounded variation

This section is devoted to the study of an important class of real-valued functions defined on a finite closed interval $I = [a, b]$. This is the class of **functions of bounded variation**. Functions in this section are all understood to be real-valued and defined on I.

For a real number α, α^+ denotes α or 0 according to whether $\alpha \geq 0$ or $\alpha < 0$, and $\alpha^- := (-\alpha)^+$. It is easily verified that $\alpha = \alpha^+ - \alpha^-$, $(\alpha + \beta)^+ \leq \alpha^+ + \beta^+$, and $(\alpha + \beta)^- \leq \alpha^- + \beta^-$ for any real numbers α and β.

Recall that a finite sequence $a = x_0 < x_1 < \cdots < x_l = b$ of points is called a partition of the interval I, where l varies from partition to partition. A generic partition of an interval will be denoted by $\mathcal{P}$.

Suppose that f is a function and $\mathcal{P} : a = x_0 < x_1 < \cdots < x_l = b$ a partition of I, let

$$P_a^b(f;\mathcal{P}) = \sum_{j=1}^{l}\{f(x_j) - f(x_{j-1})\}^+;$$

$$N_a^b(f;\mathcal{P}) = \sum_{j=1}^{l}\{f(x_j) - f(x_{j-1})\}^-;$$

and

$$V_a^b(f;\mathcal{P}) = \sum_{j=1}^{l}|f(x_j) - f(x_{j-1})|.$$

Observe that

$$V_a^b(f;\mathcal{P}) = P_a^b(f;\mathcal{P}) + N_a^b(f;\mathcal{P}).$$

Now put

$$P_a^b(f) = \sup_{\mathcal{P}} P_a^b(f;\mathcal{P});$$

$$N_a^b(f) = \sup_{\mathcal{P}} N_a^b(f;\mathcal{P});$$

and

$$V_a^b(f) = \sup_{\mathcal{P}} V_a^b(f;\mathcal{P}).$$

$P_a^b(f)$ and $N_a^b(f)$ are called respectively the **positive** and the **negative variation** of f over I, while $V_a^b(f)$ is called the **total variation** of f over I. When $a = b$, $V_a^b(f) = P_a^b(f) = N_a^b(f) = 0$, by definition. A function f is said to be **of bounded variation** on I if $V_a^b(f) < \infty$. Observe that a continuously differentiable function f is of bounded variation over I and $V_a^b(f) \le \int_a^b |f'(x)|dx$, and that a monotone function f is of bounded variation on I with $V_a^b(f) = |f(b) - f(a)|$.

Exercise 4.4.1 Show that $V_a^b(f) = P_a^b(f) + N_a^b(f)$.

Exercise 4.4.2 If $a < c < b$, show that $P_a^b(f) = P_a^c(f) + P_c^b(f)$ and similarly for negative and total variation.

Exercise 4.4.3 Show that if f and g are of bounded variation on I, then $\alpha f + \beta g$ is also of bounded variation on I for any real numbers α and β, and $V_a^b(\alpha f + \beta g) \le |\alpha| V_a^b(f) + |\beta| V_a^b(g)$.

Now suppose that f is a function of bounded variation on I. Let $x \in I$ and $\mathcal{P}$ be a partition of $[a, x]$, then

$$\begin{aligned} f(x) - f(a) &= \sum_{j=1}^{l} \{f(x_j) - f(x_{j-1})\} = P_a^x(f; \mathcal{P}) - N_a^x(f; \mathcal{P}) \\ &\leq P_a^x(f) - N_a^x(f; \mathcal{P}), \end{aligned}$$

or

$$f(x) - f(a) + N_a^x(f; \mathcal{P}) \leq P_a^x(f),$$

from which one infers that

$$f(x) \leq f(a) + P_a^x(f) - N_a^x(f).$$

Similarly, one has

$$f(x) - f(a) \geq P_a^x(f; \mathcal{P}) - N_a^x(f),$$

and hence

$$f(x) - f(a) + N_a^x(f) \geq P_a^x(f),$$

or

$$f(x) \geq f(a) + P_a^x(f) - N_a^x(f).$$

Consequently,

$$f(x) = f(a) + P_a^x(f) - N_a^x(f), \quad x \in I. \tag{4.3}$$

Since $P_a^x(f)$ and $N_a^x(f)$ are monotone increasing in x, it follows from (4.3) that f is a difference of two monotone increasing functions. Conversely, when f is a difference of two monotone increasing functions, then f is of bounded variation on I. Thus the first part of the following theorem has been shown.

Theorem 4.4.1 *A function f is of bounded variation on I if and only if f is a difference of two monotone increasing functions. Furthermore, if f is of bounded variation on I and $f = f_1 - f_2$, where f_1 and f_2 are monotone increasing and $f_1(a) = f(a)$, then there is a monotone increasing function φ on I with $\varphi(a) = 0$ such that*

$$f_1(x) = f(a) + P_a^x(f) + \varphi(x); \quad f_2(x) = N_a^x(f) + \varphi(x)$$

for $x \in I$.

Proof It remains to show the second part of the theorem. So suppose that f is of bounded variation on I and $f = f_1 - f_2$, where f_1 and f_2 are monotone increasing and $f_1(a) = f(a)$. From monotony of f_1 and f_2, one verifies that for $a \le x' < x'' \le b$,

$$\{f(x'') - f(x')\}^+ = \{f_1(x'') - f_1(x') + f_2(x') - f_2(x'')\}^+ \le f_1(x'') - f_1(x');$$
$$\{f(x'') - f(x')\}^- = \{f_1(x'') - f_1(x') + f_2(x') - f_2(x'')\}^- \le f_2(x'') - f_2(x').$$

From the preceding inequalities it then follows that for $a \le x < y \le b$ and any partition $\mathcal{P}$ of $[x, y]$,

$$P_x^y(f; \mathcal{P}) \le f_1(y) - f_1(x); \quad N_x^y(f; \mathcal{P}) \le f_2(y) - f_2(x),$$

and hence

$$P_x^y(f) \le f_1(y) - f_1(x); \quad N_x^y(f) \le f_2(y) - f_2(x). \tag{4.4}$$

In particular,

$$P_a^x(f) \le f_1(x) - f(a); \quad N_a^x(f) \le f_2(x)$$

for $x \in I$. Let $\varphi(x) = f_1(x) - \{f(a) + P_a^x(f)\}$, then $\varphi \ge 0$ and $\varphi(a) = 0$; from $f(a) + P_a^x(f) - N_a^x(f) = f(x) = f_1(x) - f_2(x)$, it follows that $f_2(x) = N_a^x(f) + \varphi(x)$ for $x \in I$. It remains to see that φ is monotone increasing. For $x < y$ in I we have

$$\varphi(y) - \varphi(x) = f_1(y) - f_1(x) - \{P_a^y(f) - P_a^x(f)\} = f_1(y) - f_1(x) - P_x^y(f) \ge 0,$$

by applying the first inequality in (4.4). This shows that φ is monotone increasing. ∎

Henceforth, a function of bounded variation on I will simply be called a **BV function** on I. For a BV function f, let functions f_P, f_N, and f_V be defined by

$$f_P(x) = P_a^x(f); \quad f_N(x) = N_a^x(f); \quad \text{and } f_V(x) = V_a^x(f),$$

then the second part of Theorem 4.4.1 could be interpreted as saying that the decomposition $f = f(a) + f_P - f_N$ is the **minimal decomposition** of f into the difference of monotone increasing functions if a partial order $\prec$ on the family of all monotone increasing functions on I is defined as follows: $f \prec g$ if and only if $g - f$ is nonnegative and monotone increasing on I.

Theorem 4.4.2 *Suppose that f is a BV function on I. If f is right(left)-continuous at $x_0 \in [a, b)$ ($x_0 \in (a, b]$), then so are f_P, f_N, and f_V.*

Proof Since $f(x) - f(a) = f_P(x) - f_N(x)$ and $f_V(x) = f_P(x) + f_N(x)$,

$$f_P(x) = \frac{1}{2}\{f_V(x) + f(x) - f(a)\} \text{ and } f_N(x) = \frac{1}{2}\{f_V(x) - f(x) + f(a)\}$$

for $x \in I$, it is therefore sufficient to show that f_V is right-continuous at x_0. For this, we have to show that $f_V(x_{0^+}) = f_V(x_0)$, or $V_{x_0}^{x_0+h}(f) \to 0$ as $h \to 0+$.

Suppose the contrary, then $\delta_0 = f_V(x_0+) - f_V(x_0) > 0$. Let $\delta = \frac{2}{3}\delta_0$, and choose $h_1 > 0$ small enough so that $x_0 + h_1 \le b$ and $V_{x_0}^{x_0+h_1}(f) < 2\delta$. Since $V_{x_0}^{x_0+h_1}(f) > \delta$, there is a partition $x_0 < x_1 < \cdots < x_l = x_0 + h_1$ such that

$$\sum_{j=1}^{l} |f(x_j) - f(x_{j-1})| > \delta.$$

As f is right-continuous at x_0, there is $h_2 > 0$ with $x_0 + h_2 < x_1$ such that $|f(x_1) - f(x_0 + h_2)| + \sum_{j=2}^{l} |f(x_j) - f(x_{j-1})| > \delta$; hence $V_{x_0+h_2}^{x_0+h_1}(f) > \delta$. Now repeat the above argument with h_1 replaced by h_2, to obtain $0 < h_3 < h_2$ such that $V_{x_0+h_3}^{x_0+h_2}(f) > \delta$. Then,

$$2\delta > V_{x_0}^{x_0+h_1}(f) \ge V_{x_0+h_3}^{x_0+h_2}(f) + V_{x_0+h_2}^{x_0+h_1}(f) > 2\delta,$$

which is absurd. Thus f_V is right-continuous at x_0. ■

Example 4.4.1 Let f be a Lebesgue integrable function on I and define

$$F(x) = \alpha + \int_a^x f(t)dt, \quad x \in I, \tag{4.5}$$

α being a constant. Then F is a BV function and

$$V_a^b(F) = \int_a^b |f(t)|dt.$$

Actually, for any partition $\mathcal{P} : a = x_0 < x_1 < \cdots < x_l = b$, we have

$$V_a^b(F; \mathcal{P}) = \sum_{j=1}^{l} \left| \int_{x_{j-1}}^{x_j} f(t)dt \right| \le \int_a^b |f(t)|dt,$$

hence,

$$V_a^b(F) \le \int_a^b |f(t)|dt < \infty. \tag{4.6}$$

Now, by Exercise 4.1.2, for any $\varepsilon > 0$ there is a step function g such that

$$\int_a^b |f(t) - g(t)|dt < \varepsilon.$$

Choose a partition $\mathcal{P} : a = x_0 < x_1 < \cdots < x_l = b$ of I such that $\{x_0, x_1, \ldots, x_l\}$ contains all the endpoints of the open intervals on which g is constant. We have then,

$$\begin{aligned}\sum_{j=1}^{l} |F(x_j) - F(x_{j-1})| &= \sum_{j=1}^{l} \left| \int_{x_{j-1}}^{x_j} f(t)dt \right| \\ &\geq \sum_{j=1}^{l} \left| \int_{x_{j-1}}^{x_j} g(t)dt \right| - \sum_{j=1}^{l} \left| \int_{x_{j-1}}^{x_j} (f(t) - g(t))dt \right| \\ &= \sum_{j=1}^{l} \int_{x_{j-1}}^{x_j} |g(t)|dt - \sum_{j=1}^{l} \left| \int_{x_{j-1}}^{x_j} (f(t) - g(t))dt \right| \\ &\geq \int_a^b |g(t)|dt - \int_a^b |f(t) - g(t)|dt \\ &\geq \int_a^b |g(t)|dt - \varepsilon \geq \int_a^b |f(t)|dt - 2\varepsilon.\end{aligned}$$

Thus,

$$V_a^b(F) \geq \int_a^b |f(t)|dt - 2\varepsilon.$$

Let $\varepsilon \to 0$; we have $V_a^b(F) \geq \int_a^b |f(t)|dt$, and hence

$$V_a^b(F) = \int_a^b |f(t)|dt,$$

by (4.6).

The function F, defined by (4.5), with f being Lebesgue integrable on I, is called an **indefinite integral** of f.

Exercise 4.4.4 Let F be an indefinite integral of f on I; show that $F_P(x) = \int_a^x f^+(t)dt$ and $F_N(x) = \int_a^x f^-(t)dt$. (Hint: use the fact that $F_P(x) = \frac{1}{2}\{F_V(x) + F(x) - F(a)\}$ and $F_N(x) = \frac{1}{2}\{F_V(x) - F(x) + F(a)\}$.)

4.5 Riemann–Stieltjes integral

The Rieman–Stieltjes integral of bounded functions on I will be defined along the same lines that the Riemann integral is defined. Suppose that g is a monotone increasing function defined on a finite closed interval $I = [a, b]$.

Given a partition $\mathcal{P} : a = x_0 < x_1 < \cdots < x_l = b$ of I and $j = 1, \ldots, l$; put

$$\mathcal{P}_j g = g(x_j) - g(x_{j-1}).$$

For a bounded function f on I, and $\mathcal{P}$ as above, let

$$\underline{f}_j = \inf_{x\in[x_{j-1},x_j]} f(x), \quad \bar{f}_j = \sup_{x\in[x_{j-1},x_j]} f(x);$$

and

$$\underline{S}_g(f,\mathcal{P}) = \sum_{j=1}^{l} \underline{f}_j \mathcal{P}_j g, \quad \bar{S}_g(f,\mathcal{P}) = \sum_{j=1}^{l} \bar{f}_j \mathcal{P}_j g.$$

Observe that for any partitions $\mathcal{P}$ and Q of I, the following sequence of inequalities holds:

$$\underline{S}_g(f,\mathcal{P}) \le \underline{S}_g(f,\mathcal{P}\vee Q) \le \bar{S}_g(f,\mathcal{P}\vee Q) \le \bar{S}_g(f,Q). \tag{4.7}$$

Now let $\underline{\int_a^b} fdg = \sup_{\mathcal{P}} \underline{S}_g(f,\mathcal{P})$ and $\overline{\int_a^b} fdg = \inf_{\mathcal{P}} \bar{S}_g(f,\mathcal{P})$; by (4.7) both $\underline{\int_a^b} fdg$ and $\overline{\int_a^b} fdg$ are finite and $\underline{\int_a^b} fdg \le \overline{\int_a^b} fdg$. In the case where $\underline{\int_a^b} fdg = \overline{\int_a^b} fdg$, f is said to be **Riemann–Stieltjes integrable** w.r.t. g and the common value, denoted $\int_a^b fdg$, is called the **Riemann–Stieltjes integral** of f w.r.t. g. From (4.7), Theorem 4.5.1 follows directly:

Theorem 4.5.1 *Let g be monotone increasing on $[a,b]$. A bounded function f on $[a,b]$ is Riemann–Stieltjes integrable w.r.t. g if and only if for any $\varepsilon > 0$, there is a partition $\mathcal{P}$ of $[a,b]$ such that*

$$\bar{S}_g(f,\mathcal{P}) - \underline{S}_g(f,\mathcal{P}) < \varepsilon.$$

Example 4.5.1 Let g be a monotone increasing function on $[a,b]$. (i) If f is continuous on $[a,b]$, then $\int_a^b fdg$ exists. (ii) If f is a BV function and g is continuous, then $\int_a^b fdg$ exists.

Clearly, (i) is an easy consequence of Theorem 4.5.1, while (ii) follows also from Theorem 4.5.1 if one notes that for any partition $\mathcal{P} : a = x_0 < x_1 < \cdots < x_n = b$ of $[a,b]$,

$$\begin{aligned}\bar{S}_g(f,\mathcal{P}) - \underline{S}_g(f,\mathcal{P}) &= \sum_{j=1}^{n} (\underline{f}_j - \bar{f}_j)\mathcal{P}_j(g) \\ &\le \sum_{j=1}^{n} V_{x_{j-1}}^{x_j}(f)\mathcal{P}_j g \le V_a^b(f) \max_{1\le j\le n} \mathcal{P}_j g.\end{aligned}$$

Example 4.5.2 Suppose that w is a nonnegative Lebesgue integrable function on $[a,b]$, and g is an indefinite integral of w (cf. Example 4.4.1), then any Riemann integrable function f on $[a,b]$ is Riemann–Stieltjes integrable w.r.t. g on $[a,b]$ and

$$\int_a^b fdg = \int_a^b f(t)w(t)dt.$$

For a partition $\mathcal{P} : a = x_0 < x_1 < \cdots < x_n = b$, define a function $\bar{f}^{\mathcal{P}}$ by $\bar{f}^{\mathcal{P}}(x) = \bar{f}_j$ if $x \in [x_{j-1}, x_j)$ and $\bar{f}^{\mathcal{P}}(b) = f(b)$; similarly define $\underline{f}^{\mathcal{P}}$ by $\underline{f}^{\mathcal{P}}(x) = \underline{f}_j$ if $x \in [x_{j-1}, x_j)$ and $\underline{f}^{\mathcal{P}}(b) = f(b)$. Now choose a sequence $\{\mathcal{P}^{(k)}\}$ of partitions so that $\|\mathcal{P}^{(k)}\| \to 0$ as $k \to \infty$, and

$$\bar{S}_g(f, \mathcal{P}^{(k)}) \to \overline{\int_a^b} fdg; \quad \underline{S}_g(f, \mathcal{P}^{(k)}) \to \underline{\int_a^b} fdg.$$

Obviously,

$$\bar{S}_g(f, \mathcal{P}^{(k)}) = \int_a^b \bar{f}^{\mathcal{P}^{(k)}}(t)w(t)dt; \quad \underline{S}_g(f, \mathcal{P}^{(k)}) = \int_a^b \underline{f}^{\mathcal{P}^{(k)}}(t)w(t)dt.$$

Since f is Riemann integrable, f is continuous at almost all points of $[a, b]$, and hence

$$\bar{f}^{\mathcal{P}^{(k)}} w \to fw \text{ a.e.}; \quad \underline{f}^{\mathcal{P}^{(k)}} w \to fw \text{ a.e.}$$

If we put $M = \sup_{t \in [a,b]} |f(t)|$, $|\bar{f}^{\mathcal{P}^{(k)}} w| \le Mw$, $|\underline{f}^{\mathcal{P}^{(k)}} w| \le Mw$, hence by LDCT

$$\lim_{k \to \infty} \bar{S}_g(f, \mathcal{P}^{(k)}) = \int_a^b f(t)w(t)dt = \lim_{k \to \infty} \underline{S}_g(f, \mathcal{P}^{(k)}),$$

and thus

$$\overline{\int_a^b} fdg = \underline{\int_a^b} fdg = \int_a^b f(t)w(t)dt,$$

i.e. f is Riemann–Stieltjes integrable w.r.t. g on $[a, b]$ and $\int_a^b fdg = \int_a^b f(t)w(t)dt$.

Exercise 4.5.1 Suppose that f is continuous on $[a, b]$ and g is monotone increasing on $[a, b]$.

(i) Show that $\int_a^b fdg = \overline{\int_a^b} fdg = \inf \bar{S}_g(f, \mathcal{P})$, where the infimum is taken over all those partitions $\mathcal{P}$, the endpoints of whose intervals other than a and b are points of continuity of g.

(ii) Show that $\int_a^b fdg = \int_a^b fd\mu_g$.

The following Lemma is a generalization of Lemma 4.2.2 when $n = 1$.

Lemma 4.5.1 *Suppose that g is a right-continuous and monotone increasing function on* $[a, b]$, *and f a bounded function on* $[a, b]$ *which is continuous wherever g is discontinuous, then*

$$\underline{\int_a^b} fdg = \int_a^b \underline{f} d\mu_g; \quad \overline{\int_a^b} fdg = \int_a^b \bar{f} d\mu_g,$$

where $\underline{f}(x) = \lim_{\delta\to 0+} \inf_{|y-x|<\delta} f(y)$ *and* $\bar{f}(x) = \lim_{\delta\to 0+} \sup_{|y-x|<\delta} f(y)$.

Proof By Lemma 4.2.1, both $\underline{f}$ and $\bar{f}$ are Lebesgue measurable. It is clear that $\underline{f} \le f \le \bar{f}$ on $[a, b]$. Choose a sequence $\{\mathcal{P}^{(k)}\}$ of partitions of $[a, b]$ such that $\|\mathcal{P}^{(k)}\| \to 0$, and

$$\underline{\int_a^b} fdg = \lim_{k\to\infty} \underline{S}_g(f, \mathcal{P}^{(k)}); \quad \overline{\int_a^b} fdg = \lim_{k\to\infty} \bar{S}_g(f, \mathcal{P}^{(k)}).$$

For each $k \in \mathbb{N}$, let $\mathcal{P}^{(k)}$ be $a = x_0^{(k)} < x_1^{(k)} < \cdots < x_{n_k}^{(k)} = b$, and define $f_k(x) = \inf_{x_{j-1}^{(k)} \le t \le x_j^{(k)}} f(t)$ if $x \in (x_{j-1}^{(k)}, x_j^{(k)}]$ and $f_k(a) = f(a)$. As we have shown in the proof of Lemma 4.2.2, $\lim_{k\to\infty} f_k(x) = \underline{f}(x)$ if $x \in [a, b]$, but is not an endpoint of intervals of the partitions $\mathcal{P}^{(k)}$, $k = 1, 2, \ldots$. Now, since f is continuous wherever g is discontinuous and g is right-continuous, we may assume that all the endpoints of the intervals of the partitions $\mathcal{P}^{(k)}$ are points of continuity of g, except possibly b. Hence $f_k(x) \to \underline{f}(x)$ for μ_g-a.e. x in $[a, b)$; but if b is a point of discontinuity of g, then f is continuous at b and hence $f_k(b) \to f(b) = \underline{f}(b)$. Thus, $\lim_{k\to\infty} f_k(x) = \underline{f}(x)$ for μ_g-a.e. x of $[a, b]$, and $\int_a^b \underline{f} d\mu_g = \lim_{k\to\infty} \int_a^b f_k d\mu_g$ by LDCT, because $|f_k(x)| \le \sup_{a\le t\le b} |f(t)|$. Since g is right-continuous, $\mu_g((c, d]) = g(d) - g(c)$ for $a \le c < d \le b$; we have $\underline{S}_g(f, \mathcal{P}^{(k)}) = \int_a^b f_k d\mu_g$. Consequently, $\int_a^b \underline{f} d\mu_g = \lim_{k\to\infty} \underline{S}_g(f, \mathcal{P}^{(k)}) = \underline{\int_a^b} fdg$. Similarly, $\int_a^b \bar{f} d\mu_g = \overline{\int_a^b} fdg$. ■

Theorem 4.5.2 *Suppose that g is a right-continuous and monotone increasing function on* $[a, b]$ *and f is a bounded function which is continuous at the* μ_g*-a.e. point of* $[a, b]$, *then f is Riemann–Stieltjes integrable w.r.t. g, and*

$$\int_a^b fdg = \int_a^b fd\mu_g.$$

Proof We claim first that f is μ_g-measurable. From $\underline{f} \le f \le \bar{f}$ and the fact that f is continuous μ_g-a.e., it follows that $\underline{f}(x) = f(x) = \bar{f}(x)$ for μ_g-a.e. x in $[a, b]$; hence f differs from $\underline{f}$ only on a set A with $\mu_g(A) = 0$. But $\underline{f}$ is Borel measurable by Lemma 4.2.1, and is therefore μ_g-measurable from the fact that μ_g is a Carathéodory measure.

Thus f is μ_g-measurable as we claim. Now, $\underline{f} = f = \bar{f}$ μ_g-a.e. implies, together with Lemma 4.5.1, that

$$\underline{\int_a^b} f dg = \int_a^b \underline{f} d\mu_g = \int_a^b f d\mu_g = \int_a^b \bar{f} d\mu_g = \overline{\int_a^b} f dg,$$

which entails that f is Riemann–Stieltjes integrable w.r.t. g and $\int_a^b f dg = \int_a^b f d\mu_g$. ∎

Theorem 4.5.3 (Integration by parts) *Suppose that f and g are monotone increasing functions on $[a, b]$ and at least one of them is continuous. Then*

$$\int_a^b f dg = f(b)g(b) - f(a)g(a) - \int_a^b g df.$$

Proof Note firstly that $\int_a^b f dg$ and $\int_a^b g df$ exist, from Example 4.5.1. Let $\mathcal{P}$: $a = x_0 < x_1 < \cdots < x_l = b$ be a partition of $[a, b]$, then

$$\begin{aligned}\bar{S}_g(f, \mathcal{P}) &= \sum_{j=1}^{l} f(x_j)[g(x_j) - g(x_{j-1})]\\ &= f(b)g(b) - f(a)g(a) - \sum_{j=1}^{l} g(x_{j-1})[f(x_j) - f(x_{j-1})]\\ &= f(b)g(b) - f(a)g(a) - \underline{S}_f(g, \mathcal{P}),\end{aligned}$$

from which, by taking a sequence $\{\mathcal{P}^{(k)}\}$ of partitions such that

$$\lim_{k\to\infty} \bar{S}_g(f, \mathcal{P}^{(k)}) = \overline{\int_a^b} f dg \quad \text{and} \quad \lim_{k\to\infty} \underline{S}_f(g, \mathcal{P}^{(k)}) = \underline{\int_a^b} g df,$$

we obtain,

$$\begin{aligned}\int_a^b f dg = \overline{\int_a^b} f dg &= f(b)g(b) - f(a)g(a) - \underline{\int_a^b} g df\\ &= f(b)g(b) - f(a)g(a) - \int_a^b g df.\end{aligned}$$ ∎

Exercise 4.5.2 Under the same assumptions as in Theorem 4.5.3, show that

$$\int_a^b f d\mu_g = f(b)g(b) - f(a)g(a) - \int_a^b g d\mu_f.$$

(Hint: cf. Exercise 4.5.1.)

Now suppose that g is a BV function on $[a, b]$ and write $g = g_1 - g_2$, where $g_1(x) = g(a) + g_P(x)$ and $g_2(x) = g_N(x)$ for $x \in [a, b]$. Recall that $g_P(x) = P_a^x(g)$ and $g_N(x) = N_a^x(g)$, $x \in [a, b]$. A bounded function f on $[a, b]$ is called Riemann–Stieltjes integrable w.r.t. g if it is Riemann–Stieltjes integrable w.r.t. g_1 and g_2, and in this case the Riemann–Stieltjes integral of f w.r.t. g, denoted $\int_a^b fdg$, is defined by

$$\int_a^b fdg = \int_a^b fdg_1 - \int_a^b fdg_2.$$

With this definition, Corollary 4.5.1 of Theorem 4.5.3 follows, by using Theorem 4.4.2.

Corollary 4.5.1 *Suppose that f and g are BV functions on $[a, b]$ and at least one of them is continuous, then*

$$\int_a^b fdg = f(b)g(b) - f(a)g(a) - \int_a^b gdf.$$

Theorem 4.5.4 (Second mean-value theorem) *Suppose that f is an integrable function on a finite interval $[a, b]$ and φ is a monotone function on $[a, b]$, then there is $c \in [a, b]$ such that*

$$\int_a^b \varphi fd\lambda = \varphi(a) \int_a^c fd\lambda + \varphi(b) \int_c^b fd\lambda.$$

Firstly we prove a lemma.

Lemma 4.5.2 *Let f and φ be as in Theorem 4.5.4 and, further φ is assumed to be nonnegative and monotone decreasing, then there is $c \in [a, b]$ such that*

$$\int_a^b \varphi fd\lambda = \varphi(a) \int_a^c fd\lambda.$$

Proof We may assume that $\varphi(a) > 0$, because otherwise $\varphi \equiv 0$ and the lemma is trivial. Define a function F on $[a, b]$ by

$$F(x) = \int_a^x fd\lambda, \quad x \in [a, b].$$

By Corollary 4.5.1 and Example 4.5.2, we have

$$\int_a^b Fd\varphi = F(b)\varphi(b) - F(a)\varphi(a) - \int_a^b \varphi dF = F(b)\varphi(b) - \int_a^b \varphi fd\lambda;$$

hence,

$$\int_a^b \varphi fd\lambda \le M\varphi(b) - M \int_a^b d\varphi = M\varphi(b) + M\{\varphi(a) - \varphi(b)\} = M\varphi(a),$$

or $\frac{1}{\varphi(a)}\int_a^b \varphi f d\lambda \le M$, where $M = \max_{x\in[a,b]} F(x)$. Similarly, if $m = \min_{x\in[a,b]} F(x)$, then $m \le \frac{1}{\varphi(a)}\int_a^b \varphi f d\lambda$; thus,

$$m \le \frac{1}{\varphi(a)}\int_a^b \varphi f d\lambda \le M,$$

from which, by the intermediate-value theorem for continuous functions, there is $c \in [a, b]$ such that $\frac{1}{\varphi(a)}\int_a^b \varphi f d\lambda = F(c) = \int_a^c f d\lambda$. ■

Proof of Theorem 4.5.4 Consider first the case where φ is monotone decreasing. Since $\varphi - \varphi(b)$ is nonnegative and monotone decreasing, by Lemma 4.5.2 there is $c \in [a, b]$ such that $\int_a^b \{\varphi - \varphi(b)\} f d\lambda = \{\varphi(a) - \varphi(b)\}\int_a^c f d\lambda$, or

$$\begin{aligned}\int_a^b \varphi f d\lambda &= \varphi(b)\int_a^b f d\lambda + \{\varphi(a) - \varphi(b)\int_a^c f d\lambda \\ &= \varphi(a)\int_a^c f d\lambda + \varphi(b)\int_c^b f d\lambda.\end{aligned}$$

If φ is monotone increasing, replacing φ by $-\varphi$ in the argument above, we also conclude that there is $c \in [a, b]$ such that

$$\int_a^b \varphi f d\lambda = \varphi(a)\int_a^c f d\lambda + \varphi(b)\int_c^b f d\lambda.$$ ■

Corollary 4.5.2 *Let f be integrable on $[a, b]$ and φ be nonnegative and monotone increasing on $[a, b]$; then there is $c \in [a, b]$ such that*

$$\int_a^b \varphi f d\lambda = \varphi(b)\int_c^b f d\lambda.$$

Proof Replace φ in Theorem 4.5.4 by $\varphi - \varphi(a)$. ■

Remark Lemma 4.5.2, Theorem 4.5.4, and Corollary 4.5.2 will all be referred to as the **second mean-value theorem**.

Exercise 4.5.3 Show that the following improper integrals exist: (i) $\int_0^\infty \frac{\sin x}{x} dx$; (ii) $\int_0^\infty \frac{\sin x}{e^x - 1} dx$.

Exercise 4.5.4 Suppose that h is an integrable function on $[a, b]$ and g is an indefinite integral of h. Show that if f is a Riemann integrable function on $[a, b]$, then f is Riemann–Stieltjes integrable w.r.t. g, and

$$\int_a^b f dg = \int_a^b f h d\lambda.$$

Exercise 4.5.5 Suppose that u and v are integrable functions on $[a, b]$ and that U and V are respectively indefinite integrals of u and v. Show that

$$\int_a^b Uv d\lambda = U(b)V(b) - U(a)V(a) - \int_a^b Vu d\lambda.$$

Exercise 4.5.6 Let f be a measurable and finite a.e. function on a measure space (Ω, Σ, μ). Suppose that $\mu(\{f \leq t\}) < \infty$ for every $t \in \mathbb{R}$ and let $F(t) = \mu(\{f \leq t\})$ for $t \in \mathbb{R}$. Define the improper Riemann–Stieltjes integral $\int_{\mathbb{R}} |t|^p dF$ by

$$\int_{\mathbb{R}} |t|^p dF = \lim_{\substack{b \to \infty \\ a \to -\infty}} \int_a^b |t|^p dF, \quad 1 \leq p < \infty.$$

Show that $\int_\Omega |f|^p d\mu = \int_{\mathbb{R}} |t|^p dF$.

A characterization of functions which are indefinite integrals will be taken up after a treatise on differentiation is given in Section 4.6.

4.6 Covering theorems and differentiation

Our purpose in this section is to establish the Lebesgue differentiation theorem for Radon measures on $\mathbb{R}^n$ and to give some of its relevant applications. To do this, we shall first exhibit a useful procedure of selecting a sequence of disjoint balls from a given collection of balls in $\mathbb{R}^n$, and deduce from it two covering theorems in $\mathbb{R}^n$; one of which is elementary but will be useful when we study the Hardy–Littlewood maximal function in Chapter 6, and the other is a Vitali type covering theorem that is the main tool for the proof of the Lebesgue differentiation theorem.

For convenience, the diameter of a set A is denoted by δA instead of diam A, for the moment, and a ball is either open or closed with positive radius unless, specified explicitly. For a ball B, we shall denote by $\widehat{B}$ the ball concentric with B and with radius five times that of B.

A collection $\mathcal{C}$ of balls in $\mathbb{R}^n$ is said to be **admissible** if $\sup_{B \in \mathcal{C}} \delta B < \infty$. Given an admissible collection $\mathcal{C}$ of balls in $\mathbb{R}^n$, we select a disjoint sequence $\{B_j\}$, finite or infinite, from $\mathcal{C}$ by the following procedure. Let $d_0 = \sup_{B \in \mathcal{C}} \delta B$, then $0 < d_0 < \infty$. Choose a ball B_1 in $\mathcal{C}$ such that $\delta B_1 \geq \frac{1}{2} d_0$. Suppose now that $B_1, \ldots, B_m$ are disjoint balls chosen from $\mathcal{C}$; if $B \cap \bigcup_{j=1}^m B_j \neq \emptyset$ for every $B \in \mathcal{C}$, stop the procedure; otherwise, let

$$d_m = \sup \left\{ \delta B : B \in \mathcal{C},\ B \cap \bigcup_{j=1}^m B_j = \emptyset \right\},$$

and choose a ball B_{m+1} from $\mathcal{C}$ which is disjoint with $\bigcup_{j=1}^m B_j$ and with $\delta B_{m+1} \geq \frac{1}{2} d_m$. Thus a disjoint sequence $\{B_j\}$, finite or infinite, is obtained by this procedure. Such a procedure of selecting $\{B_j\}$ from $\mathcal{C}$ will be referred to as **Procedure**$(\boldsymbol{S})$.

Lemma 4.6.1 *Suppose that $\mathcal{C}$ is an admissible collection of balls in $\mathbb{R}^n$ and $\{B_j\}$ is a sequence of disjoint balls selected from $\mathcal{C}$ by* Procedure(S). *Then either $\{B_j\}$ is infinite and* $\inf_j \delta B_j > 0$ *or* $\bigcup \mathcal{C} \subset \bigcup_j \widehat{B}_j$ *(recall that* $\bigcup \mathcal{C} := \bigcup_{B\in\mathcal{C}} B$*).*

Proof If $\{B_j\}$ is finite, say $\{B_j\} = \{B_1, \ldots, B_m\}$, meaning that if $B \in \mathcal{C}$, then $B \cap \bigcup_{j=1}^{m} B_j \neq \emptyset$. Let j_0 be the smallest j, $1 \leq j \leq m$, such that $B \cap B_j \neq \emptyset$. If $j_0 = 1$, then $\delta B \leq d_0 \leq 2B_1$; while if $j_0 \geq 2$, $B \cap \bigcup_{j=1}^{j_0-1} B_j = \emptyset$ and $\delta B \leq d_{j_0-1} \leq 2\delta B_{j_0}$. Hence, $\delta B \leq 2\delta B_{j_0}$ holds; this fact, together with $B \cap B_{j_0} \neq \emptyset$, implies that $B \subset \widehat{B}_{j_0}$. Thus, $\bigcup \mathcal{C} \subset \bigcup_{j=1}^{m} \widehat{B}_j$.

Now suppose that $\{B_j\}$ is infinite and $\inf_j \delta B_j = 0$. Let again $B \in \mathcal{C}$. Since $\delta B > 0$ and $\inf_j \delta B_j = 0$, there is $j_0 \in \mathbb{N}$ arbitrarily large such that $\delta B > 2\delta B_{j_0}$. But then $B \cap \bigcup_{j=1}^{j_0-1} B_j \neq \emptyset$, because otherwise $\delta B_{j_0} < \frac{1}{2}\delta B \leq \frac{1}{2} d_{j_0-1}$, contradicting the way B_{j_0} is selected by Procedure(S). Since $B \cap \bigcup_{j=1}^{j_0-1} B_j \neq \emptyset$, argue as in the first paragraph of the proof to conclude that B is contained in one of $\widehat{B}_1, \ldots, \widehat{B}_{j_0-1}$, and hence $B \subset \bigcup_j \widehat{B}_j$. Consequently, $\bigcup \mathcal{C} \subset \bigcup_j \widehat{B}_j$. ■

Lemma 4.6.1 leads immediately to the following basic **covering theorem**.

Theorem 4.6.1 *Let $\mathcal{C}$ be an admissible collection of balls in $\mathbb{R}^n$; then there is a disjoint sequence $\{B_j\}$ of balls from $\mathcal{C}$ such that*

$$\lambda^n\left(\bigcup \mathcal{C}\right) \leq 5^n \sum_j \lambda^n(B_j). \tag{4.8}$$

Proof Let $\{B_j\}$ be a sequence of disjoint balls selected from $\mathcal{C}$ by Procedure(S). By Lemma 4.6.1, either $\{B_j\}$ is infinite and $\inf \delta B_j > 0$ or $\bigcup \mathcal{C} \subset \bigcup_j \widehat{B}_j$. If $\{B_j\}$ is infinite and $\inf_j \delta B_j > 0$, then the right-hand side of (4.8) is ∞ and (4.8) holds trivially. Suppose now that $\bigcup \mathcal{C} \subset \bigcup_j \widehat{B}_j$. Then,

$$\lambda^n\left(\bigcup \mathcal{C}\right) \leq \sum_j \lambda^n(\widehat{B}_j) = 5^n \sum_j \lambda^n(B_j),$$

because $\lambda^n(\widehat{B}_j) = 5^n \lambda^n(B_j)$, by Example 4.3.1 (ii). ■

We come now to a **Vitali type** covering theorem. Let E be a subset of $\mathbb{R}^n$; a collection $\mathcal{V}$ of subsets of $\mathbb{R}^n$ is called a **Vitali cover** of E if for every x in E and any positive number ε there is V in $\mathcal{V}$, such that $\delta V < \varepsilon$ and $x \in V$. The following covering theorem is a simple version of the well-known Vitali covering theorem, but it suffices for our purpose.

Theorem 4.6.2 (Vitali) *Let E be a subset of $\mathbb{R}^n$ with $\lambda^n(E) < \infty$, and suppose that $\mathcal{V}$ is a collection of closed balls in $\mathbb{R}^n$ which forms a Vitali cover of E. Then there is a sequence $\{B_j\}$ of disjoint balls from $\mathcal{V}$ such that $\lambda^n(E \setminus \bigcup_j B_j) = 0$.*

Proof Choose an open set $G \supset E$ such that $\lambda^n(G) < \infty$, and let

$$\mathcal{C} = \{V \in \mathcal{V} : V \subset G,\ \delta V \leq 1\}.$$

Then $\mathcal{C}$ is an admissible collection of closed balls and is a Vitali cover of E. Now select a sequence $\{B_j\}$ of disjoint balls from $\mathcal{C}$ by Procedure(S). If $\{B_j\}$ is finite, say $\{B_j\} = \{B_1, \dots, B_m\}$, then $V \cap \bigcup_{j=1}^m B_j \neq \emptyset$ for every $V \in \mathcal{C}$. Take any $x \in E$ and $\varepsilon > 0$, choose $V \in \mathcal{C}$ such that $x \in V$ and $\delta V < \varepsilon$, then $\operatorname{dist}(x, \bigcup_{j=1}^m B_j) \leq \operatorname{dist}(x, V \cap \bigcup_{j=1}^m B_j) \leq \delta V < \varepsilon$. Since $\varepsilon > 0$ is arbitrary and $\bigcup_{j=1}^m B_j$ is closed, we infer that $x \in \bigcup_{j=1}^m B_j$ or $E \subset \bigcup_{j=1}^m B_j$, and hence $\lambda^n(E \setminus \bigcup_{j=1}^m B_j) = 0$. Suppose now that $\{B_j\}$ is infinite. Since $\sum_j \lambda^n(B_j) = \lambda^n(\bigcup_j B_j) \leq \lambda^n(G) < \infty$, $\inf_{j \geq l} \delta(B_j) = 0$ for any $l \in \mathbb{N}$. Observe then that for any $l \in \mathbb{N}$, $\{B_j\}_{j \geq l+1}$ is a sequence of balls selected from the admissible collection

$$\mathcal{C}^{(l)} := \left\{ V \in \mathcal{C} : V \subset G \setminus \bigcup_{j=1}^{l} B_j \right\},$$

by Procedure(S). Since $\inf_{j \geq l+1} \delta B_j = 0$, it follows from Lemma 4.6.1 that $\bigcup \mathcal{C}^{(l)} \subset \bigcup_{j \geq l+1} \widehat{B}_j$; consequently,

$$\lambda^n \left(E \setminus \bigcup_{j=1}^{l} B_j \right) \leq \lambda^n \left(\bigcup \mathcal{C}^{(l)} \right) \leq \sum_{j \geq l+1} \lambda^n(\widehat{B}_j) = 5^n \sum_{j \geq l+1} \lambda^n(B_j),$$

because $\mathcal{C}^{(l)}$ is a Vitali cover of $E \setminus \bigcup_{j=1}^{l} B_j$. Now from

$$\lambda^n \left(E \setminus \bigcup_{j} B_j \right) \leq \lambda^n \left(E \setminus \bigcup_{j=1}^{l} B_j \right) \leq 5^n \sum_{j \geq l+1} \lambda^n(B_j)$$

for $l \in \mathbb{N}$, we obtain $\lambda^n(E \setminus \bigcup_j B_j) = 0$ by letting $l \to \infty$. ■

Remark In Theorem 4.6.2, E is not required to be measurable.

Exercise 4.6.1 Show that the union of any family $\mathcal{C}$ of closed balls in $\mathbb{R}^n$ is Lebesgue measurable. (Hint: consider the Vitali cover $\mathcal{V}$ of $\bigcup \mathcal{C}$, which consists of all closed balls each of which is contained in a ball of $\mathcal{C}$).

Exercise 4.6.2 Show that Theorem 4.6.2 still holds if $\mathcal{V}$ is a Vitali cover of E consisting of open balls.

Exercise 4.6.3 Describe in $\mathbb{R}^n$ a procedure for selecting a sequence of disjoint closed cubes, from a collection $\mathcal{C}$ of closed cubes of positive bounded side lengths similar to Procedure(S), when $\mathcal{C}$ is an admissible collection of closed cubes so that Lemma 4.6.1 holds for such a collection $\mathcal{C}$. Then state the Vitali covering theorem for Vitali covers of E consisting of closed (open) cubes, where E is a subset of $\mathbb{R}^n$ with $\lambda^n(E) < \infty$.

Lebesgue differentiation of Radon measures on $\mathbb{R}^n$ is the subject we shall treat in the remaining part of this section. The differentiation is taken w.r.t. Lebesgue measure and with closed balls as base in the sense which will be defined. For the sake of simplicity in expression, a generic closed ball in $\mathbb{R}^n$ is henceforth denoted by B in this section.

Since the expression "λ^n-almost everywhere" appears often, it will hereafter be replaced by "almost everywhere". In other words, a property which holds almost everywhere w.r.t. Lebesgue measure λ^n in $\mathbb{R}^n$ will simply be said to hold almost everywhere. Accordingly, "λ^n-a.e." is often replaced by "a.e.", and λ^n-null sets will simply be called null sets.

Suppose that f is a set function (not necessarily taking only nonnegative values) defined for all closed balls inside an open set $\Omega \subset \mathbb{R}^n$ and $x \in \Omega$, define

$$\liminf_{B\to x} f(B) := \lim_{\sigma\to 0+}\left\{\inf_{\substack{\delta B<\sigma\\ x\in B}} f(B)\right\};$$

$$\limsup_{B\to x} f(B) := \lim_{\sigma\to 0+}\left\{\sup_{\substack{\delta B<\sigma\\ x\in B}} f(B)\right\}.$$

Clearly, $\liminf_{B\to x} f(B) \le \limsup_{B\to x} f(B)$; in the case $\liminf_{B\to x} f(B) = \limsup_{B\to x} f(B)$, the common value is denoted by $\lim_{B\to x} f(B)$ and we say that $\lim_{B\to x} f(B)$ exists. In the above definitions, B certainly denotes a generic closed ball B in Ω.

Exercise 4.6.4 Show that $\lim_{B\to x} f(B)$ exists and is a finite number l if and only if for any given $\varepsilon > 0$ there is $\sigma > 0$, such that

$$|f(B) - l| < \varepsilon$$

whenever $\delta B < \sigma$ and $x \in B$.

Now let μ be a Radon measure on an open set $\Omega \subset \mathbb{R}^n$; μ is said to be **differentiable** w.r.t. Lebesgue measure λ^n at $x \in \Omega$ with closed balls as base if $\lim_{B\to x} \frac{\mu(B)}{\lambda^n(B)}$ exists. Since the differentiation of Radon measures on $\mathbb{R}^n$ is always taken in this sense in what follows, if $\lim_{B\to x} \frac{\mu(B)}{\lambda^n(B)}$ exists, we simply say that μ is **differentiable at** x with derivate $\frac{d\mu}{d\lambda^n}(x) := \lim_{B\to x} \frac{\mu(B)}{\lambda^n(B)}$. We shall show that μ is differentiable with finite derivate at a.e. x of Ω, and that the function $\frac{d\mu}{d\lambda^n}$ which is defined and finite almost everywhere on Ω is measurable.

Put $\underline{D}\mu(x) = \liminf_{B\to x} \frac{\mu(B)}{\lambda^n(B)}$ and $\bar{D}\mu(x) = \limsup_{B\to x} \frac{\mu(B)}{\lambda^n(B)}$ for $x \in \Omega$. Note that $\underline{D}\mu(x) \le \bar{D}\mu(x)$ for every $x \in \Omega$.

Lemma 4.6.2 *If $\bar{D}\mu \ge \alpha$ on $S \subset \Omega$ for some $\alpha \ge 0$, then $\mu(S) \ge \alpha\lambda^n(S)$.*

Proof Clearly we may assume that $\alpha > 0$. For $l \in \mathbb{N}$, let $S_l = \{x \in S : |x| < l\}$ and let G be any open set which contains S_l and is contained in Ω. Now for any $\varepsilon > 0$ sufficiently small so that $\alpha - \varepsilon > 0$, consider the family $\mathcal{V}$ of all those closed balls $B \subset G$ such that $\mu(B) > (\alpha - \varepsilon)\lambda^n(B)$. Since $\mathcal{V}$ is a Vitali cover of S_l and $\lambda^n(S_l) < \infty$, there is a disjoint sequence $\{B_j\}$ of balls from $\mathcal{V}$ such that $\lambda^n(S_l \setminus \bigcup_j B_j) = 0$, by Vitali the covering theorem (Theorem 4.6.2). Then, $(\alpha - \varepsilon)\lambda^n(S_l) \le (\alpha - \varepsilon)\lambda^n$

$(\bigcup_j B_j) = \sum_j (\alpha - \varepsilon)\lambda^n(B_j) < \sum_j \mu(B_j) = \mu(\bigcup_j B_j) \leq \mu(G)$, and since μ is a Radon measure, it follows that $(\alpha - \varepsilon)\lambda^n(S_l) \leq \mu(S_l)$ and consequently, by letting $l \to \infty$, $(\alpha - \varepsilon)\lambda^n(S) \leq \mu(S)$ follows, as both λ^n and μ are regular measures and S_l increases to S when $l \to \infty$ (cf. Theorem 3.3.2). Finally, let $\varepsilon \searrow 0$ to conclude the proof. ■

Corollary 4.6.1 *$\bar{D}\mu < \infty$ almost everywhere on Ω.*

Proof Since $\Omega = \bigcup_{l \in \mathbb{N}} (\{x \in \Omega : \operatorname{dist}(x, \Omega^c) \geq \frac{1}{l}\} \cap \{x \in \mathbb{R}^n : |x| \leq l\})$, Ω is a countable union of compact sets; it is sufficient to show that $\lambda^n(\{x \in K : \bar{D}\mu(x) = \infty\}) = 0$ for any compact set in Ω. For such a compact set K, put $S = \{x \in K : \bar{D}\mu(x) = \infty\}$. Since for any $\alpha > 0$, $\bar{D}\mu \geq \alpha$ on S, by Lemma 4.6.2, $\lambda^n(S) \leq \frac{1}{\alpha}\mu(S) \leq \frac{1}{\alpha}\mu(K)$, which implies that $\lambda^n(S) = 0$ by letting $\alpha \to \infty$, because $\mu(K) < \infty$. ■

Lemma 4.6.3 *Suppose that $\underline{D}\mu \leq \beta$ on $S \subset \Omega$ for some $\beta \geq 0$; then there is a null set $N \subset S$ such that $\mu(S \backslash N) \leq \beta\lambda^n(S)$.*

Proof Suppose first that $\lambda^n(S) < \infty$. For $l, k \in \mathbb{N}$, take an open set G_k which contains S and is contained in Ω with $\lambda^n(G_k) < \lambda^n(S) + \frac{1}{k}$, and consider the family $\mathcal{V}$ of all those closed balls $B \subset G_k$ such that $\mu(B) < (\beta + \frac{1}{l})\lambda^n(B)$; $\mathcal{V}$ is clearly a Vitali cover of S. Since $\lambda^n(S) < \infty$, by the Vitali covering theorem there is a disjoint sequence $\{B_j\}$ of balls from $\mathcal{V}$ such that $\lambda^n(S \backslash \bigcup_j B_j) = 0$. If we let $N_{l,k} = S \backslash \bigcup_j B_j$ (observe that $\{B_j\}$ depends on l and k), $N_{l,k}$ is a null set contained in S and $(\beta + \frac{1}{l})\lambda^n(G_k) \geq (\beta + \frac{1}{l})\lambda^n(\bigcup_j B_j) > \mu(\bigcup_j B_j) \geq \mu(S \backslash N_{l,k})$. Now let $N = \bigcup_{l,k} N_{l,k}$; N is a null set in S and $(\beta + \frac{1}{l})\lambda^n(S) = (\beta + \frac{1}{l}) \inf_k \lambda^n(G_k) \geq \mu(S \backslash N)$ for each l. We simply let $l \to \infty$ to conclude that $\mu(S \backslash N) \leq \beta\lambda^n(S)$.

If $\lambda^n(S) = \infty$, for each $l \in \mathbb{N}$, put $S_l = \{x \in S : |x| \leq l\}$, then $\lambda^n(S_l) < \infty$. By the first part of the proof, for each $l \in \mathbb{N}$ there is a null set $N_l \subset S_l$ such that $\mu(S_l \backslash N_l) \leq \beta\lambda^n(S_l)$; then, $N = \bigcup_l N_l$ is a null set and $\mu(S_l \backslash N) \leq \mu(S_l \backslash N_l) \leq \beta\lambda^n(S_l) \leq \beta\lambda^n(S)$, from which $\mu(S \backslash N) \leq \beta\lambda^n(S)$ follows by letting $l \to \infty$. ■

Theorem 4.6.3 (Lebesgue) *$\frac{d\mu}{d\lambda^n}$ exists and is finite almost everywhere on Ω.*

Proof Since $\bar{D}\mu < \infty$ almost everywhere on Ω, by Corollary 4.6.1, it is only necessary to show that $\frac{d\mu}{d\lambda^n}$ exists almost everywhere on Ω. If we put $E = \{x \in \Omega : \bar{D}\mu(x) > \underline{D}\mu(x)\}$, this amounts to showing that $\lambda^n(E) = 0$; but since $\underline{D}\mu \geq 0$, $E = \bigcup_{(\alpha,\beta)} E_{(\alpha,\beta)}$, where $E_{(\alpha,\beta)} = \{x \in \Omega : \bar{D}\mu(x) \geq \alpha > \beta \geq \underline{D}\mu(x)\}$, with (α, β) being a generic pair of rational numbers α, β such that $\alpha > \beta \geq 0$, and since all such pairs (α, β) form a countable set, it suffices to show that $\lambda^n(E_{(\alpha,\beta)}) = 0$ for any such pairs of rational numbers. For such a pair (α, β), put $S = E_{(\alpha,\beta)}$. We now show that $\lambda^n(S) = 0$. Suppose the contrary that $\lambda^n(S) > 0$, then there is $l \in \mathbb{N}$ such that if we put $S_l = \{x \in S : |x| < l\}$ then $\infty > \lambda^n(S_l) > 0$. Now $\underline{D}\mu \leq \beta$ on S_l; by Lemma 4.6.3 there is a null set N inside S_l such that $\mu(S_l \backslash N) \leq \beta\lambda^n(S_l)$;

on the other hand, the fact that $\bar{D}\mu \geq \alpha$ on $S_l \backslash N$ implies, by Lemma 4.6.2, that $\alpha\lambda^n(S_l) = \alpha\lambda^n(S_l\backslash N) \leq \mu(S_l\backslash N)$. Thus,

$$\mu(S_l\backslash N) \leq \beta\lambda^n(S_l) < \alpha\lambda^n(S_l) = \alpha\lambda^n(S_l\backslash N) \leq \mu(S_l\backslash N),$$

the absurdity of which shows that $\lambda^n(S) = 0$. ∎

If we let D denote the set of all $x \in \Omega$ such that $\frac{d\mu}{d\lambda^n}(x)$ exists and is finite, D is measurable because D is the complement in Ω of a null set and null sets are measurable. We shall show in a moment that $\frac{d\mu}{d\lambda^n}$ is measurable as a function defined a.e. on Ω (cf. Section 2.5 for measurability of functions defined a.e. on Ω). For $x \in D$, $\frac{d\mu}{d\lambda^n}(x) = \lim_{B\to x} \frac{\mu(B)}{\lambda^n(B)}$, a fortiori, $\frac{d\mu}{d\lambda^n}(x) = \lim_{r\to 0} \frac{\mu(C_r(x))}{\lambda^n(C_r(x))}$, where $C_r(x)$ is the closed ball centered at x and with radius $r > 0$. Now if, as before, $B_r(x)$ denotes the open ball centered at x and with radius $r > 0$, we claim that

$$\frac{d\mu}{d\lambda^n}(x) = \lim_{r\to 0} \frac{\mu(B_r(x))}{\lambda^n(B_r(x))}. \tag{4.9}$$

To see this one needs only to observe that if $r' = r(1 - r)$ for $0 < r < 1$, then

$$(1-r)^n \frac{\mu(C_{r'}(x))}{\lambda^n(C_{r'}(x))} = \frac{\lambda^n(C_{r'}(x))}{\lambda^n(C_r(x))} \frac{\mu(C_{r'}(x))}{\lambda^n(C_{r'}(x))} \leq \frac{\mu(B_r(x))}{\lambda^n(B_r(x))} \leq \frac{\mu(C_r(x))}{\lambda^n(C_r(x))},$$

where the relation $\lambda^n(C_{r'}(x)) = (1-r)^n\lambda^n(C_r(x))$ has been used (cf. Example 4.3.1 (ii)), and (4.9) follows as $r \to 0$.

Lemma 4.6.4 *$\frac{d\mu}{d\lambda^n}$ is measurable.*

Proof For $x \in \Omega$ and $r > 0$, let $\Omega_r(x) = B_r(x) \cap \Omega$. First, we show that as a function of x, $\mu(\Omega_r(x))$ is lower semi-continuous on D (r being fixed). For $x \in D$, let I_x denote the indicator function of the set $\Omega_r(x)$, then $\mu(\Omega_r(x)) = \int_\Omega I_x d\mu$. Suppose now that $\{x_k\}$ is a sequence in D tending to x. Since $I_{x_k} \to I_x$ on $\Omega_r(x)$ and $I_x = 0$ on $\Omega\backslash\Omega_r(x)$, $I_x \leq \liminf_{k\to\infty} I_{x_k}$. It follows from the Fatou Lemma that $\mu(\Omega_r(x)) = \int_\Omega I_x d\mu \leq \liminf_{k\to\infty} \int_\Omega I_{x_k} d\mu = \liminf_{k\to\infty} \mu(\Omega_r(x_k))$. Hence, $\mu(\Omega_r(x))$ is lower semi-continuous as a function of x on D and is therefore measurable on D. Similarly, $\lambda^n(\Omega_r(x))$ is lower semi-continuous on D. By choosing a sequence of r tending to zero, we have

$$\frac{d\mu}{d\lambda^n}(x) = \lim_{r\to 0} \frac{\mu(B_r(x))}{\lambda^n(B_r(x))} = \lim_{r\to 0} \frac{\mu(\Omega_r(x))}{\lambda^n(\Omega_r(x))}$$

for $x \in D$, hence $\frac{d\mu}{d\lambda^n}$ is measurable on D (note that $\Omega_r(x) = B_r(x)$ if r is small). Since $\Omega\backslash D$ is a null set, $\frac{d\mu}{d\lambda^n}$ is measurable. ∎

$\frac{d\mu}{d\lambda^n}$ is usually extended from D to Ω by defining it to be zero on $\Omega\backslash D$. In view of Exercise 3.9.1(ii), $\frac{d\mu}{d\lambda^n}$ has a Borel measurable version and we shall henceforth take $\frac{d\mu}{d\lambda^n}$ to be a Borel measurable function on Ω.

Lemma 4.6.5 *For any Borel set $S \subset \Omega$, $\int_S \frac{d\mu}{d\lambda^n} d\lambda^n \leq \mu(S)$. In particular $\int_K \frac{d\mu}{d\lambda^n} d\lambda^n < \infty$ for compact sets $K \subset \Omega$.*

Proof Let g be a generic nonnegative and Borel measurable simple function on Ω satisfying $g \leq \frac{d\mu}{d\lambda^n} \cdot I_S$; there are disjoint Borel sets $A_1, \ldots, A_l$ in S and nonnegative numbers $\alpha_1, \ldots, \alpha_l$ such that $g = \sum_{j=1}^{l} \alpha_j I_{A_j}$. Then,

$$\int_\Omega g d\lambda^n = \sum_{j=1}^{l} \alpha_j \lambda^n(A_j).$$

But $\mu(A_j) \geq \alpha_j \lambda^n(A_j)$, $j = 1, \ldots, l$, by Lemma 4.6.2, consequently,

$$\int_\Omega g d\lambda^n \leq \sum_{j=1}^{l} \mu(A_j) = \mu\left(\sum_{j=1}^{l} A_j\right) \leq \mu(S),$$

and hence,

$$\int_S \frac{d\mu}{d\lambda^n} d\lambda^n = \int_\Omega \frac{d\mu}{d\lambda^n} \cdot I_S d\lambda^n = \sup_g \int_\Omega g d\lambda^n \leq \mu(S). \qquad \blacksquare$$

Lemma 4.6.5 implies that $\{\frac{d\mu}{d\lambda^n}\lambda^n\}^*$ is a Radon measure on Ω (cf. Example 3.8.1). Recall that $\{\frac{d\mu}{d\lambda^n}\lambda^n\}$ denotes the indefinite integral of $\frac{d\mu}{d\lambda^n}$ with respect to λ^n. Since indefinite integrals, considered later in this chapter, are always λ^n-indefinite integrals, the notation $\{\frac{d\mu}{d\lambda^n}\lambda^n\}$ is simplified to $\{\frac{d\mu}{d\lambda^n}\}$ for compactness of expression. Similarly, for a nonnegative measurable function f defined on Ω, $\{f\lambda^n\}$ will be replaced by $\{f\}$. With this notational convention, if f is **locally integrable** on Ω in the sense that f is integrable on compact sets in Ω, then $\{f\}^*$ is a Radon measure on Ω. Thus $\{\frac{d\mu}{d\lambda^n}\}^*$ is a Radon measure on Ω.

Another immediate consequence of Lemma 4.6.5 is the following.

Corollary 4.6.2 *For any $S \subset \Omega$, $\{\frac{d\mu}{d\lambda^n}\}^*(S) \leq \mu(S)$.*

Proof If S is a Borel set, then $\{\frac{d\mu}{d\lambda^n}\}^*(S) = \int_S \frac{d\mu}{d\lambda^n} d\lambda^n \leq \mu(S)$, by Lemma 4.6.5; for general S, the same inequality follows from the fact that both $\{\frac{d\mu}{d\lambda^n}\}^*$ and μ are Borel regular. $\blacksquare$

Remark As shown in Example 3.8.1, $\{f\}^*(S) = \int_S f d\lambda^n$ if S is a measurable subset of Ω and f a nonnegative measurable function. Hence, $\{\frac{d\mu}{d\lambda^n}\}^*(S) = \int_S \frac{d\mu}{d\lambda^n} d\lambda^n$ if S is a measurable subset of Ω.

Corollary 4.6.3 *If f is a nonnegative and locally integrable function on Ω, then $\frac{d}{d\lambda^n}\{f\}^* = f$ a.e. on Ω; i.e. for almost all $x \in \Omega$, $\lim_{B\to x} \frac{\int_B f d\lambda^n}{\lambda^n(B)} = f(x)$; in particular,*

$$f(x) = \lim_{r\to 0} \frac{1}{\lambda^n(B_r(x))} \int_{B_r(x)} f d\lambda^n \tag{4.10}$$

for almost every $x \in \Omega$.

Proof Put $g = \frac{d}{d\lambda^n}\{f\}^*$. By Corollary 4.6.2 and the remark following it, $\int_S g d\lambda^n = \{g\}^*(S) \le \{f\}^*(S) = \int_S f d\lambda^n$ for any measurable set $S \subset \Omega$, hence $g \le f$ a.e. on Ω.

Now, put $E = \{g < f\}$; we will show that $\lambda^n(E) = 0$ to conclude that $f = g$ a.e. For this we need only show that $\lambda^n(E') = 0$, where

$$E' = \left\{ x \in E : \lim_{B\to x} \frac{\{f\}^*(B)}{\lambda^n(B)} \text{ exists} \right\}.$$

Suppose the contrary, that $\lambda^n(E') > 0$, then there are numbers $0 < \beta < \alpha < \infty$ and $R > 0$ such that the set $S = \{x \in E' : |x| < R, f(x) > \alpha > \beta > g(x)\}$ has positive Lebesgue measure. Let G be any open set containing S and contained in Ω, and consider the family $\mathcal{V}$ of all $B \subset G$ satisfying $\beta\lambda^n(B) > \{f\}^*(B)$. $\mathcal{V}$ is a Vitali cover of S; by the Vitali covering theorem, there is a disjoint sequence $\{B_j\}$ of balls from $\mathcal{V}$ such that $\lambda^n(S \setminus \bigcup_j B_j) = 0$ (note that $\lambda^n(S) < \infty$). Then,

$$\begin{aligned}\beta\lambda^n(G) &\ge \beta\lambda^n\left(\bigcup_j B_j\right) = \sum_j \beta\lambda^n(B_j) > \sum_j \{f\}^*(B_j) = \{f\}^*\left(\bigcup_j B_j\right) \\ &= \int_{\bigcup_j B_j} f d\lambda^n \ge \int_S f d\lambda^n,\end{aligned}$$

from which it follows that $\beta\lambda^n(S) \ge \int_S f d\lambda^n$; on the other hand $\int_S f d\lambda^n \ge \alpha\lambda^n(S)$, hence,

$$\beta\lambda^n(S) \ge \int_S f d\lambda^n \ge \alpha\lambda^n(S),$$

the absurdity of which shows that $\lambda^n(E') = 0$. That (4.10) holds for almost all $x \in \Omega$ follows from (4.9). ■

Example 4.6.1 (Density and approximate continuity) Let D be a measurable subset of $\mathbb{R}^n$ with $\lambda^n(D) > 0$. For $x \in \mathbb{R}^n$, if $\lim_{B\to x} \frac{\lambda^n(B\cap D)}{\lambda^n(B)}$ exists, the limit is called the density of D at x. Certainly, the density is nonnegative and ≤ 1. If the density of D at x is 1, x is called a density point of D; while x is called a point of dispersion of D if the density of D at x is 0. A measurable function f on D is said to have **approximate limit** l at x if x is a density point of the set $\{y \in D : |f(y) - l| < \varepsilon\}$ for every $\varepsilon > 0$, and the approximate limit l will be denoted by $\operatorname{aplim}_{y\to x} f(y)$. The function f is called approximately continuous at $x \in D$ if $\operatorname{aplim}_{y\to x} f(y) = f(x)$. We claim that (i) almost every point of D is a density point of D, and almost every point of D^c is a point of dispersion of D, and (ii) a measurable function f on D is approximately continuous a.e. on D. Assertion (i) is a direct consequence of Corollary 4.6.3, by choosing f to be the indicator function of D. Observe that (i) implies that if g is a continuous function on $\mathbb{R}^n$, then f is approximately continuous at almost every point of the set $\{x \in D : f(x) = g(x)\}$. It is clear now that (ii) follows from this observation and the Lusin theorem (Theorem 4.1.1).

Exercise 4.6.5 Suppose that A is a measurable subset of $\mathbb{R}^n$. Show that $\operatorname{dist}(y, A) = o(|y - x|)$ as $y \to x$ for almost every x in A.

For a locally integrable function f on Ω, the set $L(f)$ of all those points $x \in \Omega$ such that $\lim_{B\to x} \frac{1}{\lambda^n B} \int_B |f(y) - f(x)| dy = 0$ is called the **Lebesgue set** of f.

Theorem 4.6.4 *If f is locally integrable on Ω, then $\lambda^n(\Omega \backslash L(f)) = 0$, i.e. $L(f)$ consists of almost all points of Ω.*

Proof Denote by γ the set of all rational numbers in $\mathbb{R}$. For $a \in \gamma$, there is a null set E_a in Ω such that for $x \in \Omega \backslash E_a$, the following holds, by Corollary 4.6.3:

$$\lim_{B\to x} \frac{1}{\lambda^n(B)} \int_B |f(y) - a| dy = |f(x) - a|.$$

Put $E = \bigcup_{a\in\gamma} E_a$, then $\lambda^n(E) = 0$. For $x \in \Omega \backslash E$ and $\varepsilon > 0$, there is $a \in \gamma$ such that $|f(x) - a| < \varepsilon$, hence,

$$\begin{aligned}\limsup_{B\to x} \frac{1}{\lambda^n(B)} \int_B |f(y) - f(x)| dy &\le \limsup_{B\to x} \frac{1}{\lambda^n(B)} \int_B \{|f(y) - a| + |f(x) - a|\} dy \\ &= 2|f(x) - a| < 2\varepsilon.\end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, we have $\limsup_{B\to x} \frac{1}{\lambda^n(B)} \int_B |f(y) - f(x)| = 0$, or $\lim_{B\to x} \frac{1}{\lambda^n(B)} \int_B |f(y) - f(x)| dy = 0$. ■

Theorem 4.6.5 *If f is locally integrable on Ω, then*

$$\lim_{B\to x} \frac{1}{\lambda^n(B)} \int_B f d\lambda^n = f(x)$$

for almost every $x \in \Omega$.

Proof For $x \in L(f)$,

$$\left| \frac{1}{\lambda^n(B)} \int_B f(y) dy - f(x) \right| \le \frac{1}{\lambda^n(B)} \int_B |f(y) - f(x)| dy$$

for any closed ball B containing x; then $\lim_{B\to x} \int_B f d\lambda^n = f(x)$ follows. ■

As an application of Theorem 4.6.5, we shall now prove that the space $C_c(\Omega)$ of all those continuous functions on Ω, each of which vanishes outside a compact subset of Ω, is dense in $L^p(\Omega, \mathcal{L}^n|\Omega, \lambda^n)$:

Proposition 4.6.1 *$C_c(\Omega)$ is dense in $L^p(\Omega, \mathcal{L}^n|\Omega, \lambda^n)$, $1 \le p < \infty$.*

Proof Let $f \in L^p(\Omega, \mathcal{L}^n|\Omega, \lambda^n)$ and $\varepsilon > 0$. For each $k \in \mathbb{N}$, put $F_k = \{x \in \Omega : \operatorname{dist}(x, \Omega^c) \ge \frac{1}{k}\} \cap C_k(0)$; $\{F_k\}$ is an increasing sequence of compact sets in Ω and $\Omega = \bigcup_k F_k$. Set $f_k = I_{F_k} f$, then $\lim_{k\to\infty} f_k(x) = f(x)$ for all $x \in \Omega$ and $|f_k| \le |f|$. LDCT implies that $\lim_{k\to\infty} \|f_k - f\|_p = 0$. There is then k_0 such that $\|f_{k_0} - f\|_p < \frac{\varepsilon}{3}$. Now, for each $l \in \mathbb{N}$, let $g_l(x) = f_{k_0}(x)$ if $|f_{k_0}(x)| \le l$, otherwise let $g_l(x) = 0$. By LDCT again, there is $l_0 \in \mathbb{N}$ such that $\|g_{l_0} - f_{k_0}\|_p < \frac{\varepsilon}{3}$. Put $g = g_{l_0}$; g is a bounded function and $g = 0$ outside F_{k_0}. For $0 < r < \frac{1}{2k_0}$, define $[g]_r(x) = \frac{1}{\lambda^n(B_r(x))} \int_{B_r(x)} g(y)dy$, if $x \in F_{2k_0}$; otherwise let $[g]_r(x) = 0$. Obviously, $[g]_r \in C_c(\Omega)$, $|[g]_r| \le l_{k_0}$ on F_{2k_0} and $[g]_r = 0$ outside F_{2k_0}. $[g]_r$ is therefore in $L^p(\Omega, \mathcal{L}^n|\Omega, \lambda^n)$. Since $\lim_{r\to 0}[g]_r = g$ a.e., by Theorem 4.6.5, LDCT implies $\lim_{r\to 0} \|[g]_r - g\|_p = 0$. Choose $0 < r_0 < \frac{1}{2k_0}$ such that $\|[g]_{r_0} - g\|_p < \frac{\varepsilon}{3}$. Then, $g_{r_0} \in C_c(\Omega)$ and $\|f - [g]_{r_0}\|_p \le \|f - f_{k_0}\|_p + \|f_{k_0} - g\|_p + \|g - [g]_{r_0}\|_p < \varepsilon$. ■

Theorem 4.6.6 *Suppose that D is a measurable set in $\mathbb{R}^n$ with positive measure. Then $L^p(D, \mathcal{L}^n|D, \lambda^n)$ is separable for $1 \le p < \infty$.*

Proof In the proof, we shall denote by $L^p(A)$ the space $L^p(A, \mathcal{L}|A, \lambda^n)$ if $A \in \mathcal{L}^n$. Since if $\{u_k\}_{k\in\mathbb{N}}$ is dense in $L^p(\mathbb{R}^n)$, then $\{u_k|_D\}_{k\in\mathbb{N}}$ is dense in $L^p(D)$, it is sufficient to show that $L^p(\mathbb{R}^n)$ is separable.

We call the indicator function of an oriented interval $I_1 \times \cdots \times I_n$ an elementary unit function of order k, if each I_j, $j = 1, \ldots, n$, is of the form $I_j = \left[\frac{l_j}{2^k}, \frac{l_j+1}{2^k}\right)$, $l_j \in \mathbb{Z}$. Consider now the family $\mathcal{E}$ of all finite linear combinations of elementary unit functions of all possible order with rational coefficients. It is clear that $\mathcal{E}$ is a countable set in $L^p(\mathbb{R}^n)$. Let $u \in C_c(\mathbb{R}^n)$ and $\varepsilon > 0$. As u vanishes outside $J = J_1 \times \cdots \times J_n$ with each $J_j = [-n_0, n_0]$ for some $n_0 \in \mathbb{N}$, and u is uniformly continuous on J, for any given $\varepsilon > 0$ there is $g \in \mathcal{E}$ such that $\|u - g\|_p < \varepsilon$; hence the closure of $\mathcal{E}$ in $L^p(\mathbb{R}^n)$ contains $C_c(\mathbb{R}^n)$. Thus the closure of $\mathcal{E}$ in $L^p(\mathbb{R}^n)$ is $L^p(\mathbb{R}^n)$, by Proposition 4.6.1. ■

Lemma 4.6.6 *There is a Radon measure φ on Ω such that $\mu = \{\frac{d\mu}{d\lambda^n}\}^* + \varphi$, i.e. $\mu(S) = \{\frac{d\mu}{d\lambda^n}\}^*(S) + \varphi(S)$ for all $S \subset \Omega$.*

Proof Denote by $\mathcal{K}(\Omega)$ the family of all compact sets in Ω. Both μ and $\{\frac{d\mu}{d\lambda^n}\}^*$ take finite value on $\mathcal{K}(\Omega)$; we define φ_1 on $\mathcal{K}(\Omega)$ by

$$\varphi_1(K) = \mu(K) - \left\{\frac{d\mu}{d\lambda^n}\right\}^*(K)$$

for $K \in \mathcal{K}(\Omega)$. By Corollary 4.6.2, $\varphi_1 \ge 0$. Observe that

(i) φ_1 is monotone on $\mathcal{K}(\Omega)$, i.e. for $K_1 \subset K_2$ in $\mathcal{K}(\Omega)$, $\varphi_1(K_1) \le \varphi_1(K_2)$.

(ii) For any finite number of disjoint sets $K_1, \ldots, K_l$ in $\mathcal{K}(\Omega)$,

$$\varphi_1\left(\bigcup_{j=1}^{l} K_j\right) = \sum_{j=1}^{l} \varphi_1(K_j).$$

Now define φ on $\mathcal{B}(\Omega)$ by

$$\varphi(A) = \sup \varphi_1(K)$$

for $A \in \mathcal{B}(\Omega)$, where the supremum is taken over all $K \in \mathcal{K}(\Omega)$ with $K \subset A$. Then φ is an extension of φ_1 and

$$\mu(A) = \left\{\frac{d\mu}{d\lambda^n}\right\}^* (A) + \varphi(A) \tag{4.11}$$

for $A \in \mathcal{B}(\Omega)$. That φ is an extension of φ_1 follows from (i), while (4.11) holds by taking the limit as $j \to \infty$ on both sides of

$$\mu(K_j) = \left\{\frac{d\mu}{d\lambda^n}\right\}^* (K_j) + \varphi(K_j),$$

for a sequence $\{K_j\} \subset \mathcal{K}(\Omega)$ such that $\lim_{j\to\infty} \mu(K_j) = \mu(A)$, $\lim_{j\to\infty} \{\frac{d\mu}{d\lambda^n}\}^*(K_j) = \{\frac{d\mu}{d\lambda^n}\}^*(A)$, and $\lim_{j\to\infty} \varphi(K_j) = \varphi(A)$. That such a sequence $\{K_j\}$ exists follows by applying Theorem 3.9.1 (ii) to μ and $\{\frac{d\mu}{d\lambda^n}\}^*$, and by definition of φ.

If now $\{A_j\}$ is a disjoint sequence of Borel sets in a given compact set $K \subset \Omega$, then both $\mu(\bigcup_j A_j)$ and $\{\frac{d\mu}{d\lambda^n}\}^*(\bigcup_j A_j)$ are finite, and by (4.11),

$$\begin{aligned}\varphi\left(\bigcup_j A_j\right) &= \mu\left(\bigcup_j A_j\right) - \left\{\frac{d\mu}{d\lambda^n}\right\}^* \left(\bigcup_j A_j\right) \\ &= \sum_j \left\{\mu(A_j) - \left\{\frac{d\mu}{d\lambda^n}\right\}^* (A_j)\right\} = \sum_j \varphi(A_j),\end{aligned}$$

hence we have:

(iii) For disjoint sequence $\{A_j\} \subset \mathcal{B}(\Omega)$ with $\bigcup_j A_j \subset K$ for some compact set K in Ω,

$$\varphi\left(\bigcup_j A_j\right) = \sum_j \varphi(A_j).$$

Next, we claim that φ is σ-additive on $\mathcal{B}(\Omega)$. Let $\{A_j\}$ be any disjoint sequence in $\mathcal{B}(\Omega)$. For any compact set $K \subset \bigcup_j A_j$,

$$\varphi(K) = \varphi\left(\bigcup_j \{K \cap A_j\}\right) = \sum_j \varphi(K \cap A_j) \leq \sum_j \varphi(A_j),$$

by (iii), and the obvious fact that φ is monotone on $\mathcal{B}(\Omega)$. Consequently,

$$\varphi\left(\bigcup_j A_j\right) \leq \sum_j \varphi(A_j). \tag{4.12}$$

On the other hand, fix $l \in \mathbb{N}$ and for each $j = 1, \ldots, l$ take an arbitrary compact set $K_j \subset A_j$, then

$$\varphi\left(\bigcup_j A_j\right) \geq \varphi\left(\bigcup_{j=1}^{l} K_j\right) = \sum_{j=1}^{l} \varphi(K_j), \tag{4.13}$$

by monotony of φ on $\mathcal{B}(\Omega)$ and (ii); since each K_j is an arbitrary compact set in A_j, it follows from (4.13) that

$$\varphi\left(\bigcup_j A_j\right) \geq \sum_{j=1}^{l} \varphi(A_j),$$

and hence,

$$\varphi\left(\bigcup_j A_j\right) \geq \sum_j \varphi(A_j).$$

The last inequality shows, together with (4.12), that $\varphi(\bigcup_j A_j) = \sum_j \varphi(A_j)$. Thus φ is σ-additive on $\mathcal{B}(\Omega)$.

Now construct from φ on $\mathcal{B}(\Omega)$ a measure on Ω by Method I, which is the unique $\mathcal{B}(\Omega)$-regular extension of φ by Corollary 3.4.1 and hence is a Radon measure. The Radon measure so constructed is to be denoted also by φ. That $\mu = \{\frac{d\mu}{d\lambda^n}\}^* + \varphi$ holds follows from (4.11) and Borel regularity of μ, $\{\frac{d\mu}{d\lambda^n}\}^*$, and φ. ■

Theorem 4.6.7 (Lebesgue decomposition theorem) *There is a null set $N \subset \Omega$ such that*

$$\mu = \left\{\frac{d\mu}{d\lambda^n}\right\}^* + \mu\lfloor N.$$

Proof By Lemma 4.6.6, there is a Radon measure φ on Ω such that

$$\mu = \left\{\frac{d\mu}{d\lambda^n}\right\}^* + \varphi. \tag{4.14}$$

Choose a null set $N_1 \subset \Omega$ such that, for $x \in \Omega \backslash N_1$, the derivates $\lim_{B \to x} \frac{\mu(B)}{\lambda^n(B)}$, $\lim_{B \to x} \frac{\{\frac{d\mu}{d\lambda^n}\}^*(B)}{\lambda^n(B)}$, and $\lim_{B \to x} \frac{\varphi(B)}{\lambda^n(B)}$ exist and are finite, and further, $\frac{d}{d\lambda^n}\{\frac{d\mu}{d\lambda^n}\}^*(x) = \frac{d\mu}{d\lambda^n}(x)$. That such a null set N_1 exists is a consequence of Theorem 4.6.3 and Corollary 4.6.3. From the choice of N_1 and (4.14), one concludes that the derivate $\frac{d\varphi}{d\lambda^n}(x) = 0$ for $x \in \Omega \backslash N_1$, and hence, in view of Lemma 4.6.3, there is a null set $N_2 \subset \Omega \backslash N_1$ such that $\varphi(\Omega \backslash (N_1 \cup N_2)) = 0$. Put $N = N_1 \cup N_2$; N is a null set, and for any $S \subset \Omega$,

$$\varphi(S \cap N) \leq \varphi(S) \leq \varphi(S \cap (\Omega \backslash N)) + \varphi(S \cap N) = \varphi(S \cap N),$$

or

$$\varphi(S) = \varphi(S \cap N).$$

Now,

$$\mu(S \cap N) = \left\{\frac{d\mu}{d\lambda^n}\right\}^* (S \cap N) + \varphi(S \cap N) = \varphi(S),$$

consequently,

$$\mu(S) = \left\{\frac{d\mu}{d\lambda^n}\right\}^* (S) + \varphi(S) = \left\{\frac{d\mu}{d\lambda^n}\right\}^* (S) + \mu(S \cap N)$$

for any $S \subset \Omega$; in other words,

$$\mu = \left\{\frac{d\mu}{d\lambda^n}\right\}^* + \mu \lfloor N. \qquad \blacksquare$$

The decomposition of μ into the sum $\{\frac{d\mu}{d\lambda^n}\}^* + \mu \lfloor N$ in Theorem 4.6.7 is called the **Lebesgue decomposition** of μ.

Concepts of absolute continuity and singularity for measures are introduced now for the purpose of singling out a distinguishing feature of the Lebesgue decomposition theorem. Suppose μ and ν are measures on a set Ω. The measure μ is said to be ν-**absolutely continuous** if $\mu(A) = 0$ whenever $\nu(A) = 0$; and μ is said to be ν-**singular** if $\mu = \mu \lfloor N$ where $\nu(N) = 0$. If Ω is a subset of $\mathbb{R}^n$, then a measure μ on Ω being λ^n-absolute continuous or λ^n-singular will simply be said to be absolutely continuous or singular, in this order.

In the decomposition $\mu = \{\frac{d\mu}{d\lambda^n}\}^* + \mu \lfloor N$, where $\lambda^n(N) = 0$, as claimed by Theorem 4.6.7, $\{\frac{d\mu}{d\lambda^n}\}^*$ is absolutely continuous and $\mu \lfloor N$ is singular. Thus, Theorem 4.6.7 claims that any Radon measure on Ω can be decomposed into an absolutely continuous part and a singular part. We shall see presently that such a decomposition is unique.

Lemma 4.6.7 *If μ is an absolutely continuous Radon measure on Ω, then $\mu = \{\frac{d\mu}{d\lambda^n}\}^*$.*

Proof By Theorem 4.6.7,

$$\mu = \left\{\frac{d\mu}{d\lambda^n}\right\}^* + \mu\lfloor N,$$

where $\lambda^n(N) = 0$; but absolute continuity of μ implies $\mu(N) = 0$ and hence $\mu\lfloor N = 0$. ■

Lemma 4.6.8 *If μ is a singular Radon measure on Ω, then $\frac{d\mu}{d\lambda^n} = 0$ a.e. on Ω.*

Proof There are null sets N and N' in Ω such that

$$\mu = \left\{\frac{d\mu}{d\lambda^n}\right\}^* + \mu\lfloor N$$

and

$$\mu = \mu\lfloor N',$$

by Theorem 4.6.7 and singularity of μ. For any set $S \subset \Omega\backslash N'$, we have

$$\mu(S) = \mu(S \cap N') = \mu(\emptyset) = 0,$$

and hence,

$$0 = \mu(S) = \left\{\frac{d\mu}{d\lambda^n}\right\}^*(S) + \mu(N \cap S),$$

a fortiori, $\{\frac{d\mu}{d\lambda^n}\}^*(S) = 0$. Since $\{\frac{d\mu}{d\lambda^n}\}^*(S) = \int_S \frac{d\mu}{d\lambda^n} d\lambda^n = 0$ for any measurable $S \subset \Omega\backslash N'$, $\frac{d\mu}{d\lambda^n} = 0$ a.e. on $\Omega\backslash N'$, and consequently $\frac{d\mu}{d\lambda^n} = 0$ a.e. on Ω. ■

Theorem 4.6.8 *For a Radon measure μ on Ω, the Lebesgue decomposition $\mu = \{\frac{d\mu}{d\lambda^n}\}^* + \mu\lfloor N$, where $\lambda^n(N) = 0$, is the unique decomposition of μ into a sum of an absolutely continuous and a singular Radon measure.*

Proof Let $\mu = \mu_a + \mu_s$ be a decomposition of μ into the sum of an absolutely continuous Radon measure μ_a and a singular Radon measure μ_s. Then,

$$\frac{d\mu}{d\lambda^n} = \frac{d\mu_a}{d\lambda^n} + \frac{d\mu_s}{d\lambda^n}$$

almost everywhere on Ω. Since $\frac{d\mu_s}{d\lambda^n} = 0$ a.e. on Ω, by Lemma 4.6.8, $\frac{d\mu}{d\lambda^n} = \frac{d\mu_a}{d\lambda^n}$ a.e. From Lemma 4.6.7, $\mu_a = \{\frac{d\mu_a}{d\lambda^n}\}^* = \{\frac{d\mu}{d\lambda^n}\}^*$. Let $\mu = \{\frac{d\mu}{d\lambda^n}\}^* + \mu\lfloor N$ be the Lebesgue decomposition of μ; then by what has just being proved,

$$\mu(S) = \mu_a(S) + \mu_s(S) = \left\{\frac{d\mu}{d\lambda^n}\right\}^*(S) + \mu\lfloor N(S) = \mu_a(S) + \mu\lfloor N(S);$$

in particular, if $\mu(S) < \infty$, $\mu_s(S) = \mu\lfloor N(S)$, from which $\mu_s = \mu\lfloor N$ follows by Theorem 3.3.2, because both μ_s and $\mu\lfloor N$ are regular. ■

Exercise 4.6.6 Let H^n be the n-dimensional Hausdorff measure on $\mathbb{R}^n$.

(i) Show that H^n is a Radon measure on $\mathbb{R}^n$.
(ii) Show that $\frac{dH^n}{d\lambda^n}(x) = \alpha_n$ for all $x \in \mathbb{R}^n$, where α_n is a constant depending only on the dimension n.
(iii) Show that $H^n = \alpha_n \lambda^n$.

The results in this section will be applied in Section 4.7 to study differentiability of functions of a real variable; while differentiability of measures in a general setting will be taken up in Section 5.7, where a decomposition theorem similar to Theorem 4.6.8 is established.

4.7 Differentiability of functions of a real variable and related functions

Differentiability of functions of a real variable will be studied through differentiation of Lebesgue–Stieltjes measures generated by monotone functions. An important subclass of the class of BV functions will be introduced. This is the class of absolutely continuous functions, which is much larger than the class of continuously differentiable functions, but enjoys many useful properties of the latter; in particular, the formula of integration by parts holds for absolutely continuous functions.

We start with the almost everywhere differentiability for monotone functions.

Lemma 4.7.1 *If g is a finite-valued monotone increasing function on $\mathbb{R}$, then the derivative g' exists and is finite almost everywhere on $\mathbb{R}$ and g' is measurable. Furthermore, $g' = \frac{d\mu_g}{d\lambda}$ a.e.*

Proof Let μ_g be the Lebesgue–Stieltjes measure generated by g. We know from Theorem 4.6.3 that the derivate

$$\frac{d\mu_g}{d\lambda}(x) = \lim_{I \to x} \frac{\mu_g(I)}{|I|}$$

exists and is finite for x in a subset D of $\mathbb{R}$ with $\lambda(\mathbb{R}\backslash D) = 0$, where I denotes a generic finite closed interval in $\mathbb{R}$. We claim that for $x \in D$, $g'(x)$ exists and equals $\frac{d\mu_g}{d\lambda}(x)$. Note first that points in D are necessarily points of continuity of g and $\mu_g([a, b]) = g(b) - g(a)$ if g is continuous at a and b (cf. Lemma 3.7.2). Now for $x \in D$, if $y \to x+$ through points of continuity of g, then $\lim_{y\to x+} \frac{g(y)-g(x)}{y\to x} = \frac{d\mu_g}{d\lambda}(x)$; in general, for any

$y > x$, choose points of continuity y' and y'' such that $x < y' < y < y''$ and such that $\lim_{y\to x+} \frac{y'-x}{y-x} = \lim_{y\to x+} \frac{y''-x}{y-x} = 1$, then,

$$\left(\frac{y'-x}{y-x}\right)\frac{g(y')-g(x)}{y'-x} \leq \frac{g(y)-g(x)}{y-x} \leq \left(\frac{y''-x}{y-x}\right)\frac{g(y'')-g(x)}{y''-x},$$

from which by taking the limit as $y \to x+$, we obtain $\lim_{y\to x+} \frac{g(y)-g(x)}{y-x} = \frac{d\mu_g}{d\lambda}(x)$. Similarly, $\lim_{y\to x-} \frac{g(y)-g(x)}{y-x} = \lim_{y\to x-} \frac{g(x)-g(y)}{x-y} = \frac{d\mu_g}{d\lambda}(x)$. Thus, $g'(x) = \frac{d\mu_g}{d\lambda}(x)$ for $x \in D$. This means that g' exists almost everywhere on $\mathbb{R}$. That g' is measurable follows from Lemma 4.6.4 and the fact that $g' = \frac{d\mu_g}{d\lambda}$ a.e. ■

Theorem 4.7.1 *If f is a BV function on a finite closed interval $[a, b]$, then f' exists a.e. on $[a, b]$ and is integrable. Furthermore,*

$$V_a^x(f) \geq \int_a^x |f'|d\lambda$$

for $x \in [a, b]$.

Proof Put $f_1(x) = f(a) + P_a^x(f)$ and $f_2(x) = N_a^x(f)$ for $x \in [a, b]$; then f_1 and f_2 are monotone increasing on $[a, b]$ and $f = f_1 - f_2$. That f' exists a.e. on $[a, b]$ and is measurable follows from Lemma 4.7.1 by extending f_1 and f_2 to be defined and monotone increasing on $\mathbb{R}$, as in the last paragraph of Section 3.7 and by extending f by $f = f_1 - f_2$ on $\mathbb{R}$. Then $f' = f_1' - f_2'$ a.e. on $\mathbb{R}$.

If for $i = 1, 2$, we let μ_i be the Lebesgue–Stieltjes measure on $\mathbb{R}$ generated by f_i, then from the Lebesgue decomposition theorem,

$$\mu_i = \left\{\frac{d\mu_i}{d\lambda}\right\}^* + \mu_i\lfloor N_i) \geq \left\{\frac{d\mu_i}{d\lambda}\right\}^* = \{f_i'\}^*,$$

where N_i is a null set in $\mathbb{R}$ and $\frac{d\mu_i}{d\lambda} = f_i'$ a.e., by Lemma 4.7.1. As a consequence, for $x \in [a, b]$, $P_a^x(f) = f_1(x) - f_1(a) \geq \mu_1([a, x)) \geq \int_a^x f_1' d\lambda$; similarly, $N_a^x(f) \geq \int_a^x f_2' d\lambda$. Now, $V_a^x(f) = P_a^x(f) + N_a^x(f) \geq \int_a^x (f_1' + f_2')d\lambda \geq \int_a^x |f'|d\lambda$. That f' is integrable follows from $\int_a^b |f'|d\lambda \leq V_a^b(f) < \infty$. ■

Remark Although the measurability of g' in Lemma 4.7.1 follows from that of $\frac{d\mu_g}{d\lambda}$ by Lemma 4.6.4, if a measurable function f is differentiable a.e., the measurability of f' follows from the measurability of the limit of a sequence of measurable functions. Actually,

$$f'(x) = \lim_{k\to\infty} k\left\{f\left(x + \frac{1}{k}\right) - f(x)\right\}$$

if $f'(x)$ exists, and for each $k \in \mathbb{N}$, $k\{f(x + \frac{1}{k}) - f(x)\}$ is a measurable function of x.

Exercise 4.7.1 Let f be a monotone increasing function on a finite closed interval $[a, b]$. Show that $f(x) = f(a) + \int_a^x f' d\lambda$ for all $x \in [a, b]$ if and only if the Lebesgue–Stieltjes measure μ_f generated by f is absolutely continuous.

In the remaining part of this section, functions are finite-valued and defined on a finite closed interval $[a, b]$; and for a function f and an interval I in $[a, b]$ with endpoints $c < d$, the difference $f(d) - f(c)$ will be denoted by $f(I)$.

A monotone increasing function f is said to be **absolutely continuous** if the Lebesgue–Stieltjes measure μ_f generated by f is absolutely continuous. Hence, by Exercise 4.7.1, a monotone increasing function f is absolutely continuous if and only if

$$f(x) = f(a) + \int_a^x f' d\lambda$$

holds for all $x \in [a, b]$. We shall characterize absolute continuity of a monotone increasing function by a property which can be adopted to define absolute continuity for general functions.

Lemma 4.7.2 *For a monotone increasing function f, the following two statements are equivalent:*

(I) *f is absolutely continuous.*

(II) *Given any $\varepsilon > 0$, there is $\delta > 0$ such that if $\{I_j\}$ is a disjoint sequence of intervals open in $[a, b]$ with $\sum_j |I_j| < \delta$, then $\sum_j |f(I_j)| < \varepsilon$.*

Proof For convenience, put $\mu = \mu_f$.

To show the implication (I) $\Rightarrow$ (II), note first that since $\mu(\{x\}) = 0$ for all $x \in [a, b]$, $\mu(\{x\}) = f(x+) - f(x-) = 0$, i.e. f is continuous on $[a, b]$. From Lemma 4.6.7, $\int_a^b \frac{d\mu}{d\lambda} d\lambda = \mu([a, b]) < \infty$, hence $\frac{d\mu}{d\lambda}$ is integrable. Now let $\varepsilon > 0$ be given; by Exercise 2.5.9 (iii) there is $\delta > 0$ such that if A is a measurable set in $[a, b]$ with $\lambda(A) < \delta$, then $\int_A \frac{d\mu}{d\lambda} d\lambda < \varepsilon$; if $\{I_j\}$ is a disjoint sequence of intervals open in $[a, b]$ with $\sum_j |I_j| < \delta$, then $\lambda(\bigcup_j I_j) < \delta$ and $\sum_j |f(I_j)| = \sum_j f(I_j) = \int_{\bigcup_j I_j} \frac{d\mu}{d\lambda} d\lambda < \varepsilon$. Thus (II) holds.

Suppose now that (II) holds. We will show that if N is a null set in $[a, b]$, $\mu(N) = 0$. Given $\varepsilon > 0$, choose $\delta > 0$ according to (II). There is a set G open in $[a, b]$ such that $G \supset N$ and $\lambda(G) < \delta$. But, since $G = \bigcup_j I_j$, where $\{I_j\}$ is a disjoint sequence of intervals open in $[a, b]$, $\sum_j |I_j| = \lambda(G) < \delta$, and consequently,

$$\mu(N) \leq \mu(G) = \sum_j \mu(I_j) = \sum_j f(I_j) < \varepsilon, \tag{4.15}$$

by (II), where the obvious fact that if (II) holds, f is continuous on $[a, b]$ and $\mu(I_j) = f(I_j)$, has been used. Since (4.15) holds for arbitrary $\varepsilon > 0$, $\mu(N) = 0$. ∎

Exercise 4.7.2 Show that a monotone increasing and absolutely continuous function maps null sets to null sets.

We take Lemma 4.7.2 as a hint for defining absolute continuity for general functions. A function f is said to be **absolutely continuous** if condition (II) in Lemma 4.7.2 holds for f. Condition (II) in Lemma 4.7.2 will be referred to as **condition (AC)**, and an absolutely continuous function is usually simply called an **AC function**. Immediately following, if $\mathcal{P} : x_0 = c < x_1 < \cdots < x_l = d$ is a partition of $[c, d]$, the intervals (x_{j-1}, x_j), $j = 1, \ldots, l$ are called the intervals of $\mathcal{P}$; and if f is a function defined on $[c, d]$, $\sum_{j=1}^{l} |f(x_j) - f(x_{j-1})|$ will be denoted by $|f(\mathcal{P})|$.

Lemma 4.7.3 *An AC function f is a BV function.*

Proof Since f satisfies condition (AC), there is $\delta > 0$ such that if $\{I_j\}$ is a disjoint sequence of intervals open in $[a, b]$ with $\sum_j |I_j| < \delta$, then $\sum_j |f(I_j)| < 1$. Divide $[a, b]$ into m nonoverlapping closed intervals of equal length $< \delta$, and consider one of these subintervals, say J. Let $\mathcal{P}$ be any partition of J, then $|f(\mathcal{P})| < 1$, because the intervals of $\mathcal{P}$ are in J and the sum of their lengths is smaller than δ. Since $\mathcal{P}$ is an arbitrary partition of J, the total variation of f over J is less than or equal to 1. Hence, $V_a^b(f) \leq m$. ∎

Recall that if f is a BV function, the functions f_P, f_N, and f_V are defined by

$$f_P(x) = P_a^x(f); \quad f_N(x) = N_a^x(f); \quad f_V(x) = V_a^x(f)$$

for $x \in [a, b]$.

Lemma 4.7.4 *If f is a BV function, then the following three statements are equivalent:*

(I) *f is an AC function.*

(II) *f_V is an AC function.*

(III) *Both f_P and f_N are AC functions.*

Proof The implication of (II) $\Rightarrow$ (I) and the equivalence of (II) $\Leftrightarrow$ (III) are obvious. It remains to show the implication of (I) $\Rightarrow$ (II).

Suppose (I) holds. For $\varepsilon > 0$ given, choose $\delta > 0$ according to condition (AC). We are going to show that if $\{I_j\}$ is a disjoint sequence of intervals open in $[a, b]$ with $\sum_j |I_j| < \delta$, then $\sum_j f_V(I_j) \leq \varepsilon$. For each j, let $\mathcal{P}_j$ be an arbitrary partition of I_j, and let $\{I_k^{(j)}\}_k$ be the finite family of intervals of $\mathcal{P}_j$, then $\bigcup_j \{I_k^{(j)}\}_k$ is a sequence of disjoint intervals open in $[a, b]$ and $\sum_j \sum_k |I_k^{(j)}| = \sum_j |I_j| < \delta$. From the choice of δ, $\sum_j \sum_k |f(I_k^{(j)})| = \sum_j |f(\mathcal{P}_j)| < \varepsilon$; consequently, $\sum_j f_V(I_j) \leq \varepsilon$, by taking the supremum of $\sum_j |f(\mathcal{P}_j)|$ first for all partitions $\mathcal{P}_1$ of I_1, and then for all partitions $\mathcal{P}_2$ of I_2, and so on. Thus, f_V satisfies condition (AC) and is therefore an AC function. ∎

A function f is called an **indefinite integral** if there is an integrable function g such that

$$f(x) = c + \int_a^x g d\lambda \tag{4.16}$$

for some constant c and all $x \in [a, b]$. More precisely, if (4.16) holds, f is called an indefinite integral of g.

Exercise 4.7.3 Show that if f is an indefinite integral of g, then $f' = g$ a.e.

Theorem 4.7.2 *A function f is an AC function if and only if it is an indefinite integral.*

Proof It is obvious that an indefinite integral is an AC function. Suppose now that f is an AC function. Both f_P and f_N are AC functions, by Lemma 4.7.4, hence,

$$f_P(x) = \int_a^x f_P' d\lambda; \quad f_N(x) = \int_a^x f_N' d\lambda$$

for all $x \in [a, b]$, by Exercise 4.7.1. Then,

$$f(x) = f(a) + f_P(x) - f_N(x) = f(a) + \int_a^x (f_P' - f_N') d\lambda$$

for all $x \in [a, b]$. This shows that f is an indefinite integral ($f_P' - f_N'$ is integrable because both f_P' and f_N' are integrable). ■

Exercise 4.7.4 Show that a function f is AC if and only if f' exists a.e., f' is integrable, and $f(x) = f(a) + \int_a^x f' d\lambda$ for all $x \in [a, b]$.

Corollary 4.7.1 *If f is an AC function, then $f_N' = 0$ a.e. on $\{f_P' > 0\}$ and $f_P' = 0$ a.e. on $\{f_N' > 0\}$; in other words, $(f')^+ = f_P'$ a.e. and $(f')^- = f_N'$ a.e.*

Proof Since $f' = f_P' - f_N'$ a.e., by Example 4.4.1, $V_a^b(f) = \int_a^b |f_P' - f_N'| d\lambda$; on the other hand, $V_a^b(f) = P_a^b(f) + N_a^b(f) = \int_a^b f_P' d\lambda + \int_a^b f_N' d\lambda$, since both f_P and f_N are AC, by Lemma 4.7.4. Now, $f_P' + f_N' \geq |f_P' - f_N'|$ and $\int_a^b \{f_P' + f_N' - |f_P' - f_N'|\} d\lambda = 0$ imply that $f_P' + f_N' = |f_P' - f_N'|$ a.e. From the last equality, the conclusion of the corollary follows directly. ■

Exercise 4.7.5 Suppose that f is a BV function.

(i) Show that if $V_a^b(f) = \int_a^b |f'| d\lambda$, then

$$f_V(x) = \int_a^x |f'| d\lambda; \quad f_P(x) = \int_a^x (f')^+ d\lambda; \quad f_N(x) = \int_a^x (f')^- d\lambda$$

for all $x \in [a, b]$.

(ii) Show that a BV function f is AC if and only if $V_a^b(f) = \int_a^b |f'| d\lambda$.

Exercise 4.7.6 A monotone increasing function f is said to be singular if $f' = 0$ a.e. Show that every monotone increasing function is a sum of an AC function and a singular function.

Exercise 4.7.7 Let $\{f_n\}$ be a sequence of AC functions on $[a, b]$ such that $\lim_{n\to\infty} f_n(a)$ exists and is finite, and $\{f_n'\}$ converges in $L^1[a, b]$. Show that $\{f_n\}$ converges uniformly on $[a, b]$ to an AC function.

Example 4.7.1 (Cantor's ternary function) Let $I_0 = [0, 1]$ and let $J_0 = (\frac{1}{3}, \frac{2}{3})$ be the middle third open interval of I_0. Then $I_0 \backslash J_0 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$, and call $[0, \frac{1}{3}]$ and $[\frac{2}{3}, 1]$ I_{11}, I_{12} respectively. The open middle thirds of I_{11} and I_{12} are denoted J_{11}, J_{12} respectively. Continue in this fashion; on the kth step we obtain 2^k open intervals $J_{k,1}, \ldots, J_{k,2^k}$ ordered from left to right, each of length $(\frac{1}{3})^{k+1}$. Put $G = \bigcup_{k=0}^{\infty} \bigcup_{j=1}^{2^k} J_{kj}$, then $\lambda(G) = \sum_{k=0}^{\infty} 2^k (\frac{1}{3})^{k+1} = 1$. The set $P := I_0 \backslash G$ is the intersection of a decreasing sequence of nonempty compact sets, and is therefore a nonempty compact set, called **Cantor's ternary set**. P is small in the sense that $\lambda(P) = 0$; but we shall see that P is large in the sense that cardinality of P is the same as that of $I_0 = [0, 1]$. A function f will now be defined on $[0, 1]$ as follows. For $x \in [0, 1]$, express x in ternary expansion

$$x = \sum_{j=1}^{\infty} \frac{\varepsilon_j}{3^j}, \quad \varepsilon_j \in \{0, 1, 2\},$$

and let $\zeta_j = \frac{1}{2}\varepsilon_j$ for all j. The function f is defined by

$$f(x) = \sum_{j=1}^{n-1} \frac{\zeta_j}{2^j} + \frac{1}{2^n}$$

if $\varepsilon_j \in \{0, 2\}$ for $j = 1, \ldots, n-1$, and $\varepsilon_n = 1$ for some n; otherwise, let $f(x) = \sum_{j=1}^{\infty} \frac{\zeta_j}{2^j}$. Function f is well defined, since the only situation where x has two ternary expansions that might lead to different values of $f(x)$ is when the sequence $\{\varepsilon_j\}$ of one of the expansions is of the form: for some n, $\varepsilon_1, \ldots, \varepsilon_{n-1}$ are in $\{0, 2\}$, $\varepsilon_n = 1$, and either $\varepsilon_j = 0$ for $j \geq n+1$ or $\varepsilon_j = 2$ for $j \geq n+1$; in the first case x can also be expressed as $x = \sum_{j=1}^{n-1} \frac{\varepsilon_j}{3^j} + \frac{0}{3^n} + \sum_{j\geq n+1}^{\infty} \frac{2}{3^j}$, and in either expansion

$$f(x) = \sum_{j=1}^{n-1} \frac{\zeta_j}{2^j} + \frac{1}{2^n},$$

while in the second case x can also be expanded as

$$x = \sum_{j=1}^{n-1} \frac{\varepsilon_j}{3^j} + \frac{2}{3^n} + \sum_{j\geq n+1} \frac{0}{3^j},$$

and $f(x)$ also has the value $\sum_{j=1}^{n-1} \frac{\zeta_j}{2^j} + \frac{1}{2^n}$. The function so defined is called **Cantor's ternary function**.

Exercise 4.7.8 Let f be the Cantor's ternary function.

(i) Show that f is a monotone increasing and continuous function with $f(0) = 0$ and $f(1) = 1$.

(ii) Show that each open interval J_{kj}, $k = 0, 1, 2, \ldots; j = 1, \ldots, 2^k$, defined above is of the form $\left(\sum_{j=1}^{n-1} \frac{\varepsilon_j}{3^j} + \frac{1}{3^n}, \sum_{j=1}^{n-1} \frac{\varepsilon_j}{3^j} + \frac{2}{3^n} \right)$ for some n, where $\varepsilon_1, \ldots, \varepsilon_{n-1}$ are in $\{0, 2\}$. Also show that f is constant on each such interval and find the value.

(iii) Show that if x and y in $[0, 1]$ satisfy $|x - y| \leq \frac{1}{3^n}$, then $|f(x) - f(y)| \leq \frac{1}{2^n}$.

(iv) Show that $\int_P f d\mu_f = \frac{1}{2}$.

Exercise 4.7.9 Let P be the Cantor's ternary set defined previously.

(i) Show that $x \in P$ if and only if x has a ternary expansion $x = \sum_{j=1}^{\infty} \frac{\varepsilon_j}{3^j}$, where each $\varepsilon_j \in \{0, 2\}$.

(ii) Show that Cantor's ternary function maps P onto $[0, 1]$.

(iii) A number x in $[0, 1]$ is called a ternary rational number if $x = \frac{m}{3^n}$, where m and n are nonnegative integers with $0 \leq m \leq 3^n$. Let P_0 be the set obtained by removing all those ternary rational numbers in $(0, 1)$ from P. Show that the Cantor's ternary function is 1-1 on P_0.

(iv) Show that the cardinality of P is the same as that of $[0, 1]$.

Example 4.7.1 (Continued) The Cantor's ternary function f is constant on each open interval J_{kj} and hence $f' = 0$ a.e. on $[0, 1]$. Cantor's ternary function is the most well-known singular function. Observe that $V_a^b(f) = 1$, but $\int_a^b |f'| d\lambda = 0$; hence f is not an AC function. The Cantor's ternary set P is perfect i.e. P is the set of all of its own limit points. Thus P is a perfect compact null set with cardinality that of $\mathbb{R}$.

Example 4.7.2 We now use Cantor's ternary function f on $[0, 1]$ to exhibit the fact that a measurable function of a continuous function may not be measurable.

Define a function g on $[0, 1]$ by $g(x) = f(x) + x$, where f is Cantor's ternary function. Evidently, g is strictly increasing on $[0, 1]$ and maps $[0, 1]$ continuously onto $[0, 2]$. The complement G of Cantor's ternary set P in $[0, 1]$ is an open set which is mapped by g onto an open set in $[0, 2]$ of measure 1 (note that each interval component of G is mapped by f to a point, and is hence mapped by g onto an interval of the same length); as a result, g maps the Cantor's ternary set P onto a compact set K of measure 1. By Proposition 3.11.2, K contains a nonmeasurable set W. Since $g^{-1}W \subset P$ and $\lambda(P) = 0$, $g^{-1}W$ is a null set and is therefore measurable. Put $A = g^{-1}W$ and let $h = I_A$; h is measurable. Because g is a continuous and injective map from the compact set $[0, 1]$ onto $[0, 2]$, g^{-1} is a continuous function from $[0, 2]$ onto $[0, 1]$, by Proposition 1.7.3. Now $h \circ g^{-1}$ is not measurable, because $\{h \circ g^{-1} > 0\} = W$ is nonmeasurable. Thus, a measurable function of a continuous function could be nonmeasurable.

For a right-continuous BV function g on $[a, b]$, let μ_g^+ and μ_g^- be the Lebesgue–Stieltjes measures generated by g_P and g_N respectively, and let $\mu_g = \mu_g^+ - \mu_g^-$. Note that both g_P and g_N are right-continuous, by Theorem 4.4.2. If f is both μ_g^+- and μ_g^--measurable on $[a, b]$ and is integrable w.r.t. μ_g^+ and μ_g^-, we define

$$\int_a^b f d\mu_g = \int_a^b f d\mu_g^+ - \int_a^b f d\mu_g^-.$$

The measure $|\mu_g| := \mu_g^+ + \mu_g^-$ is called the **total variational measure** generated by g, while μ_g^+ and μ_g^- are called respectively the **positive variational measure** and the **negative variational measure** generated by g. If f is a bounded function on $[a, b]$ which is continuous $|\mu_g|$-a.e., then

$$\int_a^b f dg := \int_a^b f dg_P - \int_a^b f dg_N$$

exists and is finite, by Theorem 4.5.2.

Exercise 4.7.10 Suppose that g is an AC function. Show that a Riemann integrable function f is continuous $|\mu_g|$-a.e. Then conclude that $\int_a^b f dg$ is defined and $\int_a^b f dg = \int_a^b f g' d\lambda$. (Hint: cf. Example 4.5.2.)

Theorem 4.7.3 (Integration by parts) *Let f, g be AC functions on $[a, b]$, then*

$$\int_a^b f g' d\lambda = f(b)g(b) - f(a)g(a) - \int_a^b g f' d\lambda.$$

Proof We may assume that both f and g are monotone increasing, then by Theorem 4.5.3,

$$\int_a^b f dg = f(b)g(b) - f(a)g(a) - \int_a^b g df.$$

But by Example 4.5.2,

$$\int_a^b f dg = \int_a^b f g' d\lambda; \quad \int_a^b g df = \int_a^b g f' d\lambda,$$

hence,

$$\int_a^b f g' d\lambda = f(b)g(b) - f(a)g(a) - \int_a^b g f' d\lambda. \qquad \blacksquare$$

Exercise 4.7.11 Let f and g be AC functions. Show that the product fg is AC, and (using integration by parts)

$$\int_c^d (fg)' d\lambda = \int_c^d f' g d\lambda + \int_c^d fg' d\lambda$$

for all $a \leq c < d \leq b$, and conclude that $(fg)' = f'g + fg'$ a.e.

Exercise 4.7.12 Let f be an integrable function on $[a, b]$ with the property that

$$\int_a^b fg' d\lambda = 0$$

for all AC functions g such that $g(a) = g(b) = 0$. Show that f = constant a.e. (Hint: put $c = \int_a^b f d\lambda$, and let

$$g(x) = \int_a^x \left(f - \frac{c}{b-a}\right) d\lambda$$

for $x \in [a, b]$. Observe that $g(a) = g(b) = 0$ and evaluate $\int_a^b (f - \frac{c}{b-a})^2 d\lambda$.)

Exercise 4.7.13 Let f and g be integrable functions on $[a, b]$ and suppose that

$$\int_a^b fh' d\lambda = -\int_a^b gh d\lambda$$

for all AC functions h with $h(a) = h(b) = 0$. Show that f is equivalent to an AC function $\hat{f}$ and $\hat{f}' = g$ a.e.

Theorem 4.7.4 (Change of variable) *Suppose that g is a monotone increasing AC function on $[a, b]$. Put $c = g(a)$ and $d = g(b)$. Then for any nonnegative measurable function f on $[c, d]$, the function $(f \circ g)g'$ is measurable and*

$$\int_c^d f d\lambda = \int_a^b (f \circ g) g' d\lambda.$$

Proof From $|I| = \mu_g(g^{-1}I)$, for any interval I open in $[c, d]$, it follows that $\lambda(G) = \mu_g(g^{-1}G)$ for any set G open in $[c, d]$, and hence for any Borel set B in $[c, d]$ we have (cf. Exercise 4.3.4 and recall that μ_g is absolutely continuous)

$$\lambda(B) = \mu_g(g^{-1}B) = \int_{g^{-1}B} \frac{d\mu_g}{d\lambda} d\lambda = \int_a^b I_{g^{-1}B} g' d\lambda = \int_a^b (I_B \circ g) g' d\lambda,$$

or

$$\lambda(B) = \int_H (I_B \circ g) g' d\lambda,$$

where $H = \{g' > 0\}$. Note that for a Borel set B in $[c, d]$, $I_B \circ g$ is a Borel measurable function and $(I_B \circ g)g'$ is measurable; but in general $I_A \circ g$ may not be measurable for measurable set $A \subset [c, d]$; however, we claim that $(I_A \circ g)g'$ is measurable and

$$\lambda(A) = \int_a^b (I_A \circ g)g' d\lambda.$$

To see this, first consider the case where A is a null set in $[c, d]$. Choose a Borel set B in $[c, d]$ such that $B \supset A$ and $\lambda(B) = \lambda(A) = 0$, then

$$\lambda(B) = \int_H (I_B \circ g)g' d\lambda = 0,$$

which implies that $I_B \circ g = 0$ a.e. on H and, a fortiori, $I_A \circ g = 0$ a.e. on H. Therefore $I_A \circ g$ is measurable on H; as a consequence, $(I_A \circ g)g' = 0$ a.e. on $[a, b]$ and is therefore measurable. Now, let A be any measurable set in $[c, d]$ and choose a Borel set B in $[c, d]$ such that $B \supset A$ and $\lambda(B) = \lambda(A)$; then $S := B \backslash A$ is a null set and $(I_S \circ g)g' = 0$ a.e. as we have just proved. But $(I_B \circ g)g' = (I_A \circ g + I_S \circ g)g' = (I_A \circ g)g'$ a.e. on $[a, b]$, hence $(I_A \circ g)g'$ is measurable and

$$\lambda(A) = \lambda(B) = \int_a^b (I_B \circ g)g' d\lambda = \int_a^b (I_A \circ g)g' d\lambda.$$

If $f \geq 0$ is measurable, $f = \sum_{j=1}^{\infty} \frac{1}{j} I_{A_j}$, where each A_j is a measurable set in $[c, d]$, by Theorem 2.2.1. Then,

$$\begin{aligned}
\int_c^d f d\lambda &= \int_c^d \sum_{j=1}^{\infty} \frac{1}{j} I_{A_j} d\lambda = \sum_{j=1}^{\infty} \frac{1}{j} \lambda(A_j) = \sum_{j=1}^{\infty} \frac{1}{j} \int_a^b (I_{A_j} \circ g)g' d\lambda \\
&= \int_a^b \lim_{l \to \infty} \sum_{j=1}^{l} \frac{1}{j} (I_{A_j} \circ g)g' d\lambda = \int_a^b \lim_{l \to \infty} \left\{ \left(\sum_{j=1}^{l} \frac{1}{j} I_{A_j} \right) \circ g \right\} g' d\lambda \\
&= \int_a^b (f \circ g)g' d\lambda,
\end{aligned}$$

where $(f \circ g)g' = \lim_{l \to \infty} \sum_{j=1}^{l} \frac{1}{j} (I_{A_j} \circ g)g'$ is measurable because it is the limit of measurable functions $\sum_{j=1}^{l} \frac{1}{j} (I_{A_j} \circ g)g'$. ■

Remark The change of variable formula in Theorem 4.7.4 is familiar in integral calculus. Here, it is shown under much relaxed conditions on f and g. Note that one of the delicacies in the proof is the measurability of $(f \circ g)g'$, although $f \circ g$ may not be measurable, as we see in Example 4.7.2.

4.8 Product measures and Fubini theorem

We digress in this section from the main theme of the chapter, to the construction and properties of product measures, before going to further studies of functions of several real variables. Consider measure spaces $(\Omega_i, \Sigma_i, \mu_i)$, $i = 1, 2$, and let $R = \{A_1 \times A_2 : A_i \in \Sigma_i, i = 1, 2\}$. R is a π-system. Sets in R are called **measurable rectangles**. The σ-algebra $\sigma(R)$ on $\Omega_1 \times \Omega_2$ generated by R is denoted by $\Sigma_1 \otimes \Sigma_2$. For $E \subset \Omega_1 \times \Omega_2$ and $(w_1, w_2) \in \Omega_1 \times \Omega_2$, we define sets E_{w_1} and E^{w_2} by

$$E_{w_1} = \{y \in \Omega_2 : (w_1, y) \in E\}; \quad E^{w_2} = \{x \in \Omega_1 : (x, w_2) \in E\}.$$

E_{w_1} and E^{w_2} are called respectively the w_1-**section** and w_2-**section** of E.

The lemma that follows is easily verified.

Lemma 4.8.1 *Let Σ be the family of all $E \subset \Omega_1 \times \Omega_2$ such that $E_{w_1} \in \Sigma_2, E^{w_2} \in \Sigma_1$ for all $(w_1, w_2) \in \Omega_1 \times \Omega_2$, then Σ is a σ-algebra containing R.*

Corollary 4.8.1 $\Sigma \supset \Sigma_1 \otimes \Sigma_2$.

Corollary 4.8.2 *If f is $\Sigma_1 \otimes \Sigma_2$-measurable, then for $(w_1, w_2) \in \Omega_1 \times \Omega_2$, $x \mapsto f(x, w_2)$ and $y \mapsto f(w_1, y)$ are respectively Σ_1- and Σ_2-measurable.*

Proof Since $I_E(x, w_2) = I_{E^{w_2}}(x)$ and $I_E(w_1, y) = I_{E_{w_1}}(y)$ for $E \subset \Omega_1 \times \Omega_2$, it follows from Lemma 4.8.1 and Corollary 4.8.1 that the corollary holds if f is the indicator function of a set in $\Sigma_1 \otimes \Sigma_2$. Then the corollary holds for $\Sigma_1 \otimes \Sigma_2$-measurable simple functions. For general nonnegative $\Sigma_1 \otimes \Sigma_2$-measurable functions, the corollary follows by Theorem 2.2.1; this is sufficient to conclude that the corollary holds. ■

Lemma 4.8.2 *Suppose that both $(\Omega_1, \Sigma_1, \mu_1)$ and $(\Omega_2, \Sigma_2, \mu_2)$ are σ-finite and $E \in \Sigma_1 \otimes \Sigma_2$, then $w_1 \mapsto \mu_2(E_{w_1})$ is Σ_1-measurable and $w_2 \mapsto \mu_1(E^{w_2})$ is Σ_2-measurable, and*

$$\int_{\Omega_1} \mu_2(E_{w_1}) d\mu_1(w_1) = \int_{\Omega_2} \mu_1(E^{w_2}) d\mu_2(w_2).$$

Proof Ω_1 and Ω_2 can be expressed as

$$\Omega_1 = \bigcup_{n=1}^{\infty} \Omega_n^{(1)}, \quad \Omega_2 = \bigcup_{n=1}^{\infty} \Omega_n^{(2)},$$

where $\{\Omega_n^{(1)}\} \subset \Sigma_1$, $\{\Omega_n^{(2)}\} \subset \Sigma_2$ are both disjoint and $\mu_i(\Omega_n^{(i)}) < \infty$ for $i = 1, 2$ and $n = 1, 2, \ldots$. Consider the family $\mathcal{M}$ of all those $E \in \Sigma_1 \otimes \Sigma_2$, such that the conclusions of the lemma hold if E is replaced by $E \cap (\Omega_n^{(1)} \times \Omega_m^{(2)})$ for all n and m. It is simply routine to verify that $\mathcal{M}$ is a λ-system. But it is to be noted that the only place where $E \cap (\Omega_n^{(1)} \times \Omega_m^{(2)})$ requires considering is when one verifies that if E is in $\mathcal{M}$ then E^c is in $\mathcal{M}$. Since $E \in \mathcal{M}$ is easily seen to satisfy the conclusions of the lemma, and since $\mathcal{M} \supset R$, the lemma follows from the $(\pi$-$\lambda)$ theorem. ■

Now, for $E \in \Sigma_1 \otimes \Sigma_2$, define

$$\mu_1 \times \mu_2(E) = \int_{\Omega_1} \mu_2(E_{w_1}) d\mu_1(w_1) = \int_{\Omega_2} \mu_1(E^{w_2}) d\mu_2(w_2).$$

Then $\mu_1 \times \mu_2$ is a measure on $\Sigma_1 \otimes \Sigma_2$ and $(\Omega_1 \times \Omega_2, \Sigma_1 \otimes \Sigma_2, \mu_1 \times \mu_2)$ is a measure space, called the **product space** of $(\Omega_1, \Sigma_1, \mu_1)$ and $(\Omega_2, \Sigma_2, \mu_2)$. The measure $\mu_1 \times \mu_2$ is called the **product measure** of μ_1 and μ_2. One notes that $\mu_1 \times \mu_2(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$ if $A_1 \times A_2 \in R$.

Proposition 4.8.1 *Suppose that both $(\Omega_1, \Sigma_1, \mu_1)$ and $(\Omega_2, \Sigma_2, \mu_2)$ are σ-finite, then $\mu_1 \times \mu_2$ is the unique measure on $\Sigma_1 \otimes \Sigma_2$ such that $\mu_1 \times \mu_2(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$ for all $A_1 \in \Sigma_1$ and $A_2 \in \Sigma_2$.*

Proof Let disjoint sequences $\{\Omega_n^{(1)}\} \subset \Sigma_1$ and $\{\Omega_m^{(2)}\} \subset \Sigma_2$ be as in the proof of Lemma 4.8.2, and suppose that μ is a measure on $\Sigma_1 \otimes \Sigma_2$ such that $\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$ for all $A_1 \in \Sigma_1$ and $A_2 \in \Sigma_2$. Consider the family $\mathcal{F}$ of all $E \in \Sigma_1 \otimes \Sigma_2$, such that

$$\mu(E \cap [\Omega_n^{(1)} \times \Omega_m^{(2)}]) = \mu_1 \times \mu_2(E \cap [\Omega_n^{(1)} \times \Omega_m^{(2)}])$$

for all n and m. Then $\mathcal{F}$ is a λ-system containing all measurable rectangles. Since the family R of all measurable rectangles is a π-system, it follows from the $(\pi$-$\lambda)$ theorem that $\mathcal{F} = \Sigma_1 \otimes \Sigma_2$ and thus $\mu = \mu_1 \times \mu_2$. ■

Theorem 4.8.1 (Simple version of Fubini theorem)

(i) (Tonelli) *If f is $\Sigma_1 \otimes \Sigma_2$-measurable and $f \geq 0$, then $x \mapsto \int_{\Omega_2} f(x, w_2) d\mu_2(w_2)$ is Σ_1-measurable, $y \mapsto \int_{\Omega_1} f(w_1, y) d\mu_1(w_1)$ is Σ_2-measurable, and*

$$\begin{aligned} \int_{\Omega_1 \times \Omega_2} f d\mu_1 \times \mu_2 &= \int_{\Omega_1} \left[\int_{\Omega_2} f(w_1, w_2) d\mu_2(w_2) \right] d\mu_1(w_1) \\ &= \int_{\Omega_2} \left[\int_{\Omega_1} f(w_1, w_2) d\mu_1(w_1) \right] d\mu_2(w_2). \end{aligned}$$

(ii) *If f is $\mu_1 \times \mu_2$-integrable, then conclusions in* (i) *also hold for f.*

Proof Since (ii) is an obvious consequence of (i), it is sufficient to prove (i). If $E \in \Sigma_1 \otimes \Sigma_2$ and $f = I_E$, then (i) follows from Lemma 4.8.2 and hence the lemma holds for nonnegative simple functions. If f is a nonnegative $\Sigma_1 \otimes \Sigma_2$-measurable function, by Theorem 2.2.1,

$$f = \sum_{k=1}^{\infty} \frac{1}{k} I_{A_k} = \lim_{l \to \infty} \sum_{k=1}^{l} \frac{1}{k} I_{A_k},$$

where each $A_k \in \Sigma_1 \otimes \Sigma_2$, then (i) follows from the monotone convergence theorem. ■

In general, it is not true that the product space of two σ-finite complete measure spaces is complete. For example, consider $(\mathbb{R}^2, \mathcal{L} \otimes \mathcal{L}, \lambda \times \lambda)$, where $\mathcal{L}$ is the σ-algebra of all Lebesgue measurable sets in $\mathbb{R}$ and λ the Lebesgue measure on $\mathbb{R}$. As we have shown in Section 3.11 there is a nonmeasurable set $S \subset \mathbb{R}$. Choose any nonempty null set N in $\mathbb{R}$, and consider the set $N \times S$ in $\mathbb{R}^2$. For $w \in N$, $(N \times S)_w = S$ is not in $\mathcal{L}$; hence $N \times S$ is not in $\mathcal{L} \otimes \mathcal{L}$. But $N \times S \subset N \times \mathbb{R}$ and $\lambda \times \lambda(N \times \mathbb{R}) = \lambda(N)\lambda(\mathbb{R}) = 0$, thus $N \times S$ is a $\lambda \times \lambda$-null set which is not in $\mathcal{L} \otimes \mathcal{L}$. $(\mathbb{R}^2, \mathcal{L} \otimes \mathcal{L}, \lambda \times \lambda)$ is therefore not complete and cannot be $(\mathbb{R}^2, \mathcal{L}^2, \lambda^2)$.

Exercise 4.8.1 Show that $(\mathbb{R}^{k+l}, \mathcal{L}^{k+l}, \lambda^{k+l})$ is the completion of the measure space $(\mathbb{R}^{k+l}, \mathcal{L}^k \otimes \mathcal{L}^l, \lambda^k \times \lambda^l)$ for k, l in $\mathbb{N}$. (Hint: verify first that $\mathcal{B}(\mathbb{R}^{k+l}) \subset \mathcal{L}^k \otimes \mathcal{L}^l$ and $\lambda^{k+l}(B) = \lambda^k \times \lambda^l(B)$ for $B \in \mathcal{B}(\mathbb{R}^{k+l})$.)

Suppose now that both $(\Omega_1, \Sigma_1, \mu_1)$ and $(\Omega_2, \Sigma_2, \mu_2)$ are σ-finite complete measure spaces; then corresponding to Theorem 4.8.1, the following theorem holds.

Theorem 4.8.2 (Fubini) *Let* $(\Omega_1 \times \Omega_2, \overline{\Sigma_1 \otimes \Sigma_2}, \overline{\mu_1 \times \mu_2})$ *be the completion of* $(\Omega_1 \times \Omega_2, \Sigma_1 \otimes \Sigma_2, \mu_1 \times \mu_2)$.

(i) (Tonelli) *If f is nonnegative $\overline{\Sigma_1 \otimes \Sigma_2}$-measurable, then for μ_1-a.e. w_1 in Ω_1 and μ_2-a.e. w_2 in Ω_2,*

$$v \mapsto f(w_1, v) \text{ is } \Sigma_2\text{-measurable};$$
$$u \mapsto f(u, w_2) \text{ is } \Sigma_1\text{-measurable}.$$

Furthermore,

$$w_1 \mapsto \int_{\Omega_2} f(w_1, w_2)d\mu_2(w_2) \text{ is } \Sigma_1\text{-measurable};$$
$$w_2 \mapsto \int_{\Omega_1} f(w_1, w_2)d\mu_1(w_1) \text{ is } \Sigma_2\text{-measurable},$$

and

$$\int_{\Omega_1 \times \Omega_2} f d\overline{\mu_1 \times \mu_2} = \int_{\Omega_1} \left[\int_{\Omega_2} f(w_1, w_2)d\mu_2(w_2) \right] d\mu_1(w_1)$$
$$= \int_{\Omega_2} \left[\int_{\Omega_1} f(w_1, w_2)d\mu_1(w_1) \right] d\mu_2(w_2).$$

(ii) *If f is $\overline{\mu_1 \times \mu_2}$-integrable, then the same statements in (i) hold for f.*

Lemma 4.8.3 *Suppose that $E \in \Sigma_1 \otimes \Sigma_2$ and $\mu_1 \times \mu_2(E) = 0$. Then for any subset D of E, the following statements hold:*

(1) *$D_{w_1} \in \Sigma_2$ and $\mu_2(D_{w_1}) = 0$ for μ_1-a.e. w_1 in Ω_1.*

(2) *$D^{w_2} \in \Sigma_1$ and $\mu_1(D^{w_2}) = 0$ for μ_2-a.e. w_2 in Ω_2.*

Proof Since $\mu_1 \times \mu_2(E) = \int_{\Omega_1} \mu_2(E_{w_1}) d\mu_1(w_1) = \int_{\Omega_2} \mu_1(E^{w_2}) d\mu_2(w_2) = 0$, and both $\mu_2(E_{w_1})$ and $\mu_1(E^{w_2})$ are nonnegative, $\mu_2(E_{w_1}) = 0$ for μ_1-a.e. w_1 and $\mu_1(E^{w_2}) = 0$ for μ_2-a.e. w_2. For such w_1 and w_2, D_{w_1} and D^{w_2} are in Σ_2 and Σ_1 respectively, because $D_{w_1} \subset E_{w_1}$, $D^{w_2} \subset E^{w_2}$, and both $(\Omega_1, \Sigma_1, \mu_1)$ and $(\Omega_2, \Sigma_2, \mu_2)$ are complete. Trivially, for such w_1 and w_2, $\mu_2(D_{w_1}) = \mu_1(D^{w_2}) = 0$. ■

Proof of Theorem 4.8.2 Since (ii) follows from (i) easily, it suffices to prove (i).

If $f \geq 0$ is $\overline{\Sigma_1 \otimes \Sigma_2}$-measurable, $f = \sum_{j=1}^{\infty} \frac{1}{j} I_{A_j}$, where each A_j is in $\overline{\Sigma_1 \otimes \Sigma_2}$, as claimed by Theorem 2.2.1. It is therefore sufficient to consider the case $f = I_A$ for $A \in \overline{\Sigma_1 \otimes \Sigma_2}$. There are B and C in $\Sigma_1 \otimes \Sigma_2$ such that $B \subset A \subset C$ with $\mu_1 \times \mu_2(C \backslash B) = 0$. This means that $A = B \cup D$ where $D \subset E := C \backslash B$. From Lemma 4.8.3, for μ_1-a.e. w_1 and μ_2-a.e. w_2, $D_{w_1} \in \Sigma_2$ with $\mu_2(D_{w_1}) = 0$ and $D^{w_2} \in \Sigma_1$ with $\mu_1(D^{w_2}) = 0$; for such w_1 and w_2,

$$v \mapsto I_A(w_1, v) = I_{A_{w_1}}(v) = I_{B_{w_1} \cup D_{w_1}}(v) = I_{B_{w_1}}(v) + I_{D_{w_1}}(v)$$

and

$$u \mapsto I_A(u, w_2) = I_{A^{w_2}}(u) = I_{B^{w_2} \cup D^{w_2}}(u) = I_{B^{w_2}}(u) + I_{D^{w_2}}(u)$$

are respectively Σ_2- and Σ_1-measurable. Furthermore,

$$w_1 \mapsto \int_{\Omega_2} I_A(w_1, v) d\mu_2(v) = \mu_2(B_{w_1}),$$

and

$$w_2 \mapsto \int_{\Omega_1} I_A(u, w_2) d\mu_1(u) = \mu_1(B^{w_2})$$

are respectively Σ_1- and Σ_2-measurable by Lemma 4.8.2, and hence,

$$\int_{\Omega_1} \left[\int_{\Omega_2} I_A(w_1, w_2) d\mu_2(w_2) \right] d\mu_1(w_1) = \int_{\Omega_1} \mu_2(B_{w_1}) d\mu_1(w_1);$$
$$\int_{\Omega_2} \left[\int_{\Omega_1} I_A(w_1, w_2) d\mu_2(w_2) \right] d\mu_2(w_2) = \int_{\Omega_2} \mu_1(B^{w_2}) d\mu_2(w_2).$$

Thus (i) holds for $f = I_A$, because by Lemma 4.8.2,

$$\int_{\Omega_1} \mu_2(B_{w_1}) d\mu_1(w_1) = \int_{\Omega_2} \mu_1(B^{w_2}) d\mu_2(w_2) = \mu_1 \times \mu_2(B),$$

and $\mu_1 \times \mu_2(B) = \overline{\mu_1 \times \mu_2}(A) = \int_{\Omega_1 \times \Omega_2} I_A d\overline{\mu_1 \times \mu_2}$. ■

Example 4.8.1 We use the Fubini theorem to evaluate $\int_{-\infty}^{\infty} e^{-x^2} dx$ (cf. Exercise 3.4.7 and Exercise 3.4.8). First note that since $\int_{-\infty}^{\infty} e^{-x^2} dx < \infty$ as an improper integral, $\int_{\mathbb{R}} e^{-x^2} d\lambda(x) = \int_{-\infty}^{\infty} e^{-x^2} dx$ by Exercise 3.4.7 (i). From the Fubini theorem,

$$\int_{\mathbb{R}^2} e^{-(x^2+y^2)} d\lambda^2(x,y) = \int_{\mathbb{R}} \left[\int_{\mathbb{R}} e^{-x^2} d\lambda(x) \right] e^{-y^2} d\lambda(y) = \left[\int_{\mathbb{R}} e^{-x^2} dx \right]^2.$$

Now,

$$\begin{aligned}\int_{\mathbb{R}^2} e^{-(x^2+y^2)} d\lambda^2(x,y) &= \lim_{L\to\infty} \iint_{x^2+y^2\le L^2} e^{-(x^2+y^2)} dxdy \\ &= \lim_{L\to\infty} \int_0^L \rho \int_0^{2\pi} e^{-\rho^2} d\theta d\rho = \lim_{\rho\to\infty} 2\pi \int_0^L \rho e^{-\rho^2} d\rho \\ &= \lim_{L\to\infty} \pi \int_0^L \frac{d}{d\rho}(-e^{-\rho^2}) d\rho = \pi.\end{aligned}$$

Hence $\int_{-\infty}^{\infty} e^{-x^2} dx = \left[\iint_{\mathbb{R}^2} e^{-(x^2+y^2)} dxdy \right]^{\frac{1}{2}} = \sqrt{\pi}$. By the Fubini theorem, again one finds that $\int_{\mathbb{R}^n} e^{-|x|^2} dx = \pi^{\frac{n}{2}}$.

Exercise 4.8.2

(i) Show that $\int_0^{\infty} \frac{|\sin x|}{x} dx = \infty$.

(ii) Show that $\int_0^{\infty} \frac{\sin x}{x} dx = \lim_{b\to\infty} \int_0^b \frac{\sin x}{x} dx = \frac{\pi}{2}$ by integrating $e^{-xy} \sin x$ over a suitable domain in the first quadrant of $\mathbb{R}^2$.

Exercise 4.8.3 Let (Ω, Σ, μ) be a σ-finite measure space and f a nonnegative Σ-measurable function on Ω. Put $G_f = \{(w,y) \in \Omega \times [0,\infty) : 0 < y < f(w)\}$. Show that $G_f \in \Sigma \otimes \mathcal{B}$ and $\mu \times \lambda(G_f) = \int_\Omega f d\mu$.

Exercise 4.8.4 Let $f(x,y) = \frac{xy}{(x^2+y^2)^2}$ if $(x,y) \neq (0,0)$, and $f(0,0) = 0$. Verify that $\int_{-1}^1 \left(\int_{-1}^1 f(x,y) dx \right) dy = \int_{-1}^1 \left(\int_{-1}^1 f(x,y) dy \right) dx = 0$, and decide whether f is Lebesgue integrable on $[-1,1] \times [-1,1]$ or not.

Exercise 4.8.5 Show that $\int_0^{\infty} (\sum_{j=1}^{\infty} e^{-jx} \sin x) dx = \sum_{j=1}^{\infty} \int_0^{\infty} e^{-jx} \sin x dx$ and use this fact to show that $\int_0^{\infty} \frac{\sin x}{e^x - 1} dx = \sum_{j=1}^{\infty} \frac{1}{1+j^2}$.

Exercise 4.8.6

(i) Show that $\int_0^{\infty} \frac{\tan^{-1} t}{t} dt = \infty$ by considering the double integral

$$\int_0^1 \left(\int_0^{\infty} \frac{1}{1+x^2 t^2} dt \right) dx.$$

(ii) Show that $\int_0^\infty (\frac{\tan^{-1} t}{t})^2 dt = \pi \ln 2$ by integrating the triple integral

$$\int_0^1 \left(\int_0^1 \left(\int_0^\infty \frac{1}{1+x^2t^2} \cdot \frac{1}{1+y^2t^2} dt \right) dx \right) dy.$$

Example 4.8.2 Let $\{f_n\}$ be a sequence of $L(\Omega_1 \times \Omega_2, \Sigma_1 \otimes \Sigma_2, \mu_1 \times \mu_2)$, in which it converges to f. We claim that there is a subsequence $\{f_{n_k}\}$ of $\{f_n\}$ such that for μ_1-a.e. x $\lim_{k\to\infty} \int_{\Omega_2} |f_{n_k}(x,y) - f(x,y)| d\mu_2(y) = 0$. Define F, F_n on Ω_1 by $F(x) = \int_{\Omega_2} f(x,y) d\mu_2(y)$ and $F_n(x) = \int_{\Omega_2} f_n(x,y) d\mu_2(y)$. Note that F, F_n's are measurable on $(\Omega_1, \Sigma_1, \mu_1)$ and $\int_{\Omega_1} |F_n - F| d\mu_1 = \int_{\Omega_1 \times \Omega_2} |f_n - f| d\mu_1 \times \mu_2$ by the Fubini theorem. Consequently, $\lim_{n\to\infty} \int_{\Omega_1} |F_n - F| d\mu_1 = 0$, and, by Exercise 2.7.9, $\{F_n\}$ has a subsequence $\{F_{n_k}\}$ which converges to F a.e. on Ω_1. Now, the Fatou lemma implies that $\int_{\Omega_1} \lim_{k\to\infty} |F_{n_k} - F| d\mu_1 \le \liminf_{k\to\infty} \int_{\Omega_1} |F_{n_k} - F| d\mu_1 = 0$, which means $\lim_{k\to\infty} |F_{n_k} - F| = 0$ μ_1-a.e. on Ω_1, or

$$\lim_{k\to\infty} \int_{\Omega_2} |f_{n_k}(x,y) - f(x,y)| d\mu_2(y) = 0$$

for μ_1-a.e. x in Ω_1, as we claim.

We conclude this section by applying the Fubini theorem to prove a measurability result which we shall need later. For this purpose, define first a map t from $\mathbb{R}^{2n}$ to $\mathbb{R}^n$ by $t(x,y) = x - y$, where x and y are in $\mathbb{R}^n$. If f is a Borel measurable function on $\mathbb{R}^n$, then $f \circ t$ is Borel measurable on $\mathbb{R}^{2n}$, because $\{f \circ t > \alpha\} = t^{-1}\{f > \alpha\}$, which is a Borel set in $\mathbb{R}^{2n}$. Note that for $A \subset \mathbb{R}^n$, the y-section $(t^{-1}A)^y$ of $t^{-1}A$ is $A + y := \{x + y : x \in A\}$.

Lemma 4.8.4 *If A is a null set in $\mathbb{R}^n$, then $t^{-1}A$ is a null set in $\mathbb{R}^{2n}$.*

Proof There is a Borel set $B \supset A$ with $\lambda^n(B) = 0$. Now $t^{-1}B$ is a Borel set in $\mathbb{R}^{2n}$; by the Fubini theorem,

$$\begin{aligned}\lambda^{2n}(t^{-1}B) &= \int_{\mathbb{R}^{2n}} I_{t^{-1}B} d\lambda^{2n} = \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} I_{t^{-1}B}(x,y) d\lambda^n(x) \right) d\lambda^n(y) \\ &= \int_{\mathbb{R}^n} \lambda^n((t^{-1}B)^y) d\lambda^n(y) = \int_{\mathbb{R}^n} \lambda^n(B+y) d\lambda^n(y) \\ &= \int_{\mathbb{R}^n} \lambda^n(B) d\lambda^n = \int_{\mathbb{R}^n} 0 d\lambda^n = 0,\end{aligned}$$

i.e. $t^{-1}B$ is a null set in $\mathbb{R}^{2n}$. But $t^{-1}A \subset t^{-1}B$ implies that $t^{-1}A$ is a null set. ■

Proposition 4.8.2 *If f is a measurable function on $\mathbb{R}^n$, then $f \circ t$ is a measurable function on $\mathbb{R}^{2n}$.*

Proof There is a Borel function g on $\mathbb{R}^n$ such that $f = g + h$, where $h = 0$ a.e. on $\mathbb{R}^n$. Since $f \circ t = g \circ t + h \circ t$ and $g \circ t$ is Borel measurable, $f \circ t$ is measurable if $h \circ t$ is measurable. We claim that $h \circ t = 0$ a.e. on $\mathbb{R}^{2n}$. There is a null set $A \subset \mathbb{R}^n$ such that $h = 0$ on $\mathbb{R}^n \backslash A$. Then $h \circ t = 0$ on $t^{-1}(\mathbb{R}^n \backslash A) = (t^{-1}\mathbb{R}^n) \backslash t^{-1}A = \mathbb{R}^{2n} \backslash t^{-1}A$. But, by Lemma 4.8.4, $t^{-1}A$ is a null set in $\mathbb{R}^{2n}$, hence $h \circ t = 0$ a.e. on $\mathbb{R}^{2n}$. Since $h \circ t = 0$ a.e. on $\mathbb{R}^{2n}$, it is measurable; consequently, $f \circ t$ is measurable. ∎

4.9 Smoothing of functions

Our concern in this section is the smoothing of functions and approximation of functions by smooth ones. The method we shall use is that of the Friederichs mollifier.

We define first some function spaces which will be frequently considered later. Given an open set Ω in $\mathbb{R}^n$ and a positive integer k, we shall denote by $C^k(\Omega)$ the vector space of all functions defined on Ω which have continuous partial derivatives up to order k, and denote by $C^\infty(\Omega)$ the space $\bigcap_k C^k(\Omega)$. The functions considered are either real-valued or complex-valued, as will either be clear from context or explicitly stated. For a function f defined on Ω, recall that the closure in Ω of the set $\{f \neq 0\}$ is called the support of f and is denoted by $\operatorname{supp} f$. If $\operatorname{supp} f$ is a compact set, then f is said to have compact support. The subspace of $C^k(\Omega)$, which consists of all functions in $C^k(\Omega)$ with compact support, is denoted by $C_c^k(\Omega)$; $C_c^\infty(\Omega)$ is similarly defined.

For a measurable subset Ω of $\mathbb{R}^n$, the space $L^p(\Omega, \mathcal{L}^n|\Omega, \lambda^n)$ will be simply denoted by $L^p(\Omega)$, for convenience, and accordingly the space of all those measurable functions which are in $L^p(K)$ for every compact subset K of Ω is denoted by $L^p_{\mathrm{loc}}(\Omega)$. Usually $L^1_{\mathrm{loc}}(\Omega)$ is simply denoted by $L_{\mathrm{loc}}(\Omega)$ and its elements are called locally integrable functions on Ω; correspondingly, functions in $L^p_{\mathrm{loc}}(\Omega)$ are called locally L^p functions on Ω.

Some notations regarding multi-indices are now introduced. By **multi-index**, we mean an ordered n-tuple $\alpha = (\alpha_1, \ldots, \alpha_n)$ of nonnegative integers for some integer $n > 1$ (n will be clearly implied from the context). For a multi-index $\alpha = (\alpha_1, \ldots, \alpha_n)$, the sum $\sum_{j=1}^n \alpha_j$ and the product $\prod_{j=1}^n \alpha_j!$ are denoted respectively by $|\alpha|$ and $\alpha!$; while if $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, x^α will stand for $x_1^{\alpha_1}, \ldots, x_n^{\alpha_n}$. The partial derivative symbol $\frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}$ will be abbreviated to $\frac{\partial^{|\alpha|}}{\partial x^\alpha}$ or ∂_x^α.

We are now ready to define the Friederichs mollifier. Let $\varphi \in C_c^\infty(\mathbb{R}^n)$ with $\int \varphi d\lambda^n = 1$. For definiteness, assume that $\operatorname{supp} \varphi \subset C_1(0)$, the closed ball in $\mathbb{R}^n$ centered at 0 and with radius 1. Such a function φ is called a **mollifying function**. For $\varepsilon > 0$, define $\varphi_\varepsilon(x) = \varepsilon^{-n} \varphi(\frac{x}{\varepsilon})$ for $x \in \mathbb{R}^n$; then $\operatorname{supp} \varphi_\varepsilon \subset C_\varepsilon(0)$ and $\int \varphi_\varepsilon d\lambda^n = 1$, by Example 4.3.1 (ii).

Corresponding to such a function φ and $\varepsilon > 0$, we define a linear transformation J_ε which maps functions f in $L_{\mathrm{loc}}(\Omega)$ to functions defined on $\Omega_\varepsilon = \{x \in \Omega : \operatorname{dist}(x, \Omega^c) > \varepsilon\}$, by

$$J_\varepsilon f(x) = \int_{C_\varepsilon(x)} f(y) \varphi_\varepsilon(x - y) dy, \quad x \in \Omega_\varepsilon.$$

Note that $C_\varepsilon(x) \subset \Omega$ for $x \in \Omega_\varepsilon$, hence f is integrable on $C_\varepsilon(x)$ and $J_\varepsilon f(x)$ is defined; moreover, since $\varphi_\varepsilon(x - y) = 0$ for y outside $C_\varepsilon(x)$, we may consider the defining integral for $J_\varepsilon f(x)$ as over the whole space $\mathbb{R}^n$, thus

$$J_\varepsilon f(x) = \int f(y)\varphi_\varepsilon(x - y)dy.$$

The family $\{J_\varepsilon\}_{\varepsilon>0}$, which depends on φ, is called a **Friederichs mollifier**. We often consider the case $\varphi \geq 0$, but for the moment, we do not impose this restriction.

The most well-known such nonnegative function φ is that defined as follows:

$$\varphi(x) = \begin{cases} Ce^{-\frac{1}{1-|x|^2}}, & \text{if } |x| < 1; \\ 0, & \text{if } |x| \geq 1, \end{cases}$$

where C is chosen so that $\int \varphi d\lambda^n = 1$.

Exercise 4.9.1

(i) Show that $J_\varepsilon f \in C(\Omega_\varepsilon)$.

(ii) More generally, suppose that h is a continuous function on $\mathbb{R}^n$ with $\operatorname{supp} h \subset C_\varepsilon(0)$; show that $\int f(y)h(x - y)dy$ is a continuous function of $x \in \Omega_\varepsilon$.

Exercise 4.9.2 Show that $\int f(y)\varphi_\varepsilon(x - y)dy = \int f(x - y)\varphi_\varepsilon(y)dy$, for $x \in \Omega_\varepsilon$.

Proposition 4.9.1 *If $f \in C(\Omega)$, $J_\varepsilon f(x) \to f(x)$ uniformly on any compact subset of Ω as $\varepsilon \to 0$.*

Proof Let $K \subset \Omega$ be compact. Fix $0 < \varepsilon_0 < \operatorname{dist}(K, \Omega^c)$ and let $F = \{x \in \Omega : \operatorname{dist}(x, K) \leq \varepsilon_0\}$. F is compact. Since f is uniformly continuous on F, for $\sigma > 0$, there is $\delta > 0$ with $\delta \leq \varepsilon_0$, such that $|f(x) - f(y)| \leq \sigma$ if x, y are in F and $|x - y| < \delta$. For $x \in K$, $0 < \varepsilon < \delta$, we have

$$|J_\varepsilon f(x) - f(x)| = \left| \int (f(y) - f(x))\varphi_\varepsilon(x - y)dy \right| \leq \sigma \int |\varphi_\varepsilon| d\lambda^n \leq \sigma M_\varphi,$$

where $M_\varphi = \int |\varphi| d\lambda^n$. ■

Proposition 4.9.2 *For $f \in L_{\text{loc}}(\Omega)$, $J_\varepsilon f \in C^\infty(\Omega_\varepsilon)$.*

Proof For $h \neq 0$, consider the difference quotient for $x \in \Omega_\varepsilon$,

$$\frac{1}{h}\{J_\varepsilon f(x + he_j) - J_\varepsilon f(x)\} = \int f(y)\frac{\varphi_\varepsilon(x + he_j - y) - \varphi_\varepsilon(x - y)}{h}dy,$$

where $e_j = (\delta_{j1}, \ldots, \delta_{jn})$ with δ_{jk} being 1 or zero according to whether or not $k = j$. When h is small, $\operatorname{dist}(x + he_j, \Omega^c) \geq \varepsilon_0 > \varepsilon$, and for all such small enough h, $\varphi_\varepsilon(x + he_j - y) = 0$ for y outside a compact set K in Ω; therefore,

$$\int f(y)\frac{\varphi_\varepsilon(x + he_j - y) - \varphi_\varepsilon(x - y)}{h}dy = \int_K f(y)\frac{\varphi_\varepsilon(x + he_j - y) - \varphi_\varepsilon(x - y)}{h}dy.$$

Now,

$$\left|\frac{\varphi_\varepsilon(x + he_j - y) - \varphi_\varepsilon(x - y)}{h}\right| \leq \max_{z\in\mathbb{R}^n}\left|\frac{\partial\varphi_\varepsilon}{\partial x_j}(z)\right| := M_j,$$

and hence

$$\left|f(y)\frac{\varphi_\varepsilon(x + he_j - y) - \varphi_\varepsilon(x - y)}{h}\right| \leq M_j|f(y)|$$

on K. By LDCT,

$$\frac{\partial}{\partial x_j}J_\varepsilon f(x) = \lim_{h\to 0}\frac{1}{h}\{J_\varepsilon f(x + he_j) - J_\varepsilon f(x)\} = \int f(y)\frac{\partial\varphi_\varepsilon}{\partial x_j}(x - y)dy.$$

So far we have only used the fact that $\varphi_\varepsilon \in C_c^\infty(\mathbb{R}^n)$ with $\operatorname{supp}\varphi_\varepsilon \subset C_\varepsilon(0)$. Hence, we may repeat the argument to obtain

$$\frac{\partial^{|\alpha|}}{\partial x^\alpha}J_\varepsilon f(x) = \int f(y)\frac{\partial^{|\alpha|}}{\partial x^\alpha}\varphi_\varepsilon(x - y)dy.$$

By Exercise 4.9.1 (ii), each $\frac{\partial^{|\alpha|}}{\partial x^\alpha}J_\varepsilon f$ is continuous on Ω_ε. ■

Exercise 4.9.3 If K is a compact set and G is an open set containing K, then there is C^∞ function g with $\operatorname{supp} g \subset G$ and $0 \leq g \leq 1$, such that $g = 1$ on K.

Remark When $f \in L^p(\Omega)$, $1 \leq p \leq \infty$, we may consider f as defined on $\mathbb{R}^n$ by defining f to be zero outside Ω; then $J_\varepsilon f$ is defined for $x \in \mathbb{R}^n$ and hence for $x \in \Omega$.

Theorem 4.9.1 *For $f \in L^p(\Omega)$, $p \geq 1$, we have $\|J_\varepsilon f\|_p \leq L\|f\|_p$, where $L = L(\varphi, p)$.*

Proof By the previous remark, we may assume that $\Omega = \mathbb{R}^n$.

That $\|J_\varepsilon f\|_p \leq L\|f\|_p$ when $p = 1$ or ∞ is obvious. We consider the case $1 < p < \infty$. In this case, let $q > 1$ be the exponent conjugate to p, i.e. $\frac{1}{p} + \frac{1}{q} = 1$, then,

$$
\begin{aligned}
|J_\varepsilon f(x)| &= \left| \int f(y)\varphi_\varepsilon(x-y)dy \right| \\
&\leq \int |f(y)||\varphi_\varepsilon(x-y)|dy = \int |f(x-y)||\varphi_\varepsilon(y)|dy \\
&\leq \left\{ \int |f(x-y)|^p |\varphi_\varepsilon(y)|dy \right\}^{\frac{1}{p}} \left\{ \int |\varphi_\varepsilon(y)|dy \right\}^{\frac{1}{q}} \\
&= C\left\{ \int |f(x-y)|^p |\varphi_\varepsilon(y)|dy \right\}^{\frac{1}{p}},
\end{aligned}
$$

where $C = \{\int |\varphi_\varepsilon(y)|dy\}^{1/q} = \{\int |\varphi(y)|dy\}^{\frac{1}{q}}$. In one of the steps above, we have used Hölder's inequality w.r.t. the measure ν with $d\nu = |\varphi_\varepsilon| d\lambda^n$ (cf. Exercise 2.5.7). Now the Fubini theorem implies

$$
\begin{aligned}
\|J_\varepsilon f\|_p^p &\leq C^p \int \left(\int |f(x-y)|^p |\varphi_\varepsilon(y)|dy \right) dx \\
&= C^p \int \left(\int |f(x-y)|^p |\varphi_\varepsilon(y)|dx \right) dy \\
&= C^p \|f\|_p^p \int |\varphi(y)|dy = C^p C^q \|f\|_p^p,
\end{aligned}
$$

or,

$$
\|J_\varepsilon f\|_p \leq L\|f\|_p,
$$

where $L = L(\varphi, p)$. Note that $(x, y) \mapsto |f(x-y)|^p \varphi_\varepsilon(y)$ is measurable by Proposition 4.8.2. ■

Exercise 4.9.4 Show that if $\varphi \geq 0$, the constant L in Theorem 4.9.1 can be taken to be 1.

Theorem 4.9.2 *If $f \in L^p(\Omega)$, $1 \leq p < \infty$, then* $\lim_{\varepsilon\to 0} \|J_\varepsilon f - f\|_p = 0$.

Proof We may assume that $\Omega = \mathbb{R}^n$. Let $\sigma > 0$ be given. By Proposition 4.6.1, there is $g \in C_c(\mathbb{R}^n)$ such that $\|f-g\|_p < \frac{\sigma}{2(L+1)}$, where $L = L(\varphi, p)$ is the constant in Theorem 4.9.1. Now,

$$
\begin{aligned}
\|J_\varepsilon f - f\|_p &= \|J_\varepsilon f - J_\varepsilon g + J_\varepsilon g - g + g - f\|_p \\
&\leq \|J_\varepsilon(f-g)\|_p + \|J_\varepsilon g - g\|_p + \|g-f\|_p \\
&\leq (L+1)\|f-g\|_p + \|J_\varepsilon g - g\|_p \\
&< \frac{\sigma}{2} + \|J_\varepsilon g - g\|_p,
\end{aligned}
$$

where we have used the inequality $\|J_\varepsilon(f-g)\|_p \le L\|f-g\|_p$ as asserted by Theorem 4.9.1. Let K be the support of g and put $\widehat{K} = \{x \in \mathbb{R}^n : \operatorname{dist}(x, K) \le 1\}$. $\widehat{K}$ is a compact set, outside which both g and $J_\varepsilon g$ vanish if $0 < \varepsilon \le 1$. Hence, from Proposition 4.9.1,

$$\|J_\varepsilon g - g\|_p^p = \int_{\widehat{K}} |J_\varepsilon g - g|^p d\lambda^n < \left(\frac{\sigma}{2}\right)^p,$$

or,

$$\|J_\varepsilon g - g\|_p < \frac{\sigma}{2}$$

if ε is sufficiently small, say $\varepsilon < \delta$. This means that $\|J_\varepsilon f - f\|_p < \frac{\sigma}{2} + \|J_\varepsilon g - g\|_p < \sigma$, if $\varepsilon < \delta$. ■

Corollary 4.9.1 *$C_c^\infty(\Omega)$ is dense in $L^p(\Omega)$, $1 \le p < \infty$.*

Proof Let $f \in L^p(\Omega)$, $1 \le p < \infty$, and fix $\sigma > 0$. By Proposition 4.6.1, there is $g \in C_c(\Omega)$ such that $\|f-g\|_p < \frac{\sigma}{2}$; while from Theorem 4.9.2, if $\varepsilon > 0$ is small enough, $\|J_\varepsilon g - g\|_p < \frac{\sigma}{2}$. Since g has compact support in Ω, $J_\varepsilon g$ has compact support in Ω if ε is small enough. Hence if ε is small enough, $J_\varepsilon g \in C_c^\infty(\Omega)$ and $\|J_\varepsilon g - g\|_p < \frac{\sigma}{2}$; but then $\|f - J_\varepsilon g\|_p \le \|f-g\|_p + \|g - J_\varepsilon g\|_p < \sigma$. ■

Exercise 4.9.5 Suppose that $\varphi(x) = \varphi(-x)$ for all x in $\mathbb{R}^n$ and let f, g be in $L^2(\mathbb{R}^n)$. Show that

$$\int_{\mathbb{R}^n} (J_\varepsilon f) g d\lambda^n = \int_{\mathbb{R}^n} f J_\varepsilon g d\lambda^n.$$

4.10 Change of variables for multiple integrals

A transformation formula for multiple integrals under changes of variables will be proved in this section. The changes of variables to be considered are C^1 diffeomorphisms, which we shall now describe. Let Ω be an open set in $\mathbb{R}^n$. A map $t = (t_1, \ldots, t_n)$ from Ω into $\mathbb{R}^n$ is called a **C^1 map** if its component functions t_i are continuously differentiable, i.e. first-order partial derivatives of each t_i exist and are continuous on Ω. For $x \in \Omega$, the linear map from $\mathbb{R}^n$ into $\mathbb{R}^n$ represented by the matrix $\left(\frac{\partial t_i}{\partial x_j}(x)\right)$ in reference to the standard basis of $\mathbb{R}^n$ is called the **differential** of t at x, and is denoted by $d_x t$. By the **standard basis** of $\mathbb{R}^n$ we mean the basis formed by $e_1, \ldots, e_n$, where for each j, $e_j = (\delta_{j1}, \ldots, \delta_{jn})$, with δ_{jk} being 1 or 0 according to whether $k = j$ or $k \ne j$. The symbols δ_{jk} are called **Kronecker symbols**. In this section, linear maps from $\mathbb{R}^n$ to $\mathbb{R}^n$ are represented by matrices with reference to the standard basis. The determinant of $\left(\frac{\partial t_i}{\partial x_j}(x)\right)$, called the **Jacobian** of t at x, is to be denoted by $J(t; x)$. When t is a linear map, $t_i(x) = \sum_{j=1}^n t_{ij} x_j$ for $x = (x_1, \ldots, x_n)$, where (t_{ij}) is the matrix representing t; it follows then that $\left(\frac{\partial t_i}{\partial x_j}(x)\right) = (t_{ij})$, i.e. $d_x t = t$. For a linear map t, the determinant of the matrix representing t is usually denoted by $\det t$, thus $J(t; x) = \det d_x t$ if t is a C^1 map. A C^1 map t from Ω into $\mathbb{R}^n$ is called a

C^1 **diffeomorphism** if it is injective and $d_x t$ is invertible for all $x \in \Omega$. By the **inverse function theorem**, if t is a C^1 diffeomorphism from Ω into $\mathbb{R}^n$, then t^{-1} is a C^1 diffeomorphism from $t\Omega$ onto Ω and $J(t;x)^{-1} = J(t^{-1};tx)$ for $x \in \Omega$. Note that $J(t;x) \neq 0$ for all x in Ω.

We consider first the transformation formula for integrals when changes of variables are invoked by invertible linear maps. We follow the usual practice of denoting linear maps by capital letters, and, for convenience, the matrix representing a linear map T is also denoted by T. The matrices derived from the unit matrix I by **elementary row operations** are called **elementary matrices**. They are of the following three types:

(i) A type(1) elementary matrix is one obtained from I by multiplying a row of I by a nonzero real number c;

(ii) a type(2) elementary matrix is one obtained from I by multiplying a row of I by a nonzero real number and then adding it to a different row of I;

(iii) a type(3) elementary matrix is one obtained from I by interchanging two rows of I.

Note that if T is an elementary matrix of type(1), then $\det T = c$; while $\det T = 1$ or -1, according to whether T is of type(2) or type(3). If T is an elementary matrix, the corresponding linear map T is called an **elementary linear map** of the same type.

Lemma 4.10.1 *If T is an elementary linear map and $f \geq 0$ is a measurable function on $\mathbb{R}^n$ such that $f \circ T$ is measurable, then*

$$\int_{\mathbb{R}^n} f d\lambda^n = |\det T| \int_{\mathbb{R}^n} f \circ T d\lambda^n. \tag{4.17}$$

Proof Suppose that T is of type(1), then $f \circ T(x_1, \ldots, x_n) = f(x_1, \ldots, cx_j, \ldots, x_n)$ for some $j = 1, \ldots, n$ and $c \neq 0$. By expressing $x = (x_1, \ldots, x_j, \ldots, x_n)$ as $x = (x_j, \hat{x}_j)$ and using the Fubini theorem, we have

$$\begin{aligned}
\int_{\mathbb{R}^n} f \circ T d\lambda^n &= \int_{\mathbb{R}^{n-1}} \left(\int_{\mathbb{R}} f(x_1, \ldots, cx_j, \ldots, x_n) dx_j \right) d\hat{x}_j \\
&= \frac{1}{|c|} \int_{\mathbb{R}^{n-1}} \left(\int_{\mathbb{R}} f(x_1, \ldots, x_j, \ldots, x_n) dx_j \right) d\hat{x}_j \\
&= \frac{1}{|c|} \int_{\mathbb{R}^n} f d\lambda^n,
\end{aligned}$$

where $\int_{\mathbb{R}} f(x_1, \ldots, cx_j, \ldots, x_n) dx_j = \frac{1}{|c|} \int_{\mathbb{R}} f(x_1, \ldots, x_j, \ldots, x_n) dx_j$ follows from the fact stated in Example 4.3.1 (ii). Hence,

$$\int_{\mathbb{R}^n} f d\lambda^n = |c| \int_{\mathbb{R}^n} f \circ T d\lambda^n = |\det T| \int_{\mathbb{R}^n} f \circ T d\lambda^n.$$

Similarly, (4.17) can be verified for the case when T is of type(2) or of type(3). ∎

If T is an invertible linear map from $\mathbb{R}^n$ onto $\mathbb{R}^n$ then, as is well known in elementary linear algebra, after a finite number of elementary row operations the corresponding matrix T becomes the unit matrix I, i.e.

$$I = S_1 \cdots S_k \cdot T,$$

where $S_1, \ldots, S_k$ are elementary matrices, or

$$S_k^{-1} \cdots S_1^{-1} = T,$$

where each S_j^{-1} is also elementary and of the some type as S_j; in terms of maps, this means that the invertible linear map T is a composition of a finite number of elementary linear maps, i.e.

$$T = T_1 \circ \cdots \circ T_l, \tag{4.18}$$

with each T_j being elementary.

Theorem 4.10.1 *If T is an invertible linear map from $\mathbb{R}^n$ onto $\mathbb{R}^n$ and f is a measurable function on $\mathbb{R}^n$, then $f \circ T$ is measurable; and if f is either nonnegative or integrable,*

$$\int_{\mathbb{R}^n} f d\lambda^n = |\det T| \int_{\mathbb{R}^n} f \circ T d\lambda^n. \tag{4.19}$$

Proof It is sufficient to prove (4.19) for the case $f \geq 0$. Suppose first that $f \geq 0$ is Borel measurable, then since $f \circ T$ is Borel and T is of the form (4.18), we have from Lemma 4.10.1,

$$\begin{aligned}
|\det T| \int_{\mathbb{R}^n} f \circ T d\lambda^n &= \prod_{j=1}^{l} |\det T_j| \int_{\mathbb{R}^n} f \circ T_1 \circ \cdots \circ T_l d\lambda^n \\
&= \left(\prod_{j=1}^{l-1} |\det T_j| \right) \cdot |\det T_l| \int_{\mathbb{R}^n} (f \circ T_1 \circ \cdots \circ T_{l-1}) \circ T_l d\lambda^n \\
&= \prod_{j=1}^{l-1} |\det T_j| \int_{\mathbb{R}^n} f \circ T_1 \circ \cdots \circ T_{l-1} d\lambda^n \\
&= \cdots = \int_{\mathbb{R}^n} f d\lambda^n.
\end{aligned}$$

Thus (4.19) holds when f is a nonnegative Borel function on $\mathbb{R}^n$.

Now suppose that f is nonnegative and measurable. We claim first that $f \circ T$ is measurable. Let $B \in \mathcal{B}^n$; we have to show that $(f \circ T)^{-1}B = T^{-1}(f^{-1}B)$ is measurable. As $f^{-1}B$ is measurable, $f^{-1}B = A \cup C$, where A is a Borel set and $\lambda^n(C) = 0$ (cf. Exercise 3.9.1 (i)). There is a Borel set $D \supset C$ such that $\lambda^n(D) = 0$. The indicator function I_D of D is a Borel function; by what we have proved in the first part, $|\det T|\int_{\mathbb{R}^n} I_D \circ T d\lambda^n = \lambda^n(D) = 0$; then $\int_{\mathbb{R}^n} I_D \circ T d\lambda^n = 0$, and consequently $I_D \circ T = 0$ a.e. But $I_D \circ T = I_{T^{-1}D}$ and $I_D \circ T = 0$ a.e. imply $\lambda^n(T^{-1}D) = 0$. Since $T^{-1}C \subset T^{-1}D$, $\lambda^n(T^{-1}C) = 0$. Thus $T^{-1}C$ is measurable. Now, $(f \circ T)^{-1}B = T^{-1}(A \cup C) = T^{-1}A \cup T^{-1}C$ shows that $(f \circ T)^{-1}B$ is measurable. We have shown the claim that $f \circ T$ is measurable. Since $f \circ T$ is measurable, we can repeat the first part of the proof to conclude that (4.19) holds. ∎

Corollary 4.10.1 *For a measurable set $A \subset \mathbb{R}^n$, TA is measurable and $\lambda^n(TA) = |detT|\lambda^n(A)$.*

Proof In Theorem 4.10.1, replace T by T^{-1} and consider $f = I_A$. ∎

Corollary 4.10.2 *Lebesgue measure is invariant under rotations.*

Proof Let $A \subset \mathbb{R}^n$ and T be a rotation of $\mathbb{R}^n$; we have to show that $\lambda^n(TA) = \lambda^n(A)$. By Corollary 4.10.1, $\lambda^n(TA) = |\det T|\lambda^n(A) = \lambda^n(A)$, because the matrix representing T is an orthogonal matrix and the determinant of an orthogonal matrix is 1 or −1. ∎

Now let t be a C^1 diffeomorphism from Ω into $\mathbb{R}^n$. Define a measure $\lambda^n t$ on Ω by

$$\lambda^n t(A) = \lambda^n(tA), \quad A \subset \Omega.$$

That $\lambda^n t$ measures Ω is obvious. Since t is bijective from Ω to $t\Omega$, the measure $\lambda^n t$ on Ω can be considered as a copy of λ^n on the open set $t\Omega$; actually a subset A of Ω is $\lambda^n t$-measurable if and only if tA is λ^n-measurable, and both t and t^{-1} are measure preserving (cf. Section 2.8.2). Furthermore, since a subset B of Ω is Borel if and only if tB is Borel, it follows that $\lambda^n t$ is a Radon measure on Ω.

Proposition 4.10.1 *If $f \geq 0$ is measurable on $t\Omega$, then $f \circ t$ is $\Sigma^{\lambda^n t}$-measurable on Ω and*

$$\int_{t\Omega} f d\lambda^n = \int_{\Omega} f \circ t d\lambda^n t. \tag{4.20}$$

Proof If $f = I_A$ for a measurable set A, then $f \circ t = I_{t^{-1}A}$, where $t^{-1}A$ is $\lambda^n t$-measurable; it follows that (4.20) holds in this case. For the general case, (4.20) follows from Theorem 2.2.1 and what has just been shown. ∎

Remark Since $\lambda^n = t_{\#}\lambda^n t$ on $t\Omega$, Proposition 4.10.1 follows also from Exercise 4.3.2.

Lemma 4.10.2 *$\lambda^n t$ is absolutely continuous on Ω.*

Proof Let $Q \subset \Omega$ be a nondegenerate oriented closed cube, i.e. $Q = I_1 \times \cdots \times I_n$, where $I_1, \ldots, I_n$ are finite closed intervals in $\mathbb{R}$ of the same positive length. Suppose that f is a continuously differentiable function defined on a neighborhood of Q, and consider two points x and y in Q. Let a function g on $[0, 1]$ be defined by $g(s) = f(x + s(y - x))$; then $f(y) - f(x) = g(1) - g(0) =$

$\int_0^1 g'(s)ds = \int_0^1 \left\{\sum_{j=1}^n \frac{\partial f}{\partial x_j}(x + s(y - x)) \cdot (y_j - x_j)\right\} ds = \int_0^1 \nabla f(x + s(y - x)) \cdot (y - x)ds$, where $\nabla f = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\right)$ is the gradient of f. Hence,

$$|f(y) - f(x)| \leq |y - x| \int_0^1 |\nabla f(x + s(y - x))| ds. \tag{4.21}$$

Applying (4.21) to each component function of t, we have

$$|t(y) - t(x)|^2 \leq |y - x|^2 \sum_{i=1}^n \left[\int_0^1 |\nabla t_i(x + s(y - x))| ds\right]^2 \leq |y - x|^2 M(Q)^2,$$

or

$$|t(y) - t(x)| \leq |y - x| M(Q), \tag{4.22}$$

where $M(Q)^2 = \max_{z \in Q} \sum_{i,j=1}^n \left|\frac{\partial t_i}{\partial x_j}(z)\right|^2$.

Suppose now that A is a null set in Ω. Since Ω is a countable union of open sets G, with $\overline{G}$ being a compact subset of Ω (cf. Proposition 3.9.2), to show that $\lambda^n t(A) = 0$, we may assume that A is a null set in an open set G, with $\overline{G}$ a compact set in Ω. Given that $\varepsilon > 0$, there is a sequence $\{Q_k\}$ of nondegenerate closed oriented cubes in G such that $\bigcup Q_k \supset A$ and $\sum_k \lambda^n(Q_k) < \varepsilon$, by Corollary 3.9.1. For each k, let c_k be the center of Q_k, and apply (4.22) for $x = c_k$ and $y \in Q_k$, to obtain

$$|t(y) - t(c_k)| \leq |y - c_k| M(Q_k),$$

which implies that $tQ_k - t(c_k) \subset C_r(t(c_k))$ with $r = (\frac{1}{2} \operatorname{diam} Q_k)M$, where $M^2 = \max_{z \in \overline{G}} \sum_{i,j=1}^n \left|\frac{\partial t_i}{\partial x_j}(z)\right|^2$, and consequently,

$$\lambda^n(tQ_k) = \lambda^n(tQ_k - t(c_k)) \leq \lambda^n(C_r(t(c_k))) = \left(\frac{\sqrt{n}M}{2}\right)^n \lambda^n(C_1(0))\lambda^n(Q_k),$$

by Example 4.3.1. Now,

$$\begin{aligned}\lambda^n t(A) &\leq \lambda^n t\Big(\bigcup_k Q_k\Big) \leq \sum_k \lambda^n t(Q_k) = \sum_k \lambda^n(tQ_k) \\ &\leq \left(\frac{\sqrt{n}M}{2}\right)^n \lambda^n(C_1(0)) \sum_k \lambda^n(Q_k) < \left(\frac{\sqrt{n}M}{2}\right)^n \lambda^n(C_1(0))\varepsilon,\end{aligned}$$

from which, by letting $\varepsilon \to 0$, we conclude that $\lambda^n t(A) = 0$. ■

Corollary 4.10.3 *$A \subset \Omega$ is measurable if and only if tA is measurable. Also, A is measurable if and only if it is $\lambda^n t$-measurable.*

Proof If A is measurable, then $A = B \cup N$, with B a Borel set and N a null set. By Lemma 4.10.2, $\lambda^n(tN) = \lambda^n t(N) = 0$; hence tN is a null set and is therefore measurable. Now, $tA = tB \cup tN$ implies that tA is measurable. Conversely, if tA is measurable, then A is measurable by the same argument, but with Ω replaced by $t\Omega$ and t replaced by t^{-1}.

Since $A \subset \Omega$ is $\lambda^n t$-measurable if and only if tA is measurable, the second part of the corollary follows from the first part. ∎

Lemma 4.10.3 *For a.e. x in Ω, $\frac{d\lambda^n t}{d\lambda^n}(x) = |\det d_x t|$.*

Proof It is sufficient to show that $\lim_{r\to 0} \frac{\lambda^n t(C_r(x))}{\lambda^n(C_r(x))} = |\det d_x t|$ for $x \in \Omega$, where $C_r(x)$ is the closed ball centered at x and with radius r.

Let $x \in \Omega$ and suppose first that $d_x t = I$, the identity map of $\mathbb{R}^n$. Write

$$t(y) - t(x) = d_x t(y - x) + R(x, y) = (y - x) + R(x, y). \tag{4.23}$$

Since t is differentiable at x, for each $\varepsilon > 0$, there is $\delta > 0$ such that $|R(x, y)| < \varepsilon|y - x|$ if $|y - x| < \delta$. Now if $0 < r < \delta$, we have from (4.23),

$$tC_r(x) - t(x) \subset (1 + \varepsilon)(C_r(x) - x),$$

then $\lambda^n(tC_r(x)) = \lambda^n(tC_r(x) - t(x)) \leq (1 + \varepsilon)^n \lambda^n(C_r(x) - x) = (1 + \varepsilon)^n \lambda^n(C_r(x))$, and hence

$$\limsup_{r\to 0} \frac{\lambda^n t(C_r(x))}{\lambda^n(C_r(x))} \leq (1 + \varepsilon)^n. \tag{4.24}$$

We show next that $C := C_{r(1-\varepsilon)}(t(x))$ is contained in $tC_r(x)$ if $0 < r < \delta$. Observe first that, by (4.23), $t\Gamma$ is outside C, where Γ is the boundary of $C_r(x)$. To show that $C \subset tC_r(x)$ is to show that the line segment $[t(x), z] := \{t(x) + s(z - t(x)) : 0 \leq s \leq 1\} \subset tC_r(x)$ for each $z \in \partial C$. Let $z \in \partial C$ be fixed. Define a set L of positive numbers by

$$L = \{0 < \rho \leq 1 : t(x) + s(z - t(x)) \in tC_r(x) \text{ for all } 0 \leq s \leq \rho\}.$$

By the **inverse function theorem**, t maps a neighborhood of x in $C_r(x)$ onto a neighborhood of $t(x)$; hence L is nonempty. Let $\rho_0 = \sup L$. We claim that $\rho_0 \in L$. Note first that $(0, \rho_0) \subset L$. Choose a sequence $\{s_j\}$ in $(0, \rho_0)$ such that $s_j \to \rho_0$ and let $z_j = t(x) + s_j(z - t(x))$. Then $z_j \in tC_r(x)$ for each j. Since $z_j \to z_\infty := t(x) + \rho_0(z - t(x))$ and t^{-1} is continuous, we infer that $t^{-1}z_j \to t^{-1}z_\infty$ and $t^{-1}z_\infty \in C_r(x)$ (note that each $t^{-1}z_j \in C_r(x)$). Now, $t(t^{-1}z_\infty) = z_\infty$ implies that $\rho_0 \in L$. We assert then that $\rho_0 = 1$. If $\rho_0 < 1$, $t^{-1}z_\infty \in B_r(x)$, because $t\Gamma$ is outside C; then by the inverse function theorem again, t maps a neighborhood of $t^{-1}z_\infty$ in $B_r(x)$ onto a neighborhood of z_∞; this would imply that L contains numbers larger than ρ_0, contradicting the definition of ρ_0. Now $\rho_0 = 1$ means the line segment $[t(x), z]$ is contained in

$tC_r(x)$. Thus C is contained in $tC_r(x)$, or $tC_r(x) - t(x) \supset (1-\varepsilon)(C_r(t(x)) - t(x))$. Hence,

$$\lambda^n t(C_r(x)) = \lambda^n(tC_r(x)) = \lambda^n(tC_r(x) - t(x)) \geq (1-\varepsilon)^n \lambda^n(C_r(x)),$$

or

$$\liminf_{r\to 0} \frac{\lambda^n t(C_r(x))}{\lambda^n(C_r(x))} \geq (1-\varepsilon)^n. \tag{4.25}$$

Letting $\varepsilon \to 0$ in (4.24) and (4.25), we have

$$\lim_{r\to 0} \frac{\lambda^n t(C_r(x))}{\lambda^n(C_r(x))} = 1.$$

This shows that $\lim_{r\to 0} \frac{\lambda^n t(C_r(x))}{\lambda^n(C_r(x))} = 1$, if $d_x t = I$. In general, for $x \in \Omega$, consider the map $\hat{t} = (d_x t)^{-1} \circ t$, then $d_x \hat{t} = (d_x t)^{-1} \circ d_x t = I$, hence,

$$\lim_{r\to 0} \frac{\lambda^n \hat{t}(C_r(x))}{\lambda^n(C_r(x))} = 1. \tag{4.26}$$

Now, by Corollary 4.10.1,

$$\begin{aligned}\lambda^n t(C_r(x)) &= \lambda^n(tC_r(x)) = \lambda^n(d_x t \circ (d_x t)^{-1}(tC_r(x))) \\ &= |\det d_x t| \lambda^n(\hat{t} C_r(x)),\end{aligned}$$

from which it follows that

$$\lim_{r\to 0} \frac{\lambda^n t(C_r(x))}{\lambda^n(C_r(x))} = |\det d_x t| \lim_{r\to 0} \frac{\lambda^n \hat{t}(C_r(x))}{\lambda^n(C_r(x))} = |\det d_x t|,$$

by (4.26). ■

Theorem 4.10.2 *Suppose that t is a C^1 diffeomorphism from an open set Ω in $\mathbb{R}^n$ into $\mathbb{R}^n$; then if f is a measurable function on $t\Omega$, $f \circ t$ is measurable on Ω, and if, furthermore, f is nonnegative or integrable, then,*

$$\int_{t\Omega} f d\lambda^n = \int_{\Omega} (f \circ t)(x) |J(t;x)| d\lambda^n(x). \tag{4.27}$$

Proof Since $A \subset t\Omega$ is measurable if and only if $t^{-1}A$ is measurable by Corollary 4.10.3, we infer that if $f = I_A$, then f is measurable if and only if $f \circ t = I_{t^{-1}A}$ is measurable. It follows then from Theorem 2.2.1 that a nonnegative function f is measurable if and only if $f \circ t$ is measurable; from this it follows that f is measurable on $t\Omega$ if and only if

$f \circ t$ is measurable. In particular if f is measurable on $t\Omega$, then $f \circ t$ is measurable on Ω. To verify (4.27), we need only consider the case $f \geq 0$. By Proposition 4.10.1,

$$\int_{t\Omega} f d\lambda^n = \int_{\Omega} f \circ t d\lambda^n t. \tag{4.28}$$

Since $\lambda^n t$ is absolutely continuous,

$$\lambda^n t(A) = \int_A \frac{d\lambda^n t}{d\lambda^n} d\lambda^n = \int_A |\det d_x t| d\lambda^n(x) = \int_A |J(t;x)| d\lambda^n(x)$$

for measurable $A \subset \Omega$ by Lemma 4.10.3; it follows then from Exercise 2.5.7 that $\int_\Omega f \circ t d\lambda^n t = \int_\Omega (f \circ t)(x)|J(t;x)|d\lambda^n(x)$; combining the last equality with (4.28), we conclude that (4.27) holds. ■

We illustrate the way to use Theorem 4.10.2 by an example.

Example 4.10.1 Consider the map t from the open set $\Omega := \{(\rho,\theta) : 0 < \rho < \infty, 0 < \theta < 2\pi\}$ in $\mathbb{R}^2$ into $\mathbb{R}^2$ by

$$(x_1, x_2) = t(\rho,\theta) = (\rho\cos\theta, \rho\sin\theta),$$

then, $\frac{\partial x_1}{\partial \rho} = \cos\theta$, $\frac{\partial x_1}{\partial \theta} = -\rho\sin\theta$; $\frac{\partial x_2}{\partial \rho} = \sin\theta$, $\frac{\partial x_2}{\partial \theta} = \rho\cos\theta$. Hence,

$$d_{(\rho,\theta)}t = \begin{vmatrix} \cos\theta & -\rho\sin\theta \\ \sin\theta & \rho\cos\theta \end{vmatrix} = \rho > 0.$$

t is actually a C^1 diffeomorphism from Ω onto $t\Omega = \{(x_1,x_2) \in \mathbb{R}^2 : x_2 \neq 0, \text{or } x_2 = 0 \text{ but } x_1 < 0\}$, i.e. $t\Omega$ is obtained from $\mathbb{R}^2$ by taking away the positive x_1-axis and the origin. Now if $f \geq 0$ is measurable, then, since $\lambda^2(\mathbb{R}^2 \backslash t\Omega) = 0$, we have

$$\begin{aligned} \int_{\mathbb{R}^2} f d\lambda^2 &= \int_{t\Omega} f d\lambda^n = \int_\Omega (f \circ t)(\rho,\theta)\rho d\lambda^2(\rho,\theta) \\ &= \int_0^\infty \left(\int_0^{2\pi} \rho f(\rho\cos\theta, \rho\sin\theta) d\theta \right) d\rho, \end{aligned}$$

where we have the applied the Fubini theorem in the last step.

Exercise 4.10.1 Suppose that f is a measurable function on $\mathbb{R}^3$ and is either nonnegative or integrable.

(i) Show that

$$\begin{aligned} \int_{\mathbb{R}^3} f(x,y,z) d\lambda^3(x,y,z) &= \int_G f(\rho\cos\varphi, \rho\sin\varphi, z)\rho d\lambda^3(\rho,\varphi,z) \\ &= \int_{-\infty}^\infty \int_0^{2\pi} \int_0^\infty f(\rho\cos\varphi, \rho\sin\varphi, z)\rho d\rho d\varphi dz, \end{aligned}$$

where $G = (0,\infty) \times (0,2\pi) \times \mathbb{R} = \{(\rho,\varphi,z) : 0 < \rho < \infty, 0 < \varphi < 2\pi, z \in \mathbb{R}\}$.

(ii) Show that

$$\begin{aligned}
&\int_{\mathbb{R}^3} f(x,y,z)d\lambda^3(x,y,z) \\
&= \int_H f(\rho\sin\theta\cos\varphi, \rho\sin\theta\sin\varphi, \rho\cos\theta)\rho^2\sin\theta d\lambda^3(\rho,\theta,\varphi) \\
&= \int_0^\infty\int_0^\pi\int_0^{2\pi} f(\rho\sin\theta\cos\varphi, \rho\sin\theta\sin\varphi, \rho\cos\theta)\rho^2\sin\theta d\varphi d\theta d\rho,
\end{aligned}$$

where $H = (0,\infty)\times(0,\pi)\times(0,2\pi) = \{(\rho,\theta,\varphi) : 0<\rho<\infty, 0<\theta<\pi, 0<\varphi<2\pi\}$.

4.11 Polar coordinates and potential integrals

In Example 4.10.1, ρ and θ are the **polar coordinates** of the point $(\rho\cos\theta, \rho\sin\theta)$ in $\mathbb{R}^2$, and $d\theta$ is the line element on the unit circle S^1, described by $(\cos\theta, \sin\theta)$, $0\le\theta<2\pi$; while in Exercise 4.10.1 (ii), ρ, φ, and θ are the so-called **spherical coordinates** of the point $(\rho\sin\theta\cos\varphi, \rho\sin\theta\sin\varphi, \rho\cos\theta)$ in $\mathbb{R}^3$, and $\sin\theta d\varphi d\theta$ is the surface element on the unit sphere S^2 in $\mathbb{R}^3$, described by $(\sin\theta\cos\varphi, \sin\theta\cos\varphi, \cos\theta)$, $0\le\varphi<2\pi$, $0\le\theta\le\pi$. Therefore, for nonnegative measurable function f on $\mathbb{R}^2$ or $\mathbb{R}^3$, we have

$$\int_{\mathbb{R}^2} f(x)d\lambda^2(x) = \int_0^\infty\left(\int_{S^1}\rho f(\rho x')dl(x')\right)d\rho; \tag{4.29}$$

$$\int_{\mathbb{R}^3} f(x)d\lambda^3(x) = \int_0^\infty\left(\int_{S^2}\rho^2 f(\rho x')d\sigma(x')\right)d\rho, \tag{4.30}$$

where $x=\rho x'$ with $\rho=|x|$ and $x'\in S^1$ or S^2, depending on $x\in\mathbb{R}^2$ or $\mathbb{R}^3$, dl is the line element on S^1, and $d\sigma$ the surface element on S^2. The discussion so far is formal; we shall now put it on a solid basis for $\mathbb{R}^n$ in general.

For $x\in\dot{\mathbb{R}}^n := \mathbb{R}^n\backslash\{0\}$, write $x=\rho x'$, where $\rho=|x|$ and $x'=|x|^{-1}x$ is in $S^{n-1} := \{x\in\mathbb{R}^n : |x|=1\}$; ρ and x' are called the **polar coordinates** of $x\in\dot{\mathbb{R}}^n$. The polar coordinates of a point $x\in\dot{\mathbb{R}}^n$ will be written as an ordered pair (ρ,x') and hence is represented as a point in $(0,\infty)\times S^{n-1}$. Let p be the map $x\mapsto(\rho,x')$ from $\dot{\mathbb{R}}^n$ to $(0,\infty)\times S^{n-1}$; p is obviously a bijection and both p and p^{-1} are continuous; it follows that a function f on $\dot{\mathbb{R}}^n$ is λ^n-measurable if and only if $f\circ p^{-1}$ is $p_\#\lambda^n$-measurable on $(0,\infty)\times S^{n-1}$, where $p_\#\lambda^n$ is the measure on $(0,\infty)\times S^{n-1}$, defined by $p_\#\lambda^n(A)=\lambda^n(p^{-1}A)$ for subsets A of $(0,\infty)\times S^{n-1}$ (cf. Exercise 4.3.1 and note that $\lambda^n=(p^{-1})_\#(p_\#\lambda^n)$). We then infer from Exercise 4.3.2 that if f is a nonnegative measurable or an integrable function on $\mathbb{R}^n$, then

$$\int_{\mathbb{R}^n} fd\lambda^n = \int_{\dot{\mathbb{R}}^n} fd\lambda^n = \int_{(0,\infty)\times S^{n-1}} f\circ p^{-1}dp_\#\lambda^n. \tag{4.31}$$

We shall presently show that $p_\# \lambda^n$ is a product measure. A Borel measure σ on S^{n-1} will be defined first; this measure is interpreted as measuring the surface area of sets in S^{n-1} and is therefore called the surface measure on S^{n-1}. For $E \subset S^{n-1}$ and $r > 0$, let E_r be the set $\bigcup\{\alpha E : 0 < \alpha \leq r\}$ in $\mathbb{R}^n$; clearly, $E_r = rE_1$ and E_r is a Borel set in $\mathbb{R}^n$, if $E \in \mathcal{B}(S^{n-1})$. It then follows that

$$\lambda^n(E_r) = r^n \lambda^n(E_1) \tag{4.32}$$

for $E \in \mathcal{B}(S^{n-1})$, by Example 4.3.1 (ii). Observe now that if $h > 0$, $E_{1+h}\backslash E_1$ is a spherically sliced section of the cone $\bigcup\{\alpha E : \alpha > 0\}$ of thickness h, and hence it is natural to define the surface area of $E \in \mathcal{B}(S^{n-1})$, as

$$\lim_{h \to 0+} h^{-1} \lambda^n(E_{1+h}\backslash E_1) = \lim_{h \to 0+} h^{-1}[(1+h)^n - 1]\lambda^n(E_1) = n\lambda^n(E_1),$$

where we have applied (4.32) with $r = 1 + h$. Thus we let $\sigma(E) = n\lambda^n(E_1)$ for $E \in \mathcal{B}(S^{n-1})$. It is readily verified that σ is a finite measure on $\mathcal{B}(S^{n-1})$, and the measure on S^{n-1} constructed from σ by Method I is the unique Radon measure on S^{n-1}, extending σ on $\mathcal{B}(S^{n-1})$ (this measure is also denoted by σ), and $(S^{n-1}, \Sigma^\sigma, \sigma)$ is the completion of $(S^{n-1}, \mathcal{B}(S^{n-1}), \sigma)$ (cf. Exercise 3.4.18).

From (4.32), we have

$$\lambda^n(E_r) = r^n\lambda^n(E_1) = n\lambda^n(E_1)\int_0^r \rho^{n-1}d\rho = \sigma(E)\int_0^r \rho^{n-1}d\rho$$

for $E \in \mathcal{B}(S^{n-1})$ and hence, by Borel regularity of σ, E_r is measurable and

$$\lambda^n(E_r) = \sigma(E)\int_0^r \rho^{n-1}d\rho \tag{4.33}$$

for any σ-measurable set E in S^{n-1} (see Exercise 4.11.1).

Exercise 4.11.1 Let E be a σ-measurable set in S^{n-1}; show that E_r is measurable and (4.33) holds. (Hint: there are Borel sets F and G in S^{n-1} such that $F \subset E \subset G$ and $\sigma(G\backslash F) = 0$.)

Now let γ be the unique Radon measure on $(0, \infty)$ such that $\gamma(B) = \int_B \rho^{n-1}d\rho$ for Borel sets B in $(0, \infty)$. Since $\gamma(A) = 0$ if and only if $\lambda(A) = 0$ for any $A \subset (0, \infty)$, it follows that γ-measurable sets in $(0, \infty)$ are exactly the Lebesgue measurable sets in $(0, \infty)$.

Lemma 4.11.1 *For σ-measurable sets E in S^{n-1} and measurable sets A in $(0, \infty)$,*

$$\gamma \times \sigma(A \times E) = p_\# \lambda^n(A \times E).$$

Proof For a fixed σ-measurable set E in S^{n-1}, let $\mathcal{M}$ be the family of all measurable sets A in $(0, \infty)$ such that for every positive integer n,

$$\gamma \times \sigma(A \cap (0, n] \times E) = p_{\#}\lambda^n(A \cap (0, n] \times E),$$

then, $\gamma \times \sigma(A \times E) = p_{\#}\lambda^n(A \times E)$ for $A \in \mathcal{M}$. Since $p_{\#}\lambda^n((0, r] \times E) = \lambda^n(E_r)$, we infer from (4.33) that $\mathcal{M}$ contains $\Pi = \{(0, r] : r > 0\}$, which is a π-system on $(0, \infty)$. It is routine to verify that $\mathcal{M}$ is a λ-system, and the $(\pi$-$\lambda)$ theorem implies that $\mathcal{M}$ contains all Borel sets in $(0, \infty)$. Now if A is a measurable set in $(0, \infty)$, there are Borel sets C and D in $(0, \infty)$ such that $C \subset A \subset D$ and $\lambda(D \backslash C) = \gamma(D \backslash C) = 0$, hence,

$$\begin{aligned} \gamma \times \sigma(C \times E) &= p_{\#}\lambda^n(C \times E) \leq p_{\#}\lambda^n(A \times E) \leq p_{\#}\lambda^n(D \times E) \\ &= \gamma \times \sigma(D \times E) = \gamma \times \sigma(C \times E), \end{aligned}$$

from which it follows that $\gamma \times \sigma(A \times E) = p_{\#}\lambda^n(A \times E)$. ■

Lemma 4.11.2 $\mathcal{B}((0, \infty) \times S^{n-1}) \subset \Sigma^\gamma \otimes \Sigma^\sigma \subset \Sigma^{p_{\#}\lambda^n}$.

Proof Since both $(0, \infty)$ and S^{n-1} are separable as metric space, every open set in $(0, \infty) \times S^{n-1}$ is a countable union of sets of the form $A \times B$, where A is open in $(0, \infty)$ and B is open in S^{n-1}; open sets in $(0, \infty) \times S^{n-1}$ are $\Sigma^\gamma \otimes \Sigma^\sigma$-measurable and hence $\mathcal{B}((0, \infty) \times S^{n-1}) \subset \Sigma^\gamma \otimes \Sigma^\sigma$. To show that $\Sigma^\gamma \otimes \Sigma^\sigma \subset \Sigma^{p_{\#}\lambda^n}$, it is sufficient to show that $A \times B \in \Sigma^{p_{\#}\lambda^n}$ if $A \in \Sigma^\gamma$ and $B \in \Sigma^\sigma$. There are Borel sets C and D in $(0, \infty)$ such that $C \subset A \subset D$ and $\gamma(D \backslash C) = 0$, and there are Borel sets E and F in S^{n-1} such that $E \subset B \subset F$ and $\sigma(F \backslash E) = 0$; then,

$$\gamma \times \sigma(D \times F \backslash C \times E) = 0,$$

and by Lemma 4.11.1,

$$p_{\#}\lambda^n(D \times F \backslash C \times E) = 0,$$

from which we infer that $A \times B = C \times E \cup N$, where $N \subset D \times F \backslash C \times E$ and is therefore a $p_{\#}\lambda^n$-null set. Thus N is $p_{\#}\lambda^n$-measurable and so is $A \times B$, because $C \times E$ is a Borel set in $(0, \infty) \times S^{n-1}$ and is therefore $p_{\#}\lambda$-measurable. ■

Since $\gamma \times \sigma$ is the unique measure on $\Sigma^\gamma \otimes \Sigma^\sigma$ such that $\gamma \times \sigma(A \times E) = \gamma(A)\sigma(E)$ for measurable set $A \subset (0, \infty)$, and $E \in \Sigma^\sigma$ by Proposition 4.8.1, it follows from Lemma 4.11.1 and Lemma 4.11.2 that $p_{\#}\lambda^n = \gamma \times \sigma$ on $\Sigma^\gamma \otimes \Sigma^\sigma$. Since $\mathcal{B}((0, \infty) \times S^{n-1}) \subset \Sigma^\gamma \otimes \Sigma^\sigma$, by Lemma 4.11.2, one concludes that the space $((0, \infty) \times S^{n-1}, \Sigma^{p_{\#}\lambda^n}, p_{\#}\lambda^n)$ is the completion of $((0, \infty) \times S^{n-1}, \Sigma^\gamma \otimes \Sigma^\sigma, \gamma \times \sigma)$, from the fact that $p_{\#}\lambda^n$ is Borel regular (cf. Exercise 3.4.18). That $p_{\#}\lambda^n$ is Borel regular follows from the Borel regularity of λ^n and the fact that $\mathcal{B}(\dot{\mathbb{R}}^n) = p^{-1}\mathcal{B}((0, \infty) \times S^{n-1})$.

Then, on account of (4.31), we infer immediately that if f is a nonnegative measurable function or an integrable function on $\mathbb{R}^n$, then

$$\int_{\mathbb{R}^n} f d\lambda^n = \int_{(0,\infty)\times S^{n-1}} f \circ p^{-1} d\overline{\gamma \times \sigma};$$

consequently, if we put $f(\rho,\theta) = f \circ p^{-1}(\rho,\theta)$, we have from the Fubini theorem the following theorem.

Theorem 4.11.1 (Integral in polar coordinates) *If f is a nonnegative measurable function or an integrable function on $\mathbb{R}^n$, then*

$$\begin{aligned}\int_{\mathbb{R}^n} f d\lambda^n &= \int_0^\infty \left(\int_{S^{n-1}} f(\rho,\theta) d\sigma(\theta)\right) \rho^{n-1} d\rho \\ &= \int_{S^{n-1}} \left(\int_0^\infty \rho^{n-1} f(\rho,\theta) d\rho\right) d\sigma(\theta).\end{aligned}$$

Example 4.11.1 Suppose that $0 \leq \alpha < n$ and let $\Gamma_\alpha(x,y) = |x-y|^{-\alpha}$, then for any $r > 0$,

$$\begin{aligned}\int_{B_r(x)} \Gamma_\alpha(x,y) d\lambda^n(y) &= \int_{B_r(0)} \Gamma_\alpha(0,y) d\lambda^n(y) \\ &= \int_0^r \left(\rho^{n-1} \int_{S^{n-1}} \rho^{-\alpha} d\sigma(\theta)\right) d\rho = \frac{\omega_{n-1}}{n-\alpha} r^{n-\alpha}, \qquad (4.34)\end{aligned}$$

where $\omega_{n-1} = \sigma(S^{n-1})$.

Exercise 4.11.2 Let b_n be the Lebesgue measure of the unit ball in $\mathbb{R}^n$, and let $l_n = \prod_{j=2}^n \int_0^{\frac{\pi}{2}} \cos^j \theta d\theta$ for $n \geq 2$.

(i) Show that $b_n = 2^n l_n$ for $n \geq 2$.

(ii) Show that $b_{2k} = \frac{1}{k!}\pi^k$ and $b_{2k+1} = 2^{2k+1}\frac{k!}{(2k+1)!}\pi^k$.

(Hint: express b_n in terms of b_{n-1} by using the Fubini theorem.)

Exercise 4.11.3

(i) Show that $\int_{\mathbb{R}^n} e^{-|x|^2} dx = \frac{\omega_{n-1}}{2} \int_0^\infty t^{\frac{n}{2}-1} e^{-t} dt$ and $\omega_{n-1} = \frac{n\pi^{\frac{n}{2}}}{\Gamma(\frac{1}{2}n+1)}$, where $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$.

(ii) Compare (i) and Exercise 4.11.2 (ii) to find $\Gamma(\frac{n}{2})$ for $n \in \mathbb{N}$.

In the remaining part of this section, a brief account of integral operators of potential type will be given, with an application to integral representation of C^1 functions.

For $0 < \alpha < n$, let Γ_α be the function on $\mathbb{R}^n \times \mathbb{R}^n$ defined by

$$\Gamma_\alpha(x,\xi) = \frac{1}{|x-\xi|^\alpha}, \quad (x,\xi) \in \mathbb{R}^n \times \mathbb{R}^n.$$

Given a bounded measurable set Ω with positive measure in $\mathbb{R}^n$, we denote by $\widehat{\Omega}$ the smallest closed ball centered at 0 and containing Ω, i.e. $\widehat{\Omega} = C_R(0)$, where $R = \sup_{x\in\Omega} |x|$.

Lemma 4.11.3 *For $u \in L^1(\Omega)$,*

$$\int_\Omega \Gamma_\alpha(x,\xi)|u(\xi)|d\xi < \infty$$

for a.e. x in $\mathbb{R}^n$.

Proof Let R be the radius of the ball $\widehat{\Omega}$, by the Fubini theorem and (4.34),

$$\begin{aligned}\int_{C_{2R}(0)} \left(\int_\Omega \Gamma_\alpha(x,\xi)|u(\xi)|d\xi\right) dx &= \int_\Omega |u(\xi)| \int_{C_{2R}(0)} \Gamma_\alpha(x,\xi)dxd\xi \\ &\le \int_\Omega |u(\xi)| \int_{C_{3R}(\xi)} \Gamma_\alpha(x,\xi)dxd\xi \\ &\le \frac{\omega_{n-1}}{n-\alpha}(3R)^{n-\alpha} \int_\Omega |u(\xi)|d\xi < \infty,\end{aligned}$$

i.e. $\int_\Omega \Gamma_\alpha(x,\xi)|u(\xi)|d\xi$ is an integrable function of x on $C_{2R}(0)$. Hence, $\int_\Omega \Gamma_\alpha(x,\xi)$ $|u(\xi)|d\xi < \infty$ for a.e. x in $C_{2R}(0)$; while if x is outside $C_{2R}(0)$, $\int_\Omega \Gamma_\alpha(x,\xi)|u(\xi)|d\xi$ $\le \int_\Omega |u(\xi)|d\xi < \infty$. ■

Because of Lemma 4.11.3, for $u \in L^1(\Omega)$, a function $K_\alpha u$ can be defined a.e. on $\mathbb{R}^n$ by

$$(K_\alpha u)(x) = \int_\Omega \Gamma_\alpha(x,\xi)u(\xi)d\xi, \quad x \in \mathbb{R}^n.$$

$K_\alpha u$ is a function measurable by the Fubini theorem; therefore K_α is a linear operator from $L^1(\Omega)$ into the space of measurable functions on $\mathbb{R}^n$. We call K_α an **integral operator of potential type** and Γ_α a **potential kernel.**

Theorem 4.11.2 *Suppose that Ω and D are two bounded measurable sets of positive measure in $\mathbb{R}^n$, then K_α is a bounded linear operator from $L^p(\Omega)$ into $L^p(D)$.*

Proof When $p = 1$ or ∞, the theorem is obvious. We assume that $1 < p < \infty$. Since Ω is bounded, $u \in L^1(\Omega)$ if $u \in L^p(\Omega)$, and hence $(K_\alpha u)(x) = \int_\Omega \Gamma_\alpha(x,\xi)u(\xi)d\xi$ is finite for a.e. x in $\mathbb{R}^n$. Let the radius of the ball $\widehat{\Omega \cup D}$ be R, i.e. $R = \sup_{x\in\Omega\cup D} |x|$, then for $x \in C_R(0)$,

$$\begin{aligned}
|(K_\alpha u)(x)| &\le \int_\Omega \Gamma_\alpha(x,\xi)^{\frac{1}{p}}|u(\xi)|\Gamma_\alpha(x,\xi)^{\frac{1}{q}}d\xi \\
&\le \left(\int_\Omega \Gamma_\alpha(x,\xi)|u(\xi)|^p d\xi\right)^{\frac{1}{p}}\left(\int_\Omega \Gamma_\alpha(x,\xi)d\xi\right)^{\frac{1}{q}} \\
&\le \left(\int_\Omega \Gamma_\alpha(x,\xi)|u(\xi)|^p d\xi\right)^{\frac{1}{p}}\left(\int_{C_{2R}(x)} \Gamma_\alpha(x,\xi)d\xi\right)^{\frac{1}{q}} \\
&= \left[\frac{\omega_{n-1}(2R)^{n-\alpha}}{n-\alpha}\right]^{\frac{1}{q}}\left(\int_\Omega \Gamma_\alpha(x,\xi)|u(\xi)|^p d\xi\right)^{\frac{1}{p}},
\end{aligned}$$

where q is the conjugate exponent of p and (4.34) is applied in the last step. Now, denoting $\frac{\omega_{n-1}(2R)^{n-\alpha}}{n-\alpha}$ by M, we have

$$\begin{aligned}
\|K_\alpha u\|_{p,D}^p &\le M^{\frac{p}{q}}\int_D\int_\Omega \Gamma_\alpha(x,\xi)|u(\xi)|^p d\xi\, dx \\
&= M^{\frac{p}{q}}\int_\Omega\int_D \Gamma_\alpha(x,\xi)|u(\xi)|^p dx d\xi \\
&\le M^{\frac{p}{q}}\int_\Omega |u(\xi)|^p \int_{C_{2R}(\xi)} \Gamma_\alpha(x,\xi)dx d\xi \\
&\le M^{\frac{p}{q}+1}\|u\|_{p,\Omega}^p,
\end{aligned}$$

where (4.34) is again applied in the last step, and $\|\cdot\|_{p,D}$, $\|\cdot\|_{p,\Omega}$ denote respectively the norms on $L^p(D)$ and $L^p(\Omega)$. Thus $\|K_\alpha\| \le M = \frac{\omega_{n-1}(2R)^{n-\alpha}}{n-\alpha}$. ■

It is easy to see that, more generally, if b is a bounded measurable function defined on $D\times\Omega$, the function $K_\alpha^b u$ defined for $u \in L^1(\Omega)$ by

$$(K_\alpha^b u)(x) = \int_\Omega b(x,\xi)\Gamma_\alpha(x,\xi)u(\xi)d\xi$$

is finite for a.e. x in $\mathbb{R}^n$; furthermore, K_α^b is a bounded linear operator from $L^p(\Omega)$ into $L^p(D)$, $p \ge 1$ with norm $\|K_\alpha^b\| \le C\frac{\omega_{n-1}(2R)^{n-\alpha}}{n-\alpha}$, where $C = \|b\|_\infty$ and $R = \sup_{x\in\Omega\cup D}|x|$. Of course, we assume as before that Ω and D are bounded measurable sets with positive measure in $\mathbb{R}^n$.

Theorem 4.11.3 *If Ω and D are compact sets in $\mathbb{R}^n$ with positive measure, and b is a continuous function on $D\times\Omega$, then K_α^b maps every bounded measurable function u into a continuous function on D.*

Proof Fix $x \in D$ and for $\delta > 0$, let $h \in \mathbb{R}^n$ be such that $|h| < \delta$ and $x + h \in D$; for such an h,

$$
\begin{aligned}
&|(K_\alpha^b u)(x+h) - (K_\alpha^b u)(x)| \\
&= \left| \int_\Omega \{b(x+h,\xi)\Gamma_a(x+h,\xi) - b(x,\xi)\Gamma_\alpha(x,\xi)\}u(\xi)d\xi \right| \\
&\le \|u\|_\infty \|b\|_\infty \int_{B_{2\delta}(x)} \{\Gamma_\alpha(x+h,\xi) + \Gamma_\alpha(x,\xi)\}d\xi \\
&\quad + \|u\|_\infty \int_{\Omega \setminus B_{2\delta}(x)} |b(x+h,\xi)\Gamma_\alpha(x+h,\xi) - b(x,\xi)\Gamma_\alpha(x,\xi)|d\xi \\
&\le \|u\|_\infty \|b\|_\infty \frac{\omega_{n-1}}{n-\alpha}\{(3\delta)^{n-\alpha} + (2\delta)^{n-\alpha}\} \\
&\quad + \|u\|_\infty \int_{\Omega \setminus B_{2\delta}(x)} |b(x+h,\xi)\Gamma_\alpha(x+h,\xi) - b(x,\xi)\Gamma_\alpha(x,\xi)|d\xi,
\end{aligned}
$$

because by (4.34),

$$
\int_{B_{2\delta}(x)} \Gamma_\alpha(x+h,\xi)d\xi \le \int_{B_{3\delta}(x+h)} \Gamma_\alpha(x+h,\xi)d\xi \le \frac{\omega_{n-1}(3\delta)^{n-\alpha}}{n-\alpha},
$$

$$
\int_{B_{2\delta}(x)} \Gamma_\alpha(x,\xi)d\xi \le \frac{\omega_{n-1}(2\delta)^{n-\alpha}}{n-\alpha}.
$$

Now, given $\varepsilon > 0$, choose $\delta > 0$ such that $\|u\|_\infty \|b\|_\infty \{(3\delta)^{n-\alpha} + (2\delta)^{n-\alpha}\} < \frac{\varepsilon}{2}$. Since both $\Gamma_\alpha(x+h,\xi)$ and $\Gamma_\alpha(x,\xi) \le \delta^{-\alpha}$ for $\xi \in \Omega \setminus B_{2\delta}(x)$, and

$$
\begin{aligned}
&|b(x+h,\xi)\Gamma_\alpha(x+h,\xi) - b(x,\xi)\Gamma_\alpha(x,\xi)| \\
&\le \|b\|_\infty |\Gamma_\alpha(x+h,\xi) - \Gamma_\alpha(x,\xi)| + \Gamma_\alpha(x,\xi)|b(x+h,\xi) - b(x,\xi)|,
\end{aligned}
$$

we can then choose $0 < \sigma_0 < \delta$ such that

$$
|b(x+h,\xi)\Gamma_\alpha(x+h,\xi) - b(x,\xi)\Gamma_\alpha(x,\xi)| < \{2(\|u\|_\infty \vee 1)\lambda^n(\Omega)\}^{-1}\varepsilon
$$

for all $\xi \in \Omega \setminus B_{2\delta}(x)$ whenever $|h| < \sigma_0$ and $x + h \in D$, and consequently $|(K_\alpha^b u)(x+h) - (K_\alpha^b u)(x)| < \varepsilon$ whenever $|h| < \sigma_0$ and $x + h \in D$. Thus, $K_\alpha^b u$ is continuous at $x \in D$. ■

Exercise 4.11.4 Show that if b is a continuous function on $\mathbb{R}^n \times \Omega$, then $K_\alpha^b u$ is continuous on $\mathbb{R}^n$ for $u \in L^\infty(\Omega)$, where Ω is a compact set with positive measure in $\mathbb{R}^n$.

Theorem 4.11.4 (Integral representation of C^1 functions) *Suppose that Ω is a bounded open convex domain in $\mathbb{R}^n$, then there is a bounded map A from $\Omega \times \Omega$ to $\mathbb{R}^n$ which is* *continuous off the diagonal of* $\Omega \times \Omega$*, such that if* u *is a* C^1 *function on* Ω *with* $\nabla u \in L^1(\Omega)$*, then*

$$u(x) = \frac{1}{\lambda^n(\Omega)} \int_\Omega u(\xi)d\xi - \int_\Omega A(x,\xi) \cdot \nabla u(\xi)\Gamma_{n-1}(x,\xi)d\xi \tag{4.35}$$

for $x \in \Omega$.

Proof Fix $x \in \Omega$. For $\xi \in \Omega$, let

$$g(t) = u(x + t(\xi - x)), \quad 0 \leq t \leq 1,$$

then, $g'(t) = \nabla u(x + t(\xi - x)) \cdot (\xi - x)$ and

$$u(x) = u(\xi) - \int_0^1 \nabla u(x + t(\xi - x)) \cdot (\xi - x)dt. \tag{4.36}$$

When $0 < t \leq 1$, the map $\xi \mapsto z = x + t(\xi - x)$ is an invertible affine map with Jacobian t^n at all $\xi \in \mathbb{R}^n$; we may use Theorem 4.10.2 to obtain

$$\begin{aligned}\int_\Omega |\nabla u(x + t(\xi - x)) \cdot (\xi - x)|d\xi &= \int_{x+t(\Omega - x)} \left|\nabla u(z) \cdot \frac{z - x}{t}\right| \frac{1}{t^n} dz \\ &= \int_\Omega I_{x+t(\Omega-x)}(z)|\nabla u(z) \cdot (z - x)|t^{-(n+1)}dz;\end{aligned}$$

hence,

$$\begin{aligned}&\int_0^1 \int_\Omega |\nabla u(x + t(z - x)) \cdot (\xi - x)|d\xi\, dt \\ &= \int_\Omega |\nabla u(z) \cdot (z - x)| \int_0^1 I_{x+t(\Omega-x)}(z)t^{-(n+1)}dtdz.\end{aligned}$$

But $I_{x+t(\Omega-x)}(z) = 0$, when $0 < t < \frac{|z-x|}{l(x,z)}$, where $l(x,z)$ is the length of the line segment from x to the boundary of Ω through z, thus,

$$\begin{aligned}&\int_\Omega \int_0^1 |\nabla u(x + t(\xi - x)) \cdot (\xi - x)|dtd\xi \\ &= \frac{1}{n}\int_\Omega |\nabla u(z) \cdot (z - x)| \left(\frac{l(x,z)^n}{|z - x|^n} - 1\right) dz \\ &\leq \frac{1}{n}\int_\Omega |\nabla u(z)|\{l(x,z)^n - |z - x|^n\}\Gamma_{n-1}(x,z)dz;\end{aligned}$$

now, for $0 < \rho < \operatorname{dist}(x, \Omega^c)$, we have

$$\int_{B_\rho(x)} |\nabla u(z)|\{l(x,z)^n - |z - x|^n\}\Gamma_{n-1}(x,z)dz \leq M \int_{B_\rho(x)} \Gamma_{n-1}(x,z)dz < \infty,$$

because $|\nabla u(z)|$ is bounded on $B_\rho(x)$, and consequently

$$\int_\Omega \int_0^1 |\nabla u(x+t(\xi-x))\cdot(\xi-x)|dtd\xi < \infty.$$

We have shown that $\nabla u(x+t(\xi-x))\cdot(\xi-x)$ is an integrable function of (ξ,t) on $\Omega\times[0,1]$ for $x\in\Omega$. Integrate both sides of (4.36) w.r.t. ξ over Ω to obtain (denoting $\lambda^n(\Omega)$ by $|\Omega|$),

$$\begin{aligned} u(x)|\Omega| &= \int_\Omega u(\xi)d\xi - \int_\Omega\int_0^1 \nabla u(x+t(\xi-x))\cdot(\xi-x)dtd\xi \\ &= \int_\Omega u(\xi)d\xi - \frac{1}{n}\int_\Omega \nabla u(z)\cdot(z-x)\left\{\frac{l(x,z)^n}{|z-x|^n}-1\right\}dz, \end{aligned}$$

by repeating the previous steps with $|\nabla u(x+t(\xi-x))\cdot(\xi-x)|$ replaced by $\nabla u(x+t(\xi-x))\cdot(\xi-x)$, as assured by the Fubini theorem. Now let A be the map from $\Omega\times\Omega$ to $\mathbb{R}^n$, defined by

$$A(x,\xi) = \frac{1}{n|\Omega|}\left[\frac{l(x,\xi)^n - |x-\xi|^n}{|\xi-x|}\right](\xi-x),$$

if $x\neq\xi$ and $A(x,\xi)=0$; if $x=\xi$, then

$$u(x) = \frac{1}{|\Omega|}\int_\Omega u(\xi)d\xi - \int_\Omega A(x,\xi)\cdot\nabla u(\xi)\Gamma_{n-1}(x,\xi)d\xi$$

for x in Ω. Obviously, A is continuous off the diagonal of $\Omega\times\Omega$ and $|A(x,\xi)|\le\frac{1}{n}(\operatorname{diam}\Omega)^n|\Omega|^{-1}$, since $l(x,\xi)^n - |x-\xi|^n = l(x,\xi)^n\left(1-\frac{|x-\xi|^n}{l(x,\xi)^n}\right)\le l(x,\xi)^n\le(\operatorname{diam}\Omega)^n$ if $x\neq\xi$. ■

Corollary 4.11.1 *Let $u\in C^1(\mathbb{R}^n)$. Suppose that u and all of its partial derivatives of first order are integrable. Then,*

$$u(x) = \frac{1}{nb_n}\int_{\mathbb{R}^n}\frac{(x-\xi)\cdot\nabla u(\xi)}{|x-\xi|^n}d\xi \tag{4.37}$$

for $x\in\mathbb{R}^n$, where b_n is the Lebesgue measure of the unit ball in $\mathbb{R}^n$.

Proof Observe first that $\frac{(x-\xi)\cdot\nabla u(\xi)}{|x-\xi|^n}$ is integrable on $\mathbb{R}^n$ as a function of ξ; actually for $\rho > 0$, we have

$$\begin{aligned}
&\int_{\mathbb{R}^n} \frac{|(x-\xi)\cdot\nabla u(\xi)|}{|x-\xi|^n} d\xi \\
&\leq \int_{B_\rho(x)} \frac{|\nabla u(\xi)|}{|x-\xi|^{n-1}} d\xi + \int_{\mathbb{R}^n\setminus B_\rho(x)} \frac{|\nabla u(\xi)|}{|x-\xi|^{n-1}} d\xi \\
&\leq \sup_{\xi\in B_\rho(x)} |\nabla u(\xi)| \int_{B_\rho(x)} \frac{1}{|x-\xi|^{n-1}} d\xi + \frac{1}{\rho^{n-1}} \int_{\mathbb{R}^n} |\nabla u(\xi)| d\xi \\
&< \infty,
\end{aligned}$$

by recalling that $\int_{B_\rho(x)} \frac{1}{|x-\xi|^{n-1}} d\xi = w_{n-1}\rho$.

For $x \in \mathbb{R}^n$ and $R > 0$, apply Theorem 4.11.4 with $\Omega = B_R(x)$, to obtain

$$\begin{aligned}
u(x) &= \frac{1}{R^n b_n} \int_{B_R(x)} u(\xi) d\xi \\
&\quad - \frac{1}{nR^n b_n} \int_{B_R(x)} \frac{R^n - |\xi - x|^n}{|\xi - x|} (\xi - x)\cdot\nabla u(\xi) \Gamma_{n-1}(x,\xi) d\xi \\
&= \frac{1}{R^n b_n} \int_{B_R(x)} u(\xi) d\xi - \frac{1}{nb_n} \int_{B_R(x)} \frac{(\xi - x)\cdot\nabla u(\xi)}{|\xi - x|^n} d\xi \\
&\quad + \frac{1}{nR^n b_n} \int_{B_R(x)} (\xi - x)\cdot\nabla u(\xi) d\xi,
\end{aligned}$$

because $A(x,\xi) = \frac{1}{nR^n b_n} \left[\frac{R^n - |x-\xi|^n}{|\xi - x|}\right](\xi - x)$ in this case. Now, let $R \to \infty$ to conclude that

$$u(x) = -\frac{1}{nb_n} \int_{\mathbb{R}^n} \frac{(\xi - x)\cdot\nabla u(\xi)}{|\xi - x|^n} d\xi = \frac{1}{nb_n} \int_{\mathbb{R}^n} \frac{(x-\xi)\cdot\nabla u(\xi)}{|x-\xi|^n} d\xi,$$

on noting that

$$\left|\frac{1}{R^n b_n} \int_{B_R(x)} u(\xi) d\xi\right| \leq \frac{1}{R^n b_n} \int_{\mathbb{R}^n} |u(\xi)| d\xi \to 0$$

and

$$\left|\frac{1}{nR^n b_n} \int_{B_R(x)} (\xi - x)\cdot\nabla u(\xi) d\xi\right| \leq \frac{1}{nR^{n-1} b_n} \int_{\mathbb{R}^n} |\nabla u(\xi)| \to 0$$

as $R \to \infty$; while $\int_{B_R(x)} \frac{(\xi-x)\cdot\nabla u(\xi)}{|\xi-x|^n} d\xi \to \int_{\mathbb{R}^n} \frac{(\xi-x)\cdot\nabla u(\xi)}{|\xi-x|^n} d\xi$ as $R \to \infty$ due to the fact that $\frac{(x-\xi)\cdot\nabla u(\xi)}{|x-\xi|^n}$ is integrable on $\mathbb{R}^n$ as a function of ξ. ■

Exercise 4.11.5 Suppose that $u \in C^1(\Omega)$ and $C_r(x) \subset \Omega$. Show that

$$u(x) = \frac{1}{r^n b_n}\left\{\int_{B_r(0)} u(x+\xi)d\xi - \frac{1}{n}\int_{S^{n-1}}\int_0^r (r^n - \rho^n)\frac{\partial u}{\partial \rho}(x+\rho s)d\rho d\sigma(s)\right\}.$$

Example 4.11.2 Let u be a C^1 function on the ball $B_R(x)$ in $\mathbb{R}^n$ such that ∇u is integrable on $B_R(x)$. We establish here the following estimate for the mean of the Lipschitz quotient of u at x:

$$\frac{1}{\lambda^n(B_R(x))}\int_{B_R(x)} \frac{|u(\xi) - u(x)|}{|\xi - x|} d\xi \le M(\nabla u, x), \tag{4.38}$$

where $M(\nabla u, x) = \sup_{0<r\le R} \frac{1}{\lambda^n(B_r(x))}\int_{B_r(x)} |\nabla u| d\lambda^n$.

As in the first step of the proof of Theorem 4.11.4, we have

$$\begin{aligned}
\int_{B_R(x)} \frac{|u(\xi) - u(x)|}{|\xi - x|} d\xi &\le \int_{B_R(x)} \left(\int_0^1 |\nabla u(x + t(\xi - x))| dt\right) d\xi \\
&= \int_0^1 \left(\int_{B_R(x)} |\nabla u(x + t(\xi - x))| d\xi\right) dt \\
&= \int_0^1 \left(\int_{B_{Rt(x)}} |\nabla u(z)| \frac{1}{t_n} dz\right) dt \\
&= \lambda^n(B_R(x)) \int_0^1 \frac{1}{\lambda^n(B_{Rt(x)})}\int_{B_{Rt(x)}} |\nabla u(z)| dz dt \\
&\le \lambda^n(B_R(x)) M(\nabla u, x),
\end{aligned}$$

from which (4.38) follows.

5 Basic Principles of Linear Analysis

Mathematical objects studied in linear analysis are linear transformations between vector spaces endowed with proper concepts of limit. Linear analysis, therefore, provides suitable language and framework for modeling linear phenomena, and, moreover, often suggests feasible methods for solving the corresponding problems. This is most clearly seen in the case of linear algebra when the vector spaces concerned are finite-dimensional.

This chapter is devoted to the most basic principles of linear analysis. Emphasis will be placed on the case when vector spaces are normed vector spaces, although weaker concepts of limit other than in terms of norm will occasionally be considered in view of subsequent applications.

The first basic principles are those arising from the Baire category theorem, and those from separation of sets by hyperplanes. These principles will be presented first, because they are fundamental in many constructs of linear analysis.

In the latter part of the chapter, considerable weight is laid on geometric aspects of linear analysis, with the introduction of Hilbert spaces. The main ingredients are the Riesz representation of continuous linear functionals on Hilbert spaces and Fourier expansion of elements of a Hilbert space with respect to an orthonormal basis.

Recall that vector spaces considered in our discourse are either over the complex field $\mathbb{C}$ or over the real field $\mathbb{R}$; when specification is desirable, they are called complex vector spaces or real vector spaces, according to whether they are over the complex or the real field. As usual, the smallest vector subspace containing a subset S of a vector space is called the vector space **spanned** by S and is denoted by $\langle S \rangle$.

5.1 The Baire category theorem

The Baire category theorem reveals a deep property of complete metric spaces; it is not usually applied directly, but through its consequences, such as the principle of uniform boundedness and the open mapping theorem. We shall present in this section the Baire category theorem and the principle of uniform boundedness; while the open mapping theorem and some of its consequences will be treated in Section 5.2.

Let M be a metric space. A subset S of M is said to be **nowhere dense** in M if the closure $\bar{S}$ of S contains no nonempty open balls of M. A subset A of M is said to be **of the first category** if A is a countable union of nowhere dense subsets of M. Otherwise A is said to be **of the second category**.

Theorem 5.1.1 (Baire category theorem) *A complete metric space M is of the second category.*

Proof It is required to show that if M is a union $\bigcup_{n=1}^{\infty} S_n$ of closed sets, then one of the S_n contains a nonempty open ball. Suppose the contrary, then each S_n^c has a nonempty intersection with every open ball. Thus if B_0 is an open ball with radius 1, $S_1^c \cap B_0$ contains an open ball $B_1 = B_{r_1}(x_1)$ as well as the closed ball $C_1 = C_{r_1}(x_1)$ with $r_1 < \frac{1}{2}$. Then $S_2^c \cap B_1$ contains an open ball $B_2 = B_{r_2}(x_2)$ and the closed ball $C_2 = C_{r_2}(x_2)$ with $r_2 < \frac{1}{2^2}$. Proceed in this way; a sequence of open balls $\{B_k\}$, $B_k = B_{r_k}(x_k)$, is obtained such that the closed ball $C_{k+1} := C_{r_{k+1}}(x_{k+1}) \subset S_{k+1}^c \cap B_k$ and $0 < r_k < 2^{-k}$, $k = 1, 2, \ldots$. Since $\{C_k\}$ is decreasing, $\{x_k, x_{k+1}, \ldots\} \subset C_k$, the sequence $\{x_k\}$ is a Cauchy sequence, hence $x_k \to x$ in M. But for each k, $x \in C_k \subset B_{k-1}$, or $x \in \bigcap_{k=1}^{\infty} B_k$, hence, $x \in \bigcap_{k=1}^{\infty} S_k^c = (\bigcup_{k=1}^{\infty} S_k)^c = \emptyset$, which is absurd. ■

Theorem 5.1.2 (Principle of uniform boundedness) *Let $\{f_\alpha\}$ be a family of continuous nonnegative functions defined on a Banach space X with the following properties:*

(1) $f_\alpha(x+y) \le f_\alpha(x) + f_\alpha(y)$ *for x, y in X and for each α;*
(2) $f_\alpha(\lambda x) = |\lambda| f_\alpha(x)$, *for $\lambda \in \mathbb{C}$ or $\mathbb{R}$ (depending on whether X is a complex or a real space), $x \in X$ and for each α; and*
(3) $\sup_\alpha f_\alpha(x) < \infty$ *for each $x \in X$.*

Then there is $N > 0$, such that

$$\sup_{\alpha} f_\alpha(x) \le N\|x\|$$

for all $x \in X$.

Proof For each $n \in \mathbb{N}$, let

$$S_n = \{x \in X : f_\alpha(x) \le n\ \forall \alpha\} = \bigcap_{\alpha} \{x \in X : f_\alpha(x) \le n\}.$$

Each S_n is closed and from (3), $X = \bigcup_n S_n$. By Theorem 5.1.1, for some n_0, S_{n_0} contains a ball $B = C_r(x_0)$, or

$$\sup_{\alpha; x \in B} f_\alpha(x) \le n_0.$$

Now, there is $N > 0$ such that

$$f_\alpha(x) \leq N$$

for all α if $\|x\| = 1$. To see this, for $x \in X$ with $\|x\| = 1$ and any α,

$$\begin{aligned} f_\alpha(x) = \frac{1}{r} f_\alpha(rx) &\leq \frac{1}{r}\{f_\alpha(rx + x_0) + f_\alpha(-x_0)\} \\ &\leq \frac{1}{r}\left\{n_0 + \sup_\alpha f_\alpha(-x_0)\right\} =: N. \end{aligned}$$

Now, for any $x \neq 0$ and any α,

$$f_\alpha(x) = \|x\| f_\alpha\left(\frac{x}{\|x\|}\right) \leq N\|x\|. \qquad \blacksquare$$

Actually, the principle of uniform boundedness is usually referred to the following special case of Theorem 5.1.2.

Theorem 5.1.3 *Let $\{T_\alpha\} \subset L(X, Y)$, where X is a Banach space and Y a n.v.s. Then $\sup_\alpha \|T_\alpha\| < \infty$ if and only if $\sup_\alpha \|T_\alpha x\| < +\infty$ for each $x \in X$.*

Proof That $\sup_\alpha \|T_\alpha\| < \infty$ implies that $\sup_\alpha \|T_\alpha x\| < \infty$ for all $x \in X$ is obvious; to show the other direction of implication, let $f_\alpha(x) = \|T_\alpha x\|$ and apply Theorem 5.1.2. $\blacksquare$

Theorem 5.1.4 (Banach–Steinhaus) *Let $\{T_n\} \subset L(X, Y)$, where X is a Banach space and Y a n.v.s. Suppose that $Tx = \lim_{n\to\infty} T_n x$ exists for each $x \in X$. Then $T \in L(X, Y)$ and $\|T\| \leq \liminf_{n\to\infty} \|T_n\| \leq \sup_n \|T_n\| < \infty$.*

Proof T is obviously a linear operator from X into Y. Since $\lim_{n\to\infty} T_n x$ exists, it follows that $\sup_n \|T_n x\| < \infty$ and hence $\sup_n \|T_n\| < \infty$, by Theorem 5.1.3. Now,

$$\|T\| = \sup_{\|x\|=1} \|Tx\| = \sup_{\|x\|=1} \left\|\lim_{n\to\infty} T_n x\right\|$$

$$= \sup_{\|x\|=1}\left(\lim_{n\to\infty} \|T_n x\|\right) \leq \sup_{\|x\|=1}\left(\liminf_{n\to\infty} \|T_n\| \cdot \|x\|\right)$$

$$= \liminf_{n\to\infty} \|T_n\| \leq \sup_n \|T_n\| < \infty. \qquad \blacksquare$$

Exercise 5.1.1 Let $\{T_n\} \subset L(X, Y)$, where both X and Y are Banach spaces. A necessary and sufficient condition for $\lim_{n\to\infty} T_n x$ to exist for each $x \in X$ is:

$$\begin{cases} (1) & \lim_{n\to\infty} T_n x \text{ exists for } x \text{ in a dense subset of } X; \\ (2) & \{\|T_n\|\} \text{ is bounded.} \end{cases}$$

Theorem 5.1.5 (C. Neumann) *Suppose that T is a bounded linear operator from a Banach space X into itself with $\|T\| < 1$. Then $(1 - T)^{-1}$ exists, $(1 - T)^{-1} \in L(X)$, and $(1 - T)^{-1}x = \lim_{n\to\infty} \sum_{k=0}^{n} T^k x = \sum_{k=0}^{\infty} T^k x$.*

Proof For each $x \in X$, let $x_n = \sum_{k=0}^{n} T^k x$. Since for $n > m$,

$$\|x_n - x_m\| = \left\| \sum_{k=m+1}^{n} T^k x \right\| \leq \left(\sum_{k=m+1}^{n} \|T\|^k \right) \|x\|,$$

$\{x_n\}$ is a Cauchy sequence in X. Let $Sx = \lim_{n\to\infty} x_n = \lim_{n\to\infty}(\sum_{k=0}^{n} T^k x)$. By Theorem 5.1.4, S is a bounded linear operator. Now,

$$\begin{aligned}(1 - T)Sx &= (1 - T)\left(\lim_{n\to\infty} \sum_{k=0}^{n} T^k x\right) = \lim_{n\to\infty}\left((1 - T)\sum_{k=0}^{n} T^k x\right) \\ &= \lim_{n\to\infty} (x - T^{n+1}x) = x,\end{aligned}$$

because $\|T^{n+1}x\| \leq \|T\|^{n+1}\|x\| \to 0$, implying that $T^{n+1}x \to 0$; similarly, $S(1 - T)x = x$ for $x \in X$. Hence $S = (1 - T)^{-1}$. ■

Exercise 5.1.2 Suppose that $T \in L(X)$, $T \neq 0$, where X is a Banach space. Show that for $\lambda \in \mathbb{C}$ with $|\lambda| < \|T\|^{-1}$ the operator $I - \lambda T$ is bijective. Expand $(I - \lambda T)^{-1}$ in terms of λ and T and their powers.

We now apply the Baire category theorem to show the existence of continuous functions on the finite closed interval $[a, b]$ which are nowhere differentiable on $[a, b]$.

Fix a finite closed interval $[a, b]$ and let $I = [a, c]$, where $b < c < \infty$.

Lemma 5.1.1 *Suppose that $f \in C(I)$ and let $\varepsilon > 0$ and $L > 0$ be given. Then there is a continuous and piece-wise linear function g on I such that $\max_{x\in I} |g(x) - f(x)| \leq \varepsilon$, and the absolute value of the slope of each line segment of the graph of g is greater than L.*

Proof Let $\delta > 0$ be chosen so that $|f(x) - f(y)| < \frac{\varepsilon}{4}$ if $|x - y| < \delta$. Consider a partition $a = x_0 < x_1 < \cdots < x_{k-1} < x_k = c$ of I, with $|x_j - x_{j-1}| < \delta$ for $j = 1, \ldots, k$, and let $P_0 = (x_0, f(x_0))$, $P_1 = (x_1, f(x_1) + \frac{3}{4}\varepsilon), \ldots, P_j = (x_j, f(x_j) + (-1)^{j-1}\frac{3}{4}\varepsilon), \ldots, P_k = (x_k, f(x_k))$. Let g be the piece-wise linear function whose graph consists of the line segments $[P_0, P_1], [P_1, P_2], \ldots, [P_{k-1}, P_k]$. Then g is continuous and $\max_{x\in I} |g(x) - f(x)| \leq \varepsilon$. If we choose δ small enough, then the absolute value of the slope of each $[P_{j-1}, P_j]$, $j = 1, \ldots, k$, is greater than L. ■

Theorem 5.1.6 *There is a continuous function on $[a, b]$ which is nowhere differentiable on $[a, b]$.*

Proof Let $I = [a, c]$, $b < c < \infty$. It is sufficient to show that there is $f \in C(I)$ such that f is not differentiable at every point of $[a, b]$; actually, should f be differentiable from the left at b, the function $f + g$ is differentiable nowhere on $[a, b]$ if g is a continuous function on $[a, b]$ which is differentiable on $[a, b)$, but not differentiable from the left at b. As usual, we endow $C(I)$ with sup-norm, then $C(I)$ is a complete metric space. Consider the set S of functions f in $C(I)$ such that for some $\xi \in [a, b]$, the set $\left\{\frac{f(\xi+h)-f(\xi)}{h} : 0 < h \le c-b\right\}$ is bounded. Clearly, S contains all functions in $C(I)$ which are differentiable somewhere on $[a, b]$. For $n \in \mathbb{N}$, let

$$S_n = \left\{f \in S : \sup_{0<h\le c-b} \left|\frac{f(\xi+h)-f(\xi)}{h}\right| \le n \text{ for some } \xi \in [a, b]\right\}.$$

Observe that $S = \bigcup_n S_n$. We claim first that each S_n is closed. Let $\{f_k\}$ be a sequence in S_n which converges to f in $C(I)$. To claim that S_n is closed is to show that $f \in S_n$. For each k, there is $\xi_k \in [a, b]$ such that

$$\sup_{0<h\le c-b} \left|\frac{f_k(\xi_k+h)-f_k(\xi_k)}{h}\right| \le n.$$

Since $[a, b]$ is compact, $\{\xi_k\}$ has a subsequence which converges to $\xi \in [a, b]$. If necessary, replace $\{f_k\}$ by a subsequence of itself; we may assume that $\{\xi_k\}$ converges to ξ. For $0 < h \le c-b$ and $\varepsilon > 0$, there is $N = N(h, \varepsilon) \in \mathbb{N}$ such that $k > N$ implies $\sup_{x\in I} |f_k(x) - f(x)| < \frac{\varepsilon h}{4}$. Since f is uniformly continuous on I and $\xi_k \to \xi$, there is $N_1 > N$ such that $|f(\xi_k) - f(\xi)| < \frac{\varepsilon h}{4}$ and $|f(\xi+h) - f(\xi_k+h)| < \frac{\varepsilon h}{4}$ whenever $k > N_1$. Thus, for $k > N_1$, we have

$$\begin{aligned}\left|\frac{f(\xi+h)-f(\xi)}{h}\right| &\le \frac{1}{h}\{|f_k(\xi_k+h) - f_k(\xi_k)| + |f(\xi_k) - f(\xi)| + |f_k(\xi_k) - f(\xi_k)| \\ &\quad + |f(\xi_k+h) - f_k(\xi_k+h)| + |f(\xi+h) - f(\xi_k+h)|\} \\ &< \left|\frac{f_k(\xi_k+h)-f_k(\xi_k)}{h}\right| + \varepsilon \le n + \varepsilon;\end{aligned}$$

hence, $\sup_{0<h\le c-b}\left|\frac{f(\xi+h)-f(\xi)}{h}\right| \le n+\varepsilon$ for $\varepsilon > 0$, consequently,

$$\sup_{0<h\le c-b} \left|\frac{f(\xi+h)-f(\xi)}{h}\right| \le n$$

and $f \in S_n$. This shows that S_n is closed for $n \in \mathbb{N}$.

Next we claim that each S_n is nowhere dense in $C(I)$. For this, it is sufficient to show that $C(I)\backslash S_n$ is dense in $C(I)$. Consider $f \in C(I)$ and $\varepsilon > 0$; we claim that there is $g \in C(I)\backslash S_n$ such that $\sup_{x\in I} |g(x) - f(x)| \le \varepsilon$. Let g be the continuous and piecewise linear function in Lemma 5.1.1 corresponding to ε and $L = n$, then, $g \in C(I)\backslash S_n$ and $\sup_{x\in I} |g(x) - f(x)| \le \varepsilon$. Hence, $C(I)\backslash S_n$ is dense in $C(I)$, and therefore S_n is nowhere dense in $C(I)$. Since $S = \bigcup_n S_n$ and each S_n is closed and nowhere dense in $C(I)$, S is of the first category. By Theorem 5.1.1, $C(I)$ is of the second category and therefore there is $f \in C(I)\backslash S$. Since S contains all functions which are somewhere differentiable on $[a, b]$, f is nowhere differentiable on $[a, b]$. ■

An interesting application of Theorem 5.1.3 is considered in Exercise 5.9.1, to show the existence of a continuous periodic function whose Fourier series diverges at a point.

5.2 The open mapping theorem

Theorem 5.2.1 (Banach open mapping theorem) *Suppose that T is a bounded linear map from a Banach space X onto a Banach space Y. Then T maps open sets into open sets.*

Proof Since $T(G + x_0) = TG + Tx_0$, it suffices to prove that if G is a neighborhood of 0 in X, then TG contains an open ball centered at 0 in Y.

Step 1. A weaker claim will be shown first. Here is the claim: Let B^X be an open ball in X centered at 0, then there is an open ball B^Y in Y centered at 0 such that $B^Y \subset \overline{TB^X}$. For the proof, the open ball in X centered at x with radius r will be denoted by $B_r^X(x)$; the connotation of $B_r^Y(y)$ as an open ball in Y is similarly defined. Let $B^X = B_r^X(0)$ and $U = B_{\frac{r}{2}}^X(0)$. Then, $X = \bigcup_{n=1}^{\infty}(nU)$, and $Y = TX = \bigcup_{n=1}^{\infty} nTU$. The Baire category theorem implies that there is n_0 such that $\overline{n_0 TU} = n_0\overline{TU}$ contains an open ball in Y and hence $\overline{TU}$ contains an open ball, say $B_\sigma^Y(\hat{y})$. Since $\hat{y} \in \overline{TU}$, there is $x_0 \in U$ such that $y_0 = Tx_0 \in B_{\frac{\sigma}{2}}^Y(\hat{y})$, and therefore $B_{\frac{\sigma}{2}}^Y(y_0) \subset B_\sigma^Y(\hat{y}) \subset \overline{TU}$. Now put $B^Y = B_{\frac{\sigma}{2}}^Y(0)$, then,

$$B^Y = B_{\frac{\sigma}{2}}^Y(y_0) - y_0 \subset \overline{TU} - Tx_0 \subset \overline{TU - Tx_0} \subset \overline{T(U+U)} \subset \overline{TB^X},$$

as is claimed.

Step 2. Let G be any open set containing 0 in X and let $B_r^X(0) \subset G$. Put $B_0^X = B_{\frac{r}{2}}^X(0)$. By Step 1, there is a ball $B_0^Y = B_\sigma^Y(0)$ in Y such that $B_0^Y \subset \overline{TB_0^X}$. It will be shown that $TB_r^X(0) \supset B_0^Y$. For this purpose, let $B_i^X = B_{\varepsilon_i}^X(0)$, $\varepsilon_i = \frac{r}{2^{i+1}}$, $i = 1, 2, \ldots$. By Step 1, there is a sequence $B_i^Y = B_{\eta_i}^Y(0)$ of balls in Y such that $\eta_i \to 0$ and $B_i^Y \subset \overline{TB_i^X}$. For $y \in B_0^Y$, there is $x_0 \in B_0^X$ such that $\|y - Tx_0\| < \eta_1$; then there is $x_1 \in B_1^X$ such that $\|y - Tx_0 - Tx_1\| < \eta_2$. Proceeding in this way, we find a sequence $\{x_i\}$ such that $x_i \in B_i^X$ and

$$\left\| y - \sum_{i=0}^{n} Tx_i \right\| = \left\| y - T\left(\sum_{i=0}^{n} x_i\right) \right\| < \eta_n,$$

$n = 1, 2, \ldots$. Now, $\|\sum_{i=m}^{m+l} x_i\| \leq \sum_{i=m}^{m+l} \varepsilon_i \to 0$ uniformly in l as $m \to \infty$, which implies that $\{\sum_{i=0}^{n} x_i\}$ is a Cauchy sequence. Set $x = \lim_{n\to\infty} \sum_{i=0}^{n} x_i$, then

$$Tx = \lim_{n\to\infty} T\left(\sum_{i=0}^{n} x_i\right) = \lim_{n\to\infty} \sum_{i=0}^{n} Tx_j = y.$$

But $\|x\| \leq \sum_{i=0}^{\infty} \|x_i\| < \sum_{i=0}^{\infty} \frac{r}{2^{i+1}} = r$, i.e. $x \in B_r^X(0)$, hence $y \in TB_r^X(0)$. ∎

Corollary 5.2.1 *If T is an injective continuous linear map from a Banach space onto a Banach space, then T^{-1} is a bounded linear map.*

Exercise 5.2.1 As a complement to Theorem 5.2.1, show that if l is a nonzero linear functional on a n.v.s. not necessarily continuous, then l maps open sets into open sets.

Exercise 5.2.2 Let X be a n.v.s. and F a closed vector subspace of X. For $x \in X$, let $[x] = x + F$.

(i) Show that $[x] = [y]$ if and only if $y \in [x]$.

(ii) Define $[x] + [y] = [x + y]$, $\lambda[x] = [\lambda x]$ (λ scalar). Show that both operations are well defined and $X/F := \{[x] : x \in X\}$ becomes a vector space under these operations.

(iii) For $[x] \in X/F$, define $\|[x]\| = \inf_{y\in[x]} \|y\|$. Show that $\|[x]\|$ is well defined and that it defines a norm on X/F.

(iv) Define $\tau : X \mapsto X/F$ by $\tau(x) = [x]$. Show that τ is a linear open mapping from X onto X/F. The map τ is called the **canonical map** from X onto X/F.

5.3 The closed graph theorem

For n.v.s.'s X and Y over the same field, a n.v.s. $X \oplus Y$, called the direct sum of X and Y, is constructed as follows. Let $X \oplus Y = \{[x, y] : x \in X,\ y \in Y\}$, on which vector space operations are defined by

$$[x_1, y_1] + [x_2, y_2] = [x_1 + x_2, y_1 + y_2]; \quad \alpha[x, y] = [\alpha x, \alpha y],$$

and a norm is defined by

$$\|[x, y]\| = \{\|x\|^2 + \|y\|^2\}^{\frac{1}{2}}.$$

This norm is so chosen, that when X and Y are inner product spaces (to be introduced later in Section 5.6), so is $X \oplus Y$.

That $X \oplus Y$ is a n.v.s. is a direct consequence of its definition. Observe that when both X and Y are Banach spaces, so is $X \oplus Y$.

Henceforth, by a **linear operator** T from a vector space X into a vector space Y, we shall mean that the domain of T, denoted $D(T)$, is a vector subspace of X, not necessarily the whole space X. Now, if both X and Y are n.v.s.'s over the same field, and if T is a linear operator from X into Y, T is called a **closed operator** if its graph $G(T) := \{[x, Tx] : x \in D(T)\}$ is closed in $X \oplus Y$; i.e. if $\{x_n\} \subset D(T)$ with $\lim_{n\to\infty} x_n = x$ and $\lim_{n\to\infty} Tx_n = y$, then $x \in D(T)$ and $Tx = y$. If T is a linear operator from X into Y and the closure of $G(T)$ in $X \oplus Y$ is the graph of a linear operator, then T is called closable.

Example 5.3.1 Let $X = Y = C[0, 1]$, $D(T) = \{f \in X : f' \in X\}$ and $Tf = f'$ for $f \in D(T)$. Then T is not bounded on $D(T)$, but T is a closed operator. That T is not bounded on $D(T)$ follows from

$$\|Tf_n\| = n\|f_{n-1}\|, \quad n = 1, 2, \ldots,$$

where $f_n(t) = t^n$, $t \in [0, 1]$. That T is closed is left as an exercise.

Exercise 5.3.1 Show that the linear operator T in Example 5.3.1 is closed.

Remark For a linear operator, its domain of definition has to be specified. For example, the differential operator T in Example 5.3.1 has to be considered as a different operator if its domain of definition $D(T)$ is changed to $D(T) = \{f \in X : f'' \in X\}$. Note that when defined on the new domain of definition, T is not closed, but closable.

Proposition 5.3.1 *If X and Y are n.v.s.'s, then a linear operator from X into Y is closable if and only if*

$$\{x_n\} \subset D(T), \ \lim_{n\to\infty} x_n = 0, \ \textit{and} \ \lim_{n\to\infty} Tx_n = y, \ \textit{then}\ y = 0. \tag{5.1}$$

Proof That (5.1) is necessary for T to be closable is obvious. To show that (5.1) is sufficient for T to be closable, let $[x, y] \in \overline{G(T)}$, i.e. there is $[x_n, Tx_n] \in G(T)$ such that $[x_n, Tx_n] \to [x, y]$. Define $Sx = y$. Because of (5.1), one verifies easily that S is well defined i.e. if $[x_n, Tx_n] \to [x, y]$ and $[x_n', Tx_n'] \to [x, y']$, then $y' = y$. Clearly, S is linear and $G(S) = \overline{G(T)}$. ■

Example 5.3.2 Let Ω be an open set in $\mathbb{R}^n$, $C_\alpha \in C^k(\Omega)$ for $\alpha = (\alpha_1, \ldots, \alpha_n)$, with $|\alpha| = \alpha_1 + \cdots + \alpha_n \leq k$. Define $D(A) = \{f \in L^2(\Omega) \cap C^k(\Omega) : Af \in L^2(\Omega)\}$, where $A = \sum_{|\alpha|\leq k} C_\alpha \partial^\alpha$. Then A is a closable linear operator from $L^2(\Omega)$ into $L^2(\Omega)$. If $\{f_j\} \subset D(A)$, $f_j \to 0$ in $L^2(\Omega)$, and $Af_j \to g$ in $L^2(\Omega)$, then for any $\varphi \in C_c^\infty(\Omega)$,

$$\begin{aligned}
\int_\Omega g(x)\varphi(x)dx &= \lim_{j\to\infty} \int_\Omega (Af_j)\varphi d\lambda^n \\
&= \lim_{j\to\infty} \int_\Omega \left(\sum_{|\alpha|\leq k} C_\alpha(x)\partial^\alpha f_j(x) \right) \varphi(x)dx \\
&= \lim_{j\to\infty} \int_\Omega \sum_{|\alpha|\leq k} (-1)^{|\alpha|}\partial^\alpha (C_\alpha(x)\varphi(x)) f_j(x)dx \\
&= \lim_{j\to\infty} \int_\Omega [A'\varphi](x) f_j(x)dx = 0,
\end{aligned}$$

which implies that $g = 0$. By Proposition 5.3.1, A is closable. Note that in the sequence of equalities above, the Fubini theorem and integration by parts have been used.

Exercise 5.3.2 Show that if T is a 1-1 closed operator, then T^{-1} is also closed.

Theorem 5.3.1 (Closed graph theorem) *A closed operator T with $D(T) = X$, a Banach space, and range in a Banach space Y, is bounded.*

Proof $G(T)$ is a closed subspace of $X \oplus Y$, and is therefore a Banach space. The linear operator $U : G(T) \mapsto X$ defined by

$$U[x, Tx] = x, \quad x \in X$$

is clearly one-to-one and continuous. Since $U(G(T)) = X$, by Corollary 5.2.1, U^{-1} is a continuous linear map from X onto $G(T)$, thus $T = VU^{-1}$ is continuous, where $V[x, Tx] = Tx$ is a continuous linear map from $G(T)$ to Y. ■

The following exercise is a comment on Theorem 5.3.1.

Exercise 5.3.3 Let X be the space of all sequences $(a_k)_{k\in\mathbb{N}}$ of real numbers such that $a_k \neq 0$ only for finitely many k's. X is a vector space under the usual way of defining addition and multiplication by scalars. For (a_k) in X, let $\|(a_k)\| = \max_k |a_k|$; then X is a n.v.s. Define $T : X \to X$ by $T(a_k) = (ka_k)$. Show that X is a closed operator on X, but is not bounded.

5.4 Separation principles

Consider a real vector space X; a subset E of X is said to be **convex** if $\alpha x + \beta y \in E$ whenever x and y are in E and α, β are nonnegative numbers with $\alpha + \beta = 1$. E is called a **convex cone** if it is convex and $\gamma E \subset E$ for all $\gamma > 0$. For a set $S \subset X$, there is a smallest convex set containing S. The smallest convex set containing S is called the **convex hull** of S and is usually denoted by Conv S, while the smallest convex cone containing S will be denoted by Con S. For $x \neq y$ in X, Conv$\{x, y\}$ is usually denoted by $[x, y]$ and is called the **line segment** with endpoints x and y, while, for $x \neq 0$ in X, Con$\{x\}$ is called the **half line** through x. In $\mathbb{R}^k$, the convex set $\Delta^{k-1} := \{x = (x_1, \ldots, x_k) : x_j \geq 0, j = 1, \ldots, k, \sum_{j=1}^{k} x_j = 1\}$ is called the **standard** $(k-1)$**-simplex**. Elements in X of the form $\sum_{j=1}^{k} \alpha_j x_j$ (k varies from element to element), where $x_1, \ldots, x_k$ are in X and $\alpha = (\alpha_1, \ldots, \alpha_k) \in \Delta^{k-1}$, are called **convex combinations** of $x_1, \ldots, x_k$; if $x_1, \ldots, x_k$ are in $S \subset X$, they are called convex combinations of elements in S.

For convenience, the fact that a real-valued function f assumes values $\geq \alpha$ on a set A will be expressed by $f(A) \geq \alpha$; the meaning of each of the expressions $f(A) > \alpha$, $f(A) \leq \alpha$, and $f(A) < \alpha$ is parallelly given.

Exercise 5.4.1 Let $S \subset X$. Prove the following statements:

(i) Conv S is the set of all convex combinations of elements in S.

(ii) Con $S = \{\sum_{j=1}^{k} \gamma_j x_j : k \in \mathbb{N},\ x_1, \ldots, x_k \in S,\ \gamma_j > 0,\ j = 1, \ldots, k\}$.

(iii) S is a convex cone if and only if $S + S \subset S$ and $\gamma S \subset S$ for all $\gamma > 0$.

A set $E \subset X$ is said to be **linearly open** if for any $x \in E$ and $y \in X$, $x + ty \in E$ if $|t|$ is small enough. Clearly, open sets in a n.v.s. X are linearly open. Note that if a linearly open convex cone contains the origin 0, then $E = X$.

Exercise 5.4.2 Show that a convex set $E \subset \mathbb{R}^n$ is linearly open if and only if E is open.

Exercise 5.4.3 Suppose that E is a convex cone in X, and S a convex set in X.

(i) Show that if $E \cap S = \emptyset$, then $E \cap (\operatorname{Con} S) = \emptyset$.

(ii) If S is also a convex cone, then $E + S$ and $E - S$ are convex cones and they are linearly open if one of E and S is linearly open.

Theorem 5.4.1 *If E is a nonempty linearly open convex cone not containing* 0*, then there is a hyperplane H such that $E \cap H = \emptyset$.*

Proof Denote by $\mathcal{F}$ the family of all vector subspaces F of X such that $F \cap E = \emptyset$. $\mathcal{F}$ is not empty, because $\{0\} \in \mathcal{F}$. Order $\mathcal{F}$ by set-inclusion i.e. $F_1 \leq F_2$ if $F_1 \subset F_2$ for F_1 and F_2 in $\mathcal{F}$. If $\mathcal{T}$ is a chain in $\mathcal{F}$, then $\bigcup_{F \in \mathcal{T}} F$ is in $\mathcal{F}$ and is an upper bound of $\mathcal{T}$. By Zorn's lemma (cf. Section 3.12), $\mathcal{F}$ has a maximal element H.

Let $D = H + E$; by Exercise 5.4.3, D is a linearly open convex cone. We claim that $X = D \cup H \cup (-D)$ is a disjoint union. It is obvious that $D \cap H = \emptyset$, and hence $(-D) \cap H = \emptyset$. If $h \in D \cap (-D)$, then both h and $-h$ are in D and consequently $h + (-h) = 0$ is in D, contradicting the fact that $D \cap H = \emptyset$. Thus $D \cup H \cup (-D)$ is a disjoint union. It remains to show that $X = D \cup H \cup (-D)$. Let $x \in X$, but $x \notin H$. Then $H + \langle x \rangle$ meets E, because H is a maximal element of $\mathcal{F}$. Then there is $h \in H$ and $\lambda \in \mathbb{R}$, $\lambda \neq 0$ such that $h + \lambda x \in E$; as a consequence $\lambda x \in H + E = D$, and then $x \in D$ or $(-D)$ depending on $\lambda > 0$ or $\lambda < 0$. This shows that $X = D \cup H \cup (-D)$.

It will be shown presently that H is a hyperplane. This amounts to showing that if $x \in X$, but $x \notin H$, then $H + \langle x \rangle = X$. Fix such an x and let $y \in X$, $y \notin H$. One has to show that $y \in H + \langle x \rangle$ to conclude the proof. For this purpose, one may assume that $x \in D$ and $y \in (-D)$. Since $[x, y]$ is connected (see Theorem 1.9.1) and $X = D \cup H \cup (-D)$,

$$[x, y] \cap \{D \cup (-D)\} \subsetneq [x, y] = [x, y] \cap \{D \cup H \cup (-D)\},$$

therefore there is $h \in H \cap [x, y]$. Since $h \in [x, y]$ there are $\alpha \geq 0$, $\beta \geq 0$ with $\alpha + \beta = 1$, such that $h = \alpha x + \beta y$. Now $h \in H$ implies that $h \notin D$, which forces β to be > 0 and hence $y = \frac{1}{\beta} h - \frac{\alpha}{\beta} x \in H + \langle x \rangle$. The proof of the theorem is complete. ■

A basic principle on separation of sets by linear functional is the following consequence of Theorem 5.4.1.

Corollary 5.4.1 *Suppose that E is a nonempty linearly open convex cone in X, and C is a nonempty convex set in X such that $C \cap E = \emptyset$, then there is $\ell \in X'$ such that $\ell(C) \geq 0$ and $\ell(E) < 0$.*

Proof Put $D = E - \operatorname{Con} C$. D is a linearly open convex cone and $0 \notin D$, because E and $\operatorname{Con} C$ are disjoint, by Exercise 5.4.3. By Theorem 5.4.1, there is a hyperplane H in

X such that $H \cap D = \emptyset$. Choose $\ell \in X'$ with $\ker \ell = H$ and $\ell(D) < 0$. Now, for $x \in \operatorname{Con} C$, $y \in E$, and $\gamma > 0$

$$\begin{cases} \ell(y) < \gamma \ell(x); \\ \gamma \ell(y) < \ell(x). \end{cases}$$

Let $\gamma \searrow 0$; it follows that $\ell(y) \leq 0$ for $y \in E$ and $\ell(x) \geq 0$ for $x \in \operatorname{Con} C$. In particular, $\ell(C) \geq 0$.

It remains to show that $\ell(y) < 0$ for $y \in E$. Choose $x_0 \in X$ with $\ell(x_0) > 0$, then $y + tx_0 \in E$ if $|t|$ is small enough, because E is linearly open. Since $y + tx_0 \in E$, $\ell(y + tx_0) \leq 0$, and hence $\ell(y) \leq -t\ell(x_0) < 0$ if $t > 0$ is small, as is to be shown. ■

Note that in the proof of Corollary 5.4.1 we have used the well-known fact in linear algebra that a vector subspace of X is a hyperplane in X if and only if it is the kernel of a nonzero linear functional on X.

A real-valued function φ defined on a convex set S in X is called a **convex** function if $\varphi(\alpha x + \beta y) \leq \alpha\varphi(x) + \beta\varphi(y)$ for any x, y in S and any convex pair (α, β). If φ is convex, then $\varphi(\sum_{j=1}^{k} \alpha_j x_j) \leq \sum_{j=1}^{k} \alpha_j \varphi(x_j)$ for any convex combination $\sum_{j=1}^{k} \alpha_j x_j$ of elements of S, as is easily seen by induction on k.

Consider now a convex function φ defined on an open interval I of $\mathbb{R}$. For $a < b < c$ in I, from $b = \frac{c-b}{c-a}a + \frac{b-a}{c-a}c$ it follows that $\varphi(b) \leq \frac{c-b}{c-a}\varphi(a) + \frac{b-a}{c-a}\varphi(c) = \varphi(a) - \frac{b-a}{c-a}\{\varphi(c) - \varphi(a)\}$, or

$$\frac{\varphi(b) - \varphi(a)}{b - a} \leq \frac{\varphi(c) - \varphi(a)}{c - a};$$

similarly,

$$\frac{\varphi(c) - \varphi(a)}{c - a} \leq \frac{\varphi(c) - \varphi(b)}{c - b}.$$

From the sequence of inequalities,

$$\frac{\varphi(b) - \varphi(a)}{b - a} \leq \frac{\varphi(c) - \varphi(a)}{c - a} \leq \frac{\varphi(c) - \varphi(b)}{c - b},$$

one infers that if $x \neq y$ are in I, the quotient $\frac{\varphi(y)-\varphi(x)}{y-x}$ is bounded for y near x and is an increasing function of y. Thus, both $\varphi'_-(x) := \lim_{y \to x-} \frac{\varphi(y)-\varphi(x)}{y-x}$ and $\varphi'_+(x) = lim_{y \to x+} \frac{\varphi(y)-\varphi(x)}{y-x}$ exist and are finite; furthermore, $\varphi'_-(x) \leq \varphi'_+(x)$ and $\varphi'_+(x) \leq \varphi'_-(y)$ if $x < y$ are in I. The last inequality follows from $\varphi'_+(x) \leq \frac{\varphi(z)-\varphi(x)}{z-x} \leq \frac{\varphi(y)-\varphi(z)}{y-z}$ for z strictly between x and y, by letting $z \to y$. Since the left and right derivatives of φ exist and are finite at each point of I, φ is continuous on I. Now, for $x < y$ in I, the inequalities $\varphi'_-(x) \leq \varphi'_+(x) \leq \varphi'_-(y) \leq \varphi'_+(y)$ imply that both φ'_- and φ'_+ are monotone increasing functions on I. Next, for $x < y < z$ in I, one verifies that $\varphi'_+(x) \leq \varphi'_+(y) \leq \frac{\varphi(z)-\varphi(y)}{z-y}$, from which $\varphi'_+(x) \leq \varphi'_+(x+) \leq \frac{\varphi(z)-\varphi(x)}{z-x}$ follows when $y \to x$ (note that φ is continuous); then one concludes that $\varphi'_+(x) = \varphi'_+(x+)$, by letting $z \to x$. Thus φ'_+ is a right-continuous function; similarly, one can verify that φ'_- is a left-continuous function. The following proposition has been proved.

Proposition 5.4.1 *Suppose that φ is a convex function defined on an open interval I in $\mathbb{R}$. The following statements hold:*

(i) *The left derivative $\varphi'_-(x)$ and the right derivative $\varphi'_+(x)$ exist and are finite at each point x of I; and for $x < y$ in I, $\varphi'_-(x) \leq \varphi'_+(x) \leq \varphi'_-(y)$.*
(ii) *Both φ'_- and φ'_+ are monotone increasing.*
(iii) *φ'_- is left-continuous and φ'_+ is right-continuous.*
(iv) *For $x \in I$ and $m \in [\varphi'_-(x), \varphi'_+(x)]$, $\varphi(y) \geq \varphi(x) + m(y-x)$ for all $y \in I$.*

Exercise 5.4.4 Show that if φ is a convex function on a vector space X, then, for any $t \in \mathbb{R}$, the set $\{\varphi \leq t\}$ is convex and the set $\{\varphi < t\}$ is convex and linearly open.

A real-valued function q on a real vector space X is called a **sublinear functional** on X if

(1) $q(x+y) \leq q(x) + q(y)$, x, y in X;
(2) $q(\lambda x) = \lambda q(x)$, $x \in X$, $\lambda > 0$.

Note that a sublinear functional is necessarily convex.

Exercise 5.4.5 Suppose that q is a sublinear functional on X, and put $Q = \{q < 0\}$. Show that Q is a linearly open convex cone. Also show that $q(0) = 0$ and $-q(-x) \leq q(x)$ for $x \in X$.

Exercise 5.4.6 Suppose that q is the sublinear functional on $\mathbb{R}^n$ defined by $q(x) = \max_{1 \leq j \leq n} x_j$ if $x = (x_1, \ldots, x_n)$. Show that a linear functional on $\mathbb{R}^n$ satisfies $l \leq q$ if and only if there is $\alpha \in \Delta^{n-1}$ such that $l(x) = \sum_{j=1}^n \alpha_j x_j$.

Lemma 5.4.1 *Suppose that q is a sublinear functional on X with $Q = \{q < 0\} \neq \emptyset$. Let τ be a map from a set T into X. Then there is $\ell \in X'$, $\ell \neq 0$, with $\ell \leq q$ such that $\ell(\tau(T)) \geq 0$ if and only if $q(\operatorname{Con} \tau(T)) \geq 0$.*

Proof Suppose $q(\operatorname{Con} \tau(T)) \geq 0$. Then $(\operatorname{Con} \tau(T)) \cap Q = \emptyset$. By Corollary 5.4.1, there is $\hat{\ell} \in X'$, $\hat{\ell} \neq 0$, such that $\hat{\ell}(\operatorname{Con} \tau(T)) \geq 0$ and $\hat{\ell}(Q) < 0$. It will be shown presently that there is $\sigma > 0$ such that $\sigma \hat{\ell} \leq q$.

Define a map f from X into $\mathbb{R}^2$ by

$$f(x) = (q(x), -\hat{\ell}(x)), \quad x \in X,$$

and let C be the convex hull of $f(X)$; then $C \cap \mathring{\mathbb{R}}^2_- = \emptyset$, where $\mathring{\mathbb{R}}^2_- = \{(r_1, r_2) \in \mathbb{R}^2 : r_1 < 0, r_2 < 0\}$. Actually, if $v \in C$, there are $x_1, \ldots, x_k$ in X and

$\alpha = (\alpha_1, \ldots, \alpha_k) \in \Delta^{k-1}$ such that $v = (\sum_{j=1}^k \alpha_j q(x_j), -\hat{\ell}(\sum_{j=1}^k \alpha_j x_j))$; if $\sum_{j=1}^k \alpha_j q(x_j) < 0$, then $q(\sum_{j=1}^k \alpha_j x_j) \leq \sum_{j=1}^k \alpha_j q(x_j) < 0$, implying that $\sum_{j=1}^k \alpha_j x_j \in Q$ and hence $-\hat{\ell}(\sum_{j=1}^k \alpha_j x_j) > 0$; thus $v \notin \mathring{\mathbb{R}}^2_-$. By Corollary 5.4.1, there is (α_1, α_2) in $\mathbb{R}^2$ with $\alpha_1^2 + \alpha_2^2 > 0$ such that

$$\begin{cases} \alpha_1 r_1 + \alpha_2 r_2 < 0 & \text{for } (r_1, r_2) \in \mathring{\mathbb{R}}^2_-; \\ \alpha_1 q(x) - \alpha_2 \hat{\ell}(x) \geq 0 & \text{for } x \in X. \end{cases} \tag{5.2}$$

The first inequality in (5.2) shows that $\alpha_1 \geq 0$, $\alpha_2 \geq 0$, while the second inequality shows that $\alpha_1 > 0$ and $\alpha_2 > 0$, in that $Q \neq \emptyset$. Then, $\sigma \hat{\ell}(x) \leq q(x)$ for $x \in X$ by taking $\sigma = \alpha_1^{-1}\alpha_2$. Then $\ell := \sigma \hat{\ell}$ satisfies $\ell \leq q$, $\ell \neq 0$, and $\ell(x) \geq 0$ for $x \in \tau(T)$. The other direction of the Lemma is obvious. ∎

Remark

(i) Since q is sublinear, the condition $q(\mathrm{Con}\,\tau(T)) \geq 0$ in Lemma 5.4.1 is equivalent to $q(\mathrm{Conv}\,\tau(T)) \geq 0$;

(ii) since $Q \neq \emptyset$, $\ell \neq 0$ is a consequence of $\ell \leq q$;

(iii) when $Q = \emptyset$, Lemma 5.4.1 also holds if we do not require that $\ell \neq 0$, because in this case $q(\mathrm{Con}\,\tau(T)) \geq 0$ always holds and ℓ is simply taken to be the zero functional.

It follows from the preceding remarks that Lemma 5.4.1 can be generalized to the following theorem.

Theorem 5.4.2 *Suppose that q is a sublinear functional on a real vector space X and τ a map from a set T into X. Then there is $\ell \in X'$ with $\ell \leq q$ such that $\ell(\tau(T)) \geq 0$ if and only if $q(\mathrm{Con}\,\tau(T)) \geq 0$.*

An immediate consequence of Theorem 5.4.2 is the following historically interesting result of Banach.

Corollary 5.4.2 (Banach) *If q is a sublinear functional on X, then there is $\ell \in X'$ such that $\ell \leq q$ on X.*

Proof In Theorem 4.5.2, take $\tau(t)$ to be the zero element of X for each $t \in T$. ∎

If, for a real vector space X and a sublinear functional q on X, we let $X'(q)$ be the set of all those $\ell \in X'$ such that $\ell \leq q$, then $X'(q)$ is obviously convex, and is nonempty, by Corollary 5.4.2.

From Theorem 5.4.2, there follow two important consequences.

Theorem 5.4.3 (Hahn–Banach) *Let q be a sublinear functional on a real vector space X and suppose that Y is a vector subspace of X and $\ell \in Y'(q)$. Then there is $\hat{\ell} \in X'(q)$ such that $\ell(y) = \hat{\ell}(y)$ for $y \in Y$.*

Proof Define a sublinear functional $\hat{q}$ on $X \oplus Y$ by

$$\hat{q}(x,y) = q(x) + \ell(y), \quad x \in X,\ y \in Y,$$

and a map $\hat{\tau}$ from Y into $X \oplus Y$ by

$$\hat{\tau}(y) = (y, -y), \quad y \in Y.$$

Since $\hat{\tau}$ is linear, $\operatorname{Conv}\hat{\tau}(Y) = \hat{\tau}(Y)$. Now let $v \in \operatorname{Conv}\hat{\tau}(Y) = \hat{\tau}(Y)$. Then, $v = (y, -y)$ for some $y \in Y$ and $\hat{q}(v) = q(y) + \ell(-y) \geq 0$; this means that $\hat{q}(\operatorname{Conv}\hat{\tau}(Y)) \geq 0$. By Theorem 5.4.2, there is $(\hat{\ell}, \ell_Y) \in (X \oplus Y)'$ with $(\hat{\ell}, \ell_Y) \leq \hat{q}$ such that $(\hat{\ell}, \ell_Y)(y, -y) \geq 0$ for all $y \in$ Y, where $\hat{\ell} \in X'$ and $\ell_Y \in Y'$. But $(\hat{\ell}, \ell_Y) \leq \hat{q}$ if and only if $\hat{\ell} \leq q$ on X and $\ell_Y \leq \ell$ on Y. Now, $\ell_Y \leq \ell$ implies that $\ell_Y = \ell$ and $(\hat{\ell}, \ell_Y)(y, -y) = \hat{\ell}(y) - \ell_Y(y) = \hat{\ell}(y) - \ell(y) \geq 0$ for $y \in Y$ forces $\hat{\ell}(y) = \ell(y)$ for $y \in Y$. ■

Theorem 5.4.4 (Mazur–Orlicz) *Let q be a sublinear functional on a real vector space X and τ a map from a set T into X. Suppose that θ is a map from T into $\mathbb{R}$. Then there is $\ell \in X'(q)$ such that $\theta(t) \leq \ell(\tau(t))$ for all $t \in T$ if and only if for every positive integer n,*

$$\sum_{j=1}^{n} \alpha_j \theta(t_j) \leq q\left(\sum_{j=1}^{n} \alpha_j \tau(t_j)\right) \tag{5.3}$$

for all $t_1, \ldots, t_n$ in T and $\alpha = (\alpha_1, \ldots, \alpha_n) \in \Delta^{n-1}$.

Proof Consider $\hat{X} = X \oplus \mathbb{R}$. Define $\hat{q} : \hat{X} \mapsto \mathbb{R}$ by

$$\hat{q}(x, \lambda) = q(x) + \lambda, \quad x \in X,\ \lambda \in \mathbb{R},$$

then $\hat{q}$ is a sublinear functional on $\hat{X}$. Let now

$$\hat{\tau}(t) = (\tau(t), -\theta(t)), \quad t \in T.$$

Suppose that (5.3) holds, then for $n \in \mathbb{N}$, $t_1, \ldots, t_n$ in T and $\alpha = (\alpha_1, \ldots, \alpha_n) \in \Delta^{n-1}$,

$$\begin{aligned} \hat{q}\left(\sum_{j=1}^{n} \alpha_j \tau(t_j), -\sum_{j=1}^{n} \alpha_j \theta(t_j)\right) &= \hat{q}\left(\sum_{j=1}^{n} \alpha_j \hat{\tau}(t_j)\right) \\ &= q\left(\sum_{j=1}^{n} \alpha_j \tau(t_j)\right) - \sum_{j=1}^{n} \alpha_j \theta(t_j) \geq 0, \end{aligned}$$

or $\hat{q}(\operatorname{Conv}\hat{\tau}(T)) \geq 0$. By Theorem 5.4.2, there is $\hat{\ell} \in \hat{X}'$ with $\hat{\ell} \leq \hat{q}$ on $\hat{X}$ such that $\hat{\ell}(\hat{\tau}(T)) \geq 0$. But $\hat{\ell} = (\ell, \alpha)$, $\ell \in X'$, $\alpha \in \mathbb{R}$, and $\hat{\ell}(x, \lambda) = \ell(x) + \alpha\lambda$ for $x \in X$ and $\lambda \in \mathbb{R}$. Observe then that $\hat{\ell} \leq \hat{q}$ on $\hat{X}$ means that $\ell \leq q$ on X and $\alpha = 1$; hence, $\hat{\ell}(\hat{\tau}(t)) \geq 0$ for $t \in T$ implies that $\theta(t) \leq \ell(\tau(t))$ for $t \in T$. On the other hand, if there is $\ell \in X'$ with $\ell \leq q$ and $\ell(\tau(t)) \geq \theta(t)$ for $t \in T$, then (5.3) obviously holds. ■

Corollary 5.4.3 *Let X, q, and τ be as in Theorem 5.4.4, then*

$$\max_{\ell \in X'(q)} \inf \ell(\tau(T)) = \inf q(\operatorname{Conv} \tau(T)).$$

Proof Observe firstly that $\inf \ell(\tau(T)) \leq \inf q(\operatorname{Conv} \tau(T))$ holds for any $\ell \in X'(q)$, hence $\sup_{\ell \in X'(q)} \inf \ell(\tau(T)) \leq \inf q(\operatorname{Conv} \tau(T))$, and it remains to show that there is $\ell \in X'(q)$ such that $\inf \ell(\tau(T)) = \inf q(\operatorname{Conv} \tau(T))$. In the case where $\inf q(\operatorname{Conv} \tau(T)) = -\infty$, just take any $\ell \in X'(q)$ (recall that $X'(q) \neq \emptyset$, by Corollary 5.4.2). If $\inf q(\operatorname{Conv} \tau(T)) = \beta > -\infty$, let a function θ on T be defined by $\theta(t) = \beta$ for all $t \in T$. Then (5.3) holds trivially and we may apply Theorem 5.4.4 to find $\ell \in X'(q)$ such that $\beta \leq \ell(\tau(t))$ for all $t \in T$, i.e.

$$\inf \ell(\tau(T)) \geq \beta = \inf q(\operatorname{Conv} \tau(T)).$$

But, as we observed at the beginning of the proof, $\inf \ell(\tau(T)) \leq \inf q(\operatorname{Conv} \tau(T))$, therefore $\inf \ell(\tau(T)) = \inf q(\operatorname{Conv} \tau(T))$ and the proof is complete. ■

Exercise 5.4.7 Show that if C is a convex set in a real n.v.s. X, such that $\inf_{x \in C} \|x\| = \sigma > 0$, then there is $l \in X^*$ with $\|l\| = 1$ such that $l(x) \geq \sigma$ for all $x \in C$. (Hint: apply Corollary 5.4.3.)

The conclusion of Corollary 5.4.3 is a general form of J. von Neumann's minimax theorem in game theory, as illustrated in Exercise 5.4.8.

Exercise 5.4.8 (von Neumann minimax theorem) Suppose that (a_{ij}), $1 \leq i \leq m$, $1 \leq j \leq n$, is a given $m \times n$-matrix with real entries. For each $j = 1, \ldots, n$, define a function f_j on Δ^{m-1} by

$$f_j(\alpha) = \sum_{i=1}^{m} a_{ij}\alpha_i, \quad \alpha = (\alpha_1, \ldots, \alpha_n) \in \Delta^{m-1},$$

and define a quadratic form A on $\Delta^{m-1} \times \Delta^{n-1}$ by

$$A(\alpha, \beta) = \sum_{j=1}^{n} \beta_j f_j(\alpha) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}\alpha_i\beta_j.$$

Now consider the sublinear functional q on $\mathbb{R}^n$, defined by $q(x) = \max_{1 \leq j \leq n} x_j$ for $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, and let the map τ from Δ^{m-1} to $\mathbb{R}^n$ be defined by $\tau(\alpha) = (f_1(\alpha), \ldots, f_n(\alpha))$. Use Corollary 5.4.3 and the assertion of Exercise 5.4.6 to show the following minimax equality of von Neumann:

$$\min_{\alpha \in \Delta^{m-1}} \max_{\beta \in \Delta^{n-1}} A(\alpha, \beta) = \max_{\beta \in \Delta^{n-1}} \min_{\alpha \in \Delta^{m-1}} A(\alpha, \beta).$$

Exercise 5.4.9 Let q be a sublinear functional on a real vector space X and put $Q = \{x \in X : q(x) < 0\}$. Suppose that S is a convex cone in X such that $Q \cap S = \emptyset$, and define $\hat{q}$ on X by $\hat{q}(x) = \inf_{y \in S} q(x + y)$.

(i) Show that $\hat{q}$ is a sublinear functional on X and $\hat{q} \leq q$.

(ii) Show that if $\ell \in X'(\hat{q})$, then $\ell(x) \geq 0$ for $x \in S$.

Exercise 5.4.10 Show that Theorem 5.4.2 is a consequence of Corollary 5.4.2. (Hint: apply Corollary 5.4.2 with q replaced by $\hat{q}$, as defined in Exercise 5.4.9, with $S = \mathrm{Con}(\tau(T))$.)

Exercise 5.4.11 Show that Theorem 5.4.2, Theorem 5.4.3, and Theorem 5.4.4 are equivalent to each other. (Hint: Corollary 5.4.2 is a special case of the Hahn–Banach theorem.)

Exercise 5.4.12 Let Q be a proper linearly open convex cone in a real vector space X. Fix $x_0 \in Q$.

(i) Show that the family $L = \{\ell \in X' : \ell < 0 \text{ on } Q \text{ and } \ell(x_0) = -1\}$ is nonempty and that for $x \in X$, $\sup_{\ell \in L} \ell(x)$ is finite. (Hint: for $x \in X$ there is $\sigma > 0$ such that $x_0 + \sigma x \in Q$, from which assert that $\ell(x) \leq \frac{1}{\sigma}$ for $\ell \in L$.)

(ii) Put $q(x) = \sup_{\ell \in L} \ell(x)$ for $x \in X$. Show that q is a sublinear functional on X and that $Q = \{x \in X : q(x) < 0\}$.

In this final part of the section our discussion is restricted to real normed vector spaces; and our concern is the separation of convex sets by closed affine hyperplane. By an **affine hyperplane** we mean a translation of a hyperplane in a vector space, i.e. an affine hyperplane in a vector space X is a set of the form $x + H$, where $x \in X$ and H is a hyperplane in X. We recall from elementary linear algebra that a vector subspace of a vector space X is a hyperplane if and only if it is the kernel of a nonzero linear functional on X. Note that if ℓ_1, ℓ_2 are nonzero linear functionals on X, then $\ker \ell_1 = \ker \ell_2$ if and only if $\ell_1 = \alpha \ell_2$ for some nonzero scalar α. Thus an affine hyperplane in X is a set of the form $\{x \in X : \ell(x) = \alpha\}$ for some $\ell \in X'$ $(\ell \neq 0)$ and some scalar α. If X is a normed vector space, then, since the closure of a vector subspace of X is a vector subspace of X, every hyperplane in X is either closed or dense in X. Observe that a hyperplane $H = \ker \ell$, $\ell \in X'$, in a normed vector space X is closed if and only if $\ell \in X^*$, and hence a closed affine hyperplane in X is of the form $\{x \in X : \ell(x) = \alpha\}$ for some $\ell \in X^*$ $(\ell \neq 0)$ and some scalar α.

We fix now a real n.v.s. X. Nonempty sets A and B in X are said to be separated strictly by a closed affine hyperplane if there are $\ell \in X^*$ and $\alpha \in \mathbb{R}$ such that $\ell(x) < \alpha$ for $x \in A$ and $\ell(y) > \alpha$ for $y \in B$; while they are separated strictly in the strong sense if there are $\ell \in X^*, \alpha \in \mathbb{R}$, and $\varepsilon > 0$ such that $\ell(x) \leq \alpha - \varepsilon$ for $x \in A$ and $\ell(y) \geq \alpha + \varepsilon$ for $y \in B$. Note that $\ell \in X^*$ in the above definition is necessarily nonzero, and $\{x \in X : \ell(x) = \alpha\}$ is the closed affine hyperplane in question. A closed set of the form $\{x \in X : \ell(x) \leq \alpha\}$, where $\ell \in X^*$ and $\alpha \in \mathbb{R}$, is called a closed half-space in X.

Lemma 5.4.2 *Let G be a nonempty open convex set in X not containing* 0. *Then there is $\ell \in X^*$ such that $\ell(x) < 0$ for $x \in G$.*

Proof Put $E = \bigcup_{\lambda > 0} \lambda G$. Clearly E is a nonempty open convex cone not containing 0, and we infer from Corollary 5.4.1 by taking $C = \{0\}$ that there is $\ell \in X'$ such that $\ell(x) < 0$ for $x \in E$ (and hence for $x \in G$). Since G is disjoint with the hyperplane $H := \ker \ell$, H cannot be dense in X and therefore is closed. Consequently $\ell \in X^*$. ■

Theorem 5.4.5 *Any two nonempty disjoint open convex sets A and B in X can be separated strictly by a closed affine hyperplane.*

Proof Let $G = A - B$. G is a nonempty open convex set in X not containing 0; we infer then from Lemma 5.4.2 that there is $\ell \in X^*$ such that $\ell(x - y) < 0$ for $x \in A$ and $y \in B$, and hence $\ell(A)$ is bounded above and $\ell(B)$ is bounded below. Observe that $\ell(A)$ and $\ell(B)$ are open intervals. Let $a = \sup \ell(A)$ and $b = \inf \ell(B)$; then $a \leq b$. Choose $\alpha \in [a, b]$, then $f(x) < \alpha$ for $x \in A$ and $\ell(y) > \alpha$ for $y \in B$. Thus A and B are separated strictly by the closed affine hyperplane $\{x \in X : \ell(x) = \alpha\}$. ■

Theorem 5.4.6 *Suppose that A and B are disjoint closed convex sets in X, one of which is compact. Then there is a closed affine hyperplane which separates A and B strictly in the strong sense.*

Proof We may assume that B is compact and let $G = X \backslash A$. Then G is an open set containing B. For $x \in B$, choose $r_x > 0$ such that $x + B_{r_x}(0) \subset G$. The family $\{x + B_{\frac{1}{2}r_x}(0)\}_{x \in B}$ is an open covering of B, hence there are $x_1, \ldots, x_k$ in B such that $B \subset \bigcup_{j=1}^{k}\{x_j + B_{\frac{1}{2}r_{x_j}}(0)\}$. Let $r = \min_{1 \leq j \leq k} \frac{1}{2} r_{x_j} > 0$, then $B + B_r(0) \subset G$. Therefore $\{B + B_r(0)\} \cap A = \emptyset$, and consequently

$$\left\{B + B_{\frac{1}{2}r}(0)\right\} \cap \left\{A + B_{\frac{1}{2}r}(0)\right\} = \emptyset.$$

We infer then from Theorem 5.4.5 that there are $\ell \in X^*$ and $\alpha \in \mathbb{R}$, such that

$$\ell(x + z) < \alpha, \; x \in A, \; z \in B_{\frac{1}{2}r}(0);$$
$$\ell(y + z) > \alpha, \; y \in B, \; z \in B_{\frac{1}{2}r}(0).$$

Now put $\varepsilon = \sup\{|\ell(z)| : z \in B_{\frac{1}{2}r}(0)\}$. Then, by choosing sequences $\{z_k'\}$ and $\{z_k''\}$ in $B_{\frac{1}{2}r}(0)$ such that $\ell(z_k') \to \varepsilon$ and $\ell(z_k'') \to -\varepsilon$, we conclude from $\ell(x) < \alpha - \ell(z_k')$ for $x \in A$ by letting $k \to \infty$ that $\ell(x) \leq \alpha - \varepsilon$; and conclude from $\ell(y) \geq \alpha - \ell(z_k'')$ for $y \in B$ by letting $k \to \infty$ that $\ell(y) \geq \alpha + \varepsilon$. ■

Exercise 5.4.13 Show that a set K in a real n.v.s. X is closed convex if and only if K is the intersection of a family of closed half-spaces in X.

Remark Since a complex vector space is also a real vector space, sublinear functionals are also defined on complex vector spaces. This fact is often used without being noted explicitly.

5.5 Complex form of Hahn–Banach theorem

Let X be a vector space. A **semi-norm** on X is a sublinear functional q on X such that $q(\alpha x) = |\alpha|q(x)$ for $x \in X$ and for scalar α (cf. Remark at the end of Section 5.4). Note that a semi-norm is nonnegative, because if $q(x) < 0$ for some x, then $0 = q(0) = q(x + (-x)) \leq q(x) + q(-x) = 2q(x) < 0$, which is absurd.

Theorem 5.5.1 *Let X be a vector space and q a semi-norm on X. Suppose that ℓ is a linear functional on a vector subspace Y of X such that $|\ell| \leq q$ on Y, then there is $\hat{\ell} \in X'$ with $|\hat{\ell}| \leq q$ on X such that $\hat{\ell}(y) = \ell(y)$ for $y \in Y$.*

Proof If X is a real vector space, then the theorem is a consequence of Theorem 5.4.3, as is easily verified. So we assume that X is a complex vector space. Write

$$\ell(y) = \ell_1(y) + i\ell_2(y), \quad y \in Y,$$

where $\ell_1(y) = \operatorname{Re} \ell(y)$ and $\ell_2(y) = \operatorname{Im} \ell(g)$. Then ℓ_1 and ℓ_2 are real linear functionals on Y. Since $i\ell(y) = \ell(iy)$, it follows that $\ell_2(y) = -\ell_1(iy)$, i.e.

$$\ell(y) = \ell_1(y) - i\ell_1(iy), \quad y \in Y.$$

Obviously, $|\ell_1| \leq q$ on Y. Hence there is a real linear functional $\hat{\ell}_1$ on X extending ℓ_1 such that $|\hat{\ell}_1(x)| \leq q(x)$ for $x \in X$.

Define $\hat{\ell}$ on X by

$$\hat{\ell}(x) = \hat{\ell}_1(x) - i\hat{\ell}_1(ix), \quad x \in X.$$

One can see that $\hat{\ell}$ is a linear functional on X and $\hat{\ell}$ extends ℓ. It remains only to show that $|\hat{\ell}(x)| \leq q(x)$ for $x \in X$. For any $x \in X$, there is $\beta \in \mathbb{C}$ with $|\beta| = 1$ such that $|\hat{\ell}(x)| = \beta\hat{\ell}(x)$, then,

$$\begin{aligned} |\hat{\ell}(x)| &= \beta\hat{\ell}(x) = \hat{\ell}(\beta x) = \hat{\ell}_1(\beta x) - i\hat{\ell}_1(i\beta x) \\ &= \hat{\ell}_1(\beta x) \leq q(\beta x) = |\beta|q(x) = q(x). \end{aligned}$$

■

Some relevant consequences of Theorem 5.5.1 are now considered.

Corollary 5.5.1 *Let X be a normed vector space, then for any $x_0 \in X$, there is $\ell \in X^*$, with $\|\ell\| = 1$ such that $\ell(x_0) = \|x_0\|$.*

Proof Suppose first that $x_0 \neq 0$, and let $Y = \langle\{x_0\}\rangle$ be the vector subspace of X spanned by $\{x_0\}$. Define a linear functional ℓ_1 on Y by

$$\ell_1(\alpha x_0) = \alpha\|x_0\|,$$

then $|\ell_1(\alpha x_0)| = \|\alpha x_0\|$, implying $\|\ell_1\|_{Y^*} = 1$. By Theorem 5.5.1 with q being the norm on X, there is $\ell \in X'$ extending ℓ_1 such that $|\ell(x)| \le \|x\|$. Then, $\ell(x_0) = \ell_1(x_0) = \|x_0\|$ and $\|\ell\| = 1$.

Now if $x_0 = 0$, simply take ℓ to be any $\ell \in X'$ with $\|\ell\| = 1$ (note that the first part of the proof shows that there is $\ell \in X'$ with $\|\ell\| = 1$). ■

Corollary 5.5.2 *Let X be any normed vector space. Then for any x and y in X, $x \neq y$, there is $\ell \in X^*$ such that $\ell(x) \neq \ell(y)$. i.e. X^* **separates points** of X.*

Proof Let $x_0 = x - y$. By Corollary 5.5.1, there is $\ell \in X^*$ with $\|\ell\| = 1$ such that $\ell(x_0) = \|x_0\| = \|x - y\|$. But,

$$|\ell(x) - \ell(y)| = |\ell(x - y)| = |\ell(x_0)| = \|x_0\| > 0.$$ ■

Exercise 5.5.1 Show that if $x_0 \in X$ and $x_0 \neq 0$, then there is $\ell \in X^*$ with $\|\ell\| = \|x_0\|$ and $\ell(x_0) = \|x_0\|^2$.

Exercise 5.5.2 Let $X = L^1[0, 1]$ and $Y = C[0, 1]$. Choose $x_0 \in (0, 1)$ and let $\ell(f) = f(x_0)$ for $f \in Y$. Is it possible to extend ℓ to a bounded linear functional on X?

For a normed vector space X, define a function $\langle \cdot, \cdot \rangle$ on $X \times X^*$ by

$$\langle x, x^* \rangle = x^*(x), \quad (x, x^*) \in X \times X^*,$$

$\langle \cdot, \cdot \rangle$ is called the natural pairing between X and X^*.

For $x \in X$, let $j(x) \in X^{**} := (X^*)^*$ be defined by

$$\langle x^*, j(x) \rangle = \langle x, x^* \rangle, \quad x^* \in X^*.$$

The mapping j is a linear map from X into X^{**}, and since X^* separates points of X, it is one-to-one; furthermore it is an **isometry** in the sense that $\|j(x)\| = \|x\|$ for all $x \in X$.

Theorem 5.5.2 *The mapping j is a linear isometry from X into X^{**}.*

Proof It is left only to show that $\|j(x)\| = \|x\|$, where $\|j(x)\|$ is the norm of $j(x)$ in X^{**}. From

$$\begin{aligned}\|j(x)\| &= \sup_{\substack{x^* \in X^* \\ \|x^*\|=1}} |\langle x^*, j(x) \rangle| = \sup_{\substack{x^* \in X^* \\ \|x^*\|=1}} |\langle x, x^* \rangle| \\ &\le \sup_{\substack{x^* \in X^* \\ \|x^*\|=1}} \|x\| \|x^*\| = \|x\|,\end{aligned}$$

it follows that $\|j(x)\| \le \|x\|$. On the other hand, by Corollary 5.5.1, there is $x^* \in X^*$ with $\|x^*\| = 1$ such that $\langle x, x^* \rangle = \|x\|$, hence $\|j(x)\| \ge \|x\|$. Thus, $\|j(x)\| = \|x\|$. ■

Because of Theorem 5.5.2 we shall consider X as embedded in X^{**} as a normed vector subspace through the mapping j. If $X = X^{**}$, then X is called a **reflexive** space. A reflexive normed vector space is necessarily a Banach space. In general, the closure of X in X^{**} is a Banach space, which is called the **completion** of X. Note that if x is in the completion of a n.v.s. X, then there is a Cauchy sequence $\{x_n\}$ in $X \subset X^{**}$ such that $x_n \to x$ in X^{**}.

Example 5.5.1 Let $X = L^\infty[-1, 1]$ and $Y = C[-1, 1]$, and let $\delta \in Y^*$ be defined by $\delta(f) = f(0)$ for $f \in Y$. Since δ is a bounded linear functional with norm 1 on Y, it can be extended to be a bounded linear functional on X with the same norm by the Hahn–Banach theorem; we also denote the extended functional by δ, i.e., $\delta \in L^\infty[-1, 1]^*$. It will be shown in Chapter 6 that $L^1[-1, 1]^* = L^\infty[-1, 1]$, in the sense that for $\ell \in L^1[-1, 1]^*$ there is $h \in L^\infty[-1, 1]$ such that $\ell(f) = \int_{[-1,1]} fh d\lambda$ for all $f \in L^1[-1, 1]$. We know from this fact that $\delta \in L^1[-1, 1]^{**}$. But there is no $h \in L^1[-1, 1]$ such that $\delta(f) = \int_{[-1,1]} fh d\lambda = f(0)$ for $f \in C[-1, 1]$; this means that $\delta \notin L^1[-1, 1]$, i.e., $L^1[-1, 1] \subsetneq L^1[-1, 1]^{**}$.

Exercise 5.5.3 Suppose that Y is a vector subspace of a n.v.s. X such that $\overline{Y} \neq X$, and let $Y^\perp = \{x^* \in X^* : \langle y, x^* \rangle = 0 \text{ for all } y \in Y\}$.

(i) For $x \in X \backslash \overline{Y}$, show that there is $x^* \in Y^\perp$ such that $\|x^*\| = 1$ and $\langle x, x^* \rangle = \inf_{y \in Y} \|x - y\|$. (Hint: define $l \in (\langle \{x\} \rangle + Y)^*$ by $l(\alpha x + y) = \alpha \inf_{y \in Y} \|x - y\|$ for scalar α and $y \in Y$, then extend l to be defined on X by the Hahn–Banach theorem.)

(ii) For $x \in X$, show that

$$\inf_{y \in Y} \|x - y\| = \max_{\substack{x^* \in Y^\perp \\ \|x^*\| \leq 1}} |\langle x, x^* \rangle| = \max_{\substack{x^* \in Y^\perp \\ \|x^*\| = 1}} |\langle x, x^* \rangle|.$$

Exercise 5.5.4 Let F be a closed vector subspace in a real n.v.s. X and let τ be the canonical map from X onto X/F.

(i) Suppose now that C is an open convex set with $C \cap F = \emptyset$. Show that $\tau(C)$ is an open convex set in X/F, not containing $[0]$.

(ii) Suppose that Y is a vector subspace of X and C an open convex set in X, such that $C \cap Y = \emptyset$; show that there is a closed hyperplane H such that $H \supset Y$ and $H \cap C = \emptyset$. (Hint: use Theorem 5.4.1 in $X/\overline{Y}$ and note that a hyperplane in a n.v.s. X is either closed or dense in X.)

5.6 Hilbert space

Let E be a vector space. For definiteness, it will be assumed that E is a complex space throughout this section. The case of E being a real vector space can be treated similarly.

E is called an **inner product space** if there is a map $(\cdot, \cdot) : E \times E \to \mathbb{C}$ satisfying the following conditions:

(i) $(x, x) \geq 0 \ \forall x \in E$, and $(x, x) = 0$ if and only if $x = 0$;

(ii) $(\cdot, x)$ is linear on E for each $x \in E$; and

(iii) $(x, y) = \overline{(y, x)}$ for all x, y in E (for $z \in \mathbb{C}$, $\bar{z}$ is the conjugate of z).

The map $(\cdot, \cdot)$ is called an **inner product** on E. We always consider a vector subspace F of an inner product space E as an inner product space, with the inner product inherited from that on E, i.e. the inner product on F is the restriction to $F \times F$ of that on E. Note that when E is a real vector space, condition (iii) is replaced by $(x, y) = (y, x)$. If E is an inner product space, put $\|x\| = (x, x)^{1/2}$ for $x \in E$.

Theorem 5.6.1 *If E is an inner product space, then for x, y in E, the following hold:*

(a) $\|x - y\|^2 + \|x + y\|^2 = 2(\|x\|^2 + \|y\|^2)$ (***Parallelogram identity***);

(b) $|(x, y)| \leq \|x\| \cdot \|y\|$ (***Schwarz inequality***); *and*

(c) $\|x + y\| \leq \|x\| + \|y\|$ (***Triangle inequality***).

Proof For x and y in E,

$$\|x - y\|^2 = (x - y, x - y) = \|x\|^2 - 2\,\mathrm{Re}(x, y) + \|y\|^2; \tag{5.4}$$

$$\|x + y\|^2 = (x + y, x + y) = \|x\|^2 + 2\,\mathrm{Re}(x, y) + \|y\|^2. \tag{5.5}$$

(a) follows by adding (5.4) and (5.5).

To show (b), it is sufficient to show that $|(x, y)| \leq 1$ whenever $\|x\| = \|y\| = 1$. Now if $\|x\| = \|y\| = 1$, $|\mathrm{Re}(x, y)| \leq 1$ follows from (5.4) or (5.5) according to whether $\mathrm{Re}(x, y) \geq 0$ or $R(x, y) < 0$, because the far left sides of (5.4) and (5.5) are both greater than or equal to zero. If $\theta \in \mathbb{C}$ with $|\theta| = 1$ is chosen so that $(x, \theta y) = |(x, y)|$, then

$$|(x, y)| = (x, \theta y) = \mathrm{Re}(x, \theta y) \leq 1,$$

this concludes (b). Finally,

$$\begin{aligned} \|x + y\|^2 &= (x + y, x + y) = \|x\|^2 + 2\,\mathrm{Re}(x, y) + \|y\|^2 \\ &\leq \|x\|^2 + 2\|x\| \cdot \|y\| + \|y\|^2 = (\|x\| + \|y\|)^2, \end{aligned}$$

and thus,

$$\|x + y\| \leq \|x\| + \|y\|.$$

■

From Theorem 5.6.1 (c), E is a normed vector space if the norm $\|x\|$ of x in E is defined by $\|x\| = (x, x)^{1/2}$. For an inner product space, the norm so defined is called the **norm associated with its inner product**. Unless stated otherwise, for an inner product space the norm associated with its inner product is always chosen as its norm.

An inner product space E is called a **Hilbert space** if it is complete when considered as a normed vector space. Obviously, a closed vector subspace of a Hilbert space is a Hilbert space.

The most important class of Hilbert spaces is the class of all $L^2(\Omega, \Sigma, \mu)$ with inner product (f, g), defined by $\int_\Omega f\bar{g}d\mu$ for f, g in $L^2(\Omega, \Sigma, \mu)$. The norm associated with this inner product is the L^2-norm. The space $\mathbb{C}^n$ with inner product $(z, w) = \sum_{k=1}^n z_k\bar{w}_k$ for $z = (z_1, \ldots, z_n)$ and $w = (w_1, \ldots, w_n)$ is a particular case; the norm associated with this inner product is the norm introduced for $\mathbb{C}^n$ in Section 1.4, hence $\mathbb{C}^n$ with this inner product is called the n-dimensional unitary space. Correspondingly, the Euclidean norm of $\mathbb{R}^n$ is associated with the inner product $(x, y) = \sum_{k=1}^n x_k y_k$ for $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$.

Suppose that E is a finite-dimensional vector space of dimension n and let $b_1, \ldots, b_n$ form a basis of E. For $x = \sum_{j=1}^n x_j b_j$, $y = \sum_{j=1}^n y_j b_j$ in E, where the x_j's and y_j's are scalars, define $(x, y) = \sum_{j=1}^n x_j\bar{y}_j$. E is clearly a Hilbert space with inner product so defined. Then it follows from Proposition 1.7.2 that every finite-dimensional inner product space is a Hilbert space.

An example of infinite-dimensional Hilbert space is the real space $\ell^2(\mathbb{Z})$ considered in Section 1.6 whose norm is associated with the inner product $(x, y) = \sum_{k\in\mathbb{Z}} x_k y_k$ for $x = (x_k)$ and $y = (y_k)$. We shall also use $\ell^2(\mathbb{Z})$ to denote the complex Hilbert space of all those complex sequences $(z_k)_{k\in\mathbb{Z}}$ such that $\sum_{k\in\mathbb{Z}} |z_k|^2 < \infty$, and with inner product $(z, w) := \sum_{k\in\mathbb{Z}} z_k\bar{w}_k$ for $z = (z_k)$ and $w = (w_k)$. Whether $\ell^2(\mathbb{Z})$ is a complex or real space will either be stated explicitly or occasioned by context.

As inner product on an inner product space is a generalization of the scalar product for vectors in three-dimensional Euclidean space in which two nonzero vectors are perpendicular to each other if and only if their scalar product is zero. Therefore, two elements x and y in an inner product space E are said to be **orthogonal** if $(x, y) = 0$, and, for a nonempty subset A of E, call the set $A^\perp := \{x \in E : (x, y) = 0 \ \forall y \in A\}$, the **orthogonal complement** of A in E. Obviously, $A^\perp$ is a closed vector subspace of E.

Exercise 5.6.1 Let M be a vector subspace of an inner product space E; show that $M \cap M^\perp = \{0\}$. Also show that if an element x of E can be expressed as the sum $x = y + z$ of an element y in M and an element z in $M^\perp$, then such an expression is unique.

Theorem 5.6.2 (Orthogonal projection theorem) *Suppose that E is a Hilbert space and M a closed vector subspace of E. Then for any $x \in E$, there is a unique element $y \in M$ such that*

$$\|x - y\| = \min_{z\in M} \|x - z\|. \tag{5.6}$$

Furthermore, y is characterized by

$$x - y \in M^\perp. \tag{5.7}$$

Proof Let $\alpha = \inf_{z\in M} \|x - z\|$. There is a sequence $\{y_n\}$ in M such that

$$\alpha^2 \leq \|x - y_n\|^2 \leq \alpha^2 + \frac{1}{n}, \quad n = 1, 2, \ldots.$$

By parallelogram identity,

$$\begin{aligned}&\|(y_n - x) - (y_m - x)\|^2 + \|(y_n - x) + (y_m - x)\|^2\\ &= 2(\|y_n - x\|^2 + \|y_m - x\|^2) \leq 4\alpha^2 + \frac{2}{n} + \frac{2}{m},\end{aligned}$$

or

$$\|y_n - y_m\|^2 \leq 4\alpha^2 + \frac{2}{n} + \frac{2}{m} - 4\left\|\frac{y_n + y_m}{2} - x\right\|^2 \leq 2\left(\frac{1}{n} + \frac{1}{m}\right),$$

from which it follows that $\{y_n\}$ is a Cauchy sequence in M. Since M is complete, there is $y \in M$ such that $\lim_{n\to\infty} \|y_n - y\| = 0$. Then,

$$\|x - y\|^2 = \lim_{n\to\infty} \|x - y_n\|^2 = \alpha^2,$$

i.e.

$$\|x - y\| = \alpha = \inf_{z\in M} \|x - z\| = \min_{z\in M} \|x - z\|.$$

We have shown that there is $y \in M$ such that

$$\|x - y\| = \min_{z\in M} \|x - z\|.$$

Now let y be any element of M which satisfies (5.6); then for $z \in M$ and $t \in \mathbb{R}$, we have

$$\|x - y - tz\|^2 = \|x - y\|^2 - 2\,\mathrm{Re}(x - y, z)t + t^2\|z\|^2,$$

or

$$0 \leq \|x - y - tz\|^2 - \|x - y\|^2 \leq -2\,\mathrm{Re}(x - y, z)t + t^2\|z\|^2.$$

Then for $t > 0$,

$$0 \leq -2\,\mathrm{Re}(x - y, z) + t\|z\|^2,$$

and hence,

$$\mathrm{Re}(x - y, z) \leq 0,$$

by letting $t \searrow 0$; while for $t < 0$,

$$0 \geq -2\,\mathrm{Re}(x - y, z) + t\|z\|^2$$

holds, and by letting $t \nearrow 0$, we have

$$\operatorname{Re}(x - y, z) \geq 0.$$

Hence,

$$\operatorname{Re}(x - y, z) = 0. \tag{5.8}$$

If we replace z in (5.8) by iz, then $\operatorname{Im}(x - y, z) = 0$. Thus $(x - y, z) = 0$, i.e. y satisfies (5.7). Suppose now that (5.6) holds for $y = y'$ and y'' in M, then $(x - y', y' - y'') = 0 = (x - y'', y' - y'') = 0$ by (5.7), and consequently,

$$(y' - y'', y' - y'') = (x - y'' + y' - x, y' - y'') = 0,$$

which implies that $\|y' - y''\| = 0$ or $y' = y''$. Hence, there is unique $y \in M$ that satisfies (5.6).

Finally, suppose $y \in M$ satisfies (5.7), then for $z \in M$,

$$\begin{aligned}\|x - z\|^2 &= \|(x - y) + (y - z)\|^2 = \|x - y\|^2 + 2\operatorname{Re}(x - y, y - z) + \|y - z\|^2 \\ &= \|x - y\|^2 + \|y - z\|^2 \geq \|x - y\|^2,\end{aligned}$$

or y satisfies (5.6). ■

The map that associates each $x \in X$ with the unique element y in M which satisfies (5.6) (or (5.7)) is called the **orthogonal projection** from X onto M. This map will be denoted by P_M.

Corollary 5.6.1 *Suppose that M is a closed vector subspace of a Hilbert space E; then every $x \in E$ can be expressed uniquely as $x = y + z$, where $y \in M$ and $z \in M^{\perp}$. In other words, $E = M \oplus M^{\perp}$.*

Proof For $x \in E$, let $y = P_M x$. Then $x - y \in M^{\perp}$, by (5.7), hence $x = y + (x - y) \equiv y + z$, where $y \in M$ and $z \in M^{\perp}$. The uniqueness of such an expression follows from Exercise 5.6.1. ■

Exercise 5.6.2 Let M be a closed vector subspace of a Hilbert space E.

(i) Show that P_M is linear and that the following properties hold:
(a) $P_M x = x$ if and only if $x \in M$; (b) $P_M^2 = P_M$; and (c) $\|P_M x\| \leq \|x\|$ for all $x \in E$.

(ii) Show that $1 - P_M = P_{M^{\perp}}$.

(iii) Show that $\|x\|^2 = \|P_M x\|^2 + \|P_{M^{\perp}} x\|^2$ for $x \in E$ (**Pythagoras relation**).

Theorem 5.6.3 (Riesz representation theorem) *If E is a Hilbert space, and $x^* \in E^*$, then there is a unique $y_0 \in E$ such that*

$$\langle x, x^* \rangle = (x, y_0), \quad x \in E.$$

Furthermore,

$$\|x^*\| = \|y_0\|,$$

and the map $x^* \to y_0$ is conjugate linear (an operator T from a vector space into a vector space is conjugate linear if $T(\alpha x + \beta y) = \bar{\alpha}Tx + \bar{\beta}Ty$ for all x, y in $D(T)$ and all scalars α and β).

Proof If $x^* = 0$, take $y_0 = 0$. Suppose now that $x^* \neq 0$ and let $M = \ker x^* := \{x \in E : \langle x, x^*\rangle = 0\}$. M is clearly a closed vector subspace of E. Since $x^* \neq 0$, there is $x_0 \in M^\perp$ such that $\langle x_0, x^*\rangle = 1$. Now let $x \in E$ and put $\lambda = \langle x, x^*\rangle$. By Corollary 5.6.1, $x = y + z$, where $y \in M$ and $z \in M^\perp$, hence, $\lambda = \langle x, x^*\rangle = \langle z, x^*\rangle = \langle \lambda x_0, x^*\rangle$, or $\langle z - \lambda x_0, x^*\rangle = 0$, which means that $z - \lambda x_0 \in M$. But $z - \lambda x_0$ is also in $M^\perp$, consequently $z = \lambda x_0$, by Exercise 5.6.1. Now, from $x = y + \lambda x_0$ we have $(x, x_0) = (y + \lambda x_0, x_0) = \lambda\|x_0\|^2 = \langle x, x^*\rangle\|x_0\|^2$. If we take $y_0 = \frac{x_0}{\|x_0\|^2}$, then $(x, y_0) = \langle x, x^*\rangle$ for $x \in E$. Suppose that $y_0' \in E$ also satisfies $\langle x, x^*\rangle = (x, y_0')$ for all $x \in E$, then $(y_0' - y_0, x) = 0$ for all x in E; in particular, $(y_0' - y_0, y_0' - y_0) = 0$ or $\|y_0' - y_0\| = 0$, implying $y_0' = y_0$. Hence, there is unique $y_0 \in E$ satisfying $\langle x, x^*\rangle = (x, y_0)$ for all x in E. From $\langle x, x^*\rangle = (x, y_0)$ it follows readily that $\|x^*\| \le \|y_0\|$; but $\|y_0\|^2 = (y_0, y_0) = |\langle y_0, x^*\rangle| \le \|y_0\| \cdot \|x^*\|$, hence, $\|y_0\| \le \|x^*\|$. Thus $\|y_0\| = \|x^*\|$. That $x^* \to y_0$ is conjugate linear is obvious. ■

Exercise 5.6.3

(i) Denote by R the map $x^* \mapsto y_0$ in Theorem 5.6.3. Show that E^* is a Hilbert space with inner product $(\cdot,\cdot)_*$, defined by $(x^*, y^*)_* = (Ry^*, Rx^*)$ for x^*, y^* in E^*.

(ii) Show that Hilbert spaces are reflexive.

Example 5.6.1 Define on $C[0, 1]$ an inner product by

$$(f, g) = \int_0^1 f(t)\overline{g(t)}dt, \quad f, g \in C[0, 1].$$

We claim that $C[0, 1]$ is not complete with the norm associated with this inner product. We denote this inner product space by $\hat{C}[0, 1]$ in this example. Let f be the indicator function of $[\frac{1}{2}, 1]$ on $[0, 1]$ and for each integer $n > 2$, let f_n be a continuous function such that $0 \le f_n \le 1$ and coincides with f on $[0, \frac{1}{2} - \frac{1}{n}] \cup [\frac{1}{2}, 1]$. Then $f_n \to f$ in $L^2[0, 1]$, i.e., $\|f_n - f\|_2 \to 0$. Let g be any function in $\hat{C}[0, 1]$, then $\|f_n - g\|_2 \ge \|f - g\|_2 - \|f_n - f\|_2$ and hence $\liminf_{n\to\infty} \|f_n - g\|_2 \ge \|f - g\|_2 > 0$. Thus $\{f_n\}$, which is a Cauchy sequence in $\hat{C}[0, 1]$, does not converge in $\hat{C}[0, 1]$.

The Riesz representation theorem for linear functionals on Hilbert spaces might lead to far reaching results, even when the spaces concerned are finite dimensional. We illustrate this fact by proving an interesting result of A.P. Calderón and A. Zygmund about Friederich mollifiers. Recall that from a real-valued C^∞ function φ on $\mathbb{R}^n$ with compact support in the unit closed ball $C_1(0)$ and with $\int \varphi d\lambda^n = 1$, one can construct a family

$\{J_\varepsilon\}_{\varepsilon>0}$ of operators on $L_{\text{loc}}(\mathbb{R}^n)$ in the following way (cf. Section 4.9). For $\varepsilon > 0$, let $\varphi_\varepsilon(x) = \varepsilon^{-n}(\frac{x}{\varepsilon})$ for $x \in \mathbb{R}^n$, then $\operatorname{supp}\varphi_\varepsilon \subset C_\varepsilon(0)$ and $\int \varphi_\varepsilon d\lambda^n = 1$. If $f \in L_{\text{loc}}(\mathbb{R}^n)$, define a function $J_\varepsilon f$ by

$$J_\varepsilon f(x) = \int_{\mathbb{R}^n} f(y)\varphi_\varepsilon(x-y)d\lambda^n(y), \quad x \in \mathbb{R}^n.$$

The family $\{J_\varepsilon\}_{\varepsilon>0}$ depends on φ and is called a Friederich mollifier.

Theorem 5.6.4 (Calderón–Zygmund) *For each $k \in \mathbb{N}$, there is a Friederichs mollifier $\{J_\varepsilon\}_{\varepsilon>0}$ such that $J_\varepsilon p = p$ for every polynomial p of degree $\leq k$ defined on $\mathbb{R}^n$.*

Proof Let E be the space of all real polynomials p of degree $\leq k$ on $\mathbb{R}^n$. E is a real vector space of finite dimension. Choose a nonnegative and nonzero C^∞ function η on $\mathbb{R}^n$ with $\operatorname{supp}\eta \subset C_1(0)$ and define an inner product $(\cdot,\cdot)$ on E by $(p,q) = \int_{\mathbb{R}^n} pq\eta d\lambda^n$ for p, q in E. Since $\dim E < \infty$, E is a Hilbert space. Let l be a linear functional on E defined by

$$l(p) = p(0), \quad p \in E.$$

Since $\dim E < \infty$, every linear functional on E is bounded. By Theorem 5.6.3, there is $q_0 \in E$ such that

$$p(0) = (p, q_0) = \int_{\mathbb{R}^n} pq_0\eta d\lambda^n.$$

If we choose p to be the constant polynomial 1 in the above equality, we have $\int_{\mathbb{R}^n} q_0\eta d\lambda^n = 1$. Let $\varphi = q_0\eta$ and $\{J_\varepsilon\}_{\varepsilon>0}$ the corresponding Friederich mollifier. Now for $p \in E$ and $x \in \mathbb{R}^n$,

$$\begin{aligned} J_\varepsilon p(x) &= \int_{\mathbb{R}^n} p(y)\varphi_\varepsilon(x-y)d\lambda^n(y) = \varepsilon^{-n}\int_{\mathbb{R}^n} p(y)\varphi\left(\frac{x-y}{\varepsilon}\right)d\lambda^n(y) \\ &= \int_{\mathbb{R}^n} p(x-\varepsilon y)\varphi(y)d\lambda^n(y) = \widehat{p_x}(0) = p(x), \end{aligned}$$

where $\widehat{p_x}(y) = p(x - \varepsilon y)$. ∎

Another remarkable application of the Riesz representation theorem will be presented in Section 5.7.

5.7 Lebesgue–Nikodym theorem

We consider in this section an interesting application of the Riesz representation theorem to measure theory.

Let (Ω, Σ) be a measurable space, and suppose that μ and ν are finite measures on Σ. The following theorem asserts that ν can be decomposed in a certain way relative to μ.

Theorem 5.7.1 (Lebesgue–Nikodym theorem) *Let (Ω, Σ) be a measurable space, and μ, ν finite measures on Σ. Then there is a unique $h \in L^1(\Omega, \Sigma, \mu)$ and a μ-null set N, such that*

$$\nu(A) = \int_A h d\mu + \nu(A \cap N), \ \ A \in \Sigma. \tag{5.9}$$

Proof Let $\rho = \mu + \nu$; then ρ is a finite measure on Σ. Consider the real Hilbert space $L^2(\Omega, \Sigma, \rho)$ and consider the linear functional ℓ on $L^2(\Omega, \Sigma, \rho)$, defined by

$$\ell(f) = \int f d\nu.$$

Since

$$\begin{aligned}|\ell(f)| &\leq \left(\int |f|^2 d\nu\right)^{1/2} \left(\int 1 d\nu\right)^{1/2} \leq \nu(\Omega)^{1/2} \left[\int |f|^2 d\rho\right]^{1/2} \\ &= \nu(\Omega)^{1/2} \|f\|_{L^2(\rho)},\end{aligned}$$

ℓ is a bounded linear functional on $L^2(\Omega, \Sigma, \rho)$. By the Riesz representation theorem there is unique $g \in L^2(\Omega, \Sigma, \rho)$, such that

$$\int f d\nu = \int fg d\rho = \int fg d\mu + \int fg d\nu$$

for all $f \in L^2(\Omega, \Sigma, \rho)$, or

$$\int f(1 - g) d\nu = \int fg d\mu \tag{5.10}$$

for all $f \in L^2(\Omega, \Sigma, \rho)$.

We claim first that there is a μ-null set N such that $0 \leq g(x) < 1$ for $x \in \Omega \backslash N$. Let $A_1 = \{x \in \Omega : g(x) < 0\}$ and $A_2 = \{x \in \Omega : g(x) \geq 1\}$. If we let $f = I_{A_1}$ in (5.10), then $0 \leq \nu(A_1) \leq \int_{A_1} (1 - g) d\nu = \int_{A_1} g d\mu$, from which it follows that $\mu(A_1) = 0$. Next choose $f = I_{A_2}$ in (5.10); we have $0 \geq \int_{A_2} (1 - g) d\nu = \int_{A_2} g d\mu \geq \mu(A_2)$. This implies that $\mu(A_2) = 0$. Put $N = A_1 \cup A_2$, then $\mu(N) = 0$ and $0 \leq g(x) < 1$ for $x \in \Omega \backslash N$.

We show next that (5.10) holds for every nonnegative measurable function f which vanishes on N. Suppose that f is such a function; for each positive integer n, let $f_n = f \wedge n$, i.e. $f_n(x) = f(x)$ if $f(x) \leq n$, otherwise $f_n(x) = n$. Since $1 - g > 0$ and $g \geq 0$ on $\Omega \backslash N$, $0 \leq f_n(1 - g) \nearrow f(1 - g)$, and $0 \leq f_n g \nearrow fg$, then from the monotone convergence theorem and the fact that (5.10) holds for each f_n, it follows that

$$\int f(1 - g) d\nu = \lim_{n \to \infty} \int f_n(1 - g) d\nu = \lim_{n \to \infty} \int f_n g d\mu = \int fg d\mu.$$

This shows that (5.10) holds for every such function. For $A \in \Sigma$, let $B = A \cap (\Omega \backslash N)$; then (5.10) holds for the function $f := I_B(1-g)^{-1}$ and we have $\int I_B d\nu = \int I_B \frac{g}{1-g} d\mu = \int_A I_{\Omega\backslash N} \cdot \frac{g}{1-g} d\mu$, or

$$\nu(A \cap (\Omega\backslash N)) = \int_A h d\mu$$

if we put $h = I_{\Omega\backslash N} \frac{g}{1-g}$. Note that $h \geq 0$, and, since $\int_\Omega h d\mu = \nu(\Omega\backslash N) < \infty$, $h \in L^1(\Omega, \Sigma, \mu)$. Now,

$$\nu(A) = \nu(A \cap (\Omega\backslash N)) + \nu(A \cap N) = \int_A h d\mu + \nu(A \cap N),$$

hence (5.9) holds. Now suppose that there is $h' \in L^1(\Omega, \Sigma, \mu)$ and μ-null set N', such that

$$\nu(A) = \int_A h' d\mu + \nu(A \cap N'), \quad A \in \Sigma;$$

if we put $\hat{N} = N \cup N'$, then $\int_{A\cap\hat{N}^c} h d\mu = \int_{A\cap\hat{N}^c} h' d\mu$ for all $A \in \Sigma$, and consequently $h = h'$ μ-a.e. on $\Omega\backslash\hat{N}$; but $\hat{N}$ being a μ-null set implies that $h = h'$ μ-a.e. on Ω. Thus h is unique. ■

Exercise 5.7.1 Show that Theorem 5.7.1 holds if both μ and ν are σ-finite. But in this case h may not be μ-integrable; however it is μ-integrable if ν is finite.

Measure ν is said to be **μ-absolutely continuous** on Σ, if $A \in \Sigma$ and $\mu(A) = 0$ results in $\nu(A) = 0$; while ν is **μ-singular** on Σ, if there is a μ-null set N such that $\nu(A) = (A \cap N)$ for all $A \in \Sigma$. Note that if we use μ^* and ν^* to denote the outer measures on Ω, constructed respectively from μ and ν on Σ by Method I, then the definitions given here for μ-absolute continuity and μ-singularity for ν as measure on Σ are the same as μ^*-absolute continuity and μ^*-singularity for ν^*, introduced in Section 4.6.

Corollary 5.7.1 (Radon–Nikodym) *If μ and ν are σ-finite measures on Σ and ν is μ-absolutely continuous, then there is a unique nonnegative measurable function h on Ω such that*

$$\nu(A) = \int_A h d\mu, \quad A \in \Sigma.$$

Proof We know that Theorem 5.7.1 also holds true if μ and ν are σ-finite (cf. Exercise 5.7.1). We may then apply (5.9). Since $\mu(A \cap N) = 0$ implies that $\nu(A \cap N) = 0$ for all $A \in \Sigma$ by the μ-absolute continuity of ν, the corollary follows. ■

Remark The function h in Corollary 5.7.1 is called the **Radon–Nikodym derivative** of ν w.r.t. μ, and the conclusion of the corollary is usually referred to as the Radon–Nikodym theorem and is expressed by $d\nu = h d\mu$ or $h = \frac{d\nu}{d\mu}$.

5.8 Orthonormal families and separability

Hilbert spaces considered in this section are assumed to be of infinite dimension. The finite-dimensional case can be treated similarly, but in a simpler fashion.

A family $\{e_\alpha\}_{\alpha\in I}$ of elements in a Hilbert space E is said to be **orthonormal** if $(e_\alpha, e_\beta) = \delta_{\alpha\beta} := \begin{cases} 0 & \text{if } \alpha \neq \beta \\ 1 & \text{if } \alpha = \beta \end{cases}$. It is clear that an orthonormal family is linearly independent.

Consider first a finite orthonormal family $\{e_j\}_{j=1}^n$ and let $E_n = \langle\{e_1, \dots, e_n\}\rangle$. Then E_n is a closed vector subspace of E, by Corollary 1.7.1.

Lemma 5.8.1 *Let P_n denote the orthogonal projection from E onto E_n; then $P_n x = \sum_{j=1}^n (x, e_j)e_j$ for $x \in E$.*

Proof It is clear that $P_n x = \sum_{j=1}^n (P_n x, e_j)e_j$. For each $j = 1, \dots, n$, we have $(x - P_n x, e_j) = 0$, by (5.7), hence $(P_n x, e_j) = (x, e_j)$. ■

Exercise 5.8.1 Suppose that $\{e_\alpha\}_{\alpha\in I}$ is an orthonormal family in a Hilbert space E. Show that for any $x \in E$, $\{|(x, e_\alpha)|^2\}_{\alpha\in I}$ is summable and $\sum_{\alpha\in I} |(x, e_\alpha)|^2 \leq \|x\|^2$.

Now let $\{e_k\}_{k=1}^\infty$ be an orthonormal family in E. For each $n \in \mathbb{N}$, put $E_n = \langle\{e_1, \dots, e_n\}\rangle$ and let E_∞ be the closure of $\langle\{e_k\}_{k=1}^\infty\rangle$, i.e. E_∞ is the smallest closed vector subspace containing $\{e_k\}_{k=1}^\infty$.

Theorem 5.8.1 *For $x \in E_\infty$, we have*

(i) $x = \sum_{k=1}^\infty (x, e_k)e_k$, *i.e.* $\lim_{n\to\infty} \|x - \sum_{k=1}^n (x, e_k)e_k\| = 0$.
(ii) $\|x\|^2 = \sum_{k=1}^\infty |(x, e_k)|^2$.

Proof

(i): Given that $\varepsilon > 0$, there is $y \in \langle\{e_k\}_{k=1}^\infty\rangle$ such that $\|x - y\|^2 < \varepsilon$. Now, $y = \sum_{k=1}^m \alpha_k e_k$, $\alpha_k \in \mathbb{C}$, $k = 1, \dots, m$, hence, $y \in E_m \subset E_n$ for $n \geq m$. Thus if $n \geq m$, we have

$$\|x - P_n x\|^2 \leq \|x - y\|^2 < \varepsilon,$$

or, by Lemma 5.8.1,

$$\left\| x - \sum_{k=1}^n (x, e_k)e_k \right\|^2 < \varepsilon$$

if $n \geq m$. This proves (i).

(ii): From (i),

$$\|x\|^2 = \lim_{n\to\infty} \left\| \sum_{k=1}^n (x, e_k)e_k \right\|^2.$$

But,

$$\left\|\sum_{k=1}^{n}(x, e_k)e_k\right\|^2 = \left(\sum_{j=1}^{n}(x, e_j)e_j, \sum_{k=1}^{n}(x, e_k)e_k\right)$$
$$= \sum_{j,k=1}^{n}(x, e_j)\overline{(x, e_k)}(e_j, e_k) = \sum_{k=1}^{n}|(x, e_k)|^2,$$

hence $\|x\|^2 = \sum_{k=1}^{\infty}|(x, e_k)|^2$. ■

Corollary 5.8.1 (Bessel inequality) *For $x \in E$, $\sum_{k=1}^{\infty}|(x, e_k)|^2 \leq \|x\|^2$, and the equality holds if and only if $x \in E_\infty$.*

Proof Let P be the orthogonal projection from E onto E_∞, then $\|x\|^2 = \|Px\|^2 + \|x - Px\|^2$, by Exercise 5.6.1. Hence $\|Px\|^2 \leq \|x\|^2$. But by Theorem 5.8.1,

$$\|Px\|^2 = \sum_{k=1}^{\infty}|(Px, e_k)|^2 = \sum_{k=1}^{\infty}|(x, e_k)|^2,$$

because $(x - Px, e_k) = 0$ for each k by (5.7). Hence,

$$\|x\|^2 = \|x - Px\|^2 + \sum_{k=1}^{\infty}|(x, e_k)|^2,$$

from which it follows that $\sum_{k=1}^{\infty}|(x, e_k)|^2 \leq \|x\|^2$, and that equality holds if and only if $x = Px$ or $x \in E_\infty$. ■

Exercise 5.8.2

(i) Show that for x, y in E_∞ we have

$$(x, y) = \sum_{k=1}^{\infty}(x, e_k)\overline{(y, e_k)}.$$

(ii) Show that $E = E_\infty$ if and only if $\|x\|^2 = \sum_{k=1}^{\infty}|(x, e_k)|^2$ for all $x \in E$.

(iii) Show that $E = E_\infty$ if and only if

$$x = \sum_{k=1}^{\infty}(x, e_k)e_k$$

for all $x \in E$.

Theorem 5.8.2 (Riesz–Fischer) *Let $\{e_k\}_{k\in\mathbb{N}}$ be an orthonormal family in E and $\{\alpha_k\}_{k\in\mathbb{N}}$ a sequence of scalars, then there is $x \in E$ such that $x = \sum_{k=1}^{\infty}\alpha_k e_k$ if and only if $\sum_k |\alpha_k|^2 < \infty$.*

Proof Suppose that $\sum_{k=1}^{\infty} |\alpha_k|^2 < \infty$. For each $n \in \mathbb{N}$ let $x_n = \sum_{k=1}^{n} \alpha_k e_k$. We claim that $\{x_n\}$ is a Cauchy sequence in E. Actually, for $n > m$ in $\mathbb{N}$,

$$\|x_n - x_m\|^2 = \left(\sum_{k=m+1}^{n} \alpha_k e_k, \sum_{j=m+1}^{n} \alpha_j e_j \right) = \sum_{k,j=m+1}^{n} \alpha_k \bar{e}_j (e_k, e_j)$$
$$= \sum_{k=m+1}^{n} |\alpha_k|^2 \to 0$$

as $n > m \to \infty$, so $\{x_n\}$ is a Cauchy sequence, and there is $x \in E$ such that $x = \lim_{n\to\infty} x_n$, or $x = \lim_{n\to\infty} \sum_{k=1}^{n} \alpha_k e_k = \sum_{k=1}^{\infty} \alpha_k e_k$.

Next, suppose that $x = \sum_{k=1}^{\infty} \alpha_k e_k$. This means that $x = \lim_{n\to\infty} \sum_{k=1}^{n} \alpha_k e_k$; but each $\sum_{k=1}^{n} \alpha_k e_k$ is in E_n, and hence $x \in E_\infty$. Now for each $j \in \mathbb{N}$,

$$(x, e_j) = \lim_{n\to\infty} \left(\sum_{k=1}^{n} \alpha_k e_k, e_j \right) = \alpha_j;$$

consequently,

$$\sum_{j=1}^{\infty} |\alpha_j|^2 = \sum_{j=1}^{\infty} |(x, e_j)|^2 = \|x\|^2 < \infty,$$

by Theorem 5.8.1 (ii). ■

An orthonormal family $\{e_k\}_{k=1}^{\infty}$ is called an **orthonormal basis** for E if

$$x = \sum_{k=1}^{\infty} (x, e_k) e_k$$

for all $x \in E$.

Theorem 5.8.3 *Let $\{e_k\}_{k=1}^{\infty}$ be an orthonormal family in a Hilbert space E and define E_∞ as before.*

(i) *$\{e_k\}_{k=1}^{\infty}$ is an orthonormal basis for E if and only if $E = E_\infty$.*

(ii) *$\{e_k\}_{k=1}^{\infty}$ is an orthonormal basis for E if and only if for $x \in E$, $x = 0$ whenever $(x, e_k) = 0$ for all k.*

Proof It is clear that (i) follows from Theorem 5.8.1 (i), and the fact that if $\{e_k\}_{k=1}^{\infty}$ is an orthonormal basis, then $E = E_\infty$. For the proof of (ii), in view of (i) one need only observe that for $x \in E$, $(x - Px, e_k) = 0$ for all k, where P is the orthonormal projection from E onto E_∞. ■

Exercise 5.8.3 Show that an orthonormal family $\{e_k\}_{k\in\mathbb{N}}$ in E is an orthonormal basis for E if and only if $\|x\|^2 = \sum_{k=1}^{\infty} |(x, e_k)|^2$ for all $x \in E$.

Example 5.8.1 (Hermite polynomials and Hermite functions) For nonnegative integer n and $x \in \mathbb{R}$, let

$$H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n} e^{-x^2};$$

then $H_n(x)$ is a polynomial in x of degree n with the coefficient of x^n being 2^n. The polynomials $H_n(x)$ are called **Hermite polynomials** and the functions $\psi_n(x) = e^{-\frac{x^2}{2}} H_n(x)$ are called **Hermite functions**. We have, for nonnegative integers m and n,

$$\begin{aligned}\int_{-\infty}^{\infty} \psi_n(x)\psi_m(x)dx &= \int_{-\infty}^{\infty} e^{-x^2} H_n(x)H_m(x)dx \\ &= \int_{-\infty}^{\infty} H_m(x)(-1)^n \frac{d^n}{dx^n} e^{-x^2} dx,\end{aligned}$$

from which we conclude by repeated integration by parts that

$$\begin{aligned}\int_{-\infty}^{\infty} \psi_n(x)\psi_m(x)dx &= \int_{-\infty}^{\infty} e^{-x^2} \frac{d^n}{dx^n} H_m(x)dx \\ &= \begin{cases} 0 & \text{if } m < n; \\ 2^n n!\sqrt{\pi} & \text{if } m = n. \end{cases}\end{aligned}$$

Thus $\{\psi_0, \psi_1, \psi_2, \ldots\}$ is an orthogonal family in $L^2(\mathbb{R})$. If we define **the normalized Hermite functions** $\mathcal{E}_n$ by

$$\mathcal{E}_n(x) = (2^n n!\sqrt{\pi})^{-\frac{1}{2}} \psi_n(x),$$

then $\{\mathcal{E}_0, \mathcal{E}_1, \mathcal{E}_2, \ldots\}$ is an orthonormal family in $L^2(\mathbb{R})$. Observe that $\mathcal{E}_n(x) = e^{-\frac{x^2}{2}} h_n(x)$, where $h_n(x) = (2^n n!\sqrt{\pi})^{-\frac{1}{2}} H_n(x)$; the polynomials $h_0(x)$, $h_1(x)$, $h_2(x)$, $\ldots$ are called **normalized Hermite polynomials**. Observe that since $h_n(x)$ is a polynomial of degree n, each monomial x^n is a linear combination of $h_0(x), \ldots, h_n(x)$. Let us now put $w(x) = e^{-x^2}$ and denote by $L^2_w(\mathbb{R})$ the space $L^2(\mathbb{R}, \mathcal{L}, \mu)$, where $\mu(A) = \int_A w d\lambda = \int_A e^{-x^2} dx$ for $A \in \mathcal{L}$. The space $L^2_w(\mathbb{R})$ is called the **weighted $\mathbf{L^2}$ space** on $\mathbb{R}$ with weight w. Then, Hermite polynomials form an orthogonal family in $L^2_w(\mathbb{R})$ and normalized Hermite polynomials form an orthonormal family in $L^2_w(\mathbb{R})$. We shall see in Chapter 7 that $\{\mathcal{E}_0, \mathcal{E}_1, \mathcal{E}_2, \ldots\}$ is an orthonormal basis for $L^2(\mathbb{R})$, or equivalently, $\{h_0, h_1, h_2, \ldots\}$ is an orthonormal basis for $L^2_w(\mathbb{R})$ (cf. Corollary 7.1.1).

A procedure, the **Gram–Schmidt process**, for orthonormalizing a given countable linearly independent family $\{u_k\}$ in E is now introduced. Let $e_1 = \frac{u_1}{\|u_1\|}$. Suppose now that $e_1, \ldots, e_n$ have been defined so that they form an orthonormal family and $\langle\{e_1, \ldots, e_n\}\rangle = \langle\{u_1, \ldots, u_n\}\rangle$; put $E_n = \langle\{e_1, \ldots, e_n\}\rangle$ and let z_n be the image of u_{n+1} in E_n under the orthogonal projection from E onto E_n. Since u_{n+1} is not in $\langle\{u_1, \ldots, u_n\}\rangle$, it is not in E_n and hence $u_{n+1} - z_n \neq 0$. Define $e_{n+1} = \frac{u_{n+1}-z_n}{\|u_{n+1}-z_n\|}$, then $\|e_{n+1}\| = 1$ and $e_{n+1} \in E_n^{\perp}$.

Thus $e_1,\dots,e_{n+1}$ form an orthonormal family; it is readily seen that $\langle\{u_1,\dots,u_{n+1}\}\rangle = \langle\{e_1,\dots,e_{n+1}\}\rangle$. We have therefore defined, by induction, an orthonormal family $\{e_k\}$ from $\{u_k\}$ such that $\langle\{e_1,\dots,e_n\}\rangle = \langle\{u_1,\dots,u_n\}\rangle$ for all $n \in \mathbb{N}$.

Theorem 5.8.4 *A Hilbert space E has an orthonormal basis if and only if E is separable.*

Proof If E has an orthonormal basis $\{e_k\}$, then the countable set $\bigcup_{n=1}^{\infty}\{\sum_{j=1}^{n}\alpha_j e_j : \alpha_j \in \gamma,\ j = 1,\dots,n\}$ is dense in E; hence E is separable. We have denoted by γ the countable set of rational complex numbers.

If now E is separable, say $\{x_n\}_{n=1}^{\infty}$ is dense in E. We may assume that $x_1 \neq 0$. By an obvious selection procedure, we can select a linearly independent subsequence $\{x_{n_k}\}$ of $\{x_n\}$ such that $\langle\{x_{n_k}\}\rangle = \langle\{x_n\}\rangle$. Put $x_{n_k} = y_k$. Let $\{e_k\}$ be the orthonormal family obtained from $\{y_k\}$ by the Gram–Schimdt procedure, then $\{e_k\}$ is an orthonormal family such that $\langle\{e_k\}\rangle = \langle\{y_k\}\rangle = \langle\{x_k\}\rangle$. Consequently the closure of $\langle\{e_k\}\rangle$ is E. Then $\{e_k\}_{k=1}^{\infty}$ is an orthonormal basis of E. ■

5.9 The space $L^2[-\pi,\pi]$

Historically, the most well-known orthonormal family is $\{\frac{1}{\sqrt{2\pi}}e^{ikt}\}_{k\in\mathbb{Z}}$ in $L^2[-\pi,\pi]$. It was introduced by **J. Fourier** in his study of heat conduction by means of expansion of functions as trigonometric series, and is usually referred to as the **Fourier basis**. Here $L^2[-\pi,\pi]$ stands for $L^2([-\pi,\pi],\mathcal{L}|[-\pi,\pi],\lambda)$.

For $f \in L^1[-\pi,\pi]$, the function $\hat{f}$ defined on $\mathbb{Z}$ by

$$\hat{f}(k) = \frac{1}{\sqrt{2\pi}}\int_{-\pi}^{\pi} f(t)e^{-ikt}dt$$

is called the **Fourier transform** of f, and $\hat{f}(k)$'s, $k \in \mathbb{Z}$, are called **Fourier coefficients** of f. If we put $e_k(t) = \frac{1}{\sqrt{2\pi}}e^{ikt}$, then for $f \in L^2[-\pi,\pi]$, $\hat{f}(k) = (f,e_k)$, $k \in \mathbb{Z}$, where $\int_{-\pi}^{\pi} f(t)\overline{g(t)}dt \equiv (f,g)$ is the inner product for L^2-spaces. It is easily verified that $(e_k,e_j) = \delta_{kj}$, hence $\{e_k\}$ is indeed an orthonormal family in $L^2[-\pi,\pi]$.

We shall show in this section that $\{e_k\}_{k\in\mathbb{Z}}$ is an orthonormal basis for $L^2[-\pi,\pi]$.

Let $f \in L^1[-\pi,\pi]$ and n be a nonnegative integer; define the Fourier n-th partial sum $S_n(f,t)$ of f by

$$\begin{aligned} S_n(f,t) &= \sum_{k=-n}^{n}\hat{f}(k)e_k(t) = \sum_{k=-n}^{n}\left(\int_{-\pi}^{\pi} f(s)\frac{e^{-iks}}{\sqrt{2\pi}}ds\right)\frac{1}{\sqrt{2\pi}}e^{ikt} \\ &= \frac{1}{2\pi}\sum_{k=-n}^{n}\int_{-\pi}^{\pi} f(s)e^{ik(t-s)}ds. \end{aligned}$$

We derive firstly an integral representation for $S_n(f,t)$. Define

$$D_n(t) := \frac{1}{2\pi}\left[1 + 2\sum_{k=1}^{n}\cos kt\right],$$

then,

$$\begin{aligned}
\sin\frac{1}{2}tD_n(t) &= \frac{1}{2\pi}\left[\sin\frac{1}{2}t + 2\sum_{k=1}^{n}\sin\frac{1}{2}t\cos kt\right] \\
&= \frac{1}{2\pi}\left[\sin\frac{1}{2}t + \sum_{k=1}^{n}\left\{\sin\left(k+\frac{1}{2}\right)t - \sin\left(k-\frac{1}{2}\right)t\right\}\right] \\
&= \frac{1}{2\pi}\sin\left(n+\frac{1}{2}\right)t,
\end{aligned}$$

hence if t is not an even multiple of π, we have

$$D_n(t) = \frac{1}{2\pi}\frac{\sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t}.$$

Now,

$$\begin{aligned}
S_n(f,t) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(s)\sum_{k=-n}^{n} e^{ik(t-s)}ds \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(s)\left\{1 + 2\sum_{k=1}^{n}\cos k(t-s)\right\}ds;
\end{aligned}$$

thus,

$$S_n(f,t) = \int_{-\pi}^{\pi} f(s)D_n(t-s)ds. \tag{5.11}$$

The functions D_n, $n = 0, 1, 2, \ldots$ are called **Dirichlet kernels**.

It is a common practice to extend a function on $(a, b]$ to be a periodic function on $\mathbb{R}$ with period $(b-a)$; we follow this practice by regarding f as defined on $(-\pi, \pi]$ and extend it periodically to $\mathbb{R}$ with period 2π; then,

$$\begin{aligned}
S_n(f,t) &= \int_{-\pi}^{\pi} f(s)D_n(t-s)ds = \int_{-\pi-t}^{\pi-t} f(t+s)D_n(-s)ds \\
&= \int_{-\pi-t}^{\pi-t} f(t+s)D_n(s)ds = \int_{-\pi}^{\pi} f(t+s)D_n(s)ds,
\end{aligned}$$

where the last equality follows from the fact that the function $s \mapsto f(t+s)D_n(s)$ is of period 2π (cf. Exercise 4.3.3). Thus (5.11) can be put in the form

$$S_n(f,t) = \int_{-\pi}^{\pi} f(t+s)D_n(s)ds. \tag{5.11$'$}$$

Exercise 5.9.1 Let $X = \{f \in C[-\pi, \pi] : f(-\pi) = f(\pi)\}$; X is a Banach space with sup-norm. For $n = 0, 1, 2, \ldots$ define $\ell_n(f) = S_n(f, 0)$ for $f \in X$.

(i) Show that $\ell_n \in X^*$, $n = 0, 1, 2, \ldots$ and

$$\|\ell_n\| = \int_{-\pi}^{\pi} |D_n(t)| dt;$$

(ii) Show that $\lim_{n\to\infty} \|\ell_n\| = \infty$;

(iii) Show that there is $f \in X$ such that

$$\limsup_{n\to\infty} |S_n(f, 0)| = \infty.$$

(Hint: cf. Theorem 5.1.3.)

In general, $S_n(f, t)$ is not well behaved as $n \to \infty$, so it is expedient to consider the Cesàro mean of the sequence: for $n = 0, 1, 2, \ldots$; let

$$\sigma_n(f, t) = \frac{1}{n+1} \sum_{k=0}^{n} S_k(f, t).$$

Using (5.11) we have

$$\sigma_n(f, t) = \frac{1}{n+1} \int_{-\pi}^{\pi} f(s) \sum_{k=0}^{n} D_k(t - s) ds = \int_{-\pi}^{\pi} f(s) F_n(t - s) ds, \tag{5.12}$$

where $F_n(t) = \frac{1}{n+1} \sum_{k=0}^{n} D_k(t)$. Since

$$\begin{aligned}
\sin^2 \frac{1}{2} t F_n(t) &= \frac{1}{2\pi(n+1)} \sum_{k=0}^{n} \sin(k + \frac{1}{2}) t \sin \frac{1}{2} t \\
&= \frac{1}{2\pi(n+1)} \frac{1}{2} \sum_{k=0}^{n} \{\cos kt - \cos(k+1)t\} \\
&= \frac{1}{2\pi(n+1)} \cdot \frac{1}{2} \{1 - \cos(n+1)t\} \\
&= \frac{1}{2\pi(n+1)} \sin^2 \frac{n+1}{2} t,
\end{aligned}$$

we have

$$F_n(t) = \frac{1}{2\pi(n+1)} \left(\frac{\sin \frac{n+1}{2} t}{\sin \frac{1}{2} t} \right)^2$$

if t is not an even multiple of π. $F_n(t)$, $n = 0, 1, 2, \ldots$, are called the **Féjer kernels**. Take $f = 1$ in (5.11) and (5.12), we have

$$\int_{-\pi}^{\pi} D_n(t - s) ds = \int_{-\pi}^{\pi} F_n(t - s) ds = 1, \quad t \in [-\pi, \pi]. \tag{5.13}$$

Theorem 5.9.1 (Féjer) *Suppose that f is continuous on $[-\pi, \pi]$ and $f(-\pi) = f(\pi)$. Then $\sigma_n(f, t) \to f(t)$ uniformly for $t \in [-\pi, \pi]$ when $n \to \infty$.*

Proof From (5.13),

$$\begin{aligned}|\sigma_n(f,t)-f(t)| &= \left|\int_{-\pi}^{\pi}\{f(s)-f(t)\}F_n(t-s)ds\right|\\ &\le \int_{-\pi}^{\pi}|f(s)-f(t)|F_n(t-s)ds.\end{aligned}$$

Since f is continuous on $[-\pi,\pi]$ and $f(-\pi)=f(\pi)$, for any given $\varepsilon>0$, there is $\delta>0$, such that when either $|s-t|<\delta$ or $|s-t|>2\pi-\delta$, we have $|f(s)-f(t)|\le\frac{\varepsilon}{2}$. It is obvious from the form of the function $F_n(s-t)$ that there is $N\in\mathbb{N}$ such that when $n\ge N$,

$$\sup_{\delta\le|t-s|\le 2\pi-\delta} F_n(t-s)\le\frac{\varepsilon}{8\pi M}, \tag{5.14}$$

where $M=\sup_{t\in[-\pi,\pi]}|f(t)|$. For $n\ge N$, by (5.14) and the choice of δ,

$$\begin{aligned}|\sigma_n(f,t)-f(t)| &\le \int_{\substack{|t-s|<\delta\\ \text{or } |t-s|>2\pi-\delta}}|f(s)-f(t)|F_n(t-s)ds\\ &\quad+\int_{\delta\le|t-s|\le 2\pi-\delta}|f(s)-f(t)|F_n(t-s)ds\\ &\le\frac{\varepsilon}{2}\int_{-\pi}^{\pi}F_n(t-s)ds+2M\cdot\frac{\varepsilon}{8\pi M}\cdot 2\pi=\varepsilon;\end{aligned}$$

this shows that $\sigma_n(f,t)\to f(t)$ uniformly for $t\in[-\pi,\pi]$ when $n\to\infty$, because our choice of N is independent of t. ■

Since each $\sigma_n(f,t)$ is a linear combination of $\{\frac{1}{\sqrt{2\pi}}e_k\}_{|k|\le n}$, it follows from the Féjer theorem that $\langle\{\frac{1}{\sqrt{2\pi}}e_k\}_{k\in\mathbb{Z}}\rangle$ is dense in the space of all continuous functions f on $[-\pi,\pi]$ with $f(-\pi)=f(\pi)$ w.r.t. the L^2-norm in $L^2[-\pi,\pi]$. But the latter space contains $C_c(-\pi,\pi)$ which is dense in $L^2[-\pi,\pi]$. As a consequence, the closure of $\langle\{\frac{1}{\sqrt{2\pi}}e_k\}_{k\in\mathbb{Z}}\rangle$ in $L^2[-\pi,\pi]$ is $L^2[-\pi,\pi]$. Thus we have established the following theorem.

Theorem 5.9.2 *$\{\frac{1}{\sqrt{2\pi}}e_k\}_{k\in\mathbb{Z}}$, where $e_k(t)=e^{ikt}$ is an orthonormal basis for $L^2[-\pi,\pi]$.*

Because $e^{ikt}=\cos kt+i\sin kt$, it follows from direct computation that

$$\begin{aligned}S_n(f,x) &= \frac{1}{2}\frac{1}{\pi}\int_{-\pi}^{\pi}f(t)dt\\ &\quad+\frac{1}{\pi}\sum_{k=1}^{n}\left\{\int_{-\pi}^{\pi}f(t)\cos ktdt\cos kx+\int_{-\pi}^{\pi}f(t)\sin ktdt\sin kx\right\}.\end{aligned}$$

Hence, if we put

$$
\begin{aligned}
a_n &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos nt dt, \qquad n = 0, 1, 2, \dots;\\
b_n &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin nt dt, \qquad n = 1, 2, 3, \dots,
\end{aligned} \tag{5.15}
$$

then,

$$
S_n(f, x) = \frac{1}{2}a_0 + \sum_{k=1}^{n}\{a_k \cos kx + b_k \sin kx\}. \tag{5.16}
$$

This is the traditional form of **Fourier partial sums**; the numbers a_0, a_1, b_1, a_2, b_2, ... defined by (5.15) are called the **Fourier trigonometric coefficients** of the function f and are expressed symbolically by

$$
f(x) \sim \frac{1}{2}a_0 + \sum_{n=1}^{\infty}\{a_n \cos nx + b_n \sin nx\};
$$

the series $\frac{1}{2}a_0 + \sum_{n=1}^{\infty}\{a_n \cos nx + b_n \sin nx\}$ is usually referred to as the **Fourier trigonometric series** of f. Whether or not $f(x) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty}\{a_n \cos nx + b_n \sin nx\}$ for $x \in [-\pi, \pi]$ is a well-known problem in analysis, which leads to discovery of many tools in real analysis, including the introduction of Lebesgue measure and Lebesgue integration. Since $\{\frac{1}{\sqrt{2\pi}}e_k\}_{k\in\mathbb{Z}}$ is an orthonormal basis for $L^2[a, b]$ if $b - a = 2\pi$, our discussion so far also holds on any interval of length 2π; in particular, Fourier trigonometric coefficients for integrable functions on such an interval are defined similarly.

Exercise 5.9.2 Consider $L^2[0, 2\pi]$.

(i) Show that $\{\frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{\pi}}\cos x, \frac{1}{\sqrt{\pi}}\sin x, \frac{1}{\sqrt{\pi}}\cos 2x, \frac{1}{\sqrt{\pi}}\sin 2x, \dots\}$ is an orthonormal basis for $L^2[0, 2\pi]$.

(ii) For f, g in $L^2[0, 2\pi]$, suppose that

$$
\begin{aligned}
f(x) &\sim \frac{1}{2}a_0 + \sum_{n=1}^{\infty}\{a_n \cos nx + b_n \sin nx\},\\
g(x) &\sim \frac{1}{2}c_0 + \sum_{n=1}^{\infty}\{c_n \cos nx + d_n \sin nx\}.
\end{aligned}
$$

Show that

$$
\frac{1}{\pi}\int_0^{2\pi} f\bar{g} d\lambda = \frac{1}{2}a_0\bar{c}_0 + \sum_{n=1}^{\infty}\{a_n\bar{c}_n + b_n\bar{d}_n\}.
$$

(iii) Suppose that $f \in L^2[0, 2\pi]$ and $a_n = b_n = 0$ for $n \geq k$ for some k. Show that $f(x) = \frac{1}{2}a_0 + \sum_{n=1}^{k-1}\{a_n \cos nx + b_n \sin nx\}$ for a.e. $x \in [0, 2\pi]$.

(iv) Suppose that f is AC on $[0, 2\pi]$ with $f' \in L^2[0, 2\pi]$ and satisfies $f(0) = f(2\pi)$. Show that

$$\frac{1}{\pi}\int_0^{2\pi} |f'|^2 d\lambda = \sum_{n=1}^{\infty} n^2(|a_n|^2 + |b_n|^2),$$

where a_n and b_n are as defined in (ii).

(v) Let f be as in (iv). Show that $\sum_{n=1}^{\infty}(|a_n| + |b_n|) < \infty$ and infer that the Fourier trigonometric series of f converges uniformly to f on $[0, 2\pi]$.

To give a flavor of orthonormal basis in infinite-dimensional spaces, we now prove a classical isoperimetric inequality, following A. Hurwicz.

Theorem 5.9.3 (Isoperimetric inequality) *For any piece-wise C^1 simple closed plane curve with given length L, the following inequality holds:*

$$A \le \frac{L^2}{4\pi},$$

where A is the area of the region enclosed by the curve; and equality holds when and only when the curve is a circle.

Proof Let C be such a curve and choose a parametric representation, $x = x(s), y = y(s)$, $0 \le s \le L$, with arc length as the parameter so that, when s goes from 0 to L, the curve C is traced counter clockwise. Choose the new parameter $t = 2\pi s/L$ and let

$$x(t) \sim \frac{1}{2}a_0 + \sum_{n=1}^{\infty}\{a_n \cos nt + b_n \sin nt\},$$
$$y(t) \sim \frac{1}{2}c_0 + \sum_{n=1}^{\infty}\{c_n \cos nt + d_n \sin nt\};$$

then, using the results in Exercise 5.9.2, we have

$$\frac{dx}{dt} \sim \sum_{n=1}^{\infty}\{nb_n \cos nt - na_n \sin nt\},$$
$$\frac{dy}{dt} \sim \sum_{n=1}^{\infty}\{nd_n \cos nt - nc_n \sin nt\};$$

and

$$\frac{1}{\pi}\int_0^{2\pi}\left\{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2\right\} dt = \sum_{n=1}^{\infty} n^2(a_n^2 + b_n^2 + c_n^2 + d_n^2),$$
$$\frac{1}{\pi}\int_0^{2\pi} x\frac{dy}{dt} dt = \sum_{n=1}^{\infty} n(a_n d_n - b_n c_n).$$

Since $(\frac{dx}{dt})^2 + (\frac{dy}{dt})^2 = (\frac{L}{2\pi})^2\{(\frac{dx}{ds})^2 + (\frac{dy}{ds})^2\} = (\frac{L}{2\pi})^2$ and $A = \int_0^{2\pi} x\frac{dy}{dt}dt$, we have

$$\frac{L^2}{4\pi} - A = \frac{\pi}{2}\sum_{n=1}^{\infty}\{n^2(a_n^2 + b_n^2 + c_n^2 + d_n^2) - 2n(a_nd_n - b_nc_n)\}$$
$$= \frac{\pi}{2}\sum_{n=1}^{\infty}\{(na_n - d_n)^2 + (nb_n + c_n)^2 + (n^2 - 1)(c_n^2 + d_n^2)\} \geq 0.$$

Hence $A \leq \frac{L^2}{4\pi}$. Now, $\sum_{n=1}^{\infty}\{(na_n - d_n)^2 + (nb_n + c_n)^2 + (n^2 - 1)(c_n^2 + d_n^2)\} = 0$ if and only if $a_1 = d_1$, $b_1 = -c_1$, and $a_n = b_n = c_n = d_n = 0$ for $n \geq 2$; it follows that $\frac{L^2}{4\pi} = A$ if and only if

$$x = \frac{1}{2}a_0 + a_1\cos t + b_1\sin t, \quad y = \frac{1}{2}c_0 - b_1\cos t + a_1\sin t,$$

or C is a circle. ■

Theorem 5.9.4 (Weierstrass approximation theorem) *Any continuous function on a finite closed interval $[a,b]$ can be approximated uniformly by polynomials in the interval.*

Proof We may assume without loss of generality that $[a,b] = [-\pi,\pi]$. Since any continuous function f on $[-\pi,\pi]$ can be expressed as

$$f(x) = f(-\pi) + \frac{\{f(\pi) - f(-\pi)\}}{2\pi}(x+\pi) + g(x),$$

where $g(-\pi) = g(\pi) = 0$, it is sufficient to prove the theorem for continuous functions f on $[-\pi,\pi]$ satisfying $f(-\pi) = f(\pi)$. For such a function f, $\sigma_n(f,x) \to f(x)$ uniformly for $x \in [-\pi,\pi]$, by Theorem 5.9.1. Now, $\sigma_n(f,x)$ is a finite linear combination of trigonometric functions $\cos x, \sin x, \cos 2x, \sin 2x, \ldots$; hence, each $\sigma_n(f,x)$ can be approximated uniformly by polynomials on $[-\pi,\pi]$ by Taylor's theorem. Thus, given $\varepsilon > 0$, there is n_0 such that $\sup_{x\in[-\pi,\pi]}|f(x) - \sigma_{n_0}(f,x)| \leq \frac{\varepsilon}{2}$; then let $p(x)$ be a Taylor polynomial of $\sigma_{n_0}(f,x)$ such that $\sup_{x\in[-\pi,\pi]}|\sigma_{n_0}(f,x) - p(x)| \leq \frac{\varepsilon}{2}$; therefore, $\sup_{x\in[-\pi,\pi]}|f(x) - p(x)| \leq \varepsilon$. ■

Exercise 5.9.3 Let $f_n(x) = x^n$, $n = 0,1,2,\ldots$. Show that the Gram–Schmidt process applied to the family $\{f_0, f_1, f_2, \ldots\}$ in $L^2[a,b]$ yields an orthonormal basis for $L^2[a,b]$ $(-\infty < a < b < \infty)$. When $a = -1$, $b = 1$, denote the orthonormal basis so obtained by $\{\pi_0, \pi_1, \pi_2, \ldots\}$. Show that π_n is a polynomial of degree n, $n = 0,1,2,\ldots$ and find π_0, π_1, and π_2.

Exercise 5.9.4 For $n = 0,1,2,\ldots$, let P_n be the polynomial defined by $P_n(x) = \frac{1}{2^n n!}\frac{d^n(x^2-1)^n}{dx^n}$; P_0, P_1, P_2, $\ldots$ are called **Legendre polynomials**. Show that $\{P_0, P_1, P_2, \ldots\}$ is an orthogonal family in $L^2[-1,1]$ and $\int_{-1}^{1} x^kP_n(x)dx = 0$ for $n \geq 1$ and $0 \leq k < n$.

Exercise 5.9.5 Let $\{\pi_0, \pi_1, \pi_2, \ldots\}$ and $\{P_0, P_1, P_2, \ldots\}$ be as in Exercises 5.9.3 and 5.9.4. Show that for $n = 0, 1, 2, \ldots$, there is a positive constant α_n such that $\pi_n = \alpha_n P_n$.

We digress now from the main theme of this section to discuss briefly the pointwise convergence of Fourier trigonometric series. For this we first prove the Riemann–Lebesgue lemma.

Lemma 5.9.1 (Riemann–Lebesgue) *If f is an integrable function on a finite interval $[a, b]$, then*

$$\lim_{l\to\infty} \int_a^b f(t) \sin ltdt = 0.$$

Proof If J is an interval with endpoints $c < d$ in $[a, b]$, then $\int_J \sin ltdt = -\frac{1}{l}\{\cos ld - \cos lc\} \to 0$ as $l \to \infty$; consequently, the lemma holds if f is a step function. In general, given $\varepsilon > 0$, there is a step function g on $[a, b]$ such that $\int_a^b |f(t) - g(t)|dt < \frac{\varepsilon}{2}$, and therefore,

$$\begin{aligned}\left|\int_a^b f(t) \sin ltdt\right| &\leq \left|\int_a^b \{f(t) - g(t)\} \sin ltdt\right| + \left|\int_a^b g(t) \sin ltdt\right| \\ &< \frac{\varepsilon}{2} + \left|\int_a^b g(t) \sin ltdt\right| < \varepsilon,\end{aligned}$$

if l is sufficiently large, because the lemma holds for the step function g. ■

Theorem 5.9.5 (Dini test) *Suppose that f is an integrable function on $(-\pi, \pi)$ and is extended to $\mathbb{R}$ periodically. Let $t_0 \in [-\pi, \pi]$, then,*

$$\lim_{n\to\infty} S_n(f, t_0) = f(t_0)$$

if $s \mapsto \frac{1}{s}\{f(t_0 + s) - f(t_0)\}$ is integrable in a neighborhood of 0.

Proof If $s \mapsto \frac{1}{s}\{f(t_0 + s) - f(t_0)\}$ is integrable in a neighborhood of 0, then the function g defined by

$$g(s) = \frac{1}{2\pi}\frac{f(t_0 + s) - f(t_0)}{\sin \frac{1}{2}s} = \frac{1}{2\pi}\frac{s}{\sin \frac{1}{2}s}\frac{f(t_0 + s) - f(t_0)}{s}, \quad s \in [-\pi, \pi],$$

is integrable on $[-\pi, \pi]$. Now, from (5.11)′ we have,

$$\begin{aligned}S_n(f, t_0) - f(t_0) &= \int_{-\pi}^{\pi} \{f(t_0 + s) - f(t_0)\}D_n(s)ds \\ &= \int_{-\pi}^{\pi} g(s) \sin\left(n + \frac{1}{2}\right) sds \to 0\end{aligned}$$

as $n \to \infty$, by the Riemann–Lebesgue lemma. ■

Exercise 5.9.6 Let f be an even function on $[-\pi,\pi]$ defined on $[0,\pi]$ by $f(s) = 1 - \frac{s}{\pi}$. Show that the Fourier trigonometric series of f converges uniformly to f on $[-\pi,\pi]$. In particular, verify that $\sum_{k=0}^{\infty} \frac{1}{(2k+1)^2} = \frac{\pi^2}{8}$.

Exercise 5.9.7 Suppose that f is a periodic function of period 2π on $\mathbb{R}$ and is integrable on $[-\pi,\pi]$. Show that if $f = 0$ on a neighborhood of t_0, then $S_n(f,t) \to 0$ uniformly on a neighborhood of t_0.

Exercise 5.9.8 Suppose that f is integrable on $[-\pi,\pi]$ and $f(t_0+), f(t_0-)$ exist at $t_0 \in [-\pi,\pi]$. Show that

$$\lim_{n\to\infty} S_n(f,t_0) = \frac{1}{2}\{f(t_0-) + f(t_0+)\}$$

if $\int_{-\varepsilon}^{0} \left|\frac{f(t_0+s)-f(t_0-)}{s}\right| ds < \infty$, and $\int_0^{\varepsilon} \left|\frac{f(t_0+s)-f(t_0+)}{s}\right| ds < \infty$ for some $\varepsilon > 0$. (Hint: $\int_{-\pi}^{0} D_n(s)ds = \int_0^{\pi} D_n(s)ds = \frac{1}{2}$.)

Exercise 5.9.9 Let f be a periodic function with period π on $\mathbb{R}$, and $f(s) = s$ for $0 \le s < \pi$. Find the Fourier trigonometric series for f and evaluate $\sum_{n=1}^{\infty} a_n$, where

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(s)\cos nsds, \quad n = 0, 1, 2, \ldots.$$

Lemma 5.9.2 *There is $c > 0$ s.t. $\left|\int_{\delta}^{\eta} D_n(s)ds\right| \le c$ for all $n \in \mathbb{N}$ and $0 \le \delta < \eta \le \pi$.*

Proof Let $n \in \mathbb{N}$ and $0 \le \delta < \eta \le \pi$. It will be clear from the following argument that we may assume $\delta < \frac{2}{2n+1} < \eta$; then,

$$0 \le \frac{\sin(n+\frac{1}{2})s}{\sin\frac{1}{2}s} = \frac{\frac{1}{2}s}{\sin\frac{1}{2}s} \cdot \frac{\sin(n+\frac{1}{2})s}{\frac{1}{2}s} < 1\cdot(2n+1)$$

for $0 < s < \frac{2}{2n+1}$, and hence,

$$\int_{\delta}^{\frac{2}{2n+1}} \frac{\sin(n+\frac{1}{2})s}{\sin\frac{1}{2}s} < (2n+1)\cdot\frac{2}{2n+1} = 2.$$

Thus,

$$\left|\int_{\delta}^{\eta} D_n(s)ds\right| \le \frac{1}{2\pi}\left\{2 + \left|\int_{\frac{2}{2n+1}}^{\eta} \frac{\sin(n+\frac{1}{2})s}{\sin\frac{1}{2}s}ds\right|\right\}.$$

But by the second mean-value theorem (actually, Lemma 4.5.2), there is $\frac{2}{2n+1} \leq \eta' \leq \eta$ such that

$$\begin{aligned}
\left| \int_{\frac{2}{2n+1}}^{\eta} \frac{\sin(n+\frac{1}{2})s}{\sin\frac{1}{2}s} ds \right| &= \left| \frac{1}{\sin(\frac{1}{2n+1})} \int_{\frac{2}{2n+1}}^{\eta'} \sin\left(n+\frac{1}{2}\right) sds \right| \\
&= \frac{1}{\sin(\frac{1}{2n+1})} \left| \frac{1}{n+\frac{1}{2}} \left\{ \cos 1 - \cos\left(n+\frac{1}{2}\right)\eta' \right\} \right| \\
&\leq \frac{1}{(2n+1)\sin(\frac{1}{2n+1})} \\
&= \left\{ (2n+1)\left[\frac{1}{2n+1} - \frac{1}{3!}\left(\frac{1}{2n+1}\right)^3 + \cdots \right] \right\}^{-1} \\
&\leq \left\{ 1 - \frac{1}{3!}\left(\frac{1}{2n+1}\right)^2 \right\}^{-1} = \frac{54}{53},
\end{aligned}$$

and consequently,

$$\left| \int_{\delta}^{n} D_n(s)ds \right| \leq \frac{1}{2\pi}\left(2 + \frac{54}{53}\right).$$

Thus we may take it that $c = \frac{1}{2\pi}\left(2 + \frac{54}{53}\right)$. ■

Theorem 5.9.6 (Dirichlet–Jordan) *Let f be a BV function on $[-\pi,\pi]$; then* $\frac{1}{2}a_0 + \sum_{n=1}^{\infty}\{a_n \cos nt + b_n \sin nt\} = \lim_{n\to\infty} S_n(f,t) = \frac{1}{2}\{f(t-) + f(t+)\}$.

Proof Since f is the difference of two monotone increasing functions, we may assume without loss of generality that f is monotone increasing, and consider f as defined on $(-\pi,\pi]$ and then extend f to $\mathbb{R}$ as a periodic function with period 2π. Now fix $t \in [-\pi,\pi]$. Given that $\varepsilon > 0$, there is $\delta > 0$ such that $f(t+s) - f(t+) < \frac{\varepsilon}{2c}$ for $0 < s \leq \delta$, where c is the constant in Lemma 5.9.2. We choose δ small enough so that $f(t+s)$ is monotone increasing in s on $[0,\delta]$, if $f(t+0)$ is understood to be $f(t+)$. Then, $\int_0^{\delta}\{f(t+s) - f(t+)\}D_n(s)ds = \{f(t+\delta) - f(t+)\}\int_{\delta'}^{\delta} D_n(s)ds$ for some $\delta' \in [0,\delta]$ by the second-mean value theorem, and hence

$$\left| \int_0^{\delta} \{f(t+s) - f(t+)\}D_n(s)ds \right| < \frac{\varepsilon}{2c}\cdot c = \frac{\varepsilon}{2}.$$

Now,

$$\left|\int_0^{\pi} f(t+s)D_n(s)ds - \frac{1}{2}f(t+)\right| = \left|\int_0^{\pi}\{f(t+s)-f(t+)\}D_n(s)ds\right\}$$
$$\leq \left|\int_0^{\delta}\{f(t+s)-f(t+)\}D_n(s)ds\right| + \left|\int_{\delta}^{\pi}\{f(t+s)-f(t+)\}D_n(s)ds\right|$$
$$< \frac{\varepsilon}{2} + \left|\frac{1}{2\pi}\int_{\delta}^{\pi}\frac{f(t+s)-f(t+)}{\sin\frac{1}{2}s}\cdot\sin\left(n+\frac{1}{2}\right)sds\right| < \varepsilon$$

if n is sufficiently large, by the Riemann–Lebesgue lemma, because the function $s \mapsto \frac{f(t+s)-f(t+)}{\sin\frac{1}{2}s}$ is integrable on $[\delta,\pi]$. Thus $\lim_{n\to\infty}\int_0^{\pi} f(t+s)D_n(s)ds = \frac{1}{2}f(t+)$. Similarly, $\lim_{n\to\infty}\int_{-\pi}^{0} f(t+s)D_n(s)ds = \frac{1}{2}f(t-)$. Consequently, $\lim_{n\to\infty}\int_{-\pi}^{\pi} f(\mathrm{t+s})D_n(s)ds = \lim_{n\to\infty} S_n(f,t) = \frac{1}{2}\{f(t-)+f(t+)\}$. ■

5.10 Weak convergence

The concept of limit for sequences in a metric space is defined in Section 1.4 in terms of the metric of the space. When normed vector spaces are concerned, there is a weaker form of concept of limit for sequences, towards the introduction of which we now turn.

Suppose that X is a n.v.s. and $\{x_k\}$ a sequence in X. If $x \in X$ satisfies $\langle x, x^*\rangle = \lim_{k\to\infty}\langle x_k, x^*\rangle$ for every $x^* \in X^*$, x is called a **weak limit** of the sequence $\{x_k\}$; since X^* separates points of X, if x is a weak limit of $\{x_k\}$, it is the only weak limit of $\{x_k\}$, and hence is the weak limit of $\{x_k\}$ and is denoted by $w\text{-}\lim_{k\to\infty} x_k$. We often write $x_k \rightharpoonup x$ to indicate that $x = w\text{-}\lim_{k\to\infty} x_k$. To distinguish between weak limit and limit defined in terms of the norm of X, the latter is called the limit in norm and we employ notation $x = \lim_{k\to\infty} x_k$ or $x_k \to x$ to mean that x is the limit of $\{x_k\}$ in norm. If the weak (norm) limit of a sequence exists, the sequence is said to be **weakly convergent** (**convergent in norm**) or is said to **converge weakly** (in **norm**). Clearly, in a Hilbert space E, $x = w\text{-}\lim_{k\to\infty} x_k$ if and only if $(x,y) = \lim_{x\to\infty}(x_k, y)$ for all $y \in E$, and $x_k \to x$ implies that $x_k \rightharpoonup x$.

Proposition 5.10.1 *A weakly convergent sequence in a n.v.s. X is bounded.*

Proof Let $\{x_k\}$ be a weakly convergent sequence in X. For $k \in \mathbb{N}$, let l_k be the bounded linear functional on X^*, defined by $l_k(x^*) = \langle x_k, x^*\rangle$ for $x^* \in X^*$. Note that X^* is a Banach space and by Theorem 5.5.2, $\|l_k\| = \|x_k\|$ for $k \in \mathbb{N}$. Let $x = w\text{-}\lim_{k\to\infty} x_k$, then since $\lim_{k\to\infty}|l_k(x^*)| = |\langle x, x^*\rangle|$, $\sup_k |l_k(x^*)| < \infty$ for each $x^* \in X^*$. By the principle of uniform boundedness (Theorem 5.1.3), $\sup_k \|l_k\| = \sup_k \|x_k\| < \infty$. ■

Remark Proposition 5.10.1 is actually contained in Theorem 5.1.4.

Exercise 5.10.1 Show that a bounded sequence $\{x_k\}$ converges to x weakly in a n.v.s. X if and only if there is $S \subset X^*$ such that $\langle S \rangle$ is dense in X^* and $\langle x, x^* \rangle = \lim_{k\to\infty} \langle x_k, x^* \rangle$ for $x^* \in S$.

Exercise 5.10.2 Show that a sequence $\{x_n\}$ in a finite-dimensional n.v.s. X converges weakly if and only if it converges in norm.

Theorem 5.10.1 *Every bounded sequence $\{x_k\}$ in a Hilbert space E has a subsequence which converges weakly in E.*

Proof Let F be the closure of $\langle \{x_k\} \rangle$ in E, then F is a Hilbert space with inner product inherited from E. Put $\sup_k \|x_k\| = M < \infty$. We show first that $\{x_k\}$ has a subsequence which converges weakly in F.

Since $\{(x_k, x_1)\}_k$ is a bounded sequence in $\mathbb{C}$, there is a subsequence $\{x_k^{(1)}\}$ of $\{x_k\}$ such that $\lim_{k\to\infty}(x_k^{(1)}, x_1)$ exists. Suppose now that sequences $\{x_k^{(1)}\}, \ldots, \{x_k^{(n)}\}$ have been chosen so that each of them except the first is a subsequence of the preceding one and $\lim_{k\to\infty}(x_k^{(n)}, x_j)$ exists for $j = 1, \ldots, n$. Since $\{(x_k^{(n)}, x_{n+1})\}$ is bounded, there is a subsequence $\{x_k^{(n+1)}\}$ of $\{x_k^{(n)}\}$ such that $\lim_{k\to\infty}(x_k^{(n+1)}, x_{n+1})$ exists. Clearly, $\lim_{k\to\infty}(x_k^{(n+1)}, x_j)$ exists for $j = 1, \ldots, n$, because $\{x_k^{(n+1)}\}$ is a subsequence of $\{x_k^{(n)}\}$. We have therefore obtained a sequence $\{x_k^{(1)}\}, \{x_k^{(2)}\}, \ldots, \{x_k^{(n)}\}, \ldots$ of subsequences of $\{x_k\}$ such that $\{x_k^{(n+1)}\}$ is a subsequence of $\{x_k^{(n)}\}$ for each $n \in \mathbb{N}$ and where $\lim_{k\to\infty}(x_k^{(n)}, x_j)$ exists for $j = 1, \ldots, n$. Now, $\{x_k^{(k)}\}$ is a subsequence of $\{x_k\}$ and $\lim_{k\to\infty}(x_k^{(k)}, x_j)$ exists for each $j \in \mathbb{N}$. For convenience, put $y_k = x_k^{(k)}$ for $k \in \mathbb{N}$, then $\lim_{k\to\infty}(y_k, z)$ exists for $z \in \langle \{x_k\} \rangle$. Let $l(z) = \overline{\lim_{k\to\infty}(y_k, z)}$, then l is a linear functional on $\langle \{x_k\} \rangle$; obviously, $|l(z)| \le M\|z\|$ for $z \in \langle \{x_k\} \rangle$, hence l is bounded on $\langle \{x_k\} \rangle$, and can be extended uniquely to be a bounded linear functional on F, still denoted by l. By the Riesz representation theorem, there is unique $x \in F$ such that $l(u) = (u, x)$ for $u \in F$; in particular, for $z \in \langle \{x_k\} \rangle$, $(z, x) = \overline{\lim_{k\to\infty}(y_k, z)}$ i.e. $(x, z) = \lim_{k\to\infty}(y_k, z)$. Since $\langle \{x_k\} \rangle$ is dense in F, $y_k \rightharpoonup x$ in F, by Exercise 5.10.1.

We claim now that $y_k \rightharpoonup x$ in E. Let $u \in E$, then $u = z + v$, where $z \in F$ and $v \in F^\perp$, by Corollary 5.6.1. Thus,

$$(x, u) = (x, z + v) = (x, z) = \lim_{k\to\infty}(y_k, z) = \lim_{k\to\infty}(y_k, z + v) = \lim_{k\to\infty}(y_k, u),$$

and hence $y_k \rightharpoonup x$ in E. ■

Exercise 5.10.3 Suppose that $\{e_k\}$ is an orthonormal sequence in a Hilbert space E. Show that $e_k \rightharpoonup 0$, but 0 is not a limit of $\{e_k\}$ in norm. (Hint: for $x \in E$, $\sum_{k=1}^{\infty} |(x, e_k)|^2 \le \|x\|^2$.)

Exercise 5.10.4 (Cf. Example 2.7.2) Show that if $1 < p < \infty$, then $f_n \rightharpoonup f$ in $l^p(\Omega)$ if and only if $\sup_n \|f_n\|_p < \infty$ and $f_n(\omega) \to f(\omega)$ for all $\omega \in \Omega$.

Exercise 5.10.5 Suppose that X is a reflexive Banach space and $\{x_n\}$ is a bounded sequence in X. Assume that X^* is separable and let $\{x_1^*, x_2^*, \ldots\}$ be a countable dense set in X^*. Show that $\{x_n\}$ has a subsequence which converges weakly by the following steps.

(i) Show that $\{x_n\}$ has a subsequence $\{y_n\}$ such that $\lim_{n\to\infty}\langle y_n, x_k^*\rangle$ exists and is finite for all $k \in \mathbb{N}$.

(ii) Show that $\lim_{n\to\infty}\langle y_n, x^*\rangle$ exists and is finite for all $x^* \in X^*$.

(iii) Put $l(x^*) = \lim_{n\to\infty}\langle y_n, x^*\rangle$. Show that $l \in X^{**}$, and there is $x \in X$ such that $l(x^*) = \langle x, x^*\rangle$ for all $x^* \in X^*$.

Theorem 5.10.2 (Banach–Saks) *If $\{x_k\}$ is a bounded sequence in a Hilbert space E, then it has a subsequence $\{y_k\}$ such that* $\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n y_k$ *in norm exists.*

Proof There is a subsequence $\{z_k\}$ of $\{x_k\}$ and $x \in E$ such that $z_k \rightharpoonup x$, by Theorem 5.10.1. Let $\hat{z}_k = z_k - x$, then $\hat{z}_k \rightharpoonup 0$. Choose inductively a subsequence $\{\hat{y}_k\}$ of $\{\hat{z}_k\}$ so that

$$|(\hat{y}_1, \hat{y}_{n+1})| \le \frac{1}{n}, \ \ldots, \ |(\hat{y}_n, \hat{y}_{n+1})| \le \frac{1}{n}$$

for all $n \in \mathbb{N}$. Then,

$$\begin{aligned}\left\| n^{-1}\sum_{k=1}^n \hat{y}_k\right\|^2 &= n^{-2}\sum_{i=1}^n\sum_{j=1}^n (\hat{y}_i, \hat{y}_j)\\ &= n^{-2}\left\{\sum_{i=1}^n (\hat{y}_i, \hat{y}_i) + 2\sum_{1\le i<j\le n} \mathrm{Re}(\hat{y}_i, \hat{y}_j)\right\}\\ &\le n^{-2}\left\{nC + 2\sum_{j=2}^n\sum_{i=1}^{j-1} |(\hat{y}_i, \hat{y}_j)|\right\}\\ &\le n^{-2}\{nC + 2(n-1)\} < n^{-1}\{C+2\},\end{aligned}$$

where $C = \sup_n\{\|\hat{y}_n\|^2\} \le \sup_n(\|x_n\| + \|x\|)^2 < \infty$. Thus $\lim_{n\to\infty} n^{-1}\sum_{k=1}^n \hat{y}_k = 0$. We complete the proof by letting $y_k = \hat{y}_k + x$. ∎

Theorems 5.10.1 and 5.10.2 have already shown the relevance of weak convergence, in that in terms of weak convergence, bounded sets in a Hilbert space reveal a certain compactness property. We shall now apply Theorem 5.10.1 to prove a mean ergodic theorem of F. Riesz which shows that bounded linear operators from a Hilbert space into itself of a certain kind have eigenvalue 1 whose eigenspace can be explicitly described.

In the following, we fix a bounded linear operator T from a Hilbert space E into itself, having the property that $\|T^n\| \le \alpha < \infty$ for all $n \in \mathbb{N}$ for some $\alpha > 0$. Let $T_1 = T$ and $T_n = \frac{1}{n}\{T + T^2 + \cdots + T^n\}$ for $n \ge 2$, and for $x \in E$, put $x_n = T_n x$ for $n \in \mathbb{N}$.

Lemma 5.10.1 *If $x \in \overline{(1-T)E}$, then* $\lim_{n\to\infty}\|x_n\| = 0$.

Proof If $x \in (1-T)E$, i.e. $x = y - Ty$ for some y in E, then

$$x_n = (y-Ty)_n = \frac{1}{n}\{T(y-Ty) + \cdots + T^n(y-Ty)\} = \frac{1}{n}\{Ty - T^{n+1}y\},$$

and hence, $\|x_n\| \le \frac{2\alpha}{n}\|y\|$, from which $\|x_n\| \to 0$ follows. Now suppose that $x \in \overline{(1-T)E}$. Given $\varepsilon > 0$, there is $z \in (1-T)E$ such that $\|x - z\| < \frac{\varepsilon}{2\alpha}$. It is clear that $\|(x-z)_n\| \le \alpha\|x-z\| < \frac{\varepsilon}{2}$. Since $\|z_n\| \to 0$, by the first part of the proof, there is $n_0 \in \mathbb{N}$ such that $\|z_n\| < \frac{\varepsilon}{2}$ whenever $n \ge n_0$, hence, $\|x_n\| = \|z_n + (x-z)_n\| \le \|z_n\| + \|(x-z)_n\| < \|z_n\| + \frac{\varepsilon}{2} < \varepsilon$ whenever $n \ge n_0$. Thus $\|x_n\| \to 0$. ■

Lemma 5.10.2 *If x_∞ is the weak limit of a subsequence of $\{x_n\}$, then x_∞ is a fixed point of T, i.e. $Tx_\infty = x_\infty$.*

Proof Let $\{x_{n_k}\}$ be a subsequence of $\{x_n\}$ such that $w\text{-}\lim_{k\to\infty} x_{n_k} = x_\infty$. Since $TT_nx - T_nx = T_nTx - T_nx = T_n(Tx - x) = (Tx-x)_n$, $\|TT_nx - T_nx\| \to 0$, by Lemma 5.10.1 and hence $TT_{n_k}x \rightharpoonup x_\infty$. But, since for each $y \in E$, $(Tz, y) = (z, \hat{y})$ for some $\hat{y} \in E$ and for all $z \in E$ by the Riesz representation theorem, we have $(TT_{n_k}x, y) = (x_{n_k}, \hat{y}) \to (x_\infty, \hat{y}) = (Tx_\infty, y)$ and consequently, $TT_{n_k}x \rightharpoonup Tx_\infty$. We infer from this last fact and the fact that $TT_{n_k}x \rightharpoonup x_\infty$, that $T_{x_\infty} = x_\infty$. ■

To prepare for the statement of the mean ergodic theorem of Riesz, we shall say that a sequence $\{T_n\} \subset L(X,Y)$ **converges strongly** to $T \in L(X,Y)$ if $\lim_{n\to\infty} T_nx = Tx$ for all $x \in X$, where X and Y are n.v.s.'s over the same scalar field $\mathbb{C}$ or $\mathbb{R}$. To distinguish this mode of convergence, if $\lim_{n\to\infty}\|T_n - T\| = 0$, we say that T_n **converges in operator norm** to T.

Theorem 5.10.3 (Mean ergodic theorem of Riesz) *T_n converges strongly in $L(E)$ to a linear operator T_∞ with the property that $TT_\infty = T_\infty$.*

Proof For $x \in E$, $\|x_n\| = \|\frac{1}{n}\{Tx + \cdots + T^nx\}\| \le \alpha\|x\|$, hence $\{x_n\}$ is a bounded sequence in E. $\{x_n\}$ has a subsequence $\{x_{n_k}\}$ which converges weakly to x_∞ in E. We know from Lemma 5.10.2 that $Tx_\infty = x_\infty$, and hence $(x_\infty)_n = x_\infty$. We claim that $\lim_{n\to\infty} x_n = x_\infty$, i.e. x_n converges strongly to x_∞. Now, $x_n = (x_\infty + \{x - x_\infty\})_n = (x_\infty)_n + (x-x_\infty)_n = x_\infty + (x - x_\infty)_n$, thus $\|x_n - x_\infty\| = \|(x-x_\infty)_n\|$; to verify the claim it is sufficient to show that $x - x_\infty \in \overline{(1-T)E}$, by Lemma 5.10.1. To see this, let Y be the orthogonal complement of $(1-T)E$ in E and observe that $x - x_{k_n} = \frac{1}{n_k}\{(x - Tx) + \cdots + (x - T^{n_k}x)\}$ is in $(1-T)E$, because $(x - T^mx) = (1-T)(1 + T + \cdots + T^{m-1})x \in (1-T)E$ for each $m \in \mathbb{N}$; then for $y \in Y$, we have $(x - x_{n_k}, y) = 0$, which implies that

$$(x - x_\infty, y) = \lim_{k\to\infty}(x - x_{n_k}, y) = 0,$$

i.e. $x - x_\infty \in Y^\perp = \overline{(1-T)E}$. Thus we have shown that $\|x_n - x_\infty\| \to 0$. This last fact shows in particular that all weakly convergent subsequences of $\{x_n\}$ converge weakly to the same element x_∞. Let $x_\infty = T_\infty x$, then T_∞ is a linear operator from E

into E and is the strong limit of $\{T_n\}$, i.e. $T_\infty x = \lim_{n\to\infty} T_n x$. That T_∞ is a bounded linear operator follows from the Banach–Steinhaus theorem (Theorem 5.1.4). From Lemma 5.10.2, $Tx_\infty = x_\infty$ and consequently, $TT_\infty x = T_\infty x$, or $TT_\infty = T_\infty$. ■

Corollary 5.10.1 $TT_\infty = T_\infty = T_\infty T = T_\infty^2$.

Proof From $TT_\infty = T_\infty$, it follows that $T^n T_\infty = T_\infty$ and $T_n T_\infty = T_\infty$ for all $n \in \mathbb{N}$; by letting $n \to \infty$ in the last equality, we obtain $T_\infty^2 = T_\infty$. To see that $T_\infty T = T_\infty$, note first that $T_n T - T_n = \frac{1}{n}(T^{n+1} - T)$ and hence $\|T_n Tx - T_n x\| \leq \frac{2\alpha}{n}\|x\|$ for all $x \in E$; thus $T_\infty T = T_\infty$ follows. ■

Exercise 5.10.6 Show that 1 is an eigenvalue of T and $T_\infty E$ is the eigenspace of T belonging to the eigenvalue 1.

The well-known ergodic theorem of J. von Neumann is a consequence of Theorem 5.10.3, as we shall now show.

Let (Ω, Σ, p) be a probability space. A bijective map $T : \Omega \to \Omega$ is called a **flow** on (Ω, Σ, p) if T is measurable and measure preserving.

Theorem 5.10.4 (von Neumann mean ergodic theorem) *Suppose that T is a flow on a probability space (Ω, Σ, p). Define a linear operator $\widehat{T}$ from $L^2(\Omega, \Sigma, p)$ to itself by*

$$(\widehat{T}f)(\omega) = f \circ T(\omega),\ \omega \in \Omega,\ f \in L^2(\Omega, \Sigma, p);$$

and let $\widehat{T}_n = \frac{1}{n}\{\widehat{T} + \cdots + \widehat{T}^n\}$. Then for $f \in L^2(\Omega, \Sigma, p)$, $\widehat{T}_n f \to f^*$ in $L^2(\Omega, \Sigma, p)$. Furthermore, $\widehat{T}f^* = f^*$ i.e. $f^*(T\omega) = f^*(\omega)$ for a.e. $\omega \in \Omega$.

Proof Since T is a flow on (Ω, Σ, p), $\|\widehat{T}f\| = \|f\|$ for all $f \in L^2(\Omega, \Sigma, p)$. Hence $\|\widehat{T}\| = 1$ and $\|\widehat{T}^n\| \leq \|\widehat{T}\|^n = 1$ for all $n \in \mathbb{N}$. The theorem follows from Theorem 5.10.3. ■

A flow T on (Ω, Σ, p) is called an **ergodic** flow if for each $f \in L^2(\Omega, \Sigma, p)$, the element f^* in the conclusion of Theorem 5.10.4 is constant a.e. on Ω.

Corollary 5.10.2 *Suppose that T is an ergodic flow on (Ω, Σ, p) and $\widehat{T}$, $\widehat{T}_n$, $n \in \mathbb{N}$, are defined as in Theorem 5.10.4. Then for $f \in L^2(\Omega, \Sigma, p)$, $\widehat{T}_n f \to \int_\Omega f dp$ in $L^2(\Omega, \Sigma, p)$.*

Proof For $f \in L^2(\Omega, \Sigma, p)$, let f^* be as in Theorem 5.10.4. Since $\widehat{T}_n f \to f^*$ in $L^2(\Omega, \Sigma, p)$, $\widehat{T}_n f \to f^*$ in $L^1(\Omega, \Sigma, p)$ and, a fortiori, $\lim_{n\to\infty} \int_\Omega \widehat{T}_n f dp = \int_\Omega f^* dp$; but $\int_\Omega \widehat{T}f dp = \int_\Omega \widehat{T}^2 f dp = \cdots = \int_\Omega \widehat{T}_n f dp = \cdots = \int_\Omega f dp$, from the fact that T is measure preserving, hence $\int_\Omega f^* dp = \int_\Omega f dp$. Now that $f^* =$ constant a.e. implies $f^* = \int_\Omega f dp$ a.e. ■

6 L^p Spaces

L^p spaces are the most interesting examples of Banach spaces and play a salient role in modern analysis. In this chapter basic features of L^p spaces are studied; in particular, their dual spaces are identified. Special attention is directed towards $L^p(\Omega)$ where Ω is an open set in $\mathbb{R}^n$, for example, convolution and maximal function operators in L^p, are treated. An important class of function spaces, which is related to L^p spaces and was first introduced by S.L. Sobolev in his study of equations of mathematical physics, is briefly introduced in the last section of the chapter. Further study of this class of spaces is taken up in Chapter 7 by applying the method of Fourier integrals.

Some useful inequalities for functions in L^p spaces are collected in the first section for later reference. The second section on signed and complex measures is primarily preliminary in nature for this chapter, but it also has its own merit of interest, as is shown by the Riesz representation theorem in the concluding part of the section.

6.1 Some inequalities

Some inequalities which appear frequently in studies related to L^p spaces are collected here for later reference.

6.1.1 Markov inequality

Let $f \in L^p(\Omega, \Sigma, \mu)$, $1 \le p < \infty$, then

$$\mu(\{|f| \ge \lambda\}) \le \lambda^{-p} \|f\|_p^p, \tag{6.1}$$

for all $\lambda > 0$.

The inequality (6.1), called the **Markov inequality**, follows readily from the sequence of inequalities,

$$\lambda^p \mu(\{|f| \ge \lambda) \le \int_{\{|f| \ge \lambda\}} |f|^p d\mu \le \|f\|_p^p.$$

Remark Since $\lim_{\lambda\to\infty} \mu\{|f| \geq \lambda\} = 0$ by (6.1), it follows from Exercise 2.5.9 (iii) that $\lim_{\lambda\to\infty} \int_{\{|f|\geq\lambda\}} |f|^p d\mu = 0$, and hence

$$\lim_{\lambda\to\infty} \lambda^p \mu(\{|f| \geq \lambda\}) = 0. \tag{6.2}$$

6.1.2 Chebyshev inequality

Let $f \in L^2(\Omega, \Sigma, P)$, where (Ω, Σ, P) is a probability space, then the following **Chebyshev inequality** is a special case of (6.1):

$$P(\{|f - E(f)| \geq \lambda) \leq \lambda^{-2} \operatorname{Var}(f), \tag{6.3}$$

where $E(f) = \int_\Omega f dP$ and $\operatorname{Var}(f) = \int_\Omega |f - E(f)|^2 dP$.

Remark A measurable function f on a probability space is called a random variable. If $\int_\Omega f dP$ exists, it is called the expectation of the random variable f and is denoted by $E(f)$; if $E(f)$ is finite, $\int_\Omega |f - E(f)|^2 dP$ is called the variance of f and is denoted by $\operatorname{Var}(f)$. The significance of Chebyshev inequality in probability theory will become clear when the concept of independence is introduced in Chapter 7.

6.1.3 Jensen inequality

Suppose that φ is a convex function defined on $\mathbb{R}$, and f is an integrable function on a probability space (Ω, Σ, P), then

$$\varphi(E(f)) \leq E(\varphi \circ f). \tag{6.4}$$

This inequality is referred to as the **Jensen inequality**. For the verification of (6.4), let us put $x = E(f)$ and choose $m \in [\varphi'_-(x), \varphi'_+(x)]$. By Proposition 5.4.1 (iv),

$$\varphi(x) + m(y - x) \leq \varphi(y)$$

for all $y \in \mathbb{R}$, and hence,

$$\varphi(x) + m(f(\omega) - x) \leq \varphi(f(\omega)) \tag{6.5}$$

for all $\omega \in \Omega$. It follows from (6.5) that

$$\{\varphi \circ f\}^- \leq |\varphi(x)| + |m||f| + |mx|,$$

and therefore $\{\varphi \circ f\}^-$ is integrable; consequently, $\int_\Omega \varphi \circ f dP$ exists. We can then integrate both sides of (6.5) over Ω to obtain

$$\varphi(x) + m(E(f) - x) \leq E(\varphi \circ f),$$

which reduces to (6.4), because $x = E(f)$. Thus the Jensen inequality is verified.

Remark If (Ω, Σ, μ) is a finite measure space and f is integrable on (Ω, Σ, μ), then the Jensen inequality leads to

$$\varphi\left(\frac{1}{\mu(\Omega)}\int_\Omega f d\mu\right) \le \frac{1}{\mu(\Omega)}\int_\Omega \varphi \circ f d\mu. \tag{6.6}$$

In particular, $\left|\frac{1}{\mu(\Omega)}\int_\Omega f d\mu\right|^p \le \frac{1}{\mu(\Omega)}\int |f|^p d\mu$ for $1 \le p < \infty$.

6.1.4 Extended Hölder inequality

Suppose that $f_1, \ldots, f_n$, $n \ge 3$, are measurable functions on a measure space (Ω, Σ, μ) and let $p_1 \ge 1, p_2 \ge 1, \ldots, p_n \ge 1$ be extended real numbers such that $\sum_{i=1}^n p_i^{-1} = 1$, then the following **extended Hölder inequality** holds:

$$\int_\Omega \left|\prod_{i=1}^n f_i\right| d\mu \le \prod_{i=1}^n \|f_i\|_{p_i}. \tag{6.7}$$

To see that (6.7) holds, it is sufficient to consider the case where $n = 3$; then (6.7) follows inductively. So consider the case where $n = 3$ and let $p^{-1} = \frac{1}{p_1} + \frac{1}{p_2}$. Since p and p_3 are conjugate exponents, by the Hölder inequality, we have

$$\int_\Omega |f_1 f_2 f_3| d\mu \le \|f_1 f_2\|_p \cdot \|f_3\|_{p_3}. \tag{6.8}$$

Then put $p' = \frac{p_1}{p}$, $q' = \frac{p_2}{p}$ and apply the Hölder inequality, to obtain

$$\begin{aligned}\|f_1 f_2\|_p^p &= \int_\Omega |f_1|^p |f_2|^p d\mu \le \left(\int_\Omega |f_1|^{pp'} d\mu\right)^{1/p'} \left(\int_\Omega |f_2|^{pq'} d\mu\right)^{1/q'} \\ &= \|f_1\|_{p_1}^p \|f_2\|_{p_2}^p,\end{aligned}$$

or $\|f_1 f_2\|_p \le \|f_1\|_{p_1} \cdot \|f_2\|_{p_2}$. This last inequality and (6.8) imply that (6.7) holds when $n = 3$.

Exercise 6.1.1 Suppose that Ω is a measurable subset of $\mathbb{R}^n$ with $\lambda^n(\Omega) > 0$, and f is a measurable function on Ω. Show that $f \in L^p(\Omega)$, $1 \le p < \infty$, if and only if for every $\varepsilon > 0$, there is a closed set $F \subset \Omega$ and a bounded continuous function g in $L^p(\mathbb{R}^n)$, such that $\lambda^n(\Omega \backslash F) < \varepsilon$, $f = g$ on F, and $\|f - g\|_p < \varepsilon$. (Hint: cf. (6.2) and Theorem 4.1.3.)

Exercise 6.1.2 Suppose that $\{f_n\}_{n\in\mathbb{N}}$ is an orthonormal system in $L^2(\Omega, \Sigma, \mu)$. Show that for any $\varepsilon > 0$,

$$\lim_{n\to\infty} \mu\left(\left\{\frac{\left|\sum_{k=1}^n f_k\right|}{n} \ge \varepsilon\right\}\right) = 0.$$

Exercise 6.1.3 Suppose that $\{f_n\}$ is a sequence in $L^p(\Omega, \Sigma, \mu)$, $1 \leq p < \infty$, which converges in $L^p(\Omega, \Sigma, \mu)$ to f. Show that $\{f_n\}$ has a subsequence which converges a.e. to f. (Hint: there are positive integers $n_1 < n_2 < \cdots < n_k < \cdots$ such that $\mu(\{|f_{n_k} - f| \geq \frac{1}{k}\}) \leq \frac{1}{k^2}$ for each $k \in \mathbb{N}$.)

Exercise 6.1.4 Suppose that $\sum_{n=1}^{\infty} \alpha_n = 1$, where $\alpha_n \geq 0$ for each n. Show that if $\{\beta_n\}$ is a sequence of real numbers such that $\sum_{n=1}^{\infty} \alpha_n |\beta_n| < \infty$, then

$$\left| \sum_{n=1}^{\infty} \alpha_n \beta_n \right|^p \leq \sum_{n=1}^{\infty} \alpha_n |\beta_n|^p$$

for $1 \leq p < \infty$.

6.2 Signed and complex measures

So far the integration is taken with respect to a measure on a measurable space (Ω, Σ), where a measure is understood to be a nonnegative σ-additive set function defined on Σ. But there naturally appear set functions which may take negative values, such as electric charges, and integration with respect to such set functions is a useful construct, such as the potential of the electric charge distribution. Our purpose in this section is firstly to generalize the concept of measure to cover situations when negative values might be assumed, and then to consider complex measures. In order to do this, we extend the concept of sum for systems of real numbers in Section 1.1 to systems which may contain ∞ or $-\infty$. This can be done naturally as follows. Let $\{c_\alpha\}_{\alpha \in I}$ be a system of extended real numbers; by considering $\{c_\alpha\}_{\alpha \in I}$ as a function on I, we say that the sum of $\{c_\alpha\}$ exists if its integral with respect to the counting measure on I exists. This integral is called the sum of $\{c_\alpha\}$ and is denoted by $\sum_{\alpha \in I} c_\alpha$, or $\sum_\alpha c_\alpha$ if I is clearly implied (cf. Examples 2.3.1 and 2.3.3). Note that $\{c_\alpha\}$ is summable if and only if $\sum_\alpha c_\alpha$ exists and is finite.

Let (Ω, Σ) be a measurable space; a set function $\sigma : \Sigma \to \overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ is called a **signed measure** on (Ω, Σ) if

(i) $\sigma(\emptyset) = 0$;

(ii) if $\{A_n\} \subset \Sigma$ is a disjoint sequence, then the sum $\sum_n \sigma(A_n)$ exists and

$$\sigma\left(\bigcup_n A_n\right) = \sum_n \sigma(A_n). \tag{6.9}$$

We remark first that if $\sigma(\bigcup_{n=1}^{\infty} A_n)$ is finite, then $\sum_n \sigma(A_n)$ on The right-hand side of (6.9) can be written as $\sum_{n=1}^{\infty} \sigma(A_n)$ which necessarily converges absolutely, because $\bigcup_{n=1}^{\infty} A_n$ does not depend on the order of $A_1, A_2, \ldots$. Secondly, we call attention to the fact that condition (ii) in the above definition forces σ to satisfy condition (iii);

(iii) The signed measure σ does not take both ∞ and $-\infty$ as its value.

In fact, if $\sigma(A) = -\infty$, $\sigma(B) = \infty$ for A, B in Σ, then,

$$\sigma(A \cup B) = \sigma(A \cap B) + \sigma(A \cap B^c) + \sigma(B \cap A^c)$$

does not make sense, because $-\infty$ and ∞ both appear on the right-hand side in all possible situations, as can easily be seen.

For definiteness, we shall assume that in the sequel, condition (iii)$'$ holds;

(iii)$'$ $\sigma(A) > -\infty$ for all $A \in \Sigma$.

Under this assumption, if $\{A_n\}$ is a disjoint sequence in Σ with $\sigma(\bigcup_n A_n) = \infty$, then $\sum_{n=1}^{\infty} \sigma(A_n)$ diverges to ∞.

Measures on Σ are certainly signed measures; to distinguish them from general signed measures, we shall sometimes refer to them as **positive measures**. Accordingly, if $\sigma(A) \leq 0$ for all $A \in \Sigma$, σ is called a **negative measure**.

Example 6.2.1 Let (Ω, Σ, μ) be a measure space.

(i) Suppose that $A_1, \ldots, A_k$ are disjoint sets from Σ with $\mu(A_j) < \infty$, $j = 1, \ldots, k$ and let $\alpha_1, \ldots, \alpha_k$ be real numbers. Define σ on Σ by

$$\sigma(A) = \sum_{j=1}^{k} \alpha_j \mu(A \cap A_j), \quad A \in \Sigma.$$

The set function σ is obviously a signed measure.

(ii) Suppose that f is a measurable function with $\int_\Omega f^- d\mu < \infty$, then

$$\sigma(A) = \int_A f d\mu, \quad A \in \Sigma,$$

is a signed measure.

Remark Signed measure σ, defined in Example 6.2.1 (ii), is usually referred to as the **indefinite integral** of f; but when Ω is a metric space and $\mathcal{B}(\Omega) \subset \Sigma$, the indefinite integral of f is sometimes restricted to $\mathcal{B}(\Omega)$. This should not cause any confusion, because the definite meaning of an indefinite integral will be clear from the context (cf. Example 3.8.1).

Example 6.2.2 Consider the measurable space $(\mathbb{R}, \mathcal{B})$ where $\mathcal{B}$ is the σ-algebra of all Borel sets in $\mathbb{R}$. Suppose that we order the set of all rational numbers by $\gamma_1, \gamma_2, \ldots, \gamma_n, \ldots$, and define σ on $\mathcal{B}$ by

$$\sigma(B) = \sum_{\gamma_n \in B} (-1)^n \frac{1}{n^2}, \quad B \in \mathcal{B}.$$

Then σ is a signed measure which assumes only finite values.

Exercise 6.2.1 Verify the following statements. Let σ be a signed measure on (Ω, Σ). If $\{E_n\} \subset \Sigma$ and $E_n \nearrow$, then

$$\sigma\left(\lim_{n\to\infty} E_n\right) = \sigma\left(\bigcup_n E_n\right) = \lim_{n\to\infty} \sigma(E_n);$$

if, on the other hand, $E_n \searrow$ and $\sigma(E_n) < \infty$ for some n, then

$$\sigma\left(\lim_{n\to\infty} E_n\right) = \sigma\left(\bigcap_n E_n\right) = \lim_{n\to\infty} \sigma(E_n).$$

Exercise 6.2.2 Show that if $|\sigma(E)| < \infty$, then $|\sigma(F)| < \infty$ for $F \subset E$.

We currently show that any signed measure is the difference of two positive measures, one of which is a finite measure.

In the following discussion, a fixed signed measure σ on a measurable space (Ω, Σ) is considered.

A set $E \in \Sigma$ is said to be **positive** (**negative**) if $\sigma(A \cap E) \geq 0$ (≤ 0) for all $A \in \Sigma$. Obviously, any measurable subset of a positive (negative) set is positive (negative). The empty set $\emptyset$ is both positive and negative. Certainly, if $A_1, A_2, \ldots, A_n, \ldots$ are positive (negative), then so is $\bigcup_n A_n$.

The family of all positive sets will be denoted by $\mathcal{P}_\sigma$, and that of all negative sets by $\mathcal{N}_\sigma$.

Lemma 6.2.1 *Let $\beta = \inf_{E\in\mathcal{N}_\sigma} \sigma(E)$; then $-\infty < \beta \leq 0$ and there is $B \in \mathcal{N}_\sigma$ such that $\sigma(B) = \beta$.*

Proof There is a sequence $\{B_n\}$ in $\mathcal{N}_\sigma$ such that

$$\beta = \lim_{n\to\infty} \sigma(B_n).$$

Take $B = \bigcup_n B_n$, then $B \in \mathcal{N}_\sigma$, and for each k,

$$\sigma(B) = \sigma(B_k) + \sigma(B \backslash B_k) \leq \sigma(B_k),$$

hence $\sigma(B) \leq \lim_{k\to\infty} \sigma(B_k) = \beta$. But $\sigma(B) \geq \beta$, so $\sigma(B) = \beta$. Since $\sigma(B) > -\infty$, we have $-\infty < \beta \leq 0$. ■

Theorem 6.2.1 (Hahn decomposition theorem) *There are disjoint sets A and B in Σ such that*

(i) $A \cup B = \Omega$;

(ii) $A \in \mathcal{P}_\sigma$ *and* $B \in \mathcal{N}_\sigma$.

Proof Let β and B be as in Lemma 6.2.1, and take $A = \Omega \backslash B$. It remains to show that $A \in \mathcal{P}_\sigma$. Suppose the contrary. Then there is a measurable set $E_0 \subset A$ such that $\sigma(E_0) < 0$. Naturally E_0 is not negative, because otherwise $B \cup E_0$ would be negative and $\sigma(B \cup E_0) = \sigma(B) + \sigma(E_0) < \beta$, contrary to the choice of β. Let k_1 be the smallest positive integer such that E_0 contains a measurable set E_1 with $\sigma(E_1) \geq \frac{1}{k_1}$.

Now, since $\sigma(E_0\backslash E_1) = \sigma(E_0) - \sigma(E_1) \leq \sigma(E_0) - \frac{1}{k_1} < 0$, we can repeat the above argument with E_0 replaced by $E_0\backslash E_1$. So, let k_2 be the smallest positive integer such that $E_0\backslash E_1$ contains a measurable set E_2 with $\sigma(E_2) \geq \frac{1}{k_2}$. Continue in this fashion; we obtain a sequence of mutually disjoint measurable sets $E_1, E_2, \ldots, E_n, \ldots$ in E_0 and a sequence $k_1, k_2, \ldots, k_n, \ldots$ of positive integers such that for each $n \geq 2$, k_n is the smallest positive integer such that $E_0\backslash(E_1 \cup \cdots \cup E_{n-1})$ contains a measurable set E_n with $\sigma(E_n) \geq \frac{1}{k_n}$. Since $\bigcup_{n=1}^{\infty} E_n \subset E_0$ and $|\sigma(E_0)| < \infty$, $|\sigma(\bigcup_{n=1}^{\infty} E_n)| < \infty$ (see Exercise 6.2.2), and hence,

$$\sum_{n=1}^{\infty} \frac{1}{k_n} \leq \sum_{n=1}^{\infty} \sigma(E_n) = \sigma\left(\bigcup_{n=1}^{\infty} E_n\right).$$

Thus $\sum_{n=1}^{\infty} \frac{1}{k_n}$ is a convergent series, and as a consequence,

$$\lim_{n\to\infty} \frac{1}{k_n} = 0. \tag{6.10}$$

Let $F_0 = E_0\backslash \bigcup_{n=1}^{\infty} E_n$, then $\sigma(F_0) = \sigma(E_0) - \sum_{n=1}^{\infty} \sigma(E_n) \leq \sigma(E_0) < 0$. Consider a measurable set $F \subset F_0$; we claim that $\sigma(F) \leq 0$. If $\sigma(F) > 0$, then $\sigma(F) > \frac{1}{n_0}$ for some positive integer n_0; but (6.10) implies that $n_0 < k_n$ for sufficiently large n, thus contradicting the choice of k_n for such n's, because $F \subset E_0\backslash \bigcup_{k=1}^{n-1} E_k$ for all n. Thus, $\sigma(F) \leq 0$ and consequently F_0 is a negative set. But then $F_0 \cup B$ is negative and $\sigma(F_0 \cup B) < \beta$, contrary to the choice of β. The contradiction proves the theorem. ■

The pair (A, B) in the statement of Theorem 6.2.1 is called a **Hahn decomposition** of Ω relative to the signed measure σ, or simply a **σ-decomposition** of Ω. In general, Hahn decomposition is not unique.

Exercise 6.2.3 Let σ be the signed measure of Example 6.2.2. Find two Hahn decompositions of $\mathbb{R}$ relative to σ.

Lemma 6.2.2 shows a close relation between any two Hahn decompositions of Ω relative to a signed measure σ.

Lemma 6.2.2 *Let (A_1, B_1) and (A_2, B_2) be Hahn decompositions of Ω relative to the signed measure σ; then for any $E \in \Sigma$ the following relations hold:*

$$\sigma(E \cap A_1) = \sigma(E \cap A_2); \quad \sigma(E \cap B_1) = \sigma(E \cap B_2).$$

Proof Since $A_1\backslash A_2$ is positive, $\sigma(E \cap (A_1\backslash A_2)) \geq 0$; on the other hand $E \cap (A_1\backslash A_2) \subset B_2$ implies that $\sigma(E \cap (A_1\backslash A_2)) \leq 0$. Hence $\sigma(E \cap (A_1\backslash A_2)) = 0$; similarly, $\sigma(E \cap (A_2\backslash A_1)) = 0$. Now,

$$\begin{aligned}\sigma(E \cap A_1) &= \sigma(E \cap A_1) + \sigma(E \cap (A_2\backslash A_1))\\ &= \sigma(E \cap (A_1 \cup A_2)) = \sigma(E \cap A_2) + \sigma(E \cap (A_1\backslash A_2))\\ &= \sigma(E \cap A_2).\end{aligned}$$

Similarly, $\sigma(E \cap B_1) = \sigma(E \cap B_2)$. ■

For a Hahn decomposition (A, B) of Ω relative to σ, define for $E \in \Sigma$,

$$\sigma^+(E) = \sigma(E \cap A); \quad \sigma^-(E) = -\sigma(E \cap B); \quad \text{and } |\sigma|(E) = \sigma^+(E) + \sigma^-(E).$$

Obviously, σ^+, σ^-, and $|\sigma|$ are positive measures on Σ and are independent of the chosen Hahn decomposition (A, B), by Lemma 6.2.2. The measure $|\sigma|$ is called the **total variational measure** of σ, while σ^+ and σ^- are called respectively the **positive variational measure** and the **negative variational measure** of σ. Observe that $|\sigma(E)| \leq |\sigma|(E)$ for $E \in \Sigma$. Theorem 6.2.2 speaks for itself.

Theorem 6.2.2 *The measure σ^- is a finite positive measure and $\sigma = \sigma^+ - \sigma^-$. Furthermore, if σ is finite or σ-finite then so are σ^+ and $|\sigma|$.*

The decomposition $\sigma = \sigma^+ - \sigma^-$ is called the **Jordan decomposition** *of* σ.

Integrals and indefinite integrals of functions w.r.t. a signed measure σ are only defined for functions f in $L^1(\Omega, \Sigma, |\sigma|)$ by

$$\int_\Omega f d\sigma := \int_\Omega f d\sigma^+ - \int_\Omega f d\sigma^-;$$
$$\int_E f d\sigma := \int_E f d\sigma^+ - \int_E f d\sigma^-, \quad E \in \Sigma.$$

In the above definitions, f could be a complex-valued function.

Exercise 6.2.4 Show that for $E \in \Sigma$:

(i) $\sigma^+(E) = \max_{B \in \Sigma} \sigma(B \cap E)$;
(ii) $\sigma^-(E) = -\min_{B \in \Sigma} \sigma(B \cap E)$; and
(iii) $|\sigma|(E) = \sup\{\sum_{n=1}^{\infty} |\sigma(E_n)|\}$, where the supremum is taken over all decompositions of E into countable disjoint measurable sets $E_1, E_2, \ldots$.

Exercise 6.2.5 If σ is a finite signed measure, then

$$|\sigma|(E) = \sup \left| \int_E f d\sigma \right|,$$

where the supremum is taken over all measurable functions f with $|f| \leq 1$.

Exercise 6.2.6 Let σ be the signed measure in Example 6.2.1 (ii). Show that for $E \in \Sigma$, we have

$$\sigma^+(E) = \int_E f^+ d\mu; \quad \sigma^-(E) = \int_E f^- d\mu; \quad \text{and } |\sigma|(E) = \int_E |f| d\mu.$$

Also find a Hahn decomposition of Ω relative to σ.

Exercise 6.2.7 Let σ be a signed measure and $\sigma = \sigma_1 - \sigma_2$, where σ_1 and σ_2 are positive measures with σ_2 a finite measure. Show that there is a positive finite measure μ on Σ such that $\sigma_1 = \sigma^+ + \mu$ and $\sigma_2 = \sigma^- + \mu$. (Hint: use Exercise 6.2.4)

Remark The conclusion of Exercise 6.2.7 means that the Jordan decomposition of a signed measure σ is **the minimal decomposition** of σ into the difference of two positive measures. For the corresponding fact concerning decomposition of functions of bounded variation into the difference of two monotone increasing functions, see the paragraph following Theorem 4.4.1.

Now let μ be a positive measure on (Ω, Σ). A signed measure σ on Σ is said to be **μ-absolutely continuous** if $\sigma(A) = 0$ whenever $A \in \Sigma$ and $\mu(A) = 0$. It is easily verified that σ is μ-absolutely continuous if and only if σ^+, σ^- are μ-absolutely continuous; thus, σ is μ-absolutely continuous if and only if $|\sigma|$ is μ-absolutely continuous.

Theorem 6.2.3 *If σ is a finite signed measure, then σ is μ-absolutely continuous if and only if for any given $\varepsilon > 0$ there is $\delta > 0$ such that if $A \in \Sigma$ with $\mu(A) < \delta$, then $|\sigma|(A) < \varepsilon$.*

Proof Sufficiency is obvious.

Necessity: Suppose the contrary. Then for some $\varepsilon > 0$ and for any $n \in \mathbb{N}$, there is $A_n \in \Sigma$ such that $\mu(A_n) < 2^{-n}$ and $|\sigma|(A_n) \geq \varepsilon$. Let $A = \limsup_{n\to\infty} A_n$, then for each n,

$$\mu(A) = \mu\left(\lim_{n\to\infty} \bigcup_{k\geq n} A_k\right) \leq \mu\left(\bigcup_{k\geq n} A_k\right) < \sum_{k\geq n} 2^{-k};$$

letting $n \to \infty$, we then have $\mu(A) = 0$. But,

$$|\sigma|(A) = \lim_{n\to\infty} |\sigma|\left(\bigcup_{k\geq n} A_k\right) \geq \limsup_{n\to\infty} |\sigma|(A_n) \geq \varepsilon,$$

which contradicts the fact that $|\sigma|$ is μ-absolutely continuous. ■

Theorem 6.2.4 (Radon–Nikodym) *If (Ω, Σ, μ) is a σ-finite measure space and σ is a σ-finite μ-absolutely continuous signed measure on (Ω, Σ), then there is a unique measurable function f such that $\int_\Omega f^- d\mu < \infty$, and*

$$\sigma(A) = \int_A f d\mu, \quad A \in \Sigma.$$

Proof We know that σ^+ is σ-finite and σ^- is finite on Σ. By Exercise 5.7.1, there is $f_2 \in L^1(\Omega, \Sigma, \mu)$ such that $f_2 \geq 0$, and

$$\sigma^-(A) = \int_A f_2 d\mu, \quad A \in \Sigma;$$

and there is a measurable function f_1 with $f_1 \geq 0$ such that

$$\sigma^+(A) = \int_A f_1 d\mu, \quad A \in \Sigma.$$

Let $f = f_1 - f_2$, then

$$\sigma(A) = \int_A f d\mu, \quad A \in \Sigma.$$

One can verify (cf. Exercise 6.2.6) that $f^- = f_2$ a.e., hence $\int_\Omega f^- d\mu < \infty$. That f is unique is left as an exercise. ■

Exercise 6.2.8 Show that the function f in Theorem 6.2.4 is unique.

Now complex measures are introduced. Fix a measurable space (Ω, Σ); a set function $\sigma : \Sigma \to \mathbb{C}$ is called a **complex measure** if (i) $\sigma(\emptyset) = 0$; and (ii) $\sigma(\bigcup_n A_n) = \sum_{n=1}^\infty \sigma(A_n)$ for every disjoint sequence $\{A_n\}$ in Σ. Observe that in (ii) the convergence of $\sum_{n=1}^\infty \sigma(A_n)$ does not depend on how the sequence $\{A_n\}$ is ordered, hence for any disjoint sequence $\{A_n\} \subset \Sigma$, $\sum_{n=1}^\infty |\sigma(A_n)| < \infty$. We take a hint from Exercise 6.2.4 (iii) to define the total variational measure $|\sigma|$ of a complex measure by

$$|\sigma|(E) = \sup\left\{\sum_{n=1}^\infty |\sigma(E_n)|\right\}$$

for $E \in \Sigma$, where the supremum is taken over all decompositions of E into countable disjoint measurable sets $E_1, E_2, \ldots$. When σ is a signed or complex measure on $\mathcal{B}(X)$, where X is a metric space, it is called a **Radon (Riesz) measure** if $|\sigma|^*$ is a Radon (Riesz) measure on X. Recall that $|\sigma|^*$ is the measure on X constructed from $|\sigma|$ by Method I.

Exercise 6.2.9 Show that the family of all complex Riesz measures on $\mathcal{B}(X)$ is a complex vector space.

For $A \in \Sigma$, let us put $\sigma_r(A) = \operatorname{Re}\sigma(A)$ and $\sigma_i(A) = \operatorname{Im}\sigma(A)$; then σ_r and σ_i are finite signed measures on Σ. If f is a complex-valued $|\sigma|$-integrable function on Ω, the **σ-integral** of f is defined by

$$\int_X f d\sigma := \int_X f d\sigma_r + i \int_X f d\sigma_i.$$

Suppose now that μ is a positive measure on Σ. A complex measure σ on Σ is **μ-absolutely continuous**, if $A \in \Sigma$ and $\mu(A) = 0$ implies $\sigma(A) = 0$. Obviously, σ is μ-absolutely continuous if and only if both σ_r and σ_i are μ-absolutely continuous.

Exercise 6.2.10 A complex measure σ on Σ is μ-absolutely continuous if and only if for any $\varepsilon > 0$, there is $\delta > 0$ such that if $A \in \Sigma$ with $\mu(A) < \delta$, then $|\sigma(A)| < \varepsilon$.

By applying Theorem 6.2.4 to σ_r and σ_i we obtain Theorem 6.2.5:

Theorem 6.2.5 *If (Ω, Σ, μ) is a σ-finite measure space and σ is a μ-absolutely continuous complex measure on Σ, then there is a unique μ-integrable function f on Ω such that*

$$\sigma(A) = \int_A f d\mu, \quad A \in \Sigma.$$

Henceforth, both Theorem 6.2.4 and Theorem 6.2.5 are to be referred to as the Radon–Nikodym theorem. We note in passing that the family of complex measures on Σ includes all finite signed measures on Σ.

As an application of the notion of signed (complex) measure, we present in the final part of this section the Riesz representation theorem for linear functions on $C_0(X)$; the space of all continuous functions vanishing at infinity on the locally compact metric space X. A function f on X is said to be **vanishing at infinity** if for any $\varepsilon > 0$ there is a compact set K such that $|f(x)| < \varepsilon$ for $x \in K^c$. The space $C_0(X)$ is a real or complex vector space, depending on whether the functions in question are real or complex-valued. Equipped with the norm defined by

$$\|f\| = \sup_{x \in X} |f(x)|$$

for $f \in C_0(X)$, $C_0(X)$ is a normed vector space; clearly, $\|f\| = \max_{x \in X} |f(x)|$. The norm so defined on $C_0(X)$ is usually referred to as the **uniform norm**; and unless otherwise specified, $C_0(X)$ is equipped with this norm. For definiteness, we assume that functions in $C_0(X)$ are real-valued and hence $C_0(X)$ is a real vector space.

Exercise 6.2.11

(i) Show that $C_0(X)$ is a Banach space.

(ii) Show that if $f \in C_0(X)$, then both f^+ and f^- are in $C_0(X)$.

If ℓ is a positive linear functional on $C_0(X)$, it is, a fortiori, positive on $C_c(X)$; the measure μ constructed in Section 3.10 for ℓ considered as restricted to $C_c(X)$ is also referred to as the **measure for ℓ**. As we know in Section 3.10, μ is the unique Riesz measure on X such that

$$\ell(f) = \int_X f d\mu$$

for all $f \in C_c(X)$.

Lemma 6.2.3 *Suppose that ℓ is a bounded positive linear functional on $C_0(X)$ and μ is the measure for ℓ; then $\ell(f) = \int_X f d\mu$ for $f \in C_0(X)$ and $\|\ell\| = \mu(X)$.*

Proof Since ℓ is bounded, μ is a finite measure (cf. Exercise 3.10.1).

For $f \in C_0(X)$ and $\varepsilon > 0$, there is a compact set K in X such that $|f(x)| < \varepsilon$ for $x \in K^c$. By Corollary 1.10.1, there is $g \in U_c(X)$ satisfying $g = 1$ on K. Put $h = fg$, then $h \in C_c(X)$, and

$$\begin{aligned}\ell(f) &= \ell(f-h) + \ell(h) = \ell(f-h) + \int_X h d\mu \\ &= \ell(f-h) + \int_X f d\mu + \int_X (h-f) d\mu;\end{aligned}$$

hence,

$$\left|\ell(f) - \int_X f d\mu\right| \leq \|\ell\|\varepsilon + \varepsilon\mu(X),$$

because $\|f-h\| = \|f(1-g)\| \leq \sup_{x\in K^c} |f(x)| \leq \varepsilon$. By letting $\varepsilon \searrow 0$, we obtain

$$\ell(f) = \int_X f d\mu.$$

Now, if $f \in C_0(X)$ with $\|f\| = 1$, then

$$|\ell(f)| = \left|\int_X f d\mu\right| \leq \int_X |f| d\mu \leq \mu(X),$$

and consequently $\|\ell\| \leq \mu(X)$. On the other hand, for a compact set K in X, there is a function $f \in U_c(X)$ such that $f = 1$ on K (again by Corollary 1.10.1); then,

$$\mu(K) \leq \int_X f d\mu = \ell(f) \leq \|\ell\|,$$

from which $\mu(X) \leq \|\ell\|$ follows by the inner regularity of μ. Thus $\|\ell\| = \mu(X)$. ■

Suppose now that $\ell \in C_0(X)^*$; we shall decompose ℓ as a difference of two bounded positive linear functionals on $C_0(X)$ as follows.

Denote by $C_0(X)^+$ the family $\{f \in C_0(X) : f \geq 0\}$ and define a functional ℓ^+ on $C_0(X)^+$ by

$$\ell^+(f) = \sup\{\ell(g) : g \in C_0(X)^+ \text{ and } g \leq f\}$$

for $f \in C_0(X)^+$; since $\ell^+(f) \geq \ell(0) = 0$ and

$$\ell(g) \leq \|\ell\| \cdot \|g\| \leq \|\ell\| \cdot \|f\| < \infty$$

for $g \in C_0(X)^+$ satisfying $g \leq f$, ℓ^+ is nonnegative and $\ell^+(f) \leq \|\ell\| \cdot \|f\| < \infty$. Note that ℓ^+ is **positively homogeneous** on $C_0(X)^+$ in the sense that for $f \in C_0(X)^+$ and nonnegative number α, $\ell^+(\alpha f) = \alpha\ell^+(f)$.

Lemma 6.2.4 *The functional ℓ^+ is additive, i.e. if f and g are in $C_0(X)^+$, then $\ell^+(f+g) = \ell^+(f) + \ell^+(g)$.*

Proof Let u, v in $C_0(X)^+$ be such that $u \leq f$ and $v \leq g$, then $0 \leq u + v \leq f + g$, and hence,

$$\ell^+(f+g) \geq \ell(u+v) = \ell(u) + \ell(v),$$

from which it follows that

$$\ell^+(f+g) \geq \ell^+(f) + \ell^+(g).$$

On the other hand, if $u \in C_0(X)^+$ with $u \leq f + g$, by putting $u_1 = u \wedge f$ and $u_2 = u - u_1$, one verifies easily that

$$u = u_1 + u_2,\ u_1 \leq f, \text{ and } u_2 \leq g;$$

and thus,

$$\ell(u) = \ell(u_1) + \ell(u_2) \leq \ell^+(f) + \ell^+(g),$$

implying that $\ell^+(f+g) \leq \ell^+(f) + \ell^+(g)$. ■

Now, extend ℓ^+ to $C_0(X)$ by defining

$$\ell^+(f) = \ell^+(f^+) - \ell^+(f^-)$$

for $f \in C_0(X)$. For $f \in C_0(X)$, note that both f^+ and f^- are in $C_0(X)^+$ (cf. Exercise 6.2.11 (ii)) and observe that if $f = g - h$, with g and h being in $C_0(X)^+$, then $g = f^+ + u$ and $h = f^- + u$ for some $u \in C_0(X)^+$, and hence,

$$\ell^+(f) = \ell^+(g) - \ell^+(h).$$

Therefore if f and g are in $C_0(X)$, we have

$$\begin{aligned}
\ell^+(f+g) &= \ell^+(f^+ + g^+) - \ell^+(f^- + g^-) \\
&= \ell^+(f^+) + \ell^+(g^+) - \ell^+(f^-) - \ell^+(g^-) \\
&= \ell^+(f) + \ell^+(g),
\end{aligned}$$

i.e. ℓ^+ is additive on $C_0(X)$. Obviously,

$$\ell^+(\alpha f) = \alpha \ell^+(f),$$

for $f \in C_0(X)$ and $\alpha \in \mathbb{R}$. Thus ℓ^+ is a positive linear functional on $C_0(X)$. Since

$$|\ell^+(f)| \leq \ell^+(f^+) + \ell^+(f^-) \leq \|\ell\|(\|f^+\| + \|f^-\|) \leq 2\|\ell\| \cdot \|f\|,$$

ℓ^+ is a bounded positive linear functional on $C_0(X)$.

If we let $\ell^- = \ell^+ - \ell$, then $\ell^- \in C_0(X)^*$ and $\ell = \ell^+ - \ell^-$. Since for $f \in C_0(X)^+$ we have $\ell^-(f) = \ell^+(f) - \ell(f) \geq 0$, ℓ^- is a bounded positive linear functional on $C_0(X)$. Thus, every $\ell \in C_0(X)^*$ can be decomposed as the difference $\ell^+ - \ell^-$ of two bounded positive linear functionals on $C_0(X)$. Let μ_+ and μ_- be respectively the measure for ℓ^+ and ℓ^-, and for $B \in \mathcal{B}(X)$ put $\mu(B) = \mu_+(B) - \mu_-(B)$, then μ is a finite signed measure on $\mathcal{B}(X)$ and

$$\ell(f) = \int_X f d\mu, \quad f \in C_0(X). \tag{6.11}$$

Denote as before the total variational measure of μ on $\mathcal{B}(X)$ by $|\mu|$, and let $|\mu|^*$ be the measure on X constructed from $|\mu|$ by Method I. We know from Corollary 3.4.1 that $|\mu|^*$ is the unique Borel regular measure extending $|\mu|$, and, since $|\mu|^*$ is finite, it is a Radon measure. We shall see presently that $|\mu|^*$ is a Riesz measure. For this purpose, set for the moment $\nu = \mu_+ + \mu_-$, then ν is a Riesz measure on X and $|\mu|^* \leq \nu$. Given that $B \in \mathcal{B}(X)$ and $\varepsilon > 0$, by outer regularity of ν and Proposition 3.10.1 there are $K \in \mathcal{K}$ and $G \in \mathcal{G}$ with $K \subset B \subset G$ such that $\nu(G \backslash K) < \varepsilon$ and, a fortiori, $|\mu|^*(G \backslash K) < \varepsilon$; consequently, $|\mu|^*(G) - \varepsilon < |\mu|^*(B) < |\mu|^*(K) + \varepsilon$, which in turn implies that

$$|\mu|^*(B) = \sup\{|\mu|^*(K) : K \in \mathcal{K}, K \subset B\} \tag{6.12}$$

and

$$|\mu|^*(B) = \inf\{|\mu|^*(G) : G \in \mathcal{G}, B \subset G\}.$$

Now for any $S \subset X$, there is $B \in \mathcal{B}(X)$ such that $B \supset S$ and $|\mu|^*(S) = |\mu|^*(B) = \inf\{|\mu|^*(G) : G \in \mathcal{G}, B \subset G\} \geq \inf\{|\mu|^*(G) : G \in \mathcal{G}, S \subset G\} \geq |\mu|^*(S)$; thus,

$$|\mu|^*(S) = \inf\{|\mu|^*(G) : G \in \mathcal{G}, S \subset G\},$$

i.e. $|\mu|^*$ is outer regular. Note that (6.12) implies in particular that $|\mu|^*$ is inner regular; hence $|\mu|^*$ is a Riesz measure on X and μ is a Riesz measure on $\mathcal{B}(X)$. This last fact and (6.11) prove the following Lemma 6.2.5.

Lemma 6.2.5 *For $\ell \in C_0(X)^*$ there is a finite Riesz measure μ on $\mathcal{B}(X)$ such that* (6.11) *holds.*

Lemma 6.2.6 *Suppose that μ is a finite Riesz measure on $\mathcal{B}(X)$. Define a linear functional ℓ on $C_0(X)$ by*

$$\ell(f) = \int_X f d\mu, \quad f \in C_0(X).$$

Then, $\ell \in C_0(X)^*$ and $\|\ell\| = |\mu|(X)$.

Proof For $f \in C_0(X)$,

$$|\ell(f)| = \left|\int_X f d\mu^+ - \int_X f d\mu^-\right| \leq \int_X |f| d\mu^+ + \int_X |f| d\mu^-$$
$$= \int_X |f| d|\mu| \leq \|f\| |\mu|(X),$$

hence, $\ell \in C_0(X)^*$ and $\|\ell\| \leq |\mu|(X)$.

Let (A, B) be a Hahn decomposition of X w.r.t. μ. Since $|\mu|^*$ is a finite Riesz measure on X, by Proposition 3.10.1, there are K_1 and K_2 in $\mathcal{K}$ with $K_1 \subset A$ and $K_2 \subset B$ such that $|\mu|^*(X \backslash (K_1 \cup K_2)) < \varepsilon$. Take a continuous function g on X such that $-1 \leq g \leq 1$, $g = 1$ on K_1 and $g = -1$ on K_2 according to Corollary 1.8.1, and a function $h \in U_c(X)$ such that $h = 1$ on $K_1 \cup K_2$ according to Corollary 1.10.1, and let $f = gh$; then, $f \in C_c(X)$, $-1 \leq f \leq 1$, $f = 1$ on K_1, and $f = -1$ on K_2. Now,

$$\begin{aligned}
\ell(f) = \int_X f d\mu &= \mu(K_1) - \mu(K_2) + \int_{X \backslash (K_1 \cup K_2)} f d\mu \\
&= |\mu|(K_1) + |\mu|(K_2) + \int_{X \backslash (K_1 \cup K_2)} f d\mu \\
&\geq |\mu|(K_1 \cup K_2) - \int_{X \backslash (K_1 \cup K_2)} |f| d|\mu| \\
&\geq |\mu|(X) - 2|\mu|(X \backslash (K_1 \cup K_2)) \\
&\geq |\mu|(X) - 2\varepsilon,
\end{aligned}$$

from which, since $\|f\| = 1$, it follows that $\|\ell\| \geq |\mu|(X) - 2\varepsilon$ and hence $\|\ell\| \geq |\mu|(X)$. Thus, $\ell \in C_0(X)^*$ and $\|\ell\| = |\mu|(X)$, because we already know that $\|\ell\| \leq |\mu|(X)$. ■

Theorem 6.2.6 (Riesz representation theorem) *For $\ell \in C_0(X)^*$ there is a unique finite Riesz measure μ on $\mathcal{B}(X)$ such that*

$$\ell(f) = \int_X f d\mu \tag{6.13}$$

for $f \in C_0(f)$.

Proof The existence of Riesz measure μ on $\mathcal{B}(X)$ such that (6.13) holds follows from Lemma 6.2.5. Suppose that μ_1 and μ_2 are Riesz measures on $\mathcal{B}(X)$ such that (6.13) holds, with μ replaced by either μ_1 or μ_2. Then $\mu_1 - \mu_2$ is a Riesz measure on $\mathcal{B}(X)$ (cf. Exercise 6.2.9) such that

$$\int_X f d(\mu_1 - \mu_2) = 0$$

for all $f \in C_0(X)$; it follows then from Lemma 6.2.6 that $|\mu_1 - \mu_2|(X) = 0$, i.e. $|\mu_1 - \mu_2|$ is a zero measure on $\mathcal{B}(X)$. But, for $B \in \mathcal{B}(X)$, $0 = |\mu_1 - \mu_2|(B) \geq |\mu_1(B) - \mu_2(B)|$ implies that $\mu_1(B) = \mu_2(B)$. Thus the uniqueness of μ is proved. ■

Example 6.2.3 Let ℓ be a bounded linear functional on the real space $C[0, 1]$. Then there is a BV function g on $[a, b]$ such that g is right-continuous except at 0, and

$$\ell(f) = \int_0^1 f dg, \quad f \in C[0, 1].$$

Actually, let μ be the unique Riesz measure on $\mathcal{B}([0, 1])$ such that $\int_0^1 f d\mu = \ell(f)$ for $f \in C[0, 1]$, and let $g(0) = 0$ and $g(t) = \mu([0, t])$, $t \in [0, 1]$; then g is right-continuous except at 0. Consider any partition $0 = t_0 < t_1 < \cdots < t_n = 1$; we have $\sum_{k=1}^n |g(t_k) - g(t_{k-1})| = \sum_{k=1}^n |\mu((t_{k-1}, t_k])| + |\mu(\{0\})| \leq |\mu|([0, 1])$. Therefore g is a BV function. Clearly, $\ell(f) = \int_0^1 f dg$ for $f \in C[0, 1]$.

In the above discussion we assume that $C_0(X)$ is formed from real-valued functions; a brief account will now be given of the case when $C_0(X)$ consists of complex-valued functions. Recall that a complex-valued function f can be expressed as $\mathrm{Re} f + i \,\mathrm{Im} f$, where $\mathrm{Re} f$ and $\mathrm{Im} f$ are respectively the real and imaginary parts of f. Suppose that $\ell \in C_0(X)^*$, then

$$\ell(f) = \ell_r(f) + i\ell_i(f),$$

where $\ell_r(f) = \mathrm{Re}\{\ell(f)\}$ and $\ell_i(f) = \mathrm{Im}\{\ell(f)\}$; ℓ_r and ℓ_i are bounded linear functionals on $C_0(X)$ considered as a real vector space; in particular, they are bounded linear functionals on the real vector space of all real-valued functions in $C_0(X)$. By Theorem 6.2.6 there is a unique pair (μ_r, μ_i) of finite Riesz signed measures on $\mathcal{B}(X)$ such that

$$\ell(f) = \int_X f d\mu_r + i \int_X f d\mu_i$$

for real-valued functions f in $C_0(X)$. Let us put $\mu(B) = \mu_r(B) + i\mu_i(B)$ for $B \in \mathcal{B}(X)$; then μ is a complex Riesz measure on $\mathcal{B}(X)$, and for $f \in C_0(X)$ we have

$$\begin{aligned}
\ell(f) &= \ell(\mathrm{Re} f + i \,\mathrm{Im} f) = \ell(\mathrm{Re} f) + i\ell(\mathrm{Im} f) \\
&= \ell_r(\mathrm{Re} f) + i\ell_i(\mathrm{Re} f) + i\{\ell_r(\mathrm{Im} f) + i\ell_i(\mathrm{Im} f)\} \\
&= \int_X \mathrm{Re} f d\mu_r + i \int_X \mathrm{Re} f d\mu_i + i \int_X \mathrm{Im} f d\mu_r - \int_X \mathrm{Im} f d\mu_i \\
&= \int_X \mathrm{Re} f d\mu + i \int_X \mathrm{Im} f d\mu = \int_X f d\mu.
\end{aligned}$$

We leave it as an exercise to show the uniqueness of the Riesz measure μ on $\mathcal{B}(X)$ such that $\ell(f) = \int_X f d\mu$ for $f \in C_0(X)$, as well as the fact that $\|\ell\| = |\mu|(X)$. Hence, Theorem 6.2.6 also holds when the functions in $C_0(X)$ are complex-valued.

Exercise 6.2.12 When $C_0(X)$ consists of complex-valued functions and $\ell \in C_0(X)^*$, show that there is a unique Riesz measure on X such that

$$\ell(f) = \int_X f d\mu, \quad f \in C_0(X).$$

Furthermore, show that for such a measure, $\|\ell\| = |\mu|(X)$.

6.3 Linear functionals on L^p

Let p and q be conjugate exponents i.e. $p, q \geq 1$ and $p^{-1} + q^{-1} = 1$. We shall consider a fixed measure space (Ω, Σ, μ) throughout this section, therefore measurability of sets or functions is in reference to this measure space and the measure of a set A means $\mu(A)$ with $A \in \Sigma$. The space $L^p(\Omega, \Sigma, \mu)$ will be simply denoted by L^p for $p \geq 1$, and L^p-norm of f will be denoted by $\|f\|_p$.

Our purpose in this section is to identify $(L^p)^*$ with L^q in a sense to be specified later when μ is σ-finite and $p < \infty$.

For $g \in L^q$, define a linear functional ℓ_g on L^p by

$$\ell_g(f) = \int fg d\mu, \quad f \in L^p.$$

It follows from the Hölder inequality that ℓ_g is a bounded linear functional on L^p and its norm $\|\ell_g\| \leq \|g\|_q$.

We shall actually show that $\|\ell_g\| = \|g\|_q$ if $q < \infty$; and that this equality holds for all $q \geq 1$ if (Ω, Σ, μ) is σ-finite. This means that we may consider L^q as isometrically and isomorphically embedded in $(L^p)^*$ in either case, because the map $g \mapsto l_g$ is a linear map from L^q into $(L^p)^*$.

Lemma 6.3.1 *If $q < \infty$ and $g \in L^q$, then $\|g\|_q = \|\ell_g\|$.*

Proof We may assume that $g \neq 0$ on a set of positive measure, and let

$$f = \frac{|g|^{q-1}\overline{\operatorname{sgn} g}}{\|g\|_q^{q-1}},$$

where $\operatorname{sgn} g(x) = 0$ if $g(x) = 0$, and $= g(x)/|g(x)|$ if $g(x) \neq 0$. One sees easily that $\operatorname{sgn} g$ is a measurable function and $f \in L^p$ with $\|f\|_p = 1$. Now,

$$\|\ell_g\| \geq \left| \int fg d\mu \right| = \|g\|_q^{-(q-1)} \int |g|^q d\mu = \|g\|_q,$$

This, together with $\|\ell_g\| \leq \|g\|_q$, shows that $\|\ell_g\| = \|g\|_q$. ■

Corollary 6.3.1 *If (Ω, Σ, μ) is σ-finite and $g \in L^q$, then $\|\ell_g\| = \|g\|_q$.*

Proof We need only to prove that $\|\ell_g\| \geq \|g\|_\infty$ for $g \in L^\infty$. For this purpose we may assume that $\|g\|_\infty > 0$ and for a given $0 < \varepsilon < \|g\|_\infty$, let $A = \{|g| \geq \|g\|_\infty - \varepsilon\}$. From the definition of $\|g\|_\infty$, $\mu(A) > 0$. Since μ is σ-finite, there is an increasing sequence $\{\Omega_n\} \subset \Sigma$ such that $\mu(\Omega_n) < \infty$ for each n and $\lim_{n\to\infty} \Omega_n = \Omega$. Then, $\mu(A) = \lim_{n\to\infty} \mu(A \cap \Omega_n)$ implies that $\mu(A \cap \Omega_n) > 0$ if n is large enough, say $n \geq n_0$; let $B = A \cap \Omega_{n_0}$, then $0 < \mu(B) < \infty$. Choose $f = \frac{1}{\mu(B)} I_B \overline{\operatorname{sgn} g}$, then $f \in L^1$ and $\|f\|_1 = 1$. Now,

$$\|\ell_g\| \geq \left| \int fgd\mu \right| = \frac{1}{\mu(B)} \int_B |g|d\mu \geq \|g\|_\infty - \varepsilon,$$

from which we infer that $\|\ell_g\| \geq \|g\|_\infty$ by letting $\varepsilon \searrow 0$. ■

For the statement of the next lemma (6.3.2), given a measurable function g which is finite a.e. on Ω, we denote by $S_p(g)$ the family of all those functions f such that $\|f\|_p = 1$ and fg is integrable.

Exercise 6.3.1 Suppose that (Ω, Σ, μ) is σ-finite and g is a measurable function which is finite a.e. on Ω. Show that $S_p(g)$ is nonempty.

Lemma 6.3.2 *Suppose that (Ω, Σ, μ) is a σ-finite measure space and g is measurable and finite almost everywhere. Then,*

$$\|g\|_q = \sup \left\{ \left| \int fgd\mu \right| : f \in S_p(g) \right\}.$$

Proof From the Hölder inequality, $\|g\|_q \geq \sup\{|\int fgd\mu| : f \in S_p(g)\}$, it remains to show the converse inequality. For this purpose we may assume that $g \neq 0$ on a set of positive measure.

Let the sequence $\{\Omega_n\} \subset \Sigma$ be as in the proof of Corollary 6.3.1.

Step 1. Suppose that $q < \infty$. For each $n \in \mathbb{N}$, let $A_n = \{x \in \Omega : |g(x)| \leq n\} \cap \Omega_n$. $\{A_n\}$ is an increasing sequence in Σ such that $\mu(\Omega \setminus \bigcup_n A_n) = 0$. If we let $g_n = gI_{A_n}$, then g_n is bounded and $\neq 0$ on a set of positive measure when n is sufficiently large, say $n \geq n_0$. Define, for $n \geq n_0$,

$$f_n = \frac{|g_n|^{q-1}\overline{\operatorname{sgn} g_n}}{\|g_n\|_q^{q-1}}.$$

One can verify easily that $\|f_n\|_p = 1$. Since $f_ng = \|g_n\|_q^{1-q}|g_n|^q$, f_ng is integrable and therefore $\{f_n\}_{n\geq n_0} \subset S_p(g)$. Now for $n \geq n_0$, using $f_ng = \|g_n\|_q^{1-q}|g_n|^q$, we have

$$\begin{aligned}\|g_n\|_q^q &= \int |g_n|^q d\mu \\ &= \|g_n\|_q^{q-1} \int f_ngd\mu \leq \|g_n\|_q^{q-1} \sup \left\{ \left| \int fgd\mu \right| : f \in S_p(g) \right\},\end{aligned}$$

from which it follows that $\|g_n\|_q \leq \sup\left\{\left|\int fgd\mu\right| : f \in S_p(g)\right\}$. But $\mu(\Omega \backslash \bigcup_n A_n) = 0$ implies that $|g_n|$ increases to $|g|$ a.e. on Ω, hence, on letting $n \to \infty$, we obtain $\|g\|_q \leq \sup\{|\int fgd\mu| : f \in S_p(g)\}$. Thus, $\|g\|_q = \sup\{|\int fgd\mu| : f \in S_p(g)\}$ if $q < \infty$.

Step 2. Suppose that $q = \infty$, i.e. $p = 1$. Put

$$\gamma = \sup\left\{\left|\int fgd\mu\right| : f \in S_1(g)\right\}.$$

We may assume that $\gamma < \infty$.

Given that $\varepsilon > 0$, let $A = \{x \in \Omega : |g(x)| \geq \gamma + \varepsilon\}$. We claim that $\mu(A) = 0$; otherwise, let $B_n = A \cap \Omega_n \cap \{|g| \leq n\}$, then $0 < \mu(B_n) < \infty$ if $n \geq n_0$ for some $n_0 \in \mathbb{N}$. Put $B = B_{n_0}$ and let $f = \mu(B)^{-1} I_B \overline{\operatorname{sgn} g}$. Then, $\|f\|_1 = 1$ and $\int fgd\mu = \mu(B)^{-1}\int_B |g|d\mu \leq n_0$, thus $f \in S_1(g)$; but $\int fgd\mu = \mu(B)^{-1}\int_B |g|d\mu \geq \gamma + \varepsilon$, which contradicts the definition of γ. Hence, $\mu(A) = 0$ and consequently $\|g\|_\infty \leq \gamma + \varepsilon$. Let $\varepsilon \searrow 0$; we have $\|g\|_\infty \leq \gamma$. ■

It is worthwhile noting that the proof of Lemma 6.3.2 actually shows that $\|g\|_q = \sup\{\operatorname{Re}\int_\Omega fgd\mu : f \in S_p(g)\}$, and that if $g \geq 0$, $\|g\|_q = \sup\{\int_\Omega fgd\mu : f \in S_p(g)$ and $f \geq 0\}$.

The following **integral version of the Minkowski inequality** follows from Lemma 6.3.2 with this note.

Corollary 6.3.2 *Suppose that $(\Omega_1, \Sigma_1, \mu_1)$ and $(\Omega_2, \Sigma_2, \mu_2)$ are σ-finite complete measure spaces and $f \geq 0$ is $\overline{\Sigma_1 \otimes \Sigma_2}$-measurable on $\Omega_1 \times \Omega_2$. Then for $1 \leq p < \infty$, the following inequality holds:*

$$\left\{\int_{\Omega_1}\left(\int_{\Omega_2} f(x,y)d\mu_2(y)\right)^p d\mu_1(x)\right\}^{\frac{1}{p}} \leq \int_{\Omega_2}\left(\int_{\Omega_1} f(x,y)^p d\mu_1(x)\right)^{\frac{1}{p}} d\mu_2(y). \tag{6.14}$$

Proof Put $F(x) = \int_{\Omega_2} f(x,y)d\mu_2(y), x \in \Omega_1$. F is measurable using the Fubini theorem.

Step 1. Suppose that $F(x) < \infty$ for μ_1-a.e. x. Let $h \geq 0$ be in $S_q(F) \subset L^q(\Omega_1, \Sigma_1, \mu_1)$, then

$$\begin{aligned}
\int_{\Omega_1} hFd\mu_1 &= \int_{\Omega_1}\left(\int_{\Omega_2} f(x,y)d\mu_2(y)\right) h(x)d\mu_1(x) \\
&= \int_{\Omega_2}\left(\int_{\Omega_1} f(x,y)h(x)d\mu_1(x)\right) d\mu_2(y) \\
&\leq \|h\|_q \int_{\Omega_2}\left(\int_{\Omega_1} f(x,y)^p d\mu_1(x)\right)^{1/p} d\mu_2(y) \\
&= \int_{\Omega_2}\left(\int_{\Omega_1} f(x,y)^p d\mu_1(x)\right)^{1/p} d\mu_2(y).
\end{aligned}$$

By Lemma 6.3.2, with p replaced by q, together with the note that follows it, we conclude that $\|F\|_p \leq \int_{\Omega_2} \left(\int_{\Omega_1} f(x,y)^p d\mu(x)\right)^{1/p} d\mu_2(y)$, i.e. (6.14) holds.

Step 2. Now suppose that $A = \{F = \infty\}$ has positive measure. Since μ_1 is σ-finite, there is a measurable set $A_0 \subset A$ such that $0 < \mu_1(A_0) < \infty$. Let $h = \mu_1(A_0)^{-\frac{1}{q}} I_{A_0}$ or I_{A_0} according to whether $q < \infty$ or $q = \infty$, then proceed as in Step 1; we have

$$\infty = \int_{\Omega_1} hF d\mu_1 \leq \int_{\Omega_2} \left(\int_{\Omega_1} f(x,y)^p d\mu_1(x) \right)^{1/p} d\mu_2(y).$$

Consequently (6.14) holds, because right-hand side of (6.14) is ∞. ■

Now we come to the main theorem of this section.

Theorem 6.3.1 *If (Ω, Σ, μ) is σ-finite and $1 \leq p < \infty$, then $L^q = (L^p)^*$, through the map*

$$g \mapsto \ell_g, \quad g \in L^q.$$

Proof We already know that $L^q \subset (L^p)^*$ through the map $g \mapsto \ell_g$, by Corollary 6.3.1; it remains to show that for $\ell \in (L^p)^*$, there is a unique $g \in L^q$ such that $\ell = \ell_g$.

Step 1. Suppose that $\mu(\Omega) < \infty$.

For $A \in \Sigma$, let $\nu(A) = \ell(I_A)$. Since ℓ is linear, ν is an additive set function on Σ. Now suppose that $\{A_n\}_{n=1}^{\infty} \subset \Sigma$ is disjoint, then

$$\nu\left(\bigcup_n A_n\right) = \nu\left(\bigcup_{n=1}^{N} A_n\right) + \nu\left(\bigcup_{n=N+1}^{\infty} A_n\right),$$

hence, by putting $B_N = \bigcup_{n=N+1}^{\infty} A_n$, we have

$$\begin{aligned} \left|\nu\left(\bigcup_n A_n\right) - \sum_{n=1}^{N} \nu(A_n)\right| &\leq |\nu(B_N)| \\ &\leq \|\ell\| \|I_{B_N}\|_p \\ &= \|\ell\| [\mu(B_N)]^{1/p} \to 0 \end{aligned}$$

as $N \to \infty$, because $B_N \downarrow \phi$ and $\mu(\Omega) < \infty$; consequently, $\nu(\bigcup_n A_n) = \sum_{n=1}^{\infty} \nu(A_n)$. Thus ν is a complex measure on Σ. Since ν is μ-absolutely continuous, from the Radon–Nikodym theorem, there is $g \in L^1$ such that $\nu(A) = \int_A g d\mu$, or

$$\ell(f) = \int fg d\mu \tag{6.15}$$

for simple functions f.

Suppose now that $f \in S_p(g)$. Choose a sequence $\{f_n\}$ of simple functions such that $f_n \to f$ pointwise and $|f_n| \le |f|$. Then, $|f_n g| \le |fg|$, by LDCT and (6.15),

$$\left|\int fgd\mu\right| = \lim_{n\to\infty}\left|\int f_n gd\mu\right| = \lim_{n\to\infty}|\ell(f_n)| \le \|\ell\|.$$

It then follows from Lemma 6.3.2 that $g \in L^q$ and $\|g\|_q \le \|\ell\|$.

Now let $f \in L^p$ and choose a sequence $\{\varphi_n\}$ of simple functions such that $\varphi_n \to f$ pointwise and $|\varphi_n| \le |f|$, then $\varphi_n \to f$ in L^p and by (6.15),

$$\int fgd\mu = \lim_{n\to\infty}\int \varphi_n gd\mu = \lim_{n\to\infty}\ell(\varphi_n) = \ell(f),$$

this means that $\ell = \ell_g$ and $\|\ell\| = \|\ell_g\| = \|g\|_q$.

Step 2. Suppose that (Ω, Σ, μ) is σ-finite.

Let $\{\Omega_n\} \subset \Sigma$ be as in the proof of Corollary 6.3.1. By Step 1, for each n, there is $g_n \in L^q$ with $\{g_n \ne 0\} \subset \Omega_n$ such that

$$\ell(f) = \int fg_n d\mu \tag{6.16}$$

for $f \in L^p$ with $\{f \ne 0\} \subset \Omega_n$. Define g on Ω by $g(x) = g_1(x)$ if $x \in \Omega_1$, and $g(x) = g_n(x)$ if $x \in \Omega_n \backslash \Omega_{n-1}$ for $n \ge 2$. Then, since $g_n(x) = g_{n-1}(x)$ for a.e. x in Ω_{n-1} when $n \ge 2$, $g(x) = g_n(x)$ for a.e. $x \in \Omega_n$.

Now let $f \in S_p(g)$, then $|fg_n| \le |fg|$ and $fg_n \to fg$ a.e., hence by (6.16),

$$\left|\int fgd\mu\right| = \lim_{n\to\infty}\left|\int fg_n d\mu\right| = \lim_{n\to\infty}|\ell(fI_{\Omega_n})| \le \|\ell\|.$$

From Lemma 6.3.2, $g \in L^q$ and hence for $f \in L^p$,

$$\int fgd\mu = \lim_{n\to\infty}\int fI_{\Omega_n} g_n d\mu = \lim_{n\to\infty}\ell(fI_{\Omega_n}) = \ell(f),$$

where the last equality comes from the obvious fact that $fI_{\Omega_n} \to f$ in L^p. Then $\ell = \ell_g$, and $\|\ell\| = \|_g\|_q$. That g is uniquely determined is obvious.

■

Exercise 6.3.2 shows that Theorem 6.3.1 may not hold true when $p = \infty$.

Exercise 6.3.2 Consider $L^\infty[0,1]$ and let $x_0 \in [0,1]$. Show that there is $\ell \in L^\infty[0,1]^*$ with $\|\ell\| = 1$ such that $\ell(f) = f(x_0)$ for $f \in C[0,1]$. For this ℓ show that there is no $g \in L^1[0,1]$ such that $\ell(f) = \int_{[0,1]} fgd\lambda$ for all $f \in L^\infty[0,1]$.

Exercise 6.3.3 Suppose that (Ω, Σ, μ) is σ-finite and $1 < p < \infty$. Show that L^p is reflexive.

Exercise 6.3.4 Let D be a measurable set in $\mathbb{R}^n$ with positive measure. Show that every bounded sequence in $L^p(D)$, $1 < p < \infty$, has a subsequence which converges weakly. (Hint: cf. Exercise 5.10.5.)

6.4 Modular distribution function and Hardy–Littlewood maximal function

Suppose that f is a finite a.e. measurable function on a measure space (Ω, Σ, μ). Define a function $\lambda_f : (0, \infty) \to [0, \infty]$ by

$$\lambda_f(\alpha) = \mu(\{|f| > \alpha\}). \tag{6.17}$$

Then the function λ_f enjoys the following properties:

(1) λ_f is monotone decreasing and right continuous.
(2) If $|f| \leq |g|$, then $\lambda_f \leq \lambda_g$.
(3) If $|f_n| \nearrow |f|$, then $\lambda_{f_n} \nearrow \lambda_f$.
(4) If $f = g + h$, then $\lambda_f(\alpha + \beta) \leq \lambda_g(\alpha) + \lambda_h(\beta)$ for $\alpha, \beta > 0$.

Properties (1)–(3) follow directly from the definition, while (4) is a consequence of the fact that $\{|f| > \alpha + \beta\} \subset \{|g| > \alpha\} \cup \{|h| > \beta\}$.

The function λ_f is usually called the distribution function of f; but the distribution function of a measurable function is defined differently in Section 4.3, in agreement with the distribution function of a random variable in probability theory; we shall instead call λ_f the **modular distribution function** of f.

If $\lambda_f(\alpha) < \infty$ $\forall \alpha > 0$, then λ_f generates a negative Radon measure ν on $(0, \infty)$ such that

$$\nu((a, b]) = \lambda_f(b) - \lambda_f(a), \quad 0 < a < b < \infty;$$

actually, ν is the negative of the Radon measure generated by the monotone increasing function $-\lambda_f$. We shall call ν the **Lebesque–Stieltjes measure** generated by λ_f. If φ is a Borel function on $(0, \infty)$ such that $\int_0^\infty \varphi d\nu = \int_{(0,\infty)} \varphi d\nu$ exists, then $\int_0^\infty \varphi d\nu$ will be denoted by $\int_0^\infty \varphi d\lambda_f$ or $\int_0^\infty \varphi(\alpha) d\lambda_f(\alpha)$ in this section, and called the **Lebesque–Stieltjes integral** of φ w.r.t. λ_f.

Lemma 6.4.1 *Suppose that $\lambda_f(\alpha) < \infty$ for all $\alpha > 0$ and let φ be a nonnegative Borel function on $(0, \infty)$, then*

$$\int_\Omega \varphi \circ |f| d\mu = -\int_0^\infty \varphi(\alpha) d\lambda_f(\alpha). \tag{6.18}$$

Proof We have

$$\nu((a,b]) = \lambda_f(b) - \lambda_f(a) = -\mu(\{a < |f| \leq b\}) = -\mu(|f|^{-1}(a,b]),$$

from which it follows that

$$\nu(B) = -\mu(|f|^{-1}B)$$

for Borel set B in $(\frac{1}{k}, k]$ (by the $(\pi\text{-}\lambda)$ theorem), and therefore for Borel set B in $(0,\infty)$. This means that (6.18) holds if φ is the indicator function of Borel set B in $(0,\infty)$, and consequently, if φ is a nonnegative simple Borel function on $(0,\infty)$. For a general nonnegative Borel function on $(0,\infty)$, (6.18) follows then by approximating φ pointwise by an increasing sequence of nonnegative simple Borel functions on $(0,\infty)$. ■

Exercise 6.4.1 Give the detail of the first part of the proof of Lemma 6.4.1 where the $(\pi\text{-}\lambda)$ theorem is applied.

A measurable function f on (Ω, Σ, μ) is called a **weak L^p function**; $0 < p < \infty$, if there is $0 \leq A < \infty$ depending only on f and p such that

$$\mu(\{|f| > \alpha\}) \leq \frac{A^p}{\alpha^p}, \quad \alpha > 0. \tag{6.19}$$

One sees readily that f is a weak L^p function if and only if $\sup_{\alpha>0} \alpha^p \mu\{|f| > \alpha\} < \infty$.

Exercise 6.4.2 Show that if $|f|^p$, $0 < p < \infty$, is integrable, then f is a weak L^p function.

Theorem 6.4.1 *Suppose that f is a weak L^p function, $1 \leq p < \infty$, then we have*

$$\int_\Omega |f|^p d\mu = -\int_0^\infty \alpha^p d\lambda_f(\alpha) = p\int_0^\infty \alpha^{p-1}\lambda_f(\alpha)d\alpha. \tag{6.20}$$

Proof Since f is a weak L^p function, $1 \leq p < \infty$, $\lambda_f(\alpha) < \infty$ for all $\alpha > 0$, hence the first equality in (6.20) follows from Lemma 6.4.1 by taking $\phi(\alpha) = \alpha^p$. It remains to show that

$$\int_\Omega |f|^p d\mu = p\int_0^\infty \alpha^{p-1}\lambda_f(\alpha)d\alpha.$$

We observe first that the set

$$E := \{(x,\alpha) : x \in \Omega,\ 0 < \alpha < |f(x)|\}$$

is in $\Sigma \otimes \mathcal{B}$ (cf. Exercise 4.8.3). Since I_E is $\Sigma \otimes \mathcal{B}$-measurable, from Tonelli's theorem we have

$$\begin{aligned}
&\int_{\Omega\times(0,\infty)} pI_E(x,\alpha)\alpha^{p-1}d(\mu\times\lambda)(x,\alpha)\\
&= \int_\Omega \left(\int_0^{|f(x)|} p\alpha^{p-1}d\lambda(\alpha)\right)d\mu(x)\\
&= \int_\Omega |f(x)|^p d\mu(x);
\end{aligned}$$

but we also have

$$\begin{aligned}
&\int_{\Omega\times(0,\infty)} pI_E(x,\alpha)\alpha^{p-1}d(\mu\times\lambda)(x,\alpha)\\
&= p\int_0^\infty \alpha^{p-1}\left(\int_{\{|f|>\alpha\}} d\mu\right)d\lambda(\alpha)\\
&= p\int_0^\infty \alpha^{p-1}\lambda_f(\alpha)d\alpha.
\end{aligned}$$

Hence $\int_\Omega |f|^p d\mu = p\int_0^\infty \alpha^{p-1}\lambda_f(\alpha)d\alpha$. ■

The Hardy–Littlewood maximal function will now be introduced. Let f be a locally integrable function on $\mathbb{R}^n$; the **Hardy–Littlewood maximal function** of f, denoted Mf, is defined in terms of $|f|$ as follows:

$$Mf(x) = \sup_{r>0} \frac{1}{\lambda^n(B_r(x))}\int_{B_r(x)} |f(y)|dy, \tag{6.21}$$

where $Mf(x)$ could be infinite for some $x \in \mathbb{R}^n$. Since, for each $r > 0$, the function

$$x \mapsto \frac{1}{\lambda^n(B_r(x))}\int_{B_r(x)} |f(y)|dy$$

is continuous, $\{Mf > \alpha\}$ is open for $\alpha \in \mathbb{R}$. Hence Mf is a Borel function and is therefore measurable. We shall from now on simply call Mf the **maximal function** of f.

Theorem 6.4.2 *For $f \in L^1(\mathbb{R}^n)$, Mf is a weak L^1 function. Actually there is $A > 0$, depending only on n, such that*

$$\lambda^n(\{Mf > \alpha\}) \le A\|f\|_1\alpha^{-1} \tag{6.22}$$

for $f \in L^1(\mathbb{R}^n)$ and $\alpha > 0$.

Proof For $\alpha > 0$, put $E_\alpha = \{Mf > \alpha\}$. For $x \in E_\alpha$, there is a ball $B(x)$ centered at x such that

$$\int_{B(x)} |f(y)|dy > \alpha\lambda^n(B(x)). \tag{6.23}$$

Since $\lambda^n(B(x)) < \alpha^{-1}\|f\|_1$ by (6.23), $\mathcal{C} := \{B(x) : x \in E_\alpha\}$ is an admissible collection of balls. By Theorem 4.6.1, there is a disjoint sequence $\{B_k\}$ of balls from $\mathcal{C}$ such that $\bigcup \mathcal{C} \subset \bigcup_k \widehat{B}_k$, where $\widehat{B}_k$ is concentric with B_k and has a radius five times that of B_k. Then from (6.23),

$$\begin{aligned}\lambda^n(E_\alpha) &\leq \lambda^n\left(\bigcup \mathcal{C}\right) \leq \sum_k \lambda^n(\widehat{B}_k) = 5^n \sum_k \lambda^n(B_k) \\ &< 5^n\alpha^{-1}\sum_k \int_{B_k} |f(y)|dy = 5^n\alpha^{-1}\int_{\bigcup B_k} |f(y)|dy \\ &\leq 5^n\|f\|_1\alpha^{-1},\end{aligned}$$

from which we complete the proof by taking $A = 5^n$. ■

Exercise 6.4.3 For $f \in L^p(\mathbb{R}^n)$, $1 \leq p < \infty$, show that Mf is a weak L^p function. (Hint: use Jensen's inequality to show that $Mf(x) \leq \{M|f|^p(x)\}^{1/p}$ for $x \in \mathbb{R}^n$.)

We note at this point that although Mf is a weak L^1 function, it can never be integrable except for the extreme case $f = 0$ a.e. To see this, suppose that $f \neq 0$ on a set of positive measure; then $\int_{B_R(0)} |f|d\lambda^n = c > 0$ for some $R > 0$ and hence if $|x| \geq R$, $B := B_{2|x|}(x)$ contains $B_R(0)$, from which

$$Mf(x) \geq \frac{1}{\lambda^n(B)}\int_B |f(y)|dy \geq 2^{-n}|x|^{-n}b_n^{-1}c = c_0|x|^{-n}$$

follows, where b_n is the measure of the unit ball in $\mathbb{R}^n$; thus by integrating Mf over $\mathbb{R}^n$ using polar coordinates (cf. Theorem 4.11.1), we conclude that $\int Mfd\lambda^n = \infty$. However, as the following theorem shows, $Mf \in L^p$ if $f \in L^p$ and the map $f \mapsto Mf$ is a bounded map from L^p into L^p when $p > 1$.

Theorem 6.4.3 *If* $1 < p \leq \infty$, *there is* $A_p > 0$ *such that for* $f \in L^p(\mathbb{R}^n)$ *we have*

$$\|Mf\|_p \leq A_p\|f\|_p.$$

Proof When $p = \infty$, this is obvious with $A_\infty = 1$. Consider now $1 < p < \infty$. For a fixed $\alpha > 0$, define f_1 by

$$f_1(x) = \begin{cases} f(x) & \text{if } |f(x)| \geq \frac{\alpha}{2}; \\ 0 & \text{otherwise,} \end{cases}$$

then, $f_1 \in L^1(\mathbb{R}^n)$ (see Exercise 6.4.4) and $|f(x)| \leq |f_1(x)| + \frac{\alpha}{2}$; hence $Mf \leq Mf_1 + \frac{\alpha}{2}$, which implies $\{Mf > \alpha\} \subset \{Mf_1 > \frac{\alpha}{2}\}$ and consequently by Theorem 6.4.2 (note that A can be taken to be 5^n),

$$\lambda_{Mf}(\alpha) \leq \frac{2 \cdot 5^n}{\alpha} \|f_1\|_1 = \frac{2 \cdot 5^n}{\alpha} \int_{\{|f| \geq \frac{\alpha}{2}\}} |f(x)| dx.$$

Now by (6.20),

$$\begin{aligned}
\|Mf\|_p^p &= p \int_0^\infty \alpha^{p-1} \lambda_{Mf}(\alpha) d\alpha \\
&\leq p \int_0^\infty \alpha^{p-1} \left(\frac{2 \cdot 5^n}{\alpha} \int_{\{|f| \geq \frac{\alpha}{2}\}} |f(x)| dx \right) d\alpha \\
&= 2 \cdot 5^n \cdot p \int_{\mathbb{R}^n} |f(x)| \int_0^{2|f(x)|} \alpha^{p-2} d\alpha dx \\
&= \frac{2 \cdot 5^n \cdot p}{p-1} \int_{\mathbb{R}^n} 2^{p-1} |f(x)|^p dx \\
&= 2^p \cdot 5^n \frac{p}{p-1} \|f\|_p^p = A_p^p \|f\|_p^p,
\end{aligned}$$

where $A_p = 2\left(\frac{5^n p}{p-1}\right)^{1/p}$. ∎

Exercise 6.4.4 Show that the function f_1 defined at the beginning of the proof of Theorem 6.4.3 is integrable.

As an application of maximal function, a direct proof of Theorem 4.6.4 will now be given using Theorem 6.4.2 together with the Markov inequality (6.1). An application of Theorem 6.4.3 to the study of Sobolev space is presented in Section 6.6. Actually, we shall prove that if f is a locally integrable function on an open set $\Omega \subset \mathbb{R}^n$, then

$$\lim_{r \to 0} \frac{1}{\lambda^n(B_r(x))} \int_{B_r(x)} |f(y) - f(x)| dy = 0 \tag{6.24}$$

for a.e. $x \in \Omega$, and leave the proof for the general statement as an exercise. Because of the local nature of (6.24), we may assume that f is an integrable function on $\mathbb{R}^n$. Put $\theta(f, x) = \limsup_{r \to 0} \frac{1}{\lambda^n(B_r(x))} \int_{B_r(x)} |f(y) - f(x)| dy$; our aim is to show that $\theta(f, x) = 0$ for a.e. x in $\mathbb{R}^n$, or, equivalently, to show that $\lambda^n(\{\theta(f, \cdot) > \alpha\}) = 0$ for every $\alpha > 0$. Now, given that $\varepsilon > 0$, there is a continuous function g on $\mathbb{R}^n$ such that $\|f - g\|_1 < \varepsilon$ (cf. Exercise 6.1.1), then,

$$\theta(f, x) = \theta(f - g + g, x) \leq \theta(f - g, x) + \theta(g, x) = \theta(f - g, x),$$

because $\theta(g,x) = 0$; but $\theta(f-g,x) \le M(f-g)(x) + |f(x) - g(x)|$ and consequently,

$$\{\theta(f,\cdot) > \alpha\} \subset \{\theta(f-g,\cdot) > \alpha\} \subset \left\{M(f-g) > \frac{\alpha}{2}\right\} \cup \left\{|f-g| > \frac{\alpha}{2}\right\}.$$

Hence,

$$\lambda^n(\{\theta(f,\cdot) > \alpha\}) \le (A\|f-g\|_1 + \|f-g\|_1)\frac{2}{\alpha} \le 2(A+1)\frac{\varepsilon}{\alpha};$$

by letting $\varepsilon \to 0$, we have $\lambda^n(\{\theta(f,\cdot) > \lambda\}) = 0$. Thus, $\theta(f,x) = 0$ for a.e. x in $\mathbb{R}^n$ and (6.24) is established.

Exercise 6.4.5 Show that $\lim_{B\to x} \frac{1}{\lambda^n(B)} \int_B |f(y) - f(x)| dy = 0$ follows from (6.24). (Hint: if $x \in B$, then $B \subset B_{2r}(x)$, where r is the radius of B.)

6.5 Convolution

The operation of taking convolution was used in Section 4.9 when introducing the Friederichs mollifier for the purpose of smoothing functions. An account of general features of convolution for functions on $\mathbb{R}^n$ will be given in this section; its connection with the Fourier integral will be seen in Chapter 7. Referring to Exercises 1.6.6 and 1.6.7, we note in passing that convolution can be introduced for functions on groups with a measure invariant under translations w.r.t. the group operation and is often proved to be a useful operation.

We first state Proposition 4.8.2 as a lemma for later reference.

Lemma 6.5.1 *Let f be a measurable function on $\mathbb{R}^n$, then $F(x,y) := f(x-y)$, x, y in $\mathbb{R}^n$, is a measurable function on $\mathbb{R}^{2n} = \mathbb{R}^n \oplus \mathbb{R}^n$.*

Let f and g be measurable functions on $\mathbb{R}^n$. The **convolution** of f and g is the function $f * g$ defined for all those x for which the following integral exists and is finite:

$$f * g(x) = \int f(x-y)g(y)dy.$$

Exercise 6.5.1

(i) Show that if $f * g(x)$ exists and is finite, then $g * f(x)$ exists and is finite, and $g * f(x) = f * g(x)$.

(ii) Show that if $f * g$ exists and is finite for a.e. x, then $f * g$ is measurable. (Hint: apply Lemma 6.5.1 and the Fubini theorem.)

Exercise 6.5.2 Suppose that $[a,b]$ and $[c,d]$ are finite closed intervals of equal length. Find $I_{[a,b]} * I_{[c,d]}$; in particular, show that $I_{[-\frac{\alpha}{2},\frac{\alpha}{2}]} * I_{[-\frac{\alpha}{2},\frac{\alpha}{2}]}(x) = \alpha(1 - \frac{|x|}{\alpha})^+, \alpha > 0$.

Theorem 6.5.1 (Young inequality) *Suppose that $f \in L^p$, $p \geq 1$, and $g \in L^1$, then $f * g$ exists and is finite a.e., and*

$$\|f * g\|_p \leq \|f\|_p \|g\|_1.$$

Proof The case where $p = \infty$ is obvious. We consider the case where $1 \leq p < \infty$. Let $h(x, y) = f(x-y)g(y)$; h is measurable by Lemma 6.5.1. Using the integral version of the Minkowski inequality (Corollary 6.3.2), we have

$$\begin{aligned}\left(\int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n} |f(x-y)g(y)|dy\right)^p dx\right)^{\frac{1}{p}} &\leq \int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n} |f(x-y)g(y)|^p dx\right)^{\frac{1}{p}} dy \\ &= \|f\|_p\|g\|_1;\end{aligned}$$

then, $\left(\int_{\mathbb{R}^n}\left|\int_{\mathbb{R}^n} f(x-y)g(y)dy\right|^p dx\right)^{\frac{1}{p}} \leq \|f\|_p\|g\|_1$, and consequently $f * g$ exists and is finite a.e., and $\|f * g\|_p \leq \|f\|_p\|g\|_1$. ■

Example 6.5.1 We give here another proof of Theorem 5.6.1 without recourse to the integral version of the Minkowski inequality. Since

$$\begin{aligned}\int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n} |f(x-y)|^p|g(y)|dy\right) dx &= \int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n} |f(x-y)|^p|g(y)|dx\right) dy \\ &= \|f\|_p^p \int_{\mathbb{R}^n} |g(y)|dy = \|f\|_p^p\|g\|_1,\end{aligned}$$

therefore, $\int_{\mathbb{R}^n} |f(x-y)|^p|g(y)|dy < \infty$ for a.e. x, and hence

$$\int_{\mathbb{R}^n} |f(x-y)||g(y)|dy \leq \left(\int_{\mathbb{R}^n} |f(x-y)|^p|g(y)|dy\right)^{\frac{1}{p}} \cdot \|g\|_1^{\frac{1}{q}} < \infty \text{ for a.e. } x,$$

which implies that $f * g$ exists and is finite a.e., and

$$\begin{aligned}\|f * g\|_p^p &\leq \int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n} |f(x-y)||g(y)|dy\right)^p dx \\ &\leq \|g\|_1^{\frac{p}{q}} \int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n} |f(x-y)|^p|g(y)|dy\right) dx \\ &= \|g\|_1^{\frac{p}{q}}\|f\|_p^p\|g\|_1 = \|f\|_p^p\|g\|_1^p,\end{aligned}$$

or

$$\|f * g\|_p \leq \|f\|_p\|g\|_1.$$

Lemma 6.5.2 *For $f \in L^p$, $1 \leq p < \infty$, and $y \in \mathbb{R}^n$, let $f^y(x) = f(x-y)$. Then, $\lim_{y\to 0}\|f^y - f\|_p = 0$.*

Proof Given $\varepsilon > 0$, there is a continuous function g with compact support such that $\|f - g\|_p < \frac{\varepsilon}{3}$, by Proposition 4.6.1. Then,

$$\begin{aligned}\|f^y - f\|_p &= \|f^y - g^y + g^y - g + g - f\|_p \\ &\leq \|f^y - g^y\|_p + \|g - f\|_p + \|g^y - g\|_p \\ &< \frac{2}{3}\varepsilon + \|g^y - g\|_p,\end{aligned}$$

but since g is continuous with compact support, $\|g^y - g\|_p < \frac{\varepsilon}{3}$ when $|y|$ is small. Thus, $\|f^y - f\|_p < \varepsilon$ when $|y|$ is small. ■

We shall denote by $C_0(\mathbb{R}^n)$ the space of all those continuous functions f on $\mathbb{R}^n$ with the property that for any $\varepsilon > 0$, there is $R > 0$ such that $|f(x)| < \varepsilon$ whenever $|x| > R$. Functions in $C_0(\mathbb{R}^n)$ are functions **vanishing at infinity**, introduced in Section 6.2. Clearly, $C_c(\mathbb{R}^n) \subset C_0(\mathbb{R}^n)$.

Theorem 6.5.2 *If p and q are conjugate exponents, $f \in L^p$ and $g \in L^q$, then $f * g(x)$ exists and is finite for all x, and $f * g$ is bounded and uniformly continuous on $\mathbb{R}^n$. Furthermore, $\|f * g\|_\infty \leq \|f\|_p \|g\|_q$; and if $1 < p < \infty$, then $f * g \in C_0(\mathbb{R}^n)$.*

Proof From the Hölder inequality, $|f * g(x)| \leq \|f\|_p \|g\|_q$ for all x, hence $f * g(x)$ exists and is finite for all x, and $\|f * g\|_\infty \leq \|f\|_p \|g\|_q$.

To show that $f * g$ is uniformly continuous, we may assume that $1 \leq p < \infty$ (otherwise interchange p and q). Now,

$$|f * g(x - y) - f * g(x)| = |(f^y - f) * g(x)| \leq \|f^y - f\|_p \|g\|_q,$$

hence $f * g$ is uniformly continuous on $\mathbb{R}^n$, by Lemma 6.5.2.

Finally, suppose that $1 < p < \infty$ (then $1 < q < \infty$). Choose sequences $\{f_k\}, \{g_k\}$ in $C_c(\mathbb{R}^n)$ so that $\|f_k - f\|_p \to 0$ and $\|g_k - g\|_q \to 0$ as $k \to \infty$; this is possible by Proposition 4.6.1. Then, $\{f_k * g_k\}$ is a sequence of continuous functions with compact support, and

$$\begin{aligned}\sup_{x \in \mathbb{R}^n} |f_k * g_k(x) - f * g(x)| &= \sup_{x \in \mathbb{R}^n} |f_k * (g_k - g)(x) + (f_k - f) * g(x)| \\ &\leq \|f_k\|_p \|g_k - g\|_q + \|f_k - f\|_p \|g\|_q \to 0\end{aligned}$$

as $k \to \infty$, because $\{f_k\}$, being a convergent sequence in L^p, is bounded in L^p.

Now given $\varepsilon > 0$, from what we have just shown choose k_0 large enough so that $\sup_{x \in \mathbb{R}^n} |f_{k_0} * g_{k_0}(x) - f * g(x)| < \varepsilon$, and then choose $R > 0$ such that $f_{k_0} * g_{k_0}(x) = 0$ when $|x| > R$; thus $|f * g(x)| < \varepsilon$, when $|x| > R$. This shows that $f * g \in C_0(\mathbb{R}^n)$. ■

Remark Theorem 6.5.2 is an example showing the smoothing effect of convolution.

Exercise 6.5.3 Show that for f, g, and h in L^1, $(f * g) * h = f * (g * h)$.

Example 6.5.2 The Friederich mollifier $\{J_\varepsilon\}_{\varepsilon>0}$ constructed from a mollifying function φ introduced in Section 4.9 can be expressed as

$$J_\varepsilon f(x) = f * \varphi_\varepsilon(x), \quad x \in \mathbb{R}^n,$$

for $f \in L^{\text{loc}}(\mathbb{R}^n)$. By Proposition 4.9.2 and Theorem 6.5.2, $J_\varepsilon f \in C^\infty(\mathbb{R}^n) \cap C_0(\mathbb{R}^n)$ if $f \in L^p$, $1 < p < \infty$.

Exercise 6.5.4 Show that there is no $u \in L^1$ such that $u * f = f$ for all $f \in L^1$. (Hint: if there is such a u, then $u * \varphi_\varepsilon = \varphi_\varepsilon$ for all $\varepsilon > 0$, where φ is a mollifying function.)

Example 6.5.3 Suppose that f, g are in L^1 and $f \in C^1(\mathbb{R}^n)$ with bounded partial derivatives. Since $f \in C^1(\mathbb{R}^n)$ with bounded partial derivatives, f is uniformly continuous; consequently if $f * g(x)$ exists and is finite, then $f * g(x')$ exists and is finite if $|x' - x| < \delta$, where $\delta > 0$ is chosen so that $|f(z) - f(z')| < 1$ if $|z - z'| < \delta$. This, together with the known fact that $f * g$ exists and is finite a.e., shows that $f * g$ exists and is finite everywhere and is uniformly continuous on $\mathbb{R}^n$. Now for any x, y in $\mathbb{R}^n$, $\frac{|f(x)-f(y)|}{|x-y|} \leq M$ for a fixed $M > 0$, because partial derivatives of f are bounded. We can then apply LDCT to infer that

$$\frac{\partial}{\partial x_j} f * g(x) = \frac{\partial f}{\partial x_j} * g(x), \quad x \in \mathbb{R}^n, \, j = 1, \ldots, n.$$

But from Theorem 6.5.2, $\frac{\partial f}{\partial x_j} * g$ is bounded and continuous. Hence, $f * g \in C^1(\mathbb{R}^n)$ and its partial derivatives are bounded.

By the Young inequality (Theorem 6.5.1), L^1 is closed under the binary operation of convolution, which is associative (cf. Exercise 6.5.3) and clearly distributive w.r.t. the addition of elements in L^1. Thus with the introduction of the binary operation $*$ into L^1, L^1 becomes a commutative algebra; it is an example of the so-called Banach algebras, in that it is a Banach space which is also an algebra that satisfies the inequality $\|f * g\|_1 \leq \|f\|_1 \|g\|_1$ for f, g in L^1. Because of the conclusion of Exercise 6.5.4, there exists no identity element in L^1 w.r.t. the multiplication operation $*$. However, if φ is a mollifying function (cf. Example 6.5.2), $\lim_{\varepsilon\to 0} \varphi_\varepsilon * f = f$ in L^1, by Theorem 4.9.2; such a family $\{\varphi_\varepsilon\}_{\varepsilon>0}$ is called an **approximate identity** for L^1. Just as we construct the approximate identity $\{\varphi_\varepsilon\}_{\varepsilon>0}$ from a mollifying function φ, starting from an integrable function h on $\mathbb{R}^n$ with $\int h d\lambda^n = 1$, we define for each $t > 0$ a function h_t by

$$h_t(x) = t^{-n} h\left(\frac{x}{t}\right), \quad x \in \mathbb{R}^n,$$

then, $\int h_t d\lambda^n = 1$. We shall see that $\{h_t\}_{t>0}$ is an approximate identity for L^1.

Lemma 6.5.3 *For $\varepsilon > 0$ and $\delta > 0$, there is $t_0 > 0$ such that*

$$\int_{|y|\geq\delta} |h_t(y)|dy < \varepsilon,$$

whenever $0 < t \leq t_0$.

Proof Since $h \in L^1$, there is $R > 0$ such that $\int_{|y|\geq R} |h(y)|dy < \varepsilon$. Then,

$$\int_{|y|\geq\delta} |h_t(y)|dy = \int_{|y|\geq\frac{\delta}{t}} |h(y)|dy < \varepsilon$$

if $\frac{\delta}{t} \geq R$. We choose $t_0 = \frac{\delta}{R}$ to complete the proof. ∎

Theorem 6.5.3 *$\{h_t\}_{t>0}$ is an approximate identity for L^1, i.e.*

$$\lim_{t\to 0} \|h_t * f - f\|_1 = 0, \quad f \in L^1.$$

Proof For $f \in L^1$ and $\varepsilon > 0$, there is $\delta > 0$ such that

$$\int_{\mathbb{R}^n} |f(x-y) - f(x)|dx < \frac{\varepsilon}{2\|h\|_1} \tag{6.25}$$

if $|y| < \delta$. Since we may assume that $\|f\|_1 > 0$, there is $t_0 > 0$ such that

$$\int_{|y|\geq\delta} |h_t(y)|dy < \frac{\varepsilon}{4\|f\|_1} \tag{6.26}$$

whenever $0 < t \leq t_0$, by Lemma 6.5.3. Now,

$$\begin{aligned}
&\int_{\mathbb{R}^n} |h_t * f - f|d\lambda^n \\
&= \int_{\mathbb{R}^n} \left| \int_{\mathbb{R}^n} \{f(x-y) - f(x)\}h_t(y)dy \right| dx \\
&\leq \int_{\mathbb{R}^n} |h_t(y)| \int_{\mathbb{R}^n} |f(x-y) - f(x)|dxdy \\
&\leq \int_{|y|<\delta} |h_t(y)| \int_{\mathbb{R}^n} |f(x-y) - f(x)|dxdy + 2\|f\|_1 \int_{|y|\geq\delta} |h_t(y)|dy \\
&< \frac{\varepsilon}{2\|h\|_1} \int_{|y|<\delta} |h_t(y)|dy + \frac{\varepsilon}{2} \leq \varepsilon
\end{aligned}$$

if $0 < t \leq t_0$ by (6.25) and (6.26). ∎

We know from Theorem 6.5.2 that $f * h_t$ is a bounded and uniformly continuous function for each $t > 0$ if $f \in L^\infty$; we show now, as a supplement to Theorem 6.5.3, that $f * h_t$ converges to f uniformly on every compact set of $\mathbb{R}^n$ as $t \to 0$ if $f \in L^\infty \cap C(\mathbb{R}^n)$.

Theorem 6.5.4 *If f is a bounded continuous function on $\mathbb{R}^n$, then $\lim_{t\to 0} f * h_t = f$ uniformly on every compact set K of $\mathbb{R}^n$.*

Proof For a compact set K in $\mathbb{R}^n$ and $\varepsilon > 0$, there is $\delta > 0$ such that $|f(x-y) - f(x)| < \frac{\varepsilon}{2\|h\|_1}$ whenever $x \in K$ and $|y| < \delta$. Then by Lemma 6.5.3, there is $t_0 > 0$ such that $\int_{|y|\ge\delta} |h_t(y)|dy < \frac{\varepsilon}{4\|f\|_\infty}$ if $0 < t \le t_0$. Now for $x \in K$ and $0 < t \le t_0$, we have

$$
\begin{aligned}
|h_t * f(x) - f(x)| &= \left| \int_{\mathbb{R}^n} h_t(y)\{f(x-y) - f(x)\}dy \right| \\
&\le \int_{\mathbb{R}^n} |h_t(y)||f(x-y) - f(x)|dy \\
&= \int_{|y|<\delta} |h_t(y)||f(x-y) - f(x)|dy + \int_{|y|\ge\delta} |h_t(y)||f(x-y) - f(x)|dy \\
&< \frac{\varepsilon}{2\|h\|_1} \int_{|y|<\delta} |h_t(y)|dy + 2\|f\|_\infty \int_{|y|\ge\delta} |h_t(y)|dy \\
&< \frac{\varepsilon}{2} + 2\|f\|_\infty \frac{\varepsilon}{4\|f\|_\infty} = \varepsilon,
\end{aligned}
$$

which means that $h_t * f(x) \to f(x)$ uniformly for $x \in K$ as $t \to 0$. ■

Exercise 6.5.5 Let $p(x) = \frac{1}{\pi}\frac{1}{1+x^2}$ and write $p_t(x) = \frac{t}{\pi}\frac{1}{t^2+x^2}$ as $p(x,t)$ for $x \in \mathbb{R}$ and $t > 0$. The function $(x,t) \mapsto p(x,t)$ on $\mathbb{R} \times (0,\infty)$ is called the **Poisson kernel**.

(i) For $f \in L^1(\mathbb{R})$, let

$$\Pi(x,t) = p_t * f = \int_{\mathbb{R}} p(x-y,t)f(y)dy, \quad (x,t) \in \mathbb{R} \times (0,\infty).$$

Show that

$$
\begin{aligned}
\frac{\partial^2 \Pi}{\partial x^2}(x,t) &= \int_{\mathbb{R}} \frac{\partial^2 p}{\partial x^2}(x-y,t)f(y)dy; \\
\frac{\partial^2 \Pi}{\partial t^2}(x,t) &= \int_{\mathbb{R}} \frac{\partial^2 p}{\partial t^2}(x-y,t)f(y)dy.
\end{aligned}
$$

(Hint: $\frac{\partial p}{\partial x}(x,t)$, $\frac{\partial^2 p}{\partial x^2}(x,t)$, $\frac{\partial p}{\partial t}(x,t)$, $\frac{\partial^2 p}{\partial t^2}(x,t)$ are bounded on $\mathbb{R} \times (t_0,\infty)$ for any $t_0 > 0$.)

(ii) Let f and Π be as in (i). Show that Π is harmonic on $\mathbb{R} \times (0,\infty)$. Furthermore, if f is bounded and continuous, show that Π can be extended continuously to $\mathbb{R} \times [0,\infty)$ and that $\Pi(x,0) = f(x)$ for $x \in \mathbb{R}$.

6.6 The Sobolev space $W^{k,p}(\Omega)$

A brief account of Sobolev spaces, which are fundamental in modern theory of partial differential equations and calculus of variations, will now be given.

A locally integrable function u defined on an open set $\Omega \subset \mathbb{R}^n$ is said to be **weakly differentiable up to order** k on Ω, k being a positive integer, if for any multi-index $\alpha = (\alpha_1, \ldots, \alpha_n)$ with $|\alpha| = \alpha_1 + \cdots + \alpha_n \leq k$ there is a locally integrable function g_α on Ω, such that

$$\int_\Omega u\partial^\alpha \varphi d\lambda^n = (-1)^{|\alpha|} \int_\Omega g_\alpha \varphi d\lambda^n, \tag{6.27}$$

for all $\varphi \in C_c^\infty(\Omega)$. Observe that g_α is uniquely determined by u in the sense that any two such functions are equivalent. We therefore denote g_α by u_α . Note that $u_0 = u_{(0,\ldots,0)} = u$. Clearly, functions u in $C^k(\Omega)$ are weakly differentiable up to order k on Ω with $u_\alpha = \partial^\alpha u$. For $p \geq 1$, let $W^{k,p}(\Omega)$ be the equivalence class of all such functions u in $L^p(\Omega)$ which is weakly differentiable up to order k on Ω such that $u_\alpha \in L^p(\Omega)$ for all α with $|\alpha| \leq k$. $W^{k,p}(\Omega)$ is a vector space with the usual definition of addition and multiplication by scalar. On $W^{k,p}(\Omega)$ a norm $\|\cdot\|_{k,p}$ is defined by

$$\begin{aligned} \|u\|_{k,p} &= \left(\sum_{|\alpha| \leq k} \|u_\alpha\|_p^p \right)^{\frac{1}{p}} \quad \text{if } p < \infty; \\ &= \sum_{|\alpha| \leq k} \|u_\alpha\|_\infty \quad \text{if } p = \infty. \end{aligned} \tag{6.28}$$

To see that $\|u\|_{k,p}$ is actually a norm, we need only verify that triangle inequality holds when $1 \leq p < \infty$: $\|u+v\|_{k,p} \leq (\sum_{|\alpha|\leq k}\{\|u_\alpha\|_p + \|v_\alpha\|_p\}^p)^{\frac{1}{p}} \leq (\sum_{|\alpha|\leq k} \|u_\alpha\|_p^p)^{\frac{1}{p}} + (\sum_{|\alpha|\leq k} \|v_\alpha\|_p^p)^{\frac{1}{p}} = \|u\|_{k,p} + \|v\|_{k,p}$, where we have used the Minkowski inequality for $l^p(S)$ with S a finite set. Of course, there are equivalent norms for $W^{k,p}(\Omega)$; for example, we may also define $\|u\|_{k,p}$ as $\sum_{|\alpha|\leq k} \|u_\alpha\|_p$. We prefer the norm defined in (6.28), because when $p = 2$, the norm comes from an inner product on $W^{k,2}(\Omega)$, defined by

$$(u, v)_k = \sum_{|\alpha| \leq k} \int_\Omega u_\alpha \bar{v}_\alpha d\lambda^n. \tag{6.29}$$

If u is weakly differentiable to certain order, u_α's are called generalized partial derivatives of u, and often u_α is denoted by $\partial^\alpha u$ or $\frac{\partial^{|\alpha|}u}{\partial x_1^{\alpha_1}\cdots\partial x_n^{\alpha_n}}$; many notations related to smooth functions are also borrowed to be applied to weakly differentiable functions, for example, if u is weakly differentiable to first order, ∇u is used to denote $(\frac{\partial u}{\partial x_1}, \ldots, \frac{\partial u}{\partial x_n})$ and is called the **generalized gradient** of u.

In what follows in this section, p and q are conjugate exponents.

Theorem 6.6.1 *$W^{k,p}(\Omega)$ is a Banach space.*

Proof Let $\{u^{(j)}\}$ be a Cauchy sequence in $W^{k,p}(\Omega)$. For each α with $|\alpha| \le k$, $\{u_\alpha^{(j)}\}$ is a Cauchy sequence in $L^p(\Omega)$, hence, $\lim_{j\to\infty} \|u_\alpha^{(j)} - g_\alpha\|_p = 0$ for some $g_\alpha \in L^p(\Omega)$. If we put $u = g_0$, we shall show that $u \in W^{k,p}(\Omega)$ and $\lim_{j\to\infty} \|u^{(j)} - u\|_{k,p} = 0$. For any given $\varphi \in C_c^\infty(\Omega)$,

$$\begin{aligned}
&\left|\int_\Omega u\partial^\alpha \varphi d\lambda^n - (-1)^{|\alpha|}\int_\Omega g_\alpha \varphi d\lambda^n\right| \\
&= \left|\int_\Omega (u - u^{(j)})\partial^\alpha \varphi d\lambda^n + (-1)^{|\alpha|}\int_\Omega (u_\alpha^{(j)} - g_\alpha)\varphi d\lambda^n\right| \\
&\le \|u - u^{(j)}\|_p \|\partial^\alpha \varphi\|_q + \|g_\alpha - u_\alpha^{(j)}\|_p \|\varphi\|_q,
\end{aligned}$$

from which by letting $j \to \infty$, we have

$$\int_\Omega u\partial^\alpha \varphi d\lambda^n = (-1)^\alpha \int_\Omega g_\alpha \varphi d\lambda^n,$$

and hence u is weakly differentiable up to order k with $u_\alpha = g_\alpha$. Thus $u \in W^{k,p}(\Omega)$. That $\lim_{j\to\infty} \|u - u^{(j)}\|_{k,p} = 0$ follows from $\lim_{j\to\infty} \|u_\alpha - u_\alpha^{(j)}\|_p = 0$, for each α with $|\alpha| \le k$. ∎

Theorem 6.6.1 implies in particular that $W^{k,2}(\Omega)$ is a Hilbert space with inner product defined by (6.29).

Exercise 6.6.1 A locally integrable function u defined on an open set Ω in $\mathbb{R}^n$ is in $W^{k,p}(\Omega)$, $p > 1$, if and only if for each multi-index α with $|\alpha| \le k$, there is a constant $C_\alpha > 0$ such that

$$\left|\int_\Omega u\partial^\alpha \varphi d\lambda^n\right| \le C_\alpha \|\varphi\|_q$$

for all $\varphi \in C_c^\infty(\Omega)$, where p, q are conjugate exponents.

Exercise 6.6.2 Let $\{J_\varepsilon\}_{\varepsilon>0}$ be a Friederich mollifier and suppose that u is weakly differentiable up to order k on an open set $\Omega \subset \mathbb{R}^n$. Show that for any multi-index α with $|\alpha| \le k$, we have

$$\partial^\alpha (J_\varepsilon u)(x) = J_\varepsilon u_\alpha(x), \quad x \in \Omega_\varepsilon,$$

where $\Omega_\varepsilon = \{x \in \Omega : \operatorname{dist}(x, \Omega^c) > \varepsilon\}$.

Exercise 6.6.3 Let $u \in W^{k,p}(\Omega)$, $1 \le p < \infty$. Show that there is a sequence $\{v_j\} \subset C^\infty(\mathbb{R}^n)$ such that for every $\varepsilon > 0$, $v_j \in W^{k,p}(\Omega_\varepsilon)$ when j is large and $v_j \to u$ in $W^{k,p}(\Omega_\varepsilon)$. Note that $v_j \in W^{k,p}(\Omega_\varepsilon)$ implicitly implies that the restriction of v_j to Ω_ε is also denoted by v_j.

Exercise 6.6.4 Let I be an open interval in $\mathbb{R}$. Show that a locally integrable function f on I is in $W^{1,1}(I)$ if and only if it is equivalent to a function g which is absolutely continuous on every finite closed interval in I and $g' \in L^1(I)$.

Theorem 6.6.2 *Suppose that $u \in W^{1,1}(\mathbb{R}^n)$, then*

$$u(x) = \frac{1}{nb_n} \int_{\mathbb{R}^n} \frac{(x-\xi) \cdot \nabla u(\xi)}{|x-\xi|^n} d\xi$$

for a.e. x in $\mathbb{R}^n$, where $b_n = \lambda^n(B_1(0))$ and $\nabla u = \left(\frac{\partial u}{\partial x_1}, \ldots, \frac{\partial u}{\partial x_n}\right)$.

Proof We know from Exercises 6.1.3 and 6.6.3 that there is a sequence $\{u_j\}$ in $C^\infty(\mathbb{R}^n) \cap W^{1,1}(\mathbb{R}^n)$ such that $\lim_{j\to\infty} \|u_j - u\|_{1,1} = 0$ and $u_j(x) \to u(x)$ for a.e. x in $\mathbb{R}^n$. Apply Corollary 4.11.1 to each u_j; we have

$$u_j(x) = \frac{1}{nb_n} \int_{\mathbb{R}^n} \frac{(x-\xi) \cdot \nabla u_j(\xi)}{|x-\xi|^n} d\xi, \quad x \in \mathbb{R}^n. \tag{6.30}$$

Fix $R > 0$. Let $\Omega = B_{R+1}(0)$, $D = B_R(0)$, and put

$$g_j(x) = \int_\Omega \frac{|\nabla u_j(\xi) - \nabla u(\xi)|}{|x-\xi|^{n-1}} d\xi, \quad x \in D.$$

By Theorem 4.11.2, $\|g_j\|_1 \to 0$ as $j \to \infty$; hence, $\{g_j\}$ has a subsequence $\{g_{j'}\}$ such that $g_{j'}(x) \to 0$ as $j' \to \infty$ for a.e. x in D, by Exercise 6.1.3. Now,

$$u_{j'}(x) = \frac{1}{nb_n} \int_{\mathbb{R}^n} \frac{(x-\xi) \cdot (\nabla u_{j'}(\xi) - \nabla u(\xi))}{|x-\xi|^n} d\xi + \frac{1}{nb_n} \int_{\mathbb{R}^n} \frac{(x-\xi) \cdot \nabla u(\xi)}{|x-\xi|^n} d\xi;$$

if we show that $\int_{\mathbb{R}^n} \frac{(x-\xi)\cdot(\nabla u_{j'}(\xi)-\nabla u(\xi))}{|x-\xi|^n} d\xi \to 0$ for a.e. x in D as $j' \to \infty$, then $u(x) = \frac{1}{nb_n} \int_{\mathbb{R}^n} \frac{(x-\xi)\cdot\nabla(\xi)}{|x-\xi|^n} d\xi$ for a.e. x in D. But, for $x \in D$, we have

$$\begin{aligned}
\left| \int_{\mathbb{R}^n} \frac{(x-\xi) \cdot (\nabla u_{j'}(\xi) - \nabla u(\xi))}{|x-\xi|^n} d\xi \right| &\le \int_{\mathbb{R}^n} \frac{|\nabla u_{j'}(\xi) - \nabla u(\xi)|}{|x-\xi|^{n-1}} d\xi \\
&= g_{j'}(x) + \int_{\mathbb{R}^n \setminus \Omega} \frac{|\nabla u_{j'}(\xi) - \nabla u(\xi)|}{|x-\xi|^{n-1}} d\xi \\
&\le g_{j'}(x) + \int_{\mathbb{R}^n} |\nabla u_{j'}(\xi) - \nabla u(\xi)| d\xi \to 0
\end{aligned}$$

as $j' \to \infty$ for those x where $g_{j'}(x) \to 0$. Thus $u(x) = \frac{1}{nb_n} \int_{\mathbb{R}^n} \frac{(x-\xi)\cdot\nabla u(\xi)}{|x-\xi|^n} d\xi$ for a.e. x in D. Since $R > 0$ is arbitrary, the theorem is proved. ∎

The closure of $C_c^\infty(\Omega)$ in $W^{k,p}(\Omega)$ is denoted by $\overset{\circ}{W}{}^{k,p}(\Omega)$; functions in $\overset{\circ}{W}{}^{k,p}(\Omega)$ are said to **vanish on $\partial\Omega$ in a generalized sense**.

Exercise 6.6.5 Suppose that Ω is bounded and let $u \in \overset{\circ}{W}{}^{k,\infty}(\Omega)$. Show that u is equivalent to a function $v \in C^k(\Omega)$ which can be continuously extended to be zero on $\partial\Omega$, together with all its partial derivatives up to order k.

Exercise 6.6.6 Show that if $u \in W^{k,p}(\Omega)$, then

$$\int_\Omega u\partial^\alpha v d\lambda^n = (-1)^{|\alpha|} \int_\Omega u_\alpha v d\lambda^n$$

for all $v \in \overset{\circ}{W}{}^{k,q}(\Omega)$ if $|\alpha| \leq k$.

Exercise 6.6.7 Let g be in $C^\infty(\mathbb{R}^n)$ satisfying $0 \leq g \leq 1, g = 0$ outside $B_2(0)$, and $g = 1$ on $B_1(0)$. For $j \in \mathbb{N}$, let g_j be the function defined on $\mathbb{R}^n$ by

$$g_j(x) = g(j^{-1}x), \quad x \in \mathbb{R}^n.$$

(i) Suppose that $u \in C^k(\mathbb{R}^n) \cap W^{k,p}(\mathbb{R}^n)$, $1 \leq p < \infty$. Show that $\lim_{j\to\infty} \|g_j u - u\|_{k,p} = 0$.

(ii) Show that $\overset{\circ}{W}{}^{k,p}(\mathbb{R}^n) = W^{k,p}(\mathbb{R}^n)$ if $1 \leq p < \infty$.

Theorem 6.6.3 (Poincaré) *If Ω is a bounded open set in $\mathbb{R}^n$, then on $\overset{\circ}{W}{}^{k,p}(\Omega)$ the norm $\|\cdot\|_{k,p}$ is equivalent to the norm $|\cdot|_{k,p}$, defined for $u \in \overset{\circ}{W}{}^{k,p}(\Omega)$ by*

$$|u|_{k,p} = \left(\sum_{|\alpha|=k} \|u_\alpha\|_p^p\right)^{1/p}, \quad p < \infty;$$
$$|u|_{k,\infty} = \sum_{|\alpha|=k} \|u_\alpha\|_\infty.$$

Proof We prove the theorem for $k = 1$ and $p < \infty$; the proof for the general case will be clear from the proof of this particular case.

For $u \in \overset{\circ}{W}{}^{1,p}(\Omega)$, we are going to show that there is $C > 0$, independent of u, such that $\|u\|_{1,p} \leq C|u|_{1,p}$. From the definition of $\overset{\circ}{W}{}^{1,p}(\Omega)$, we may assume that $u \in C_c^\infty(\Omega)$. By letting $u = 0$ outside Ω, we may further assume that $u \in C_c^\infty(I)$, where I is an open oriented cube containing Ω and with side-width $= l$. Express I as $I = I_1 \times \hat{I}_1$, where $I_1 = (a, b) \subset \mathbb{R}$ and $\hat{I}_1 \subset \mathbb{R}^{n-1}$; then for $x \in I$, x can be expressed as $(x_1, \hat{x}_1)$ with $x_1 \in (a, b)$ and $\hat{x}_1 \in \hat{I}_1$. Now, $u(x) = \int_a^{x_1} \frac{\partial u}{\partial x_1}(t, \hat{x}_1)dt$ implies that $|u(x)|^p \leq (x_1 - a)^{p/q} \int_a^b |\frac{\partial u}{\partial x_1}(t, \hat{x}_1)|^p dt$ and hence,

$$\|u\|_p^p \le (b-a)^{p/q}(b-a)\int_I \left|\frac{\partial u}{\partial x_1}(x)\right|^p dx = (b-a)^p \left\|\frac{\partial u}{\partial x_1}\right\|_p^p$$
$$\le (b-a)^p \sum_{j=1}^n \left\|\frac{\partial u}{\partial x_j}\right\|_p^p,$$

from which it follows that

$$\|u\|_{1,p}^p \le \{1+(b-a)^p\}|u|_{1,p}^p;$$

therefore $\|u\|_{1,p} \le C|u|_{1,p}$, where $C = \{1+(b-a)^p\}^{1/p}$. Then,

$$|u|_{1,p} \le \|u\|_{1,p} \le C|u|_{1,p},$$

implying that $\|\cdot\|_{1,p}$ and $|\cdot|_{1,p}$ are equivalent. ∎

Remark Since $|u|_{k,p} \le \|u\|_{k,p}$ for $u \in \overset{\circ}{W}{}^{k,p}(\Omega)$, Theorem 6.6.3 is equivalent to the statement that there is $C > 0$ such that

$$\|u\|_{k,p} \le C|u|_{k,p} \tag{6.31}$$

for all $u \in \overset{\circ}{W}{}^{k,p}(\Omega)$. Inequality (6.31) is called the **Poincaré inequality**; and Theorem 6.6.3 is usually referred to as the Poincaré inequality.

The following lemma is a generalization of Example 4.11.2.

Lemma 6.6.1 *Let $u \in W^{1,p}(\mathbb{R}^n)$, $1 \le p < \infty$, then,*

$$\int_{B_R(x)} \frac{|u(\xi)-u(x)|}{|\xi - x|} d\xi \le M|\nabla u|(x)$$

for x in $\mathbb{R}^n$, where $M|\nabla u|$ is the maximal function of ∇u.

Proof Fix a Friederichs mollifier $\{J_\varepsilon\}_{\varepsilon>0}$, and let $u_\varepsilon = J_\varepsilon u$ (cf. Section 4.9), then $\lim_{\varepsilon\to 0}\|J_\varepsilon u - u\|_{1,p} = 0$, by Exercise 6.6.2 and Theorem 4.9.2; hence $u_\varepsilon \to u$, $\nabla u_\varepsilon \to \nabla u$ in $L^p(\mathbb{R}^n)$. Fix $x \in \mathbb{R}^n$ and $R > 0$, in terms of polar coordinates of $y - x$; we have

$$\int_{B_R(x)} |u_\varepsilon(y) - u(y)|dy = \int_0^R \rho^{n-1}\int_{S^{n-1}} |u_k(\rho,\theta) - u(\rho,\theta)|d\sigma(\theta)d\rho$$
$$= \int_{S^{n-1}}\int_0^R \rho^{n-1}|u_k(\rho,\theta) - u(\rho,\theta)|d\rho d\sigma(\theta) \to 0$$

as $\varepsilon \searrow 0$. We infer then from Example 4.8.2 that there is a sequence $\varepsilon_k \searrow 0$ such that $\int_0^R \rho^{n-1}|u_{\varepsilon_k}(\rho,\theta) - u(\rho,\theta)|d\rho \to 0$ as $k \to \infty$ for σ-a.e. $\theta \in S^{n-1}$.

Then for any $0 < \delta < R$, $\int_\delta^R |u_{\varepsilon_k}(\rho,\theta) - u(\rho,\theta)|d\rho \to 0$ as $k \to \infty$. Similarly, $\int_\delta^R |\nabla u_{\varepsilon_k'}(\rho,\theta) - \nabla u(\rho,\theta)|d\rho \to 0$ as $k \to \infty$ for σ-a.e. $\theta \in S^{n-1}$. Since we may choose ε_k', a subsequence of ε_k, we conclude that there is a sequence $\{u_k\}$ in $C^\infty(\mathbb{R}^n) \cap W^{1,p}(\mathbb{R}^n)$ such that for a.e. y in $B_R(x)\backslash B_\delta(x)$,

$$\int_\delta^R |u_k(x + t(y - x)) - u(x + t(y - x))|dt \to 0$$

and

$$\int_\delta^R |\nabla u_k(x + t(y - x)) - \nabla u(x + t(y - x))|dt \to 0$$

as $k \to \infty$. Therefore for a.e. y in $B_R(x)\backslash B_\delta(x)$, $u(x + t(y - x))$ is AC on $[\delta, R]$ and $\frac{d}{dt}u(x + t(y - x)) = \nabla u(x + t(y - x)) \cdot (y - x)$ for a.e. t on $[\delta, R]$. Then, as in Example 4.11.2,

$$\begin{aligned}\int_{B_R(x)\backslash B_\delta(x)} \frac{|u(\xi) - u(x)|}{|\xi - x|}d\xi &\leq \int_0^1 \frac{1}{t^n} \int_{B_{Rt}(x)\backslash B_{\delta(t)}(x)} |\nabla u(z)|dz \\ &\leq \int_0^1 \frac{1}{t^n} \int_{B_{Rt}(x)} |\nabla u(z)|dz \\ &\leq \lambda^n(B_R(x)) \cdot M|\nabla u|(x).\end{aligned}$$

We conclude the proof by letting $\delta \searrow 0$. ■

Theorem 6.6.4 *There is a positive constant $\theta = \theta(n,p)$, $1 < p < \infty$ with the property that if $u \in W^{1,p}(\mathbb{R}^n)$, then for $\varepsilon > 0$ there is a closed set $F \subset \mathbb{R}^n$ such that $u|_F$, the restriction of u to F, is Lipschitz with Lipschitz constant* $\mathrm{Lip}(u|_F)$, *satisfying*

$$\mathrm{Lip}(u|_F)^p \lambda^n(\mathbb{R}^n\backslash F) < \theta(n,p)\varepsilon.$$

Proof For x, y in $\mathbb{R}^n$, put $q(x,y) = \frac{|u(y)-u(x)|}{|y-x|}$, then

$$\frac{1}{\sigma_R} \int_{B_R(x)} q(x,y)dy \leq M|\nabla u|(x), \tag{6.32}$$

from Lemma 6.6.1, where $\sigma_R = \lambda^n(B_R(x))$. For $x \in \mathbb{R}^n$ and $\lambda > 0$, let $W_R(x,\lambda) = \{y \in B_R(x) : q(x,y) \leq \lambda\}$; we have from (6.1) and (6.32),

$$\lambda^n(B_R(x)\backslash W_R(x,\lambda)) \leq \frac{1}{\lambda} \int_{B_R(x)} q(x,y)dy \leq \frac{\sigma_R}{\lambda}M|\nabla u|(x). \tag{6.33}$$

Now put $Z_\delta = \{x \in \mathbb{R}^n : M|\nabla u|(x) \leq \delta\}$, and choose $k_0 > 1$ such that

$$\lambda^n(B_R(x) \cap B_R(y)) > \frac{2}{k_0}\sigma_R, \quad R = |x-y|. \tag{6.34}$$

Consider now x, y in Z_δ; we have from (6.33),

$$\lambda^n(B_R(z) \backslash W_R(z, k_0\delta)) \leq \frac{M|\nabla u|(z)}{k_0\delta}\sigma_R \leq \frac{1}{k_0}\sigma_R, \tag{6.35}$$

for $z = x$ or y and $R = |x-y|$. It follows from (6.34) and (6.35) that $W_R(x, k_0\delta) \cap W_R(y, k_0\delta) \neq \emptyset$; choose $z_0 \in W_R(x, k_0\delta) \cap W_R(y, k_0\delta)$, then

$$q(x,y) \leq q(x, z_0) + q(y, z_0) \leq 2k_0\delta. \tag{6.36}$$

Given that $\varepsilon > 0$, by (6.2) there is $\delta > 0$ such that $\delta^p\lambda^n(\{M|\nabla u| > \delta\}) < \varepsilon$. Choose then a closed set F in Z_δ with $\lambda^n(\mathbb{R}^n \backslash F) < 2\lambda^n(\{M|\nabla u| > \delta\})$. The restriction of u to F is a Lipschitz function with Lipschitz constant $\leq 2k_0\delta$, by (6.36); therefore $(\frac{\mathrm{Lip}(u|_F)}{2k_0})^p\lambda^n(\mathbb{R}^n \backslash F) < 2\varepsilon$. We choose $\theta = \theta(n,p) = 2^{p+1}k_0^p$ to complete the proof. ■

Remark If $M(|u| + |\nabla u|)$ is substituted for $M|\nabla u|$ in Theorem 6.6.4, the closed set F can be chosen so that $\|u|_F\|_\infty + \mathrm{Lip}(u|_F) \leq 2\,\mathrm{Lip}(u|_F)$; this observation, together with the known fact that $u|_F$ can be extended to a Lipschitz function v on $\mathbb{R}^n$ such that $\|v\|_\infty + \mathrm{Lip}(v) \leq A(\|u|_F\|_\infty + \mathrm{Lip}(u|_F))$, where A is a constant depending only on n (cf. [St, Chapter VI]), shows that Theorem 6.6.4 can be formulated as follows. A function $u \in L^p(\mathbb{R}^n)$ is in $W^{1,p}(\mathbb{R}^n)$ if and only if for any given $\varepsilon > 0$ there is a Lipschitz function v on $\mathbb{R}^n$, and a closed set F such that $u = v$ on F, $\lambda^n(\mathbb{R}^n \backslash F) < \varepsilon$, and $\|u - v\|_{1,p} < \varepsilon$.

Besides, Theorem 6.6.4 also holds when $p = 1$, because in the last paragraph of the proof of the theorem, δ can be chosen so that $\delta\lambda^n(\{M|\nabla u| > \delta\}) < \varepsilon$ follows from the improved form of Theorem 6.4.2:

$$\lambda^n(\{Mf > \alpha\}) \leq 2A\alpha^{-1}\int_{\{|f| > \frac{\alpha}{2}\}} |f| d\lambda^n,$$

of which we refer to [St, P.7].

Since $W^{k,2}(\Omega)$ is a Hilbert space, it will be denoted by $H^k(\Omega)$; accordingly, $\overset{\circ}{W}{}^{k,2}(\Omega)$ is denoted by $\overset{\circ}{H}{}^k(\Omega)$. By Exercise 6.6.7 (ii), $\overset{\circ}{H}{}^k(\mathbb{R}^n) = H^k(\mathbb{R}^n)$; $H^k(\mathbb{R}^n)$ is usually abbreviated to H^k. In Chapter 7, with the help of the Fourier integral, H^s will also be defined for fractional number s.

7 Fourier Integral and Sobolev Space H^s

The Fourier integral is a useful construct in analysis which is based on an idea of **J. Fourier** for resolving functions into basic harmonics in his treatment of conduction of heat. When functions are periodic, say of period 2π, they are resolved as Fourier series (see Section 5.9). For nonperiodic functions on $\mathbb{R}$, the idea leads to a Fourier integral. The Fourier integral for L^1 functions on $\mathbb{R}^n$ can be defined straight away, and is treated in Section 7.1. Since L^2 is a Hilbert space, it is desirable to define a Fourier integral for L^2 functions; but a straightforward definition for L^2 functions is lacking; some variation is therefore necessary for the purpose. We shall get around this through the Fourier integral for rapidly decreasing functions, introduced in Section 7.2. Applications to Sobolev spaces H^s and to partial differential equations are provided in later sections of the chapter. The Fourier integral of probability distributions is introduced in Section 7.5, and is applied to prove the central limit theorem of probability theory.

A Fourier integral is also called a Fourier transform.

For the convenience of expressing certain functions defined on $\mathbb{R}^n$, the function $x \mapsto f(x)$ will sometimes be expressed by $f(x)$. For example, $x \mapsto x^\alpha$ is simply denoted by x^α, and if f is a function on $\mathbb{R}^n$, the function $x \mapsto x^\alpha f(x)$ is denoted by $x^\alpha f$.

7.1 Fourier integral for L^1 functions

For $f \in L^1 := L^1(\mathbb{R}^n)$, define the **Fourier integral** Ff of f by

$$(Ff)(\xi) = \frac{1}{(2\pi)^{\frac{n}{2}}} \int_{\mathbb{R}^n} f(x) e^{-i\xi \cdot x} dx, \quad \xi \in \mathbb{R}^n.$$

Since $|f(x)e^{-i\xi\cdot x}| = |f(x)|$, Ff is defined and is finite for every $\xi \in \mathbb{R}^n$. One verifies readily that

(1) $\|Ff\|_\infty \le \frac{1}{(2\pi)^{\frac{n}{2}}}\|f\|_1$;

(2) Ff is uniformly continuous on $\mathbb{R}^n$ (note that this follows from LDCT).

Exercise 7.1.1 Let $f(x) = e^{-|x|}$, $x \in \mathbb{R}$. Find Ff.

Exercise 7.1.2 Suppose that $f_1, \ldots, f_n$ are in $L^1(\mathbb{R})$ and let $f(x) = \prod_{j=1}^n f_j(x_j)$ for $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Show that $f \in L^1(\mathbb{R}^n)$ and $(Ff)(\xi) = \prod_{j=1}^n (Ff_j)(\xi_j)$ for $\xi = (\xi_1, \ldots, \xi_n)$.

Example 7.1.1

(i) For $\alpha > 0$, consider the function $f = I_{[-\alpha,\alpha]}$ on $\mathbb{R}$; then $(Ff)(\xi) = \frac{2\sin\alpha\xi}{\sqrt{2\pi}\xi}$, $\xi \in \mathbb{R}$. For $n > 1$, $f = I_{[-\alpha,\alpha]\times\cdots\times[-\alpha,\alpha]}$, $(Ff)(\xi) = \frac{2^n}{(2\pi)^{\frac{n}{2}}}\prod_{j=1}^n \frac{\sin\alpha\xi_j}{\xi_j}$.
This follows from Exercise 7.1.2.

(ii) For $n = 1$, consider the function $f(x) = e^{-\frac{x^2}{2}}$. We have

$$(Ff)(\xi) = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{-\frac{x^2}{2}} e^{-i\xi x} dx = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{-\frac{x^2}{2}} \cos\xi x dx,$$

and

$$\begin{aligned}(Ff)'(\xi) &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{-\frac{x^2}{2}}(-x\sin\xi x)dx \\ &= \frac{1}{\sqrt{2\pi}}\xi\int_{\mathbb{R}} e^{-\frac{x^2}{2}}\cos\xi x dx = \xi(Ff)(\xi).\end{aligned}$$

The first equality follows by LDCT and the second by integration by parts. Then $(Ff)(\xi) = Ce^{-\frac{\xi^2}{2}}$ with C being a constant. But $(Ff)(0) = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{-\frac{x^2}{2}} dx = 1 = C$. Thus $(Ff)(\xi) = e^{-\frac{\xi^2}{2}}$. For $n > 1$, if $f(x) = e^{-\frac{|x|^2}{2}}$, then $(Ff)(\xi) = e^{-\frac{|\xi|^2}{2}}$.

Exercise 7.1.3 Consider the function $f(x) = e^{-\frac{1}{2}x^2}$ in Example 7.1.1 (ii). Use a contour integral to show that

$$\int_{\mathbb{R}} e^{-\frac{1}{2}(x+i\xi)^2} dx = \int_{\mathbb{R}} e^{-\frac{1}{2}x^2} dx = \sqrt{2\pi}$$

and give a direct verification that

$$Ff(\xi) = e^{-\frac{1}{2}\xi^2}.$$

Theorem 7.1.1 *If $f, g \in L^1(\mathbb{R}^n)$, $(F\{f * g\})(\xi) = (2\pi)^{\frac{n}{2}}(Ff)(\xi)(Fg)(\xi)$.*

Proof Observe first that $f * g \in L^1(\mathbb{R}^n)$, by the Young inequality (Theorem 6.5.1). Then,

$$\begin{aligned}(F\{f * g\})(\xi) &= \frac{1}{(2\pi)^{\frac{n}{2}}} \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} f(x-y)g(y)dy \right) e^{-i\xi\cdot x}dx \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}} \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} f(x-y)e^{-i\xi\cdot(x-y)}dx \right) g(y)e^{-i\xi\cdot y}dy \\ &= \int_{\mathbb{R}^n} (Ff)(\xi)g(y)e^{-i\xi\cdot y}dy = (2\pi)^{\frac{n}{2}}(Ff)(\xi)(Fg)(\xi).\end{aligned}$$

It is to be noted that since $f(x-y)g(y)e^{-i\xi\cdot x}$ is an integrable function of (x, y) in $\mathbb{R}^{2n}$, it is legitimate to use the Fubini theorem in the above argument. ■

Example 7.1.2 Let $\alpha > 0$. It is readily verified that $\frac{1}{\alpha}I_{[-\frac{\alpha}{2},\frac{\alpha}{2}]} * I_{[-\frac{\alpha}{2},\frac{\alpha}{2}]}(x) = \left(1 - \frac{|x|}{\alpha}\right)^+$ (cf. Exercise 6.5.2), it then follows than

$$\left(F\left(1 - \frac{|x|}{\alpha}\right)^+\right)(\xi) = \frac{\sqrt{2\pi}}{\alpha}\left(\frac{2\sin\frac{\alpha}{2}\xi}{\sqrt{2\pi}\xi}\right)^2 = \frac{1}{\alpha}\cdot\frac{2(1-\cos\alpha\xi)}{\sqrt{2\pi}\xi^2}.$$

For $f \in L^1(\mathbb{R}^n)$, the **inverse Fourier integral** $\check{F}f$ of f is defined by

$$(\check{F}f)(\xi) = \frac{1}{(2\pi)^{\frac{n}{2}}} \int_{\mathbb{R}^n} f(x)e^{i\xi\cdot x}dx, \quad \xi \in \mathbb{R}^n.$$

For $f \in L^1$, Ff and $\check{F}f$ are often denoted by $\hat{f}$ and $\check{f}$ respectively.

Exercise 7.1.4 Recall that for $a \in \mathbb{R}^n$, $\sigma > 0$, and a function f on $\mathbb{R}^n$, $f^a(x) = f(x-a)$, $f_\sigma(x) = \sigma^{-n}f(\frac{x}{\sigma})$ for $x \in \mathbb{R}^n$.

(i) Show that $\widehat{f^a}(\xi) = e^{-i\xi\cdot a}\hat{f}(\xi)$ for $f \in L^1$.

(ii) Show that $\widehat{f_\sigma}(\xi) = \hat{f}(\sigma\xi)$.

Exercise 7.1.5 Let f, g be in L^1. Show that

$$\int_{\mathbb{R}^n} f\hat{g}d\lambda^n = \int_{\mathbb{R}^n} \hat{f}gd\lambda^n.$$

Theorem 7.1.2 (Riemann–Lebesgue) *If $f \in L^1$, then $\hat{f} \in C_0(\mathbb{R}^n)$.*

Proof If f is the function considered in Example 7.1.1 (i), then $\lim_{|\xi|\to\infty}\hat{f}(\xi) = 0$; hence the theorem holds for indicator functions of cubes, by Exercise 7.1.4 (i); as a consequence the theorem holds for finite linear combinations of indicator functions of cubes. But, as $C_c(\mathbb{R}^n)$ is dense in L^1, one verifies easily that the family of all finite linear combinations of indicator functions of cubes is dense in L^1. Thus for $f \in L^1$ and $\varepsilon > 0$, there is a finite linear combination φ of indicator functions of cubes such that $\|f - \varphi\|_1 < \frac{\varepsilon}{2}$, then $|\hat{f}(\xi)| \le |\widehat{(f-\varphi)}(\xi)| + |\hat{\varphi}(\xi)| < \frac{\varepsilon}{2} + |\hat{\varphi}(\xi)|$, from which it follows that $|\hat{f}(\xi)| < \varepsilon$ if $|\xi|$ is large enough, because $\hat{\varphi} \in C_0(\mathbb{R}^n)$. ∎

Theorem 7.1.3 *For $f \in L^1$, f is uniquely determined by $\hat{f}$; in other words, the map $f \mapsto \hat{f}$ is injective on L^1.*

Proof Take h to be the function defined on $\mathbb{R}^n$ by $h(x) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{|x|^2}{2}}$ (note $\int h d\lambda^n = 1$), then $\{h_\sigma\}_{\sigma>0}$ is an approximate identity for L^1. Put $m_\sigma f = f * h_\sigma$. Then, since $\hat{h}(\xi) = h(\xi)$, as in Example 7.1.1 (ii), we have

$$\begin{aligned}
(m_\sigma f)(x) &= \int_{\mathbb{R}^n} f(y) h_\sigma(x-y)dy = \int_{\mathbb{R}^n} f(y) h_\sigma(y-x)dy \\
&= \sigma^{-n} \int_{\mathbb{R}^n} f(y) h\left(\frac{y-x}{\sigma}\right) dy \\
&= \frac{\sigma^{-n}}{(2\pi)^{\frac{n}{2}}} \int_{\mathbb{R}^n} \left(f(y) \int_{\mathbb{R}^n} e^{-i\frac{y-x}{\sigma}\cdot z} h(z)dz \right) dy \\
&= \frac{\sigma^{-n}}{(2\pi)^{\frac{n}{2}}} \int_{\mathbb{R}^n} \left(f(y) \int_{\mathbb{R}^n} e^{-i(y-x)\cdot\frac{z}{\sigma}} h(z)dz \right) dy \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}} \int_{\mathbb{R}^n} \left(f(y) \int_{\mathbb{R}^n} e^{-i(y-x)\cdot z} h(\sigma z)dz \right) dy \\
&= \int_{\mathbb{R}^n} e^{ix\cdot z} \hat{f}(z) h(\sigma z)dz;
\end{aligned}$$

this means that the function $m_\sigma f$ is uniquely determined by $\hat{f}$. But as $m_\sigma f \to f$ in L^1 as $\sigma \to 0$, f is uniquely determined by $\hat{f}$. ∎

Theorem 7.1.4 (L^1 inversion theorem) *If both f and $\hat{f}$ are in L^1, then $f = (\hat{f})^\vee$, i.e. f is the inverse Fourier integral of $\hat{f}$.*

Proof Let h and $\{m_\sigma\}_{\sigma>0}$ be as in the proof of Theorem 7.1.3. There we have shown that

$$(m_\sigma f)(x) = \int_{\mathbb{R}^n} e^{ix\cdot\xi} \hat{f}(\xi) h(\sigma\xi) d\xi;$$

since $|e^{ix\cdot\xi}\hat{f}(\xi)h(\sigma\xi)| \le \frac{1}{(2\pi)^{n/2}} |\hat{f}(\xi)|$ and $\lim_{\sigma\to 0} e^{ix\cdot\xi}\hat{f}(\xi)h(\sigma\xi) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{ix\cdot\xi}\hat{f}(\xi)$, it follows from LDCT that $\lim_{\sigma\to 0}(m_\sigma f)(x) = (\hat{f})^\vee(x)$ for each $x \in \mathbb{R}^n$. Now, $|(m_\sigma f)(x)| \le \frac{1}{(2\pi)^{\frac{n}{2}}} \|\hat{f}\|_1$ implies that $\lim_{\sigma\to 0} \int_{B_R(0)} |m_\sigma f - (\hat{f})^\vee| d\lambda^n = 0$ for any $R > 0$, again by LDCT; this, together with $\lim_{\sigma\to 0} \int_{B_R(0)} |m_\sigma f - f| d\lambda^n = 0$ (cf. Theorem 6.5.3), shows that $f = (\hat{f})^\vee$ a.e. on $B_R(0)$ for any $R > 0$, and consequently $f = (\hat{f})^\vee$ a.e. on $\mathbb{R}^n$. ∎

As an application of the L^1 inversion theorem, we establish the fact that the family $\{\mathcal{E}_0, \mathcal{E}_1, \mathcal{E}_2, \ldots\}$ of normalized Hermite functions introduced in Example 5.8.1 is an orthonormal basis for $L^2(\mathbb{R})$, or equivalently, that the family $\{h_0, h_1, h_2, \ldots\}$ of normalized Hermite polynomials is an orthonormal basis for $L^2_w(\mathbb{R})$ where $w(x) = e^{-x^2}$.

Corollary 7.1.1 *The family $\{\mathcal{E}_0, \mathcal{E}_1, \mathcal{E}_2, \ldots\}$ of normalized Hermite functions is an orthonormal basis for $L^2(\mathbb{R})$.*

Proof By Theorem 5.8.3, we need to show that if $f \in L^2(\mathbb{R})$ is such that

$$\text{(A)} \qquad \int_{-\infty}^{\infty} f(x)\mathcal{E}_n(x)dx = 0, \quad n = 0, 1, 2, \ldots,$$

then $f = 0$ a.e. Recall from Example 5.8.1 that $\mathcal{E}_n(x) = e^{-\frac{x^2}{2}} h_n(x)$, where $h_n(x)$ is a polynomial of degree n and that each monomial x^n is a linear combination of $h_0(x), \ldots, h_n(x)$; hence if $f \in L^2(\mathbb{R})$ satisfies the condition (A), then it satisfies the condition

$$\text{(B)} \qquad \int_{-\infty}^{\infty} f(x)e^{-\frac{x^2}{2}} x^n dx = 0, \quad n = 0, 1, 2, \ldots.$$

Therefore, it suffices to show that if $f \in L^2(\mathbb{R})$ satisfies the condition (B), then $f = 0$ a.e. Now let $f \in L^2(\mathbb{R})$ satisfy the condition (B). Put $g(x) = f(x)e^{-\frac{x^2}{2}}$, then $g \in L^1(\mathbb{R})$, by the Schwarz inequality and

$$\begin{aligned} \hat{g}(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-itx} f(x)e^{-\frac{x^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left(\sum_{n=0}^{\infty} \frac{(-itx)^n}{n!} \right) f(x)e^{-\frac{x^2}{2}} dx; \end{aligned}$$

but for $N \in \mathbb{N}$,

$$\left| \sum_{n=0}^{N} \frac{(-itx)^n}{n!} f(x)e^{-\frac{x^2}{2}} \right| \le |f(x)|e^{|tx|}e^{-\frac{x^2}{2}},$$

of which the function on the right-hand side is integrable because

$$\int_{-\infty}^{\infty} |f(x)|e^{|tx|}e^{-\frac{x^2}{2}} dx \le \|f\|_2 \left\{ \int_{-\infty}^{\infty} e^{2|tx|}e^{-x^2} dx \right\}^{\frac{1}{2}} < \infty.$$

It follows then from LDCT that

$$\begin{aligned}
\hat{g}(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \lim_{N\to\infty} \left(\sum_{n=0}^{N} \frac{(-itx)^n}{n!} f(x) e^{-\frac{x^2}{2}} \right) dx \\
&= \lim_{N\to\infty} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left(\sum_{n=0}^{N} \frac{(-it)^n}{n!} x^n f(x) e^{-\frac{x^2}{2}} \right) dx \\
&= \lim_{N\to\infty} \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{N} \frac{(-it)^n}{n!} \int_{-\infty}^{\infty} f(x) e^{-\frac{x^2}{2}} x^n dx = 0
\end{aligned}$$

by condition (B). Thus, $\hat{g} = 0$ and by Theorem 7.1.4, $g(t) = (\hat{g})^{\vee}(t) = 0$ a.e. and hence $f = 0$ a.e. ■

A remarkable application of the Fourier integral is the **Poisson summation formula**, which states that

$$\sum_{n=-\infty}^{\infty} f(2n\pi) = \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} \hat{f}(n)$$

for integrable functions f satisfying certain condition. Usually, the Poisson summation formula is established for $f \in C^2(\mathbb{R})$ such that

$$|f(x)| + |f'(x)| + |f''(x)| \leq C(1 + x^2)^{-1}, \quad x \in \mathbb{R},$$

for some constant $C > 0$. We shall prove the formula under weaker conditions. For an integrable function f on $\mathbb{R}$ and $n \in \mathbb{Z}$, let

$$f_n(x) = f(x + 2\pi n), \quad x \in \mathbb{R}.$$

We first claim that $\{f_n(x)\} = \{f_n(x)\}_{n\in\mathbb{Z}}$ is summable for a.e. x in $\mathbb{R}$. For this purpose it is sufficient to show that $\{f_n(x)\}$ is summable for a.e. x in $[-\pi, \pi]$, because if $\{f_n(x)\}$ is summable, then $\{f_n(x + 2\pi m)\} = \{f_{n+m}(x)\} = \{f_n(x)\}$ for any $m \in \mathbb{Z}$, and hence $\sum_{n\in\mathbb{Z}} f_n(x + 2\pi m) = \sum_{n\in\mathbb{Z}} f_n(x)$. Now,

$$\int_{-\pi}^{\pi} \sum_{n\in\mathbb{Z}} |f_n(x)| dx = \sum_{n\in\mathbb{Z}} \int_{-\pi}^{\pi} |f_n(x)| dx = \int_{\mathbb{R}} |f| d\lambda < \infty$$

implies that $\sum_{n\in\mathbb{Z}} |f_n(x)| < \infty$ for a.e. x in $[-\pi, \pi]$. Hence $\{f_n(x)\}$ is summable for a.e. x in $\mathbb{R}$ and if we put $[f](x) = \sum_{n\in\mathbb{Z}} f_n(x)$, if $\{f_n(x)\}$ is summable and $[f](x) = 0$ otherwise, $[f]$ is defined on $\mathbb{R}$ and periodic with period 2π. Furthermore, $[f]$ is integrable on $[-\pi, \pi]$. The function $[f]$ is called the **stacked function** of f. If we define for $j \in \mathbb{N}$ the function $[f]_j$ on $\mathbb{R}$ by

$$[f]_j(x) = \sum_{|n|\leq j} f_n(x),$$

then $[f]_j \to [f]$ a.e. and $|[f]_j| \leq [|f|]$ a.e. Since $[|f|]$ is integrable on $[-\pi, \pi]$, it follows from LDCT that $[f]_j \to [f]$ in $L^1[-\pi, \pi]$. We have proved the following lemma (7.1.1).

Lemma 7.1.1 *Suppose that $f \in L^1 = L^1(\mathbb{R})$. Then $[f]_j \to [f]$ a.e. as well as in $L^1[-\pi, \pi]$ as $j \to \infty$.*

In the immediate following, for $f \in W^{1,1}(\mathbb{R})$ we always take a version of f which is AC on every finite closed interval of $\mathbb{R}$ (note that since $W^{1,1}(\mathbb{R}) = \overset{\circ}{W}{}^{1,1}(\mathbb{R})$, $f(x) = \int_{-\infty}^{x} f'(x)dx$ for a.e. x).

Lemma 7.1.2 *If $f \in W^{1,1}(\mathbb{R})$, then $[f]$ is an AC function on $[-\pi, \pi]$ and satisfies $[f](-\pi) = [f](\pi)$. Furthermore, $[f]' = [f']$ a.e.*

Proof We take a version of f which is AC on every finite closed interval of $\mathbb{R}$. Then f' exists a.e. and is integrable on $\mathbb{R}$; and since

$$
\begin{aligned}
f_n(x) &= f(x + 2n\pi) = f(-\pi + 2n\pi) + \int_{-\pi+2n\pi}^{x+2n\pi} f'(s)ds \\
&= f_n(-\pi) + \int_{-\pi}^{x} f_n'(s)ds \qquad (7.1)
\end{aligned}
$$

for $x \in [-\pi, \pi]$, we have

$$
[f]_j(x) = [f]_j(-\pi) + \int_{-\pi}^{x} [f']_j(s)ds, \quad x \in [-\pi, \pi]. \qquad (7.2)
$$

As $[f']_j \to [f']$ in $L^1[-\pi, \pi]$ as $j \to \infty$, by Lemma 7.1.1, $\lim_{j\to\infty} \int_{-\pi}^{x} [f']_j(s)ds = \int_{-\pi}^{x} [f'](s)ds$ for $x \in [-\pi, \pi]$; we conclude that $\lim_{j\to\infty}[f]_j(-\pi)$ exists and is finite by letting $j \to \infty$ in (7.2) for x such that $\lim_{j\to\infty}[f]_j(x) = [f](x)$. Because $|f| \in W^{1,1}(\mathbb{R})$, we also know that $\lim_{j\to\infty}[|f|]_j(-\pi)$ exists and is finite, from which follows that $\{f_n(-\pi)\}$ is summable and hence $[f](-\pi) = \lim_{j\to\infty}[f]_j(-\pi)$. Now for any finite subset F of $\mathbb{Z}$ and $x \in [-\pi, \pi]$,

$$
\sum_{n\in F} \left| \int_{-\pi}^{x} f_n'(s)ds \right| \leq \int_{-\pi}^{x} \sum_{n\in F} |f_n'(s)|ds \leq \int_{-\pi}^{x} [|f'|](s)ds < \infty,
$$

implying that $\{\int_{-\pi}^{x} f_n'(s)ds\}$ is summable for each $x \in [-\pi, \pi]$. We then infer from (7.1) that $\{f_n(x)\}$ is summable and $[f](x) = \lim_{j\to\infty}[f]_j(x)$ for each $x \in [-\pi, \pi]$. Now let $j \to \infty$ in (7.2); we have

$$
[f](x) = [f](-\pi) + \int_{-\pi}^{x} [f'](s)ds, \quad x \in [-\pi, \pi];
$$

consequently $[f]$ is AC on $[-\pi, \pi]$ and $[f]' = [f']$ a.e. on $[-\pi, \pi]$. Finally, $[f](-\pi) = \sum_{n\in\mathbb{Z}} f_n(-\pi) = \sum_{n\in\mathbb{Z}} f_{n+1}(-\pi) = \sum_{n\in\mathbb{Z}} f_n(\pi) = [f](\pi)$. ■

Lemma 7.1.3 *If $f \in W^{1,1}(\mathbb{R})$, then*

$$\sum_{k\in\mathbb{Z}} f(2\pi k) = \lim_{j\to\infty} \frac{1}{\sqrt{2\pi}} \sum_{|n|\le j} \hat{f}(n).$$

Proof Since $[f]$ is an AC function on $[-\pi, \pi]$, by Lemma 7.1.2, from Theorem 5.9.6 we know that

$$[f](0) = \lim_{j\to\infty} S_j([f], 0) = \lim_{j\to\infty} \frac{1}{\sqrt{2\pi}} \sum_{|k|\le j} \widehat{[f]}(k),$$

where $\widehat{[f]}(k)$, $k \in \mathbb{Z}$, are the Fourier coefficients of $[f]$; but

$$\begin{aligned}\widehat{[f]}(k) &= \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} [f](x)e^{-ikx}dx = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} \sum_{n\in\mathbb{Z}} f(x+2\pi n)e^{-ikx}dx \\ &= \frac{1}{\sqrt{2\pi}} \sum_{n\in\mathbb{Z}} \int_{-\pi}^{\pi} f(x+2\pi n)e^{-ikx}dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x)e^{-ikx}dx = \hat{f}(k),\end{aligned}$$

hence, $[f](0) = \sum_{n\in\mathbb{Z}} f(2\pi n) = \lim_{j\to\infty} \frac{1}{\sqrt{2\pi}} \sum_{|k|\le j} \hat{f}(k)$. ■

Theorem 7.1.5 (Poisson summation formula) *If $f \in W^{2,1}(\mathbb{R})$, then $\{\hat{f}(n)\}$ is summable and*

$$\sum_{n\in\mathbb{Z}} f(2\pi n) = \frac{1}{\sqrt{2\pi}} \sum_{n\in\mathbb{Z}} \hat{f}(n). \tag{7.3}$$

Proof In view of Lemma 7.1.3, it is sufficient to show that $\{\hat{f}(n)\}$ is summable.

Since $f \in W^{2,1}(\mathbb{R})$, $f' \in W^{1,1}(\mathbb{R})$. Then $[f'] = [f]'$ is AC and is therefore in $L^2[-\pi, \pi]$. Now,

$$\widehat{[f]}(k) = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} [f](x)e^{-ikx} = \frac{i}{k}\frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} [f]'(x)e^{-ikx}dx = \frac{i}{k}\widehat{[f]'}(k),$$

if $k \neq 0$ (note $[f](-\pi) = [f](\pi)$), hence,

$$\begin{aligned}\sum_{k\in\mathbb{Z}} \left|\widehat{[f]}(k)\right| &= \left|\widehat{[f]}(0)\right| + \sum_{k\neq 0} \frac{1}{|k|} \left|\widehat{[f]'}(k)\right| \\ &\le \left|\widehat{[f]}(0)\right| + \left(\sum_{k\neq 0} \frac{1}{k^2}\right)^{\frac{1}{2}} \left(\sum_{k\in\mathbb{Z}} \left|\widehat{[f]'}(k)\right|^2\right)^{\frac{1}{2}} < \infty,\end{aligned}$$

because $\sum_{k\in\mathbb{Z}} |\widehat{[f]'}(k)|^2 = \|[f]'\|_2^2$. Thus $\{\widehat{[f]}(n)\}$ is summable. We have shown in the proof of Lemma 7.1.3 that $\hat{f}(n) = \widehat{[f]}(n)$ for $n \in \mathbb{Z}$, hence $\{\hat{f}(n)\}$ is summable. ■

Example 7.1.3 Let $g(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, and for $t > 0$ let $g_t(x) = \frac{1}{\sqrt{2\pi}t} e^{-\frac{x^2}{2t^2}}$, $x \in \mathbb{R}$. The family $\{g_t\}$ is called the **Gauss kernel**. From Example 7.1.1 (ii) and Exercise 7.1.4 (ii), $\hat{g}_t(\xi) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2\xi^2}{2}}$. Using (7.3), we conclude that

$$\frac{1}{\sqrt{2\pi}\,t} \sum_{n\in\mathbb{Z}} e^{-\frac{(2\pi n)^2}{2t^2}} = \frac{1}{2\pi} \sum_{n\in\mathbb{Z}} e^{-\frac{n^2t^2}{2}},$$

from which on replacing t by $2\pi\sqrt{t}$, we have

$$\frac{1}{\sqrt{2\pi t}} \sum_{n\in\mathbb{Z}} e^{-\frac{n^2}{2t}} = \sum_{n\in\mathbb{Z}} e^{-2\pi^2 n^2 t}. \tag{7.4}$$

The relation (7.4) is Jacobi's identity for the theta function θ,

$$\theta(t) = t^{-\frac{1}{2}} \theta\left(\frac{1}{t}\right), \quad t > 0, \tag{7.5}$$

where $\theta(t) = \sum_{n\in\mathbb{Z}} e^{-\pi n^2 t}$.

Example 7.1.4 Consider the **Poisson kernel** $P_t(x) = \frac{1}{\pi} \frac{t}{t^2+x^2}$, $t > 0$, $x \in \mathbb{R}$. From the Cauchy integral formula, if $\xi > 0$ and $y < -t$,

$$\int_{\mathbb{R}} \frac{e^{-i\xi x} dx}{(x - it)(x + it)} - \int_{\mathbb{R}} \frac{e^{-i\xi(x+iy)} dx}{(x + iy - it)(x + iy + it)} = 2\pi i \left(\frac{e^{-\xi t}}{-2it}\right) = \frac{\pi}{t} e^{-\xi t},$$

where $\frac{e^{-\xi t}}{-2it}$ is the value of the function $\frac{e^{-i\xi z}}{z-it}$ at $z = -it$. But,

$$\int_{\mathbb{R}} \frac{e^{-i\xi(x+iy)} dx}{(x + iy)^2 + t^2} = e^{\xi y} \int_{\mathbb{R}} \frac{e^{-i\xi x}}{(x + iy)^2 + t^2} dx \to 0$$

as $y \to -\infty$. Hence $\int_{\mathbb{R}} \frac{e^{-i\xi x}}{x^2+t^2} dx = \frac{\pi}{t} e^{-\xi t}$ if $\xi > 0$.

If $\xi < 0$, take $y > t$ and then let $y \to \infty$; we obtain $\int_{\mathbb{R}} \frac{e^{-i\xi x}}{x^2+t^2} dx = \frac{\pi}{t} e^{\xi t}$ by the same argument. Thus $\widehat{P}_t(\xi) = \frac{1}{\sqrt{2\pi}} \frac{t}{\pi} \int_{\mathbb{R}} \frac{e^{-i\xi x}}{x^2+t^2} dx = \frac{1}{\sqrt{2\pi}} e^{-|\xi|t}$. Apply (7.3); we have

$$\sum_{n\in\mathbb{Z}} P_t(2n\pi) = \frac{t}{\pi} \sum_{n\in\mathbb{Z}} \frac{1}{t^2 + (2n\pi)^2} = \frac{1}{2\pi} \sum_{n\in\mathbb{Z}} e^{-|n|t} = \frac{1}{2\pi} \frac{1 + e^{-t}}{1 - e^{-t}}, \tag{7.6}$$

or

$$\sum_{n\in\mathbb{Z}} \frac{1}{t^2 + n^2} = \frac{\pi}{t} \frac{1 + e^{-2\pi t}}{1 - e^{-2\pi t}}$$

on replacing t by $2\pi t$. When $t \to 0+$, (7.6) becomes $\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$.

Exercise 7.1.6 Show that $\int_{\mathbb{R}} e^{-|\xi|t} e^{i\xi x} d\xi = 2\pi P_t(x)$ and verify that $\widehat{P}_t(\xi) = \frac{1}{\sqrt{2\pi}} e^{-|\xi|t}$.

7.2 Fourier integral on L^2

The Fourier integral for L^2 functions will be defined by using properties of the Fourier integral operator on the space of rapidly decreasing functions.

Denote by $\mathcal{S}$ the space of all complex-valued functions f in $C^\infty(\mathbb{R}^n)$ such that for all multi-indices α and β

$$P_{\alpha\beta}(f) := \sup_{x\in\mathbb{R}^n} |x^\alpha \partial^\beta f(x)| < \infty,$$

where $x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ and $\partial^\beta f(x) = \frac{\partial^{|\beta|} f}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}(x)$.

$\mathcal{S}$ is called the **Schwartz space** in $\mathbb{R}^n$, and functions in $\mathcal{S}$ are usually referred to as **rapidly decreasing** functions. For each pair α, β of multi-indices, $P_{\alpha\beta}(\cdot)$ is a semi-norm on $\mathcal{S}$. Note that (i) $\mathcal{D} := C_c^\infty(\mathbb{R}^n) \subset \mathcal{S}$; and (ii) the function $e^{-\frac{|x|^2}{2}}$ is in $\mathcal{S}$.

Define a metric ρ on $\mathcal{S}$ by

$$\rho(f, g) = \sum_{\alpha,\beta} \frac{1}{e^{|\alpha|} e^{|\beta|}} \cdot \frac{P_{\alpha\beta}(f - g)}{1 + P_{\alpha\beta}(f - g)} \cdot \frac{1}{n^{|\alpha|} n^{|\beta|}}. \tag{7.7}$$

Since $\{\frac{1}{e^{|\alpha|}e^{|\beta|}} \cdot \frac{P_{\alpha\beta}(f-g)}{1+P_{\alpha\beta}(f-g)} \cdot \frac{1}{n^{|\alpha|}n^{|\beta|}}\}_{\alpha,\beta}$ is summable with sum $\le \sum_{j,k\ge 0} \frac{1}{e^j e^k}$, $\rho(f, g)$ is a nonnegative finite number.

Exercise 7.2.1 Show that ρ is actually a metric on $\mathcal{S}$.

We observe first the following elementary inequalities:

$$\begin{aligned} (1 + |x|)^N &\le 2^N(1 + |x|^N), \quad N \ge 0,\ x \in \mathbb{R}^n; \\ |x|^N &\le \delta^{-1} \sum_{j=1}^n |x_j|^N, \quad N \ge 0,\ x \in \mathbb{R}^n, \end{aligned} \tag{7.8}$$

where $\delta = \min_{|x|=1} \sum_{j=1}^n |x_j|^N$. For the first one, we may assume that $|x| > 1$, then $(1 + |x|)^N \le (2|x|)^N < 2^N(1 + |x|^N)$; while the second inequality follows by first considering the case $|x| = 1$ and then reducing the general case to this particular case.

Proposition 7.2.1 *For $f \in \mathcal{S}$, $x^\alpha \partial^\beta f \in L^1$ for any multi-indices α and β.*

Proof

$$\begin{aligned} \int_{\mathbb{R}^n} |x^\alpha \partial^\beta f(x)| dx &= \int_{\mathbb{R}^n} |x^\alpha|(1 + |x|^{n+1})|\partial^\beta f(x)| \frac{1}{1 + |x|^{n+1}} dx \\ &\le \int_{\mathbb{R}^n} |x^\alpha| \left(1 + \delta^{-1} \sum_{j=1}^n |x_j|^{n+1}\right) |\partial^\beta f(x)| \frac{1}{1 + |x|^{n+1}} dx \\ &\le M \int_{\mathbb{R}^n} \frac{1}{1 + |x|^{n+1}} dx < \infty, \end{aligned}$$

for some $M > 0$, where $\delta = \min_{|x|=1} \sum_{j=1}^n |x_j|^{n+1}$ (cf. (7.8)). ■

Now let $f \in \mathcal{S}$, then $f \in L^1$ by Proposition 7.2.1, and $\hat{f}$ is defined. We show the existence of $\frac{\partial}{\partial \xi_j}\hat{f}(\xi)$ as follows. Consider for $h \neq 0$ the difference quotient

$$\frac{\hat{f}(\xi_1, \dots, \xi_j + h, \dots, \xi_n) - \hat{f}(\xi)}{h} = \frac{1}{(2\pi)^{n/2}} \int f(x) e^{-i\xi \cdot x} \frac{(e^{-ihx_j} - 1)}{h} dx.$$

Since $\left|\frac{e^{-ihx_j}-1}{h}\right| \leq |x_j|$ and $|x_j||f| \in L^1$, by Proposition 7.2.1, it follows from LDCT that $\frac{\partial}{\partial \xi_j}\hat{f}(\xi)$ exists and

$$\frac{\partial}{\partial \xi_j}\hat{f}(\xi) = (-i)\widehat{x_j f}(\xi).$$

By Proposition 7.2.1, we can repeat the above argument with f replaced by $x_j f$, and obtain for any multi-index α the following formula:

$$\partial_\xi^\alpha \hat{f}(\xi) = (-i)^{|\alpha|}\widehat{x^\alpha f}(\xi). \tag{7.9}$$

Since $x^\alpha f \in L^1$ and $\widehat{x^\alpha f}$ is uniformly continuous, Proposition 7.2.2 then follows from (7.9).

Proposition 7.2.2 *If $f \in \mathcal{S}$, then $\hat{f} \in C^\infty(\mathbb{R}^n)$.*

Using the Fubini theorem and integration by parts, one asserts

$$\widehat{\partial^\beta f}(\xi) = (i)^{|\beta|}\xi^\beta \hat{f}(\xi) \tag{7.10}$$

for any multi-index β. Combining (7.9) and (7.10), one obtains

$$(i)^{|\alpha+\beta|}\xi^\beta \partial_\xi^\alpha \hat{f}(\xi) = \widehat{\partial_x^\beta (x^\alpha f)}(\xi) \tag{7.11}$$

for any multi-indices α and β.

Theorem 7.2.1 *$F\mathcal{S} \subset \mathcal{S}$, and F is a continuous map with respect to the metric ρ on $\mathcal{S}$ defined by* (7.7).

Proof That $f \in \mathcal{S}$ implies that $\hat{f} \in \mathcal{S}$ follows directly from (7.11):

$$\sup_{\xi \in \mathbb{R}^n} |\xi^\beta \partial_\xi^\alpha \hat{f}(\xi)| \leq \|\widehat{\partial_x^\beta (x^\alpha f)}\|_\infty \leq \|\partial_x^\beta (x^\alpha f)\|_1 < \infty.$$

To see that F is continuous, first observe that a sequence $\{f_k\} \subset \mathcal{S}$ converges to $f \in \mathcal{S}$ in the metric ρ defined by (7.7) if and only if $\lim_{k\to\infty} P_{\alpha\beta}(f_k - f) = 0$ for each pair α, β of multi-indices. Now from (7.11),

$$P_{\beta\alpha}(\hat{f}_k - \hat{f}) \leq \|\widehat{\partial_x^\beta [x^\alpha (f_k - f)]}\|_\infty \leq \|\partial_x^\beta [x^\alpha (f_k - f)]\|_1;$$

observe that if $\rho(f_k, f) \to 0$, then $\partial_x^\beta[x^\alpha(f_k(x) - f(x))] \to 0$ uniformly on $\mathbb{R}^n$ and

$$\begin{aligned}|\partial_x^\beta[x^\alpha(f_k(x) - f(x))]| &\leq |(1 + |x|^{n+1})\partial_x^\beta[x^\alpha(f_k(x) - f(x))]|\frac{1}{1 + |x|^{n+1}} \\ &\leq M\frac{1}{1 + |x|^{n+1}}.\end{aligned}$$

LDCT can be applied to obtain $\lim_{k\to\infty} \|\partial_x^\beta[x^\alpha(f_k - f)]\|_1 = 0$, implying that $\lim_{k\to\infty} P_{\beta\alpha}(\hat{f}_k - \hat{f}) = 0$ and consequently $\rho(\hat{f}_k, \hat{f}) \to 0$. ■

Since $\check{F}f = F\tilde{f}$, where $\tilde{f}(x) = f(-x)$, $\check{F}$ is also a continuous map from $\mathcal{S}$ to $\mathcal{S}$ w.r.t. the metric defined by (7.7).

Taking into account Theorem 7.1.4 and the fact that $\mathcal{S} \subset L^1$, we conclude that Theorem 7.2.2 holds.

Theorem 7.2.2 (Fourier inversion theorem) *Both F and $\check{F}$ are continuous and bijective from $\mathcal{S}$ to $\mathcal{S}$ and $\check{F}(Ff) = f = F(\check{F}f)$ for $f \in \mathcal{S}$.*

Theorem 7.2.3 (Parseval relations) *For f, g in $\mathcal{S}$ the following relations hold:*

(i) $\int \hat{f} g d\lambda^n = \int f\hat{g} d\lambda^n$;

(ii) $\int f\bar{g} d\lambda^n = \int \hat{f}\bar{\hat{g}} d\lambda^n$.

Proof (i) is the conclusion of Exercise 7.1.5; (ii) follows from (i) by replacing f and g by $\check{f}$ and $\bar{g}$ respectively. ■

Exercise 7.2.2 Let $(f, g) = \int f\bar{g} d\lambda^n$ be the L^2 inner product of f and g in $\mathcal{S}$. Show that (i) and (ii) in Theorem 7.2.3 are equivalent and are equivalent to any of the following relations:

(a) $(\hat{f}, g) = (f, \check{g})$;

(b) $(f, g) = (\hat{f}, \hat{g})$.

We are ready to define the Fourier integral for functions in L^2. Since $C_c^\infty(\mathbb{R}^n)$ is dense in L^p, $1 \leq p < \infty$, and $C_c^\infty(\mathbb{R}^n) \subset \mathcal{S}$, $\mathcal{S}$ is dense in L^2. For $f \in L^2$, there is a sequence $\{f_k\}$ in $\mathcal{S}$ such that $\lim_{k\to\infty} \|f_k - f\|_2 = 0$; a fortiori, $\{f_k\}$ is a Cauchy sequence in L^2. By relation (b) in Exercise 7.2.2, $\|f_k - f_l\|_2^2 = \|\hat{f}_k - \hat{f}_l\|_2^2$ for all k, l in $\mathbb{N}$, therefore $\{\hat{f}_k\}$ is a Cauchy sequence in L^2 and converges in L^2 to $g \in L^2$. We claim that g is independent of the sequence $\{f_k\}$ in $\mathcal{S}$, which converges to f in L^2. Suppose that $\{g_k\}$ is another sequence in $\mathcal{S}$ that converges to f in L^2; then $\lim_{k\to\infty} \|f_k - g_k\|_2 = 0$, but $\|\hat{f}_k - \hat{g}_k\|_2 = \|f_k - g_k\|_2$ implies that $\lim_{k\to\infty} \hat{g}_k = \lim_{k\to\infty} \hat{f}_k = g$ in L^2. Thus g is uniquely determined by f in the way we specify; we then denote g by $\hat{f}'$ for the moment. From the definition, one verifies readily that $(f, g) = (\hat{f}', \hat{g}')$ for f, g in L^2.

Lemma 7.2.1 *If $f \in L^1 \cap L^2$, then $\hat{f} = \hat{f}'$.*

Proof Fix a Friederich mollifier $\{J_\varepsilon\}_{\varepsilon>0}$ constructed from a mollifying function $\varphi \geq 0$. For $\varepsilon > 0$, let $f_\varepsilon = f I_{B_{1/\varepsilon}(0)}$. Then $J_\varepsilon f_\varepsilon \in C_c^\infty(\mathbb{R}^n) \subset \mathcal{S}$. We claim that $J_\varepsilon f_\varepsilon \to f$ in both L^1 and L^2. Actually for $p = 1$ or 2, we have

$$\begin{aligned}\|J_\varepsilon f_\varepsilon - f\|_p &\leq \|J_\varepsilon(f_\varepsilon - f)\|_p + \|J_\varepsilon f - f\|_p \\ &\leq \|f_\varepsilon - f\|_p + \|J_\varepsilon f - f\|_p \to 0\end{aligned}$$

as $\varepsilon \to 0$. From $\|J_\varepsilon f_\varepsilon - f\|_1 \to 0$, as $\varepsilon \to 0$, we infer that $\widehat{J_\varepsilon f_\varepsilon} \to \hat{f}$ uniformly on $\mathbb{R}^n$; while from $\|J_\varepsilon f_\varepsilon - f\|_2 \to 0$, as $\varepsilon \to 0$, we conclude that $\widehat{J_\varepsilon f_\varepsilon} \to \hat{f}'$ in L^2 and, consequently, there is a sequence of ε tending to zero such that $\widehat{J_\varepsilon f_\varepsilon} \to \hat{f}'$ a.e. on $\mathbb{R}^n$. Hence $\hat{f} = \hat{f}'$ a.e. ■

Because of Lemma 7.2.1, it is natural to call $\hat{f}'$ the Fourier integral of f in L^2 and also denote $\hat{f}'$ by $\hat{f}$. Similarly $\check{f}$ is also defined for $f \in L^2$. We shall also use F and $\check{F}$ to denote the maps $f \mapsto \hat{f}$ and $f \mapsto \check{f}$ respectively from L^2 onto L^2. Note that $(f,g) = (\hat{f},\hat{g}) = (\check{f},\check{g})$ for f,g in L^2.

Exercise 7.2.3 Show that both F and $\check{F}$ are linear bijective isometries from L^2 onto itself and $\check{F} = F^{-1}$.

Exercise 7.2.4 Suppose that $f \in W^{k,2}(\mathbb{R}^n)$, $k \in \mathbb{N}$. Show that $\widehat{\partial^\alpha f}(\xi) = (i)^{|\alpha|}\xi^\alpha \hat{f}(\xi)$ for a.e. $\xi \in \mathbb{R}^n$ if $|\alpha| \leq k$. (Hint: $W^{k,2}(\mathbb{R}^n) = \overset{\circ}{W}{}^{k,2}(\mathbb{R}^n)$.)

7.3 The Sobolev space H^s

For each $s \in \mathbb{R}$, an inner product $(\cdot,\cdot)_s$ on $\mathcal{S}$ is defined by

$$(f,g)_s = \int (1+|\xi|^2)^s \hat{f}(\xi)\overline{\hat{g}(\xi)}d\xi;$$

and the associated norm on $\mathcal{S}$ is denoted by $|\cdot|_s$. Thus,

$$|f|_s = \left(\int (1+|\xi|^2)^s |\hat{f}(\xi)|^2 d\xi\right)^{\frac{1}{2}}.$$

As usual, $(f,g) = \int f\bar{g}d\lambda^n$ is the inner product of f and g in L^2.

A few basic properties of inner products $(\cdot,\cdot)_s$ are now listed.

(1) $(f,g) = (f,g)_0$.

(2) $|(f,g)_0| \leq |f|_s|g|_{-s}$. This follows directly from

$$(f,g)_0 = \int (1+|\xi|^2)^{s/2}\hat{f}(\xi)(1+|\xi|^2)^{-s/2}\hat{g}(\xi)d\xi,$$

by Schwarz's inequality.

(3) $|f|_s = \max_{\substack{g\in\mathcal{S}\\ g\neq 0}} \frac{|(f,g)_0|}{|g|_{-s}}$. To see this, one observes first from (2) that

$$|f|_s \geq \sup_{\substack{g\in\mathcal{S}\\ g\neq 0}} \frac{|(f,g)_0|}{|g|_{-s}};$$

now, since $(1+|\xi|^2)^s\hat{f}(\xi) \in \mathcal{S}$, there is $h \in \mathcal{S}$ such that $\hat{h}(\xi) = (1+|\xi|^2)^s\hat{f}(\xi)$, and hence,

$$\begin{aligned}
|h|_{-s}^2 &= \int (1+|\xi|^2)^{-s}(1+|\xi|^2)^{2s}|\hat{f}(\xi)|^2 d\xi \\
&= \int (1+|\xi|^2)^{s}|\hat{f}(\xi)|^2 d\xi = |f|_s^2; \\
(f,h)_0 &= \int (1+|\xi|^2)^{s}|\hat{f}(\xi)|^2 d\xi = |f|_s^2,
\end{aligned}$$

resulting in $\frac{|(f,h)_0|}{|h|_{-s}} = |f|_s$.

(4) $|\partial^\alpha f|_s \leq |f|_{s+|\alpha|}$. This is obvious by (7.10).

The Sobolev space H^s is the completion of $\mathcal{S}$ under the norm $|\cdot|_s$. The Sobolev space H^s is a Hilbert space for each $s \in \mathbb{R}$. Observe that in the case $s \geq 0$, if $\{f_k\}$ is a Cauchy sequence in $\mathcal{S}$ in the norm $|\cdot|_s$, then it is a Cauchy sequence in L^2, hence it is legitimate to identify each element of H^s, $s \geq 0$, with an element of L^2. Those elements of L^2 which belong to H^s can be characterized as follows.

Theorem 7.3.1 *An element f of L^2 is in H^s, $s \geq 0$, if and only if there is a sequence $\{f_k\} \subset \mathcal{S}$ such that $\|f_k - f\|_2 \to 0$ as $k \to \infty$ and $\sup_k |f_k|_s < \infty$.*

Proof If $f \in H^s$, there is $\{f_k\} \subset \mathcal{S}$ such that $|f_k - f|_s \to 0$ as $k \to \infty$, a fortiori, $\|f_k - f\|_2 \to 0$ as $k \to \infty$ and $\sup_k |f_k|_s < \infty$.

Conversely, suppose that there is a sequence $\{f_k\} \subset \mathcal{S}$ such that $\|f_k - f\|_2 \to 0$ and $\sup_k |f_k|_s < \infty$. By the Banach–Saks theorem (Theorem 5.10.2), there is a subsequence $\{g_k\}$ of $\{f_k\}$ and g in H^s such that $|\frac{1}{N}\sum_{k=1}^N g_k - g|_s \to 0$ as $N \to \infty$, a fortiori, $\|\frac{1}{N}\sum_{k=1}^N g_k - g\|_2 \to 0$. But $\|g_k - f\|_2 \to 0$ implies that $\|\frac{1}{N}\sum_{k=1}^N g_k - f\|_2 \to 0$, and consequently $f = g$. Thus $f \in H^s$. ∎

Exercise 7.3.1 Show that if k is a nonnegative integer, then $W^{k,2}(\mathbb{R}^n) = H^k$, in the sense that $W^{k,2}(\mathbb{R}^n) = H^k$ as set and the norms $\|\cdot\|_{k,2}$ and $|\cdot|_k$ are equivalent.

We will now show that in tempo with s becoming larger, elements of H^s become smoother. This is the content of the Sobolev lemma.

A preliminary lemma is shown first.

Lemma 7.3.1 *Suppose that $s \in \mathbb{R}$ and k is a nonnegative integer such that $s - k > \frac{n}{2}$; then there is $C > 0$ such that*

$$\max_{x\in\mathbb{R}^n} \sum_{|\alpha|\le k} |\partial^\alpha f(x)| \le C|f|_s$$

for $f \in \mathcal{S}$.

Proof Since $\widehat{\partial^\alpha f}(\xi) = (i)^{|\alpha|}\xi^\alpha \hat{f}(\xi)$, $\widehat{\partial^\alpha f}$ is in $\mathcal{S}$; it follows from Fourier's Inversion theorem (Theorem 7.2.2) that

$$\begin{aligned}\partial^\alpha f(x) &= (2\pi)^{-\frac{n}{2}} \int e^{ix\cdot\xi}(i)^{|\alpha|}\xi^\alpha \hat{f}(\xi)d\xi \\ &= (2\pi)^{-\frac{n}{2}}(i)^{|\alpha|} \int e^{ix\cdot\xi}\xi^\alpha(1+|\xi|^2)^{\frac{(s-k)}{2}}\hat{f}(\xi)(1+|\xi|^2)^{\frac{(k-s)}{2}}d\xi,\end{aligned}$$

and hence, when $|\alpha| \le k$,

$$\begin{aligned}|\partial^\alpha f(x)|^2 &\le (2\pi)^{-n}\int |\xi^\alpha|^2(1+|\xi|^2)^{s-k}|\hat{f}(\xi)|^2 d\xi \cdot \int (1+|\xi|^2)^{k-s}d\xi \\ &\le C'\int \frac{|\xi|^{2|\alpha|}}{(1+|\xi|^2)^k}(1+|\xi|^2)^s|\hat{f}(\xi)|^2 d\xi \\ &\le C|f|_s^2,\end{aligned}$$

where we have used the obvious fact that $\int (1+|\xi|^2)^{k-s}d\xi < \infty$. Thus,

$$\max_{x\in\mathbb{R}^n} \sum_{|\alpha|\le k} |\partial^\alpha f(x)| \le C|f|_s,$$

with $C > 0$ depending only on s, k, and n. ■

Theorem 7.3.2 (Sobolev lemma) *Suppose that $s \in \mathbb{R}$ and k is a nonnegative integer such that $s - k > \frac{n}{2}$; then $H^s \subset C^k(\mathbb{R}^n)$.*

Proof Consider f in H^s. There is a sequence $\{f_k\} \subset \mathcal{S}$ such that $|f_k - f|_s \to 0$ as $k \to \infty$; $\{f_k\}$ is therefore a Cauchy sequence in H^s. From Lemma 7.3.1, there is $C > 0$ such that

$$\max_{x\in\mathbb{R}^n} \sum_{|\alpha|\le k} |\partial^\alpha(f_m(x) - f_l(x))| \le C|f_m - f_l|_s \to 0$$

as $m, l \to \infty$, which means that $\{f_k\}$ converges uniformly on $\mathbb{R}^n$ to a function g in $C^k(\mathbb{R}^n)$. Then, $\int_{|x|\le R} |f_k(x) - g(x)|^2 dx \to 0$ as $k \to \infty$ for any $R > 0$;

but $\lim_{k\to\infty} |f_k - f|_s = 0$ implies that $\lim_{k\to\infty} \|f_k - f\|_2 = 0$, and therefore that $\int_{|x|\le R} |f_k(x) - f(x)|^2 dx \to 0$ as $k \to \infty$. Now,

$$\left\{\int_{|x|\le R} |f(x) - g(x)|^2 dx\right\}^{\frac{1}{2}} \le \left\{\int_{|x|\le R} |f(x) - f_k(x)|^2 dx\right\}^{\frac{1}{2}} + \left\{\int_{|x|\le R} |f_k(x) - g(x)|^2 dx\right\}^{\frac{1}{2}} \to 0$$

as $k \to \infty$, hence, $\int_{|x|\le R} |f(x) - g(x)|^2 dx = 0$ and consequently $f = g$ a.e. on $B_R(0)$. Since $R > 0$ is arbitrary, $f = g$ a.e. on $\mathbb{R}^n$. ■

7.4 Weak solutions of the Poisson equation

We illustrate the use of Sobolev space in this section by considering the existence and regularity of weak solutions of the Poisson equation,

$$\Delta u = f. \tag{7.12}$$

A **classical solution** u of (7.12) on an open domain Ω of $\mathbb{R}^n$ is a function u, defined on Ω such that $\Delta u(x) = f(x)$ for a.e. x of Ω. If f is continuous and u is a C^2 classical solution of (7.12) on Ω, then for any $v \in C_c^\infty(\Omega)$, we have

$$\int_\Omega fv d\lambda^n = \int v\Delta u d\lambda^n = \int u\Delta v d\lambda^n.$$

Therefore, when f is locally integrable on Ω, a locally integrable function u on Ω is called a **weak solution** of (7.12) if

$$\int_\Omega fv d\lambda^n = \int u\Delta v d\lambda^n$$

for all $v \in C_c^\infty(\Omega)$.

Exercise 7.4.1 Show that a C^2 function u on Ω is a classical solution of (7.12) if and only if it is a weak solution of (7.12).

We shall first prove the following regularity result for weak solutions of (7.12).

Theorem 7.4.1 *Suppose that $f \in C^\infty(\Omega)$. Then any locally L^2 weak solution of* (7.12) *is in $C^\infty(\Omega)$.*

The proof of Theorem 7.4.1 is preceded by some preliminaries relating to Friederich mollifiers. We fix a Friederich mollifier $\{J_\varepsilon\}_{\varepsilon>0}$ with a mollifying function φ which is nonnegative and satisfies the symmetry property: $\varphi(-x) = \varphi(x)$ for all x in $\mathbb{R}^n$. For example, we may take φ to be the function defined by $\varphi(x) = c\exp\{-\frac{1}{1-|x|^2}\}$ if $|x| < 1$ and $\varphi(x) = 0$ if $|x| \geq 1$, where c is a positive constant chosen so that $\int \varphi d\lambda^n = 1$.

Lemma 7.4.1 *Let $\{J_\varepsilon\}$ be a Friederich mollifier as previously specified.*

(i) $\|J_\varepsilon f\|_p \leq \|f\|_p$ *for* $f \in L^p$, $1 \leq p < \infty$.
(ii) $(J_\varepsilon f, g) = (f, J_\varepsilon g)$ *for* $f, g \in L^2$.
(iii) *If* $f \in C^1(\mathbb{R}^n)$, *then* $\frac{\partial}{\partial x_j} J_\varepsilon f(x) = J_\varepsilon \frac{\partial f}{\partial x_j}(x)$ *for all* $x \in \mathbb{R}^n$ *and* $j = 1, \ldots, n$.
(iv) *If* $f \in \mathcal{S}$, *then* $J_\varepsilon f \in \mathcal{S}$ *and* $|J_\varepsilon f|_s \leq |f|_s$.

Proof (i) is known in Section 4.10; (ii) follows directly from the definition of J_ε and the assumption that $\varphi(-x) = \varphi(x)$; while (iii) is a consequence of applying LDCT to the difference quotient involved in the definition of partial derivatives; it remains to show (iv). Since $J_\varepsilon f = f * \varphi_\varepsilon$, $\widehat{J_\varepsilon f} = (2\pi)^{\frac{n}{2}} \hat{f} \cdot \hat{\varphi}_\varepsilon$, which implies immediately that $\widehat{J_\varepsilon f} \in \mathcal{S}$, but by the Fourier inversion theorem, $J_\varepsilon f = (\widehat{J_\varepsilon f})^{\vee}$ and hence $J_\varepsilon f \in \mathcal{S}$. Now,

$$\begin{aligned}|J_\varepsilon f|_s^2 &= \int (1 + |\xi|^2)^s |\widehat{J_\varepsilon f}(\xi)|^2 d\xi = (2\pi)^n \int (1 + |\xi|^2)^s |\hat{f}(\xi)|^2 |\hat{\varphi}_\varepsilon(\xi)|^2 d\xi \\ &\leq \|\varphi_\varepsilon\|_1 \int (1 + |\xi|^2)^s |\hat{f}(\xi)|^2 d\xi = |f|_s^2.\end{aligned}$$

Hence $|J_\varepsilon f|_s \leq |f|_s$. ■

Lemma 7.4.2 *There is a constant $C > 0$ such that*

$$|v|_s \leq C(|\Delta v|_{s-2} + |v|_{s-1})$$

for all $v \in \mathcal{S}$.

Proof For $\xi \in \mathbb{R}^n$, we have

$$(1 + |\xi|^2)^2 = 1 + 2|\xi|^2 + |\xi|^4 < |\xi|^4 + 2(1 + |\xi|^2) < 2\{|\xi|^4 + (1 + |\xi|^2)\},$$

hence,

$$\begin{aligned}|v|_s^2 &= \int (1 + |\xi|^2)^s |\hat{v}(\xi)|^2 d\xi \\ &< 2\int (1 + |\xi|^2)^{s-2} \{|\xi|^4 + (1 + |\xi|^2)\} |\hat{v}(\xi)|^2 d\xi \\ &= 2\left\{\int (1 + |\xi|^2)^{s-2} |\widehat{\Delta v}(\xi)|^2 d\xi + \int (1 + |\xi|^2)^{s-1} |\hat{v}(\xi)|^2 d\xi\right\} \\ &= 2(|\Delta v|_{s-2}^2 + |v|_{s-1}^2) \\ &\leq 2(|\Delta v|_{s-2} + |v|_{s-1})^2,\end{aligned}$$

and consequently,

$$|v|_s \leq \sqrt{2}(|\Delta v|_{s-2} + |v|_{s-1}).$$

■

Proof of Theorem 7.4.1 For $x \in \Omega$, there is $g \in C_c^\infty(\Omega)$, which takes a constant value in the neighborhood of x; it is therefore sufficient to prove that $gu \in C^\infty(\mathbb{R}^n)$ for each $g \in C_c^\infty(\Omega)$.

Consider now any $g \in C_c^\infty(\Omega)$. In order to show that $gu \in C^\infty(\mathbb{R}^n)$, it is sufficient to show that $gu \in H^s$ for all $s \in \mathbb{N}$, by the Sobolev lemma (Theorem 7.3.2); but since $\|J_\varepsilon(gu) - gu\|_2 \to 0$ as $\varepsilon \to 0$, from Theorem 7.3.1, it is sufficient to show that given $g \in C_c^\infty(\Omega)$, for each $s \in \mathbb{N}$, there is a constant $C_s > 0$ such that

$$|J_\varepsilon(gu)|_s \leq C_s, \quad \varepsilon > 0. \tag{7.13}$$

When $s = 0$, (7.13) is a consequence of $\|J_\varepsilon(gu)\|_2 \leq \|gu\|_2$ (cf. Lemma 7.4.1 (i)). Suppose that (7.13) holds for $s - 1$, we are going to show that (7.13) holds for s. Using the Fubini theorem and integration by parts, we have for $v \in \mathcal{S}$,

$$\begin{aligned}(\Delta(J_\varepsilon(gu)), v) &= (J_\varepsilon(gu), \Delta v) = (gu, \Delta J_\varepsilon v) = (u, g(\Delta J_\varepsilon v))\\ &= (u, \Delta(gJ_\varepsilon v)) - \left(u, 2\sum_{j=1}^n \frac{\partial g}{\partial x_j}\frac{\partial J_\varepsilon v}{\partial x_j} + J_\varepsilon v \cdot \Delta g\right)\\ &= (f, gJ_\varepsilon v) - 2\sum_{j=1}^n \left(J_\varepsilon\left(u\frac{\partial g}{\partial x_j}\right), \frac{\partial v}{\partial x_j}\right) + (J_\varepsilon(u\Delta g), v),\end{aligned}$$

where Lemma 7.4.1 has been applied. Hence,

$$\begin{aligned}&|(\Delta J_\varepsilon(gu), v)|\\ &\leq \left\{|J_\varepsilon(gf)|_{s-2}|v|_{2-s} + 2\sum_{j=1}^n \left|J_\varepsilon\left(u\frac{\partial g}{\partial x_j}\right)\right|_{s-1} \cdot \left|\frac{\partial v}{\partial x_j}\right|_{1-s} + |J_\varepsilon(u\Delta g)|_{s-1}|v|_{1-s}\right\}\\ &\leq |v|_{2-s}\left\{|J_\varepsilon(gf)|_{s-2} + 2\sum_{j=1}^n \left|J_\varepsilon\left(u\frac{\partial g}{\partial x_j}\right)\right|_{s-1} + |J_\varepsilon(u\Delta g)|_{s-1}\right\},\end{aligned}$$

where (2) and (4) in Section 7.3 are used. Thus, by (3) in Section 7.3, we conclude that

$$|\Delta(J_\varepsilon(gu))|_{s-2} \leq |J_\varepsilon(gf)|_{s-2} + 2\sum_{j=1}^n \left|J_\varepsilon\left(u\frac{\partial g}{\partial x_j}\right)\right|_{s-1} + |J_\varepsilon(u\Delta g)|_{s-1}. \tag{7.14}$$

Now from Lemma 7.4.2,

$$|J_\varepsilon(gu)|_s \leq C(|\Delta J_\varepsilon(gu)|_{s-2} + |J_\varepsilon(gu)|_{s-1}). \tag{7.15}$$

Substitute (7.14) into (7.15); we have

$$|J_\varepsilon(gu)|_s \le C'\left(|J_\varepsilon(gf)|_{s-2} + 2\sum_{j=1}^{n}\left|J_\varepsilon\left(u\frac{\partial g}{\partial x_j}\right)\right|_{s-1} + |J_\varepsilon(u\Delta g)|_{s-1} + |J_\varepsilon(gu)|_{s-1}\right).$$

But $|J_\varepsilon(gf)|_{s-2} \le |gf|_{s-2}$, by Lemma 7.4.1 (iv), and

$$2\sum_{j=1}^{n}\left|J_\varepsilon\left(u\frac{\partial g}{\partial x_j}\right)\right|_{s-1} + |J_\varepsilon(u\Delta g)|_{s-1} + |J_\varepsilon(gu)|_{s-1} \le C'_s,$$

by the assumption that (7.13) holds for $(s-1)$. Therefore,

$$|J_\varepsilon(gu)|_s \le C_s = |gf|_{s-2} + C'_s. \qquad \blacksquare$$

Regarding the existence of weak solutions of the Poisson equation (7.12), we now establish the existence and uniqueness of a weak solution of (7.12) in $\overset{\circ}{W}{}^{1,2}(\Omega)$ when Ω is bounded and $f \in L^2(\Omega)$.

Theorem 7.4.2 *Suppose that Ω is bounded and $f \in L^2(\Omega)$, then there is a unique weak solution of (7.12) in $\overset{\circ}{W}{}^{1,2}(\Omega)$.*

Proof It is only necessary to consider the case that f and solutions to be sought are real-valued; therefore $\overset{\circ}{W}{}^{1,2}(\Omega)$ is assumed to consist of real-valued functions. By the Poincaré inequality (Theorem 6.6.3), $\overset{\circ}{W}{}^{1,2}(\Omega)$ can be considered as a Hilbert space with the inner product

$$(u,v)'_1 = \sum_{j=1}^{n}\int_\Omega \frac{\partial u}{\partial x_j}\frac{\partial v}{\partial x_j}d\lambda^n$$

for u, v in $\overset{\circ}{W}{}^{1,2}(\Omega)$. Since $|(f,v)| \le \|f\|_2\|v\|_2 \le \|f\|_2\|v\|_{1,2} \le C\|f\|_2|v|_{1,2}$ for all $v \in \overset{\circ}{W}{}^{1,2}(\Omega)$, by (6.31), the linear functional $v \mapsto -\int_\Omega fv d\lambda^n$ is a bounded linear functional on $\overset{\circ}{W}{}^{1,2}(\Omega)$; it then follows from the Riesz representation theorem that there is $u \in \overset{\circ}{W}{}^{1,2}(\Omega)$, such that

$$-\int_\Omega fv d\lambda^n = (v,u)'_1 = \sum_{j=1}^{n}\int_\Omega \frac{\partial v}{\partial x_j}\frac{\partial u}{\partial x_j}d\lambda^n$$

for $v \in \overset{\circ}{W}{}^{1,2}(\Omega)$ and therefore for $v \in C_c^\infty(\Omega)$ in particular. But if $v \in C_c^\infty(\Omega)$,

$$\int_\Omega \frac{\partial v}{\partial x_j}\frac{\partial u}{\partial x_j}d\lambda^n = -\int_\Omega \frac{\partial^2 v}{\partial x_j^2}u d\lambda^n$$

for each $j = 1, \ldots, n$; thus we have

$$\int_\Omega fv d\lambda^n = \int_\Omega u \Delta v d\lambda^n$$

for $v \in C_c^\infty(\Omega)$. Hence u is a weak solution of (7.12).

Suppose now that $w \in \overset{\circ}{W}{}^{1,2}(\Omega)$ is also a weak solution of (7.12). Then,

$$\int_\Omega (u-w)\Delta v d\lambda^n = -\sum_{j=1}^n \int_\Omega \frac{\partial(u-w)}{\partial x_j}\frac{\partial v}{\partial x_j} d\lambda^n = 0$$

for all $v \in C_c^\infty(\Omega)$. We claim now that

$$\sum_{j=1}^n \int_\Omega \frac{\partial(u-w)}{\partial x_j}\frac{\partial v}{\partial x_j} d\lambda^n = 0$$

for all $v \in \overset{\circ}{W}{}^{1,2}(\Omega)$. Let $v \in \overset{\circ}{W}{}^{1,2}(\Omega)$; choose a sequence $\{v_k\}$ in $C_c^\infty(\Omega)$ such that $\lim_{k\to\infty} |v - v_k|_{1,2} = 0$; then,

$$\begin{aligned}
&\sum_{j=1}^n \int_\Omega \frac{\partial(u-w)}{\partial x_j}\frac{\partial v}{\partial x_j} d\lambda^n \\
&= \sum_{j=1}^n \int_\Omega \frac{\partial(u-w)}{\partial x_j}\frac{\partial v_k}{\partial x_j} d\lambda^n + \sum_{j=1}^n \int_\Omega \frac{\partial(u-w)}{\partial x_j}\frac{\partial(v-v_k)}{\partial x_j} d\lambda^n \\
&= \sum_{j=1}^n \int_\Omega \frac{\partial(u-w)}{\partial x_j}\frac{\partial(v-v_k)}{\partial x_j} d\lambda^n,
\end{aligned}$$

and consequently from Schwarz inequality,

$$\left|\sum_{j=1}^n \int_\Omega \frac{\partial(u-w)}{\partial x_j}\frac{\partial v}{\partial x_j} d\lambda^n\right| = \left|(u-w, v-v_k)_1'\right| \le |u-w|_{1,2} \cdot |v-v_k|_{1,2} \to 0$$

as $k \to \infty$. Hence,

$$\sum_{j=1}^n \int_\Omega \frac{\partial(u-w)}{\partial x_j}\frac{\partial v}{\partial x_j} d\lambda^n = 0$$

for $v \in \overset{\circ}{W}{}^{1,2}(\Omega)$. Since $u - w \in \overset{\circ}{W}{}^{1,2}(\Omega)$, we have

$$0 = \sum_{j=1}^n \int_\Omega \left[\frac{\partial(u-w)}{\partial x_j}\right]^2 d\lambda^n = |u-w|_{1,2}^2,$$

implying that $u = w$. Therefore, (7.12) has a unique weak solution in $\overset{\circ}{W}{}^{1,2}(\Omega)$. ■

7.5 Fourier integral of probability distributions

The Fourier integral of probability distributions will be discussed in this section, with an application to the central limit theorem in probability theory. This is preceded by a very brief introduction of the necessary basic notions, terminology, and notations in the probability theory, as formulated by **A.N. Kolmogoroff**.

Kolmogoroff's formulation of probability theory is based on measure theory. A measure space (Ω, Σ, P) with $P(\Omega) = 1$ is called a **probability space**, of which Ω is called the **sample space**; and sets in the σ-algebra Σ are called **events** (more precisely, measurable events); and for $A \in \Sigma$, $P(A)$ is referred to as the probability of event A. A measurable function on the sample space Ω is called a **random variable** (often abbreviated as r.v.). Random variables are usually denoted by capital Roman letters, such as $X, Y, Z, \ldots$ etc. It should be noted that a probability space is usually a construct suggested by first observations of outcomes of experiments on a random phenomenon; these outcomes are referred to as sample points and form the sample space Ω. Such a construct provides a solid mathematical framework to discuss questions related to the random phenomenon; such questions are usually addressed in terms of random variables. Our construction of the Bernoulli sequence space, starting with Section 1.3 and through Examples 1.7.1, 2.1.1, and 3.4.6, illustrates revealingly the point we just made. Henceforth, random variables are assumed to take finite real value P-almost everywhere and hence for a random variable, we always consider a finite real-valued version (for a probability space, P-almost everywhere is expressed as P-**almost surely** and is abbreviated as a.s.). Suppose that X is a random variable; if $\int_\Omega XdP$ exists, it is called the **expectation** of X and is denoted by $E(X)$; if $E(X)$ is finite, $\int_\Omega |X - E(X)|^2 dP$ is called the **variance** of X and is denoted by $\mathrm{Var}(X)$. The σ-algebra $\sigma(X) := \{X^{-1}(B) : B \in \mathcal{B}\}$ is the smallest subσ-algebra of Σ relative to which X is measurable; as implied by Exercise 2.5.10 (ii), if $E(X)$ is finite, the family $\{\int_A XdP : A \in \sigma X\}$ characterizes the r.v. X, or intuitively, $\{\int_A XdP : A \in \sigma(X)\}$ is the information one obtains by observing the r.v. X. This suggests considering $\sigma(X)$ as where the information regarding X resides. Accordingly, the σ-algebra Σ is where information on all random variables resides. As we know, in Example 4.3.2, the Bernoulli sequence space and $([0, 1], \mathcal{B}|[0, 1], \lambda)$ are measure-theoretically the same space, hence the choice of probability space is for convenience, and not of primary importance.

The most simple but fundamental notion in probability theory is that of independence. We shall discuss independence at some length to give a touch of the flavor of a basic aspect of probabilistic argument; however the notion of conditioning, basic and fundamental as it is, will not be touched upon here.

In the following, random variables are in reference to a fixed probability space (Ω, Σ, P) and σ-algebras on Ω are always subσ-algebras of Σ. A finite family $\{\Sigma_1, \ldots, \Sigma_k\}$ of σ-algebras on Ω is said to be **independent** if for any choice of $A_j \in \Sigma_j$, $j = 1, \ldots, k$, $P(\bigcap_{j=1}^k A_j) = \prod_{j=1}^k P(A_j)$ holds. A family $\{\Sigma_\alpha\}$ of σ-algebras on Ω is said to be **independent** if all of its finite subfamilies are independent. If $\{\Sigma_\alpha\}$ is independent, then $\Sigma_\alpha' s$ are said to be independent. For a family $\{A_\alpha\}$ of events, the σ-algebra $\sigma(\{A_\alpha\})$ is abbreviated to $\sigma(A_\alpha' s)$; in particular, if $A \in \Sigma$, $\sigma(A) = \{\emptyset, A, A^c, \Omega\}$. Events

A_α, $\alpha \in I$, are said to be independent if $\{\sigma(A_\alpha)\}_{\alpha\in I}$ is independent. It is readily verified that events $A'_\alpha s$ are independent if and only if for any finite set of indices $\alpha_1, \ldots, \alpha_k$, $P(\bigcap_{l=1}^k A_{\alpha_l}) = \prod_{l=1}^k P(A_{\alpha_l})$.

Given a sequence $A_1, A_2, \ldots$ of events, let $\mathcal{T} = \mathcal{T}(A_1, A_2, \ldots) = \bigcap_{n=1}^\infty \sigma(A_n, A_{n+1}, \ldots)$. Events in $\mathcal{T}$ are referred to as **tail events** of the sequence $\{A_n\}$. It is evident that $\liminf_{n\to\infty} A_n$ and $\limsup_{n\to\infty} A_n$ are tail events of the sequence $\{A_n\}$. The following zero-one law of Kolmogoroff is a far-reaching consequence of the notion of independence.

Theorem 7.5.1 (Kolmogoroff's zero–one law) *If $A_1, A_2, A_3, \ldots$ are independent events, then every tail event of $\{A_n\}$ has probability zero or one.*

Proof Suppose that A is a tail event of the sequence $\{A_n\}$. For $n \geq 2$, let $\mathcal{L}$ be the family of all such $B \in \Sigma$, with the property that

$$P(B_1 \cap \cdots \cap B_{n-1} \cap B) = P(B_1) \cdots P(B_{n-1}) P(B),$$

where for each $j = 1, \ldots, n-1$, $B_j = A_j$ or Ω; then $\mathcal{L}$ is a λ-system. Next, let $\mathcal{P}$ be the family of all finite intersections of $A_n, A_{n+1}, \ldots$; $\mathcal{P}$ is then a π-system and $\mathcal{P} \subset \mathcal{L}$. Hence $\sigma(\mathcal{P}) \subset \mathcal{L}$ by the (π-λ) theorem. But $A \in \sigma(A_n, A_{n+1}, \ldots) = \sigma(\mathcal{P}) \subset \mathcal{L}$; this means that $A, A_1, \ldots, A_{n-1}$ are independent.

We now claim that $P(A \cap B) = P(A)P(B)$ for $B \in \sigma(A_1, A_2, \ldots)$. For this purpose, let $\mathcal{L}' = \{B \in \Sigma : P(A \cap B) = P(A)P(B)\}$ and $\mathcal{P}'$ be the family of all finite intersections of $A_1, A_2, \ldots$; clearly, $\mathcal{L}'$ is a λ-system and $\mathcal{P}'$ a π-system. The fact that $A, A_1, \ldots, A_{n-1}$ are independent for each $n \geq 2$ implies that $\mathcal{P}' \subset \mathcal{L}'$. Thus, $\sigma(\mathcal{P}') = \sigma(A_1, A_2, \ldots) \subset \mathcal{L}'$, by the ($\pi$-$\lambda$) theorem, which means that $P(A \cap B) = P(A)P(B)$ for $B \in \sigma(A_1, A_2, \ldots)$; but since $A \in \sigma(A_1, A_2, \ldots)$, $P(A) = P(A)^2$. Hence $P(A) = 0$ or 1. ∎

Exercise 7.5.1 Let $\mathcal{T} = \bigcap_n \sigma(A_n, A_{n+1}, \ldots)$, where $A_1, A_2, \ldots$ are independent events. Show that if X is a $\mathcal{T}$-measurable random variable, then X = constant a.s.

In accord with notations for certain sets introduced in the second paragraph of Section 2.2, if T is a map from a set Ω to a set S, the set $T^{-1}A$, $A \subset S$, will be denoted by $\{T \in A\}$; and if $T_\alpha : \Omega \to S_\alpha$, $\alpha \in I$, then $\bigcap_{\alpha\in I} T_\alpha^{-1} A_\alpha$, $A_\alpha \subset S_\alpha$, is denoted by $\{T_\alpha \in A_\alpha, \alpha \in I\}$; in particular, if $X_1, \ldots, X_k$ are random variables, then $\bigcap_{j=1}^k \{X_j \in B_j\} = \{X_1 \in B_1, \ldots, X_k \in B_k\}$. When a probability measure P is concerned, $P(\{\cdots\})$ will be abbreviated to $P(\cdots)$.

Given a family $\{X_\alpha\}$ of r.v.'s, the smallest σ-algebra relative to which every X_α is measurable is denoted by $\sigma(X'_\alpha s)$; in particular, $\sigma(X_1, \ldots, X_k)$ is the smallest σ-algebra relative to which $X_1, \ldots, X_k$ are measurable.

Exercise 7.5.2 If $X_1, \ldots, X_k$ are r.v.'s, let $X = (X_1, \ldots, X_k)$ be the map from Ω to $\mathbb{R}^k$ defined by $X(\omega) = (X_1(\omega), \ldots, X_k(\omega))$ for $\omega \in \Omega$. Show that $\sigma(X_1, \ldots, X_k) = \{X^{-1}B : B \in \mathcal{B}^k\}$.

We shall call a map $X : \Omega \to \mathbb{R}^k$, $k \geq 2$ a **random vector** if $X^{-1}B \in \Sigma$ for all $B \in \mathcal{B}^k$; in other words, X is a random vector if X is $\Sigma|\mathcal{B}^k$-measurable. Put $X = (X_1, \ldots, X_k)$, where $X_1, \ldots, X_k$ are the component functions of X. Since $\{X^{-1}B : B \in \mathcal{B}^k\} \supset \bigcup_{j=1}^k \{X_j^{-1}B_j : B_j \in \mathcal{B}\}$, we conclude that if X is a random vector, then $X_1, \ldots, X_k$ are r.v.'s; on the other hand, if $X_1, \ldots, X_k$ are r.v.'s, then X is a random vector, by Exercise 7.5.2. Thus, $X = (X_1, \ldots, X_k)$ is a random vector if and only if $X_1, \ldots, X_k$ are r.v.'s.

A family $\{X_\alpha\}$ of r.v.'s is said to be **independent** if $\{\sigma(X_\alpha)\}$ is independent; then we also say that $X_\alpha' s$ are independent.

Exercise 7.5.3 Suppose that $\{X_\alpha\}$ is an independent family of r.v.'s and that $\{g_\alpha\}$ is a family of Borel functions on $\mathbb{R}$. Show that $\{g_\alpha \circ X_\alpha\}$ is an independent family of r.v.'s.

Lemma 7.5.1 *If $X_1, \ldots, X_n$, $n \geq 2$, are independent r.v.'s then for integer j, $1 \leq j < n$, $\sigma(X_1, \ldots, X_j)$ and $\sigma(X_{j+1}, \ldots, X_n)$ are independent.*

Proof Put $\widehat{X} = (X_1, \ldots, X_j)$ and $\widehat{Y} = (X_{j+1}, \ldots, X_n)$. In view of Exercise 7.5.2, we need to show that

$$P(\widehat{X} \in B, \widehat{Y} \in C) = P(\widehat{X} \in B) \cdot P(\widehat{Y} \in C) \tag{7.16}$$

for all $B \in \mathcal{B}^j$ and $C \in \mathcal{B}^{n-j}$. Consider $B_l \in \mathcal{B}$, $l = 1, \ldots, n$; we have

$$P(\widehat{X} \in B_1 \times \cdots \times B_j, \widehat{Y} \in B_{j+1} \times \cdots \times B_n) = P(X_1 \in B_1, \ldots, X_n \in B_n)$$
$$= \prod_{l=1}^n P(X_l \in B_l) = P(\widehat{X} \in B_1 \times \cdots \times B_j) \cdot P(\widehat{Y} \in B_{j+1} \times \cdots \times B_n),$$

hence, (7.16) holds for $B = B_1 \times \cdots \times B_j$ and $C = B_{j+1} \times \cdots \times B_n$. Fix $B_{j+1}, \ldots, B_n$ and let $\mathcal{N} = \{B \in \mathcal{B}^j : P(\widehat{X} \in B, \widehat{Y} \in B_{j+1} \times \cdots \times B_n) = P(\widehat{X} \in B)P(\widehat{Y} \in B_{j+1} \times \cdots \times B_n)\}$. Evidently, $\mathcal{N}$ is a λ-system containing the family $\mathcal{P}$ of all sets of the form $B_1 \times \cdots \times B_j$, where $B_1, \ldots, B_j$ are in $\mathcal{B}$. Now $\mathcal{P}$ is a π-system and $\sigma(\mathcal{P}) = \mathcal{B}^j$, therefore $B^j \supset \mathcal{N} \supset \sigma(\mathcal{P}) = \mathcal{B}^j$. Thus $\mathcal{N} = \mathcal{B}^j$. This means that

$$P(\widehat{X} \in B, \widehat{Y} \in B_{j+1} \times \cdots \times B_n) = P(\widehat{X} \in B) \cdot P(\widehat{Y} \in B_{j+1} \times \cdots \times B_n)$$

for $B \in B^j$ and $B_{j+1}, \ldots, B_n$ in $\mathcal{B}$. Next fix $B \in \mathcal{B}^j$ and let

$$\mathcal{N}' = \{C \in \mathcal{B}^{n-j} : P(\widehat{X} \in B, \widehat{Y} \in C) = P(\widehat{X} \in B) \cdot P(\widehat{Y} \in C)\}.$$

Argue as in the immediately preceding part of the proof, we infer that $\mathcal{N}' = \mathcal{B}^{n-j}$ and finish the proof. ■

Lemma 7.5.2 *Suppose that $X_1, \ldots, X_n$, $n \geq 2$, are independent r.v.'s, and let $1 \leq j < n$ be an integer. Then $g_1 \circ (X_1, \ldots, X_j)$ and $g_2 \circ (X_{j+1}, \ldots, X_n)$ are independent if g_1 and g_2 are Borel functions on $\mathbb{R}^j$ and $\mathbb{R}^{n-j}$ respectively.*

Proof Let B and C be Borel sets of $\mathbb{R}$. Since $\{g_1 \circ (X_1, \ldots, X_j) \in B\} = \{(X_1, \ldots, X_j) \in g_1^{-1}B\}$ and $\{g_2 \circ (X_{j+1}, \ldots, X_n) \in C\} = \{(X_{j+1}, \ldots, X_n) \in g_2^{-1}C\}$, and since $g_1^{-1}B$

and $g_2^{-1}C$ are in $\mathcal{B}^j$ and B^{n-j} respectively, we know from Exercise 7.5.2 that $\{g_1 \circ (X_1, \dots, X_j) \in B\}$ and $\{g_2 \circ (X_{j+1}, \dots, X_n) \in C\}$ are in $\sigma(X_1, \dots, X_j)$ and $\sigma(X_{j+1}, \dots, X_n)$ respectively. It then follows from Lemma 7.5.1 that

$$\begin{aligned} &P(g_1 \circ (X_1, \dots, X_j) \in B, g_2 \circ (X_{j+1}, \dots, X_n)) \\ &= P(g_1 \circ (X_1, \dots, X_j) \in B) \cdot P(g_2 \circ (X_{j+1}, \dots, X_n) \in C). \end{aligned}$$ ■

Theorem 7.5.2 *If X and Y are independent integrable r.v.'s, then XY is integrable and $E(XY) = E(X) \cdot E(Y)$.*

Proof By Exercise 7.5.3, X^{ε_1} and X^{ε_2} are independent, where each of the symbols ε_1 and ε_2 is either + or −. We may therefore assume that both X and Y are nonnegative. Observe then that if S_1 and S_2 are simple functions measurable w.r.t. $\sigma(X)$ and $\sigma(Y)$ respectively, then $E(S_1S_2) = E(S_1) \cdot E(S_2)$. Now, choose increasing sequences $\{S_n^{(1)}\}$ and $\{S_n^{(2)}\}$ of simple functions such that each $S_n^{(1)}$ is $\sigma(X)$-measurable and each $S_n^{(2)}$ is $\sigma(Y)$-measurable; and furthermore $S_n^{(1)} \nearrow X$ and $S_n^{(2)} \nearrow Y$ pointwise. Using the monotone convergence theorem, we have

$$\begin{aligned} E(X) \cdot E(Y) &= \left[\lim_{n\to\infty} E(S_n^{(1)})\right]\left[\lim_{n\to\infty} E(S_n^{(2)})\right] = \lim_{n\to\infty}[E(S_n^{(1)}) \cdot E(S_n^{(2)})] \\ &= \lim_{n\to\infty} E(S_n^{(1)}S_n^{(2)}) = E(XY). \end{aligned}$$ ■

Corollary 7.5.1 *If $X_1, \dots, X_n$, $n \geq 2$ are independent integrable r.v.'s, then $X_1 \cdots X_n$ is integrable and $E(X_1 \cdots X_n) = E(X_1) \cdots E(X_n)$.*

Proof When $n = 2$, this is Theorem 7.5.2. Suppose now that $n \geq 3$; then $X_1 \cdots X_{n-1}$ and X_n are independent, by Lemma 7.5.2, and the corollary follows by induction on n. ■

Corollary 7.5.2 *If $X_1, \dots, X_n$ are independent and integrable, then*

$$\operatorname{Var}\left(\sum_{j=1}^n X_j\right) = \sum_{j=1}^n \operatorname{Var}(X_j).$$

Proof

$$\begin{aligned} \operatorname{Var}\left(\sum_{j=1}^n X_j\right) &= E\left(\left[\sum_{j=1}^n \{X_j - E(X_j)\}\right]^2\right) \\ &= E\left(\sum_{j=1}^n \{X_j - E(X_j)\}^2 + \sum_{j\neq k}\{X_j - E(X_j)\}\{X_k - E(X_k)\}\right) \\ &= \sum_{j=1}^n \operatorname{Var}(X_j) + \sum_{j\neq k} E(\{X_j - E(X_j)\}\{X_k - E(X_k)\}) \\ &= \sum_{j=1}^n \operatorname{Var}(X_j), \end{aligned}$$

because $X_j - E(X_j)$ and $X_k - E(X_k)$ are independent, by Exercise 7.5.3 if $j \neq k$, and hence $E(\{X_j - E(X_j)\}\{X_k - E(X_k)\}) = E(\{X_j - E(X_j)\}) \cdot E(\{X_k - E(X_k)\}) = 0$, by Theorem 7.5.2. ∎

A probability measure μ on $\mathcal{B}$ is called a **probability distribution** and the distribution $X_\# P$ of a r.v. X is called the **probability distribution** of X (recall that $X_\# P(B) = P(X \in B)$ for $B \in \mathcal{B}$). A family of r.v.'s is said to be **identically distributed** if random variables of the family have identical probability distribution. For $p > 0$, $E(|X|^p)$ is called the p-th **absolute moment** of the r.v. X; while if $m \in \mathbb{N}$, $E(X^m)$ is referred to as the m-th **moment** of X.

Example 7.5.1 A r.v. X is said to be normally distributed with mean m and variance σ^2 if for $B \in \mathcal{B}$,

$$X_\# P(B) = P(X \in B) = \frac{1}{\sqrt{2\pi}\sigma} \int_B \exp\left\{\frac{-(x-m)^2}{2\sigma^2}\right\} dx,$$

where as usual we write $\exp\{\beta\}$ for e^β if the expression for β is complicated. If X is normally distributed with mean m and variance σ^2, then

$$\begin{aligned}
E(X) &= \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} \exp\left\{\frac{-(x-m)^2}{2\sigma^2}\right\} x dx \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} e^{-t^2} (\sqrt{2}\sigma t + m)\sqrt{2}\sigma dt \\
&= \frac{m}{\sqrt{\pi}} \int_{\mathbb{R}} e^{-t^2} dt = m; \\
\operatorname{Var}(X) &= \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} \exp\left\{\frac{-(x-m)^2}{2\sigma^2}\right\} (x-m)^2 dx \\
&= \frac{2\sigma^2}{\sqrt{\pi}} \int_{\mathbb{R}} e^{-t^2} t^2 dt = \sigma^2.
\end{aligned}$$

Thus X actually has m as its expectation and σ^2 its variance. The probability distribution μ, defined by

$$\mu(B) = \frac{1}{\sqrt{2\pi}\sigma} \int_B \exp\left\{\frac{-(x-m)^2}{2\sigma^2}\right\} dx, \quad B \in \mathcal{B},$$

is called the **normal distribution with mean m and variance σ^2** and is denoted by $N(m, \sigma^2)$. The distribution $N(0, 1)$ is called the **standard normal distribution**.

Example 7.5.2 Consider the Bernoulli sequence space $(\Omega, \sigma(\mathcal{Q}), P)$ of Example 3.4.6. Recall that $\Omega = \{\omega = (\omega_k) : \omega_k \in \{0, 1\}, k \in \mathbb{N}\}$; $\mathcal{Q}$ is the smallest algebra on Ω that contains all sets of the form $E(\varepsilon_1, \ldots, \varepsilon_n) = \{\omega = (\omega_k) : \omega_1 = \varepsilon_1, \ldots, \omega_n = \varepsilon_n\}$, $n \in \mathbb{N}$ and $\varepsilon_j \in \{0, 1\}, j = 1, \ldots, n$, and P is the unique probability measure on $\sigma(\mathcal{Q})$ such that $P(E(\varepsilon_1, \ldots, \varepsilon_n)) = 2^{-n}$. If for $j \in \mathbb{N}$ and $\varepsilon \in \{0, 1\}$ let $E_\varepsilon^j = \{\omega = (\omega_k) : \omega_j = \varepsilon\}$, then we know from Exercise 1.3.2 that

$$P\left(E_{\varepsilon_1}^{j_1} \cap \cdots \cap E_{\varepsilon_k}^{j_k}\right) = \prod_{l=1}^{k} P(E_{\varepsilon_l}^{j_l}) = 2^{-k} \tag{7.17}$$

if $1 \leq j_1 < \cdots < j_k$ is any finite sequence in $\mathbb{N}$. Now for $j \in \mathbb{N}$, define a r.v. X_j by $X_j(\omega) = \omega_j$, then $\sigma(X_j) = \{\emptyset, E_0^j, E_1^j, \Omega\}$; and therefore we infer from (7.17) that $\{\sigma(X_j)\}$ is independent and consequently the r.v.'s $X_1, \ldots, X_j, \ldots$ are independent. Clearly, the probability distribution of each X_j is the measure μ on $\mathcal{B}$ such that $\mu(\{0\}) = \mu(\{1\}) = \frac{1}{2}$. Hence the sequence $\{X_j\}$ is identically distributed; furthermore $E(X_j) = \frac{1}{2}$, $\mathrm{Var}(X_j) = \frac{1}{8}$, and m-th moment of X_j is $\frac{1}{2}$ for $m \in \mathbb{N}$.

Return now to the general discussion and consider an independent and identically distributed sequence $\{X_j\}$ of r.v.'s. Such a sequence is usually referred to as an **i.i.d.** sequence. Suppose that the common probability distribution of $X_j's$ is μ, then for any Borel function g on $\mathbb{R}$ such that $\int_{\mathbb{R}} g d\mu$ exists, we know from (4.1) that $\int_{\mathbb{R}} g d\mu = \int_{\Omega} g \circ X_j dP$; in particular, the m-th moment is the same for all $X_j's$ if it exists for one of them. Thus $E(X_j^2) = E(X_1^2)$ for all j. Assume now that $E(X_1^2) < \infty$ and let $S_n = \sum_{j=1}^n X_j$, $n \in \mathbb{N}$. Then, $E(S_n) = nE(X_1)$ or $E(\frac{S_n}{n}) = E(X_1)$, and hence from the Chebyshev inequality (6.3), we have

$$P\left(\left|\frac{S_n}{n} - E(X_1)\right| \geq \varepsilon\right) \leq \varepsilon^{-2}\,\mathrm{Var}\left(\frac{S_n}{n}\right) = \frac{1}{n\varepsilon^2}\,\mathrm{Var}(X_1)$$

for any given $\varepsilon > 0$. This is stated as a theorem.

Theorem 7.5.3 (Weak law of large numbers) *Suppose that $\{X_j\}$ is an i.i.d. sequence of r.v.'s with finite second moment, then*

$$P\left(\left|\frac{S_n}{n} - E(X_1)\right| \geq \varepsilon\right) \leq \frac{1}{n\varepsilon^2}\,\mathrm{Var}(X_1) \tag{7.18}$$

for any given $\varepsilon > 0$, where $S_n = \sum_{j=1}^n X_j$.

A sequence $\{Y_j\}$ of r.v.'s is said to **converge in probability** to a r.v. Y if $\lim_{j\to\infty} P(|Y_j - Y| \geq \varepsilon) = 0$ for every $\varepsilon > 0$; the notation $Y_j \to Y[P]$ is used to mean that $\{Y_j\}$ converges to Y in probability. Apparently, convergence of Y_j to Y a.s. or in L^p-norm as $j \to \infty$ implies that $Y_j \to Y[P]$, hence convergence in probability is weaker than convergence a.s. and convergence in L^p-norm. Since Theorem 7.5.3 implies that $\frac{S_n}{n} \to E(X_1)[P]$, it is usually referred to as the **weak law** of large numbers.

Theorem 7.5.4 (Strong law of large numbers) *Suppose that $\{X_j\}$ is an independent sequence of r.v.'s such that $E(X_j) = 0$ and $E(X_j^4) \leq C < \infty$ for $j \in \mathbb{N}$. Let $S_n = \sum_{j=1}^n X_j$, then $\frac{S_n}{n} \to 0$ a.s. as $n \to \infty$.*

Proof Observe that (cf. Exercise 7.5.3):

(i) $E(X_i X_j^3) = E(X_i)E(X_j^3) = 0$ if $i \neq j$;

(ii) $E(X_i X_j^2 X_k) = 0$ if i, j, k are different from one another; and

(iii) $E(X_iX_jX_kX_l) = 0$ if i, j, k, l are different from one another;
and note that

(iv) $\{E(X_j^2)\}^2 \leq E(X_j^4) \leq C$ for all j by Jensen's inequality (6.4).

Now since $E(S_n^4) = \sum_{i,j,k,l} E(X_iX_jX_kX_l)$, we conclude from (i), (ii), and (iii) that

$$\begin{aligned} E(S_n^4) &= \sum_{j=1}^{n} E(X_j^4) + \binom{4}{2} \sum_{1\leq i<j\leq n} E(X_i^2X_j^2) \\ &\leq nC + 6 \sum_{1\leq i<j\leq n} E(X_i^2)E(X_j^2); \end{aligned}$$

but $E(X_i^2)E(X_j^2) \leq \frac{1}{2}\{E(X_i^2)^2 + E(X_j^2)^2\} \leq \frac{1}{2}\{E(X_i^4) + E(X_j^4)\}$, by (iv), for each pair $i < j$, and consequently

$$E(S_n^4) \leq nC + 6\frac{n(n-1)}{2}C \leq 3Cn^2,$$

or

$$E\left(\left(\frac{S_n}{n}\right)^4\right) \leq \frac{3C}{n^2}.$$

The last inequality implies that $E(\sum_{n=1}^{\infty}(\frac{S_n}{n})^4) \leq 3C\sum_{n=1}^{\infty}\frac{1}{n^2} < \infty$, and hence $\sum_{n=1}^{\infty}(\frac{S_n}{n})^4 < \infty$ a.s. Then, $\lim_{n\to\infty}\frac{S_n}{n} = 0$ a.s. follows. ■

Corollary 7.5.3 *Let $\{X_j\}$ be an independent sequence of r.v.'s with bounded fourth moment such that $E(X_j) = E(X_1)$ for all $j \in \mathbb{N}$; then $\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n} X_j = E(X_1)$ a.s.*

Proof Put $Y_j = X_j - E(X_j)$; then $E(Y_j) = 0$ for all j and $\{E(Y_j^4)\}$ is bounded. We then apply Theorem 7.5.4 to conclude the proof. ■

Now apply Corollary 7.5.3 to the sequence $\{X_j\}$ of Example 7.5.2; we have

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n} X_j = \frac{1}{2} \quad \text{a.s.}$$

i.e. the event $\{\omega \in \Omega : \lim_{n\to\infty}\frac{S_n(\omega)}{n} = \frac{1}{2}\}$ occurs with probability one, where $S_n = \sum_{j=1}^{n} X_j$; in other words, if we interpret $\{X_j\}$ as a sequence of tossing of a fair coin, the relative frequency with which heads appears in the first n tosses approaches $\frac{1}{2}$ as $n \to \infty$ almost certainly. This is what we proclaim in the last paragraph of Section 1.3.

As we know in Example 4.3.2, the Bernoulli sequence space $(\Omega, \sigma(\mathcal{Q}), P)$ and $([0,1], \mathcal{B}|[0,1], \lambda)$ are measure-theoretically the same space; it is therefore worthwhile considering the counterpart of the sequence $\{X_j\}$ of Example 7.5.2 in the space $([0,1], \mathcal{B}|[0,1], \lambda)$. For $x \in [0,1]$, let $0.x_1 \ldots x_k \ldots$ be the binary expansion of x with the convention that in case where two expansions are possible, the expansion with infinitely many $1's$ is chosen, and for $j \in \mathbb{N}$, define a r.v. Z_j by $Z_j(x) = x_j$. From the discussion in Example 3.4.6, one verifies readily from the independence of the sequence $\{X_j\}$ of Example 7.5.2 that $\{Z_j\}$ is independent, $E(Z_j) = E(Z_1) = \frac{1}{2}$, and $E(Z_j^4) = \frac{1}{2}$. Then,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} Z_j = \frac{1}{2} \quad \text{a.s.}$$

i.e.

$$\lim_{n\to\infty} \frac{1}{n}\{\text{number of } 1's \text{ in } x_1, \ldots, x_n\} = \frac{1}{2} \tag{7.19}$$

for almost every x of $[0, 1]$. We call a number x in $[0, 1]$ **a normal number** if (7.19) holds. Then (7.19) can be stated as follows.

Theorem 7.5.5 (Borel) *Almost all numbers in* $[0, 1]$ *are normal.*

We now come to introduce the Fourier integral for probability distributions. The Fourier integral φ of a probability distribution μ is a function on $\mathbb{R}$, defined by

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx} d\mu(x), \quad t \in \mathbb{R}. \tag{7.20}$$

We call attention to inconsistency in the definition of Fourier integral for functions and for probability distributions; should consistency of definition be preferred, the function φ, defined by (7.20), would be called the Fourier inverse integral of μ. It is readily seen that $\varphi(0) = 1$, $|\varphi(t)| \le 1$, and φ is uniformly continuous on $\mathbb{R}$. In probability theory, φ is called the **characteristic function** of μ; and if a r.v. X has μ as its probability distribution, φ is also referred to as the **characteristic function of** X. Note that if φ is the characteristic function of X, then,

$$\varphi(t) = E(e^{itX}), \quad t \in \mathbb{R}.$$

Exercise 7.5.4 Let φ be the characteristic function of the r.v. X, and suppose that $E(|X|) < \infty$. Show that $\varphi \in C^1(\mathbb{R})$ and $\varphi'(t) = E(iXe^{itX})$.

Exercise 7.5.5 Show that the characteristic function φ of $N(0, 1)$ is given by $\varphi(t) = e^{-\frac{t^2}{2}}$.

Exercise 7.5.6 Suppose that φ is the characteristic function of a probability distribution μ. Show that for $u > 0$,

$$\mu\left(\left(-\infty, \frac{-2}{u}\right] \cup \left[\frac{2}{u}, \infty\right)\right) \le \frac{1}{u} \int_{-u}^{u} (1 - \varphi(t))dt.$$

(Hint: $\frac{1}{u}\int_{-u}^{u}(1 - \varphi(t))dt = 2\int_{-\infty}^{\infty}(1 - \frac{\sin ux}{ux})d\mu(x) \ge 2\int_{|x|\ge\frac{2}{\mu}}(1 - \frac{1}{|ux|})d\mu(x)$.)

Theorem 7.5.6 *Suppose that X_1 and X_2 are independent random variables with characteristic functions φ_1 and φ_2 respectively, and let φ be the characteristic function of $X_1 + X_2$, then $\varphi = \varphi_1\varphi_2$.*

Proof Since e^{itX_1} and e^{itX_2} are independent, we have

$$\varphi(t) = E(e^{it(X_1+X_2)}) = E(e^{itX_1} \cdot e^{itX_2})$$

$$= E(e^{itX_1}) \cdot E(e^{itX_2}) = \varphi_1(t)\varphi_2(t). \qquad \blacksquare$$

Theorem 7.5.7 (Inversion formula) *Let μ be a probability distribution with characteristic function φ; then for $a < b$ in $\mathbb{R}$,*

$$\mu((a,b]) = \lim_{L\to\infty} \frac{1}{2\pi} \int_{-L}^{L} \frac{e^{-ita} - e^{-itb}}{it} \varphi(t)dt$$

if $\mu(\{a\}) = \mu(\{b\}) = 0$.

Proof Put $S(L) = \int_0^L \frac{\sin t}{t} dt$, $L > 0$. Then $\int_0^L \frac{\sin \theta t}{t} dt = \operatorname{sgn} \theta S(L|\theta|)$ and $\lim_{L\to\infty} S(L) = \frac{\pi}{2}$. Now consider the integral

$$\begin{aligned}\mathcal{I}(L) &= \frac{1}{2\pi} \int_{-L}^{L} \left[\frac{e^{-ita} - e^{-itb}}{it}\right] \varphi(t)dt \\ &= \frac{1}{2\pi} \int_{-L}^{L} \left(\int_{-\infty}^{\infty} \frac{e^{it(x-a)} - e^{it(x-b)}}{it} d\mu(x) \right) dt.\end{aligned}$$

From the elementary inequality $|e^{i\theta} - 1| \le |\theta|$, we have

$$\left| \frac{e^{it(x-a)} - e^{it(x-b)}}{it} \right| = \frac{1}{|t|} |e^{it(b-a)} - 1| \le b - a$$

for any $x \in \mathbb{R}$. We may therefore apply the Fubini theorem to the integral defining $\mathcal{I}(L)$:

$$\begin{aligned}\mathcal{I}(L) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\int_{-L}^{L} \frac{e^{it(x-a)} - e^{it(x-b)}}{it} dt \right) d\mu(x) \\ &= \int_{-\infty}^{\infty} \left(\int_0^L \left[\frac{\sin t(x-a)}{\pi t} - \frac{\sin t(x-b)}{\pi t} \right] dt \right) d\mu(x) \\ &= \int_{-\infty}^{\infty} \left[\frac{\operatorname{sgn}(x-a)}{\pi} S(L|x-a|) - \frac{\operatorname{sgn}(x-b)}{\pi} S(L|x-b|) \right] d\mu(x).\end{aligned}$$

Let us denote the integrand of this last integral by $\theta_L(a,b;x)$ and put $\theta_{ab}(x) = \lim_{L\to\infty}\theta_L(a,b;x)$; then,

$$\theta_{ab}(x) = \begin{cases} 0 & \text{if } x < a \text{ or } x > b; \\ \dfrac{1}{2} & \text{if } x = a \text{ or } x = b; \\ 1 & \text{if } a < x < b. \end{cases}$$

From the second mean-value theorem, $\{\int_0^\alpha \frac{\sin t}{t}dt\}_{\alpha>0}$ is bounded and therefore $|\theta_L(a,b;x)| \le M < \infty$ for all $L > 0$ and $x \in \mathbb{R}$. Hence by LDCT, we conclude that

$$\lim_{L\to\infty}\mathcal{I}(L) = \int_{-\infty}^{\infty}\theta_{ab}(x)d\mu(x)$$

$$= \frac{1}{2}\mu(\{a\}) + \mu((a,b)) + \frac{1}{2}\mu(\{b\}) = \mu((a,b]).$$ ■

Exercise 7.5.7 Let μ be the probability measure on $\mathcal{B}$ concentrated at 0. Find the characteristic function of μ and use Theorem 7.5.7 to show that

$$\lim_{L\to\infty}\int_0^L \frac{\sin at}{t}dt = \int_0^\infty \frac{\sin at}{t}dt = \frac{\pi}{2}$$

for all $a > 0$.

Corollary 7.5.4 *If the probability distributions μ and ν have the same characteristic function, then $\mu = \nu$.*

Proof Let $\Pi = \{(a,b] : \mu(\{a\}) = \mu(\{b\}) = \nu(\{a\}) = \nu(\{b\}) = 0\} \cup \{\emptyset\}$, and $\mathcal{N} = \{B \in \mathcal{B} : \mu(B) = \nu(B)\}$. Theorem 7.5.7 implies that $\mathcal{N} \supset \Pi$. But Π is a π-system, $\mathcal{N}$ is a λ-system, and $\sigma(\Pi) = \mathcal{B}$, hence it follows from the $(\pi$-$\lambda)$ theorem that $\mathcal{N} = \mathcal{B}$. ■

Corollary 7.5.4 means that the characteristic function of a probability distribution μ uniquely determines μ and is therefore named the characteristic function of μ.

We are ready to state and prove the central limit theorem in probability theory. Suppose that $\{X_j\}$ is an i.i.d. sequence of random variables such that $E(X_j) = 0$, $\text{Var}(X_j) = E(X_j^2) = 1$, and $E(|X_j|^3) < \infty$. For $n \in \mathbb{N}$, put $Y_n = \frac{1}{\sqrt{n}}\sum_{j=1}^n X_j$.

Theorem 7.5.8 (Central limit theorem) *The characteristic function of Y_n converges to the characteristic function of $N(0,1)$ uniformly on any given finite interval.*

Proof Denote by φ the common characteristic function of X_j's and by μ the common distribution of X_j's. Using the fundamental theorem of calculus repeatedly, we have

$$\begin{aligned}
e^{itx} &= 1 + i\int_0^{tx} e^{i\theta}d\theta = 1 + i\int_0^{tx}\left(1 + i\int_0^{\theta} e^{is}ds\right)d\theta \\
&= 1 + itx - \int_0^{tx}\left(\int_0^{\theta} e^{is}ds\right)d\theta \\
&= 1 + itx - \int_0^{tx}\left(\int_0^{\theta}\left(1 + i\int_0^{s} e^{i\tau}d\tau\right)ds\right)d\theta \\
&= 1 + itx - \frac{1}{2}t^2x^2 - i\int_0^{tx}\left(\int_0^{\theta}\left(\int_0^{s} e^{i\tau}d\tau\right)ds\right)d\theta \\
&= 1 + itx - \frac{1}{2}t^2x^2 + h(tx),
\end{aligned}$$

where $|h(tx)| \leq \frac{1}{6}|tx|^3$; consequently,

$$\varphi(t) = \int_{\mathbb{R}} e^{itx}d\mu(x) = 1 - \frac{1}{2}t^2 + \int_{\mathbb{R}} h(tx)d\mu(x) = 1 - \frac{1}{2}t^2 + H(t). \tag{7.21}$$

Note that $\int_{\mathbb{R}} itxd\mu(x) = itE(X_j) = 0$ and $\int_{\mathbb{R}} t^2x^2d\mu(x) = t^2E(X_j^2) = t^2$ have been used in deriving (7.21), and that

$$|H(t)| \leq \frac{1}{6}E(|X_j|^3)|t|^3 \equiv C|t|^3. \tag{7.22}$$

Now let I be a finite interval in $\mathbb{R}$; then for some $b > 0$, $|t| \leq b$ for $t \in I$, and hence there is $n_0 \in \mathbb{N}$, such that

$$\left(1 - \frac{1}{2}\frac{t^2}{n}\right) \geq \frac{1}{2}, \quad t \in I \tag{7.23}$$

if $n \geq n_0$. Denote now by φ_n the characteristic function of Y_n. We know from Theorem 7.5.6 that

$$\begin{aligned}
\varphi_n(t) &= E\left(\exp\left\{\frac{it}{\sqrt{n}}\sum_{j=1}^{n} X_j\right\}\right) = \left[\varphi\left(\frac{t}{\sqrt{n}}\right)\right]^n \\
&= \left[1 - \frac{1}{2}\frac{t^2}{n} + H\left(\frac{t}{\sqrt{n}}\right)\right]^n \\
&= \left(1 - \frac{1}{2}\frac{t^2}{n}\right)^n\left\{1 + \left(1 - \frac{1}{2}\frac{t^2}{n}\right)^{-1}H\left(\frac{t}{\sqrt{n}}\right)\right\}^n \\
&= \left(1 - \frac{1}{2}\frac{t^2}{n}\right)^n(1 + G(t,n))^n,
\end{aligned}$$

of which for $n \geq n_0$ and $t \in I$, we have from (7.22) and (7.23),

$$|G(t,n)| \leq 2Cb^3 n^{-\frac{3}{2}}.$$

Observe now from the mean-value theorem in differential calculus that, for $\alpha \in \mathbb{R}$ and $n \in \mathbb{N}$,

$$\begin{aligned}\left(1+\frac{\alpha}{n}\right)^n &= \exp\left\{\ln\left(1+\frac{\alpha}{n}\right)^n\right\}\\ &= \exp\left\{n[\ln(n+\alpha)-\ln n]\right\} = \exp\left\{\frac{n\alpha}{n+\alpha_n}\right\},\end{aligned}$$

where $n+\alpha_n$ is between n and $n+\alpha$, and consequently

$$\lim_{n\to\infty}\left(1+\frac{\alpha}{n}\right)^n = e^{\alpha}$$

uniformly for $|\alpha| \leq B$ if $B > 0$ is fixed. As a consequence,

$$\lim_{n\to\infty}\left(1-\frac{1}{2}\frac{t^2}{n}\right)^n = e^{-\frac{t^2}{2}}$$

uniformly for $t \in I$; and since $|nG(t,n)| \leq 2Cb^3 n^{-\frac{1}{2}}$ for $n \geq n_0$ and $t \in I$, for any given $\varepsilon > 0$ there is $n_1 \geq n_0$ in $\mathbb{N}$ such that if $n \geq n_1$ and $t \in I$, then,

$$\left|\left(1+\frac{nG(t,n)}{n}\right)^n - e^{nG(t,n)}\right| < \frac{\varepsilon}{2}. \tag{7.24}$$

We may choose n_1 sufficiently large so that, if $n \geq n_1$ and $t \in I$, then $|nG(t,n)|$ will be small enough so that $|nG(t,n)| < \frac{\varepsilon}{4}$, and

$$1-|nG(t,n)| \leq e^{nG(t,n)} \leq 1+2|nG(t,n)|. \tag{7.25}$$

Finally, using (7.24) and (7.25), we have for $n \geq n_1$ and $t \in I$,

$$\begin{aligned}(1+G(t,n))^n - 1 &> e^{nG(t,n)} - \frac{\varepsilon}{2} - 1 \geq 1-|nG(t,n)| - \frac{\varepsilon}{2} - 1 > -\varepsilon;\\ (1+G(t,n))^n - 1 &< e^{nG(t,n)} + \frac{\varepsilon}{2} - 1 \leq 1+2|nG(t,n)| + \frac{\varepsilon}{2} - 1 \leq \varepsilon.\end{aligned}$$

Thus, $|(1+G(t,n))^n - 1| < \varepsilon$ if $n \geq n_1$ and $t \in I$ i.e. $\lim_{n\to\infty}(1+G(t,n))^n = 1$ uniformly for $t \in I$. Summing up, we have shown that $\varphi_n(t)$ converges to $e^{-\frac{t^2}{2}}$ uniformly for $t \in I$. But $e^{-\frac{t^2}{2}}$ is the characteristic function of $N(0,1)$ (cf. Exercise 7.5.5). ■

The following exercise illustrates the relevance of the central limit theorem.

Exercise 7.5.8 Let Y_n be as in Theorem 7.5.8 and μ_n the probability distribution of Y_n; and let ν be $N(0,1)$. Furthermore, put $F_n(x) = \mu_n((-\infty, x])$ and $F(x) = \nu((-\infty, x])$ for $x \in \mathbb{R}$.

(i) Given that $\varepsilon > 0$. Show that there is $a > 0$ such that

$$\nu(\{|x| \geq a\}) < \varepsilon;$$
$$\mu_n(\{|x| \geq a\}) < \varepsilon, \ n = 1, 2, 3, \ldots,$$

(Hint: cf. Exercise 7.5.6 and central limit theorem.)

(ii) Show that if f is a bounded continuous function on $\mathbb{R}$, then

$$\lim_{n\to\infty} \int_{\mathbb{R}} f d\mu_n = \int_{\mathbb{R}} f d\nu.$$

(Hint: use (i) and Theorem 7.5.7.)

(iii) For $-\infty < \alpha < \beta < \infty$, define a continuous function $f_{\alpha,\beta}$ as follows:

$$f_{\alpha,\beta}(t) = \begin{cases} 1, & t \leq \alpha; \\ 0, & t \geq \beta; \\ \frac{\beta - t}{\beta - \alpha}, & \alpha < t < \beta. \end{cases}$$

Now let $-\infty < u < x < y < \infty$. By applying (ii) for $f = f_{x,y}$ and $f_{u,x}$ in this order, show that

$$\limsup_{n\to\infty} F_n(x) \leq F(y); \quad F(x-) \leq \liminf_{n\to\infty} F_n(x),$$

and then conclude that $\lim_{n\to\infty} F_n(x) = F(x)$ for $x \in \mathbb{R}$.

(iv) Show that for any finite interval I in $\mathbb{R}$,

$$\lim_{n\to\infty} \mu_n(I) = \frac{1}{\sqrt{2\pi}} \int_I e^{-\frac{t^2}{2}} dt.$$

Postscript

Although the general basic principles of real analysis are few, because of their wide applicability and their proven relevance over time in the development of mathematical analysis for its own purpose or for applications, manifold variations and derived principles have emerged whose scope is seldom matched by those of other subjects in mathematics. Therefore to write a book of reasonable size on real analysis which provides all the variations and derived principles is deemed to be impossible. I have, no matter how unwillingly, had to choose for discussion only those topics which are necessary for the understanding of those modern methods in analysis which apply the so-called real variables techniques.

Some brief treatment of Housdorff measures on Euclidean n-space and a more systematic discussion of real variables methods in harmonic analysis would be desirable. To do this sufficiently well to reveal the merit of these topics would not only increase the size of the book beyond a reasonable range, but would not really be in the reach of my capabilities. In this regard, I can do no better than to refer the interested reader to the masterful works [EG] and [St], listed in the bibliography.