# 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 $\mu$ defined for all subsets $A$ of a given set $\Omega$ is called an **outer measure on** $\Omega$ if it is monotone and $\sigma$-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 $\Omega$. 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 $\sigma$-subadditivity. A nonnegative set function $\tau$ 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 $\mu$ on $\Omega$ is usually simply called a **measure on $\boldsymbol{\Omega}$**. Sometimes we also say that $\mu$ measures $\Omega$. We emphasize that a measure on a set $\Omega$ and a measure on a $\sigma$-algebra on $\Omega$ are different objects; the former is an outer measure which is in general not $\sigma$-additive on $2^{\Omega}$.

Let $\mu$ be an outer measure on $\Omega$. Following Carathéodory, we say that a subset $A$ of $\Omega$ is $\mu$-**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 $\Omega$ is $\mu$-measurable and that if $\mu(A) = 0$, then $A$ is $\mu$-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, $\mu$ is an outer measure on $\Omega$ (recall that $\mu$ is called the counting measure on $\Omega$), and that every subset of $\Omega$ is $\mu$-measurable. It happens that $\mu$ is a measure on $2^\Omega$.

**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 $\mu$ is an outer measure on $\Omega$. What are the $\mu$-measurable subsets of $\Omega$? 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 $\nu$ is an outer measure on $\Omega$. What are the $\nu$-measurable subsets of $\Omega$?

**Exercise 3.1.2** Let $(\Omega, \Sigma, \mu)$ 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 $\mu_w$ measures $\Omega$ and every set in $\Sigma$ is $\mu_w$-measurable.

Suppose that $\mu$ is an outer measure on $\Omega$ and $A \subset \Omega$, then the **restriction** of $\mu$ 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 $\mu$ measure $\Omega$.

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

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

**Exercise 3.1.4** Suppose that $\mu$ measures $\Omega$ and that $A$ is a $\mu$-measurable subset of $\Omega$. 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 $\mu$-measurability for $A$.)

**Exercise 3.1.5** Let $\mu$ be an outer measure on $\Omega$. 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 $\mu$-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 $\lambda$ 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

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

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

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

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

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

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),

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

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 $\varepsilon$ 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 $\lambda$ is an outer measure on $\mathbb{R}$. ■

The measure $\lambda$ is called the **Lebesgue** measure on $\mathbb{R}$. We shall show later that $\lambda$ admits a fairly large class of $\lambda$-measurable sets, but, for the moment, we content ourselves by showing that every finite open interval $I$ is $\lambda$-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 $\lambda$-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 $\lambda$-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 $\lambda$-measurable and $\lambda(J) = |J|$. (Hint: $\lambda(\{x\}) = 0$ for $x \in \mathbb{R}$.)

**Exercise 3.2.2** Show that sets of the form $(a, \infty)$ or $(-\infty, a)$ are $\lambda$-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 $\Sigma$-algebra of measurable sets

Suppose that $\mu$ is an outer measure on $\Omega$ 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 $\mu$-measurable, then so is $\Omega \setminus A = A^c$.*

***Proof*** Obvious. ■

**Proposition 3.3.2** *If $A_1, A_2$ are $\mu$-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 $\mu$-measurable sets is $\mu$-measurable. This fact, together with Proposition 3.3.1, implies that the intersection of finitely many $\mu$-measurable sets is $\mu$-measurable.

**Proposition 3.3.3** *If $\{A_j\}_{j=1}^{\infty}$ is a disjoint sequence of $\mu$-measurable sets in $\Omega$ 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 $\mu$-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 $\sigma$-subadditivity of outer measures. ■

**Proposition 3.3.4** *If $\{A_j\}_{j=1}^{\infty}$ is a disjoint sequence of $\mu$-measurable sets, then $\bigcup_{j=1}^{\infty} A_j$ is $\mu$-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 $\mu(B) \geq \mu\big(B \cap \bigcup_{j=1}^{\infty} A_j\big) = \sum_{j=1}^{\infty} \mu(B \cap A_j) = \infty$; hence (3.3) also holds. ■

If we denote by $\Sigma^{\mu}$ the family of all $\mu$-measurable sets, it follows from Propositions 3.3.1, 3.3.2, and 3.3.4 that $\Sigma^{\mu}$ is both a $\pi$-system and a $\lambda$-system and is therefore a $\sigma$-algebra; while $\mu$ is $\sigma$-additive on $\Sigma^{\mu}$ by Proposition 3.3.3. Since $\Sigma^{\mu}$ contains all those subsets $A$ of $\Omega$ such that $\mu(A) = 0$, we have shown the following theorem.

**Theorem 3.3.1** *$\Sigma^{\mu}$ is a $\sigma$-algebra and $(\Omega, \Sigma^{\mu}, \mu)$ is a complete measure space.*

For later reference, $(\Omega, \Sigma^{\mu}, \mu)$ is called the **measure space** for $\mu$; and $\Sigma^{\mu}$-measurable functions are sometimes said to be $\mu$-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 $\sigma$-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 $\mu$ on $\Omega$ is said to be ***regular*** if for each $B \subset \Omega$, there is a $\mu$-measurable set $A \supset B$ such that $\mu(A) = \mu(B)$; more generally, if $\Sigma$ is a sub$\sigma$-algebra of $\Sigma^\mu$, we say that $\mu$ is $\Sigma$-**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 $\Omega$ and $\mu$ is a regular measure on $\Omega$, 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 $\mu$-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 $\lambda$ on $\mathbb{R}$ is a regular measure. This follows from Exercise 3.2.3.

**Exercise 3.3.1** Suppose that $\mu$ is a regular measure on $\Omega$ 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 $\Omega$ be a nonempty set, $\mathcal{G}$ a class of subsets of $\Omega$ containing $\emptyset$, and $\tau\colon \mathcal{G} \to [0, +\infty]$ satisfy $\tau(\emptyset) = 0$. Recall that such a set function $\tau$ is called a premeasure.

For a premeasure $\tau$, 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 $\tau^*$ measures $\Omega$ and is called the (outer) measure on $\Omega$ **constructed** from $\tau$ by **Method I**. That $\tau^*$ is an outer measure on $\Omega$ follows from the same arguments as in the proof of Theorem 3.2.1 to show that $\lambda$ 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 $\tau^*$ on $\mathbb{R}^n$ is called the **Lebesgue measure** on $\mathbb{R}^n$. The Lebesgue measure on $\mathbb{R}^n$ will be denoted by $\lambda^n$ and $\lambda^n$-measurable sets are called **Lebesgue measurable** sets. In conformity with the notation for Lebesgue measure on $\mathbb{R}$, introduced in Section 3.2, $\lambda^1$ will be replaced by $\lambda$. We shall denote by $\mathcal{L}^n$ the $\sigma$-algebra of all $\lambda^n$-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, $\lambda^n$-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 $\lambda^n$.

**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 $\tau_\varepsilon(I)$ = volume of $I$ for $I \in \mathcal{G}_\varepsilon$. Show that the measure $\tau_\varepsilon^*$ on $\mathbb{R}^n$ is the Lebesgue measure.

**Exercise 3.4.2** Let $\lambda^n$ 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 $\lambda^n(I)$ = volume of I if I is an open oriented rectangle. (Hint: use Lemma 1.7.2 to show $\lambda^n(I) \geq$ volume of I.)

(iii) Show that every open oriented rectangle is $\lambda^n$-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 $\varphi$ 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 $\lambda$-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 $\lambda$-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 $\lambda$-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 $\lambda$-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 $\lambda$-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 $\frac{k}{n} \le y < \frac{k+1}{n}$, $k = 0, 1, \ldots, n-1$. 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 $\frac{k}{n} \le y_0 < \frac{k+1}{n}$. 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 $\Omega$, 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 $\tau$ 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 $\tau^*$ 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 $\tau^*$ that $\tau^*(\bigcup_n G_n^{(k)}) \le \sum_n \tau(G_n^{(k)})$, and consequently that $\tau^*(A) \le \inf_k \tau^*(\bigcup_n G_n^{(k)}) \le \liminf_{k\to\infty} \sum_n \tau(G_n^{(k)}) = \tau^*(B)$. But $B \subset A$ implies $\tau^*(B) \le \tau^*(A)$, 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 $\lambda^n(G) \le \lambda^n(B) + \varepsilon$.

(ii) Show that for any $B \subset \mathbb{R}^n$, there is a $G_\delta$-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 $\tau$ is a $\sigma$-additive set function on an algebra $\mathcal{A}$ on $\Omega$, and let $\tau^*$ be the measure on $\Omega$ constructed from $\tau$ by Method I. Then $\sigma(\mathcal{A}) \subset \Sigma^{\tau^*}$ and $\tau(A) = \tau^*(A)$ for $A \in \mathcal{A}$. Furthermore, if $\tau$ is $\sigma$-finite, then the restriction of $\tau^*$ to $\sigma(\mathcal{A})$ is the unique measure on $\sigma(A)$ extending $\tau$.*

***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 $\sigma$-subadditivity of $\tau$ (cf. Exercise 2.1.1. (iv)), concluding that $\tau^*(A) \geq \tau(A)$.

Suppose now that $\nu$ 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 $\tau$ is $\sigma$-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 $\Omega$, the algebra $\mathcal{Q}$ of all cylinders in $\Omega$, and the set function $P$, defined in Section 1.3. We know from Example 2.1.1 that $P$ is $\sigma$-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 $\Omega$ 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 $(\Omega, \Sigma, \mu)$, the measure $\mu^*$ on $\Omega$ constructed from $\mu$ (considered as defined on $\Sigma$) by Method I is the unique $\Sigma$-regular measure on $\Omega$ 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 $\mu^*$ is $\Sigma$-regular.

To prove uniqueness, let $\nu$ be a $\Sigma$-regular measure on $\Omega$ 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 $\Sigma$ 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 $(\Omega, \Sigma, \mu)$ is $\sigma$-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 $(\Omega, \Sigma, \mu)$ is $\sigma$-finite, show that $(\Omega, \Sigma^{\mu^*}, \mu^*)$ is the completion of $(\Omega, \Sigma, \mu)$ (cf. Section 2.8.3).

(iii) If $\mu$ measures $\Omega$ and $\Sigma = \Sigma^{\mu}$, show that $\mu^* = \mu$ if and only if $\mu$ is regular.

**Remark** Because of Corollary 3.4.1, we may consider any measure space $(\Omega, \Sigma, \mu)$ as obtained by restricting to $\Sigma$ the $\Sigma$-regular measure $\mu^*$ on $\Omega$. Note that if $\mu$ is a measure on $\Omega$, the measure $\mu^*$ on $\Omega$ constructed from $\mu$ as a measure on $\Sigma^\mu$ by Method I is in general different from the original measure $\mu$ on $\Omega$ (cf. Exercise 3.4.18 (iii)).

**Theorem 3.4.2** *Let $\mathcal{A}$, $\tau$ be as in Theorem 3.4.1. Then $\Sigma^{\tau^*}$ is the largest $\sigma$-algebra containing $\mathcal{A}$ on which $\tau^*$ is $\sigma$-additive.*

***Proof*** Let $\Sigma'$ be a $\sigma$-algebra containing $\mathcal{A}$ on which $\tau^*$ is $\sigma$-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 $\sum_n \tau(A_n) \le \tau^*(B) + \varepsilon$. Put $H = \bigcup_n A_n$, then $H, H \cap A, H \cap A^c$ are in $\Sigma'$, and

$$\begin{aligned}\tau^*(B) + \varepsilon \ge \sum_n \tau(A_n) = \sum_n \tau^*(A_n) \ge \tau^*(H)\\ &= \tau^*(H \cap A) + \tau^*(H \cap A^c) \ge \tau^*(B \cap A) + \tau^*(B \cap A^c)\\ &\ge \tau^*(B).\end{aligned}$$

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 $(\pi$-$\lambda)$ 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 $\lambda^n$ to $\mathcal{B}^n$ by Method I is $\lambda^n$.

(ii) Show that $\lambda^n$ is not $\sigma$-additive on any $\sigma$-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 $\rho$ 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 $\rho$, $\rho(A, B)$ is usually denoted by $\operatorname{dist}(A, B)$ and is called the **distance** between $A$ and $B$. Recall that for a metric space $X$, we use $\mathcal{B}(X)$ to denote the $\sigma$-algebra generated by the family of all open sets of $X$ and that sets in $\mathcal{B}(X)$ are called Borel sets.

Let $\mu$ be a measure on $X$, with $X$ being a metric space, $\mu$ 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 $\mu$ is a Carathéodory measure on a metric space $X$, then every closed subset of $X$ is $\mu$-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,

$$\begin{aligned} \mu(D_1 \cup D_3 \cup \cdots \cup D_{2k-1}) &= \mu(D_1) + \mu(D_3) + \cdots + \mu(D_{2k-1})\,; \\ \mu(D_2 \cup D_4 \cup \cdots \cup D_{2k}) &= \mu(D_2) + \mu(D_4) + \cdots + \mu(D_{2k}) \end{aligned}$$

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,

$$\begin{aligned} \mu\left(\bigcup_{j=1}^{\infty} A_j\right) &= \mu\left(A_n \cup \bigcup_{j=n+1}^{\infty} A_j\right) = \mu\left(A_n \cup \bigcup_{j=n+1}^{\infty} D_j\right) \\ &\leq \mu(A_n) + \sum_{j=n+1}^{\infty} \mu(D_j), \end{aligned}$$

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 $(B, \rho)$),

$$\sup_n \mu(B_n) = \mu\left(\bigcup_{n=1}^{\infty} B_n\right) = \mu(B),$$

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,

$$\mu(A \cup B) \geq \mu(A) + \sup_n \mu(B_n) = \mu(A) + \mu(B).$$ ■

**Corollary 3.5.1** *If $\mu$ is a Carathéodory measure on a metric space $X$, then all Borel subsets of $X$ are $\mu$-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 $\tau_\varepsilon$ 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 $i$; $\tau_\varepsilon$ is the measure constructed from the restriction of $\tau$ to $\mathcal{G}_\varepsilon = \{\mathcal{C} \in \mathcal{G} : \operatorname{diam} \mathcal{C} \leq \varepsilon\}$ by Method I. Since $\tau_\varepsilon(A)$ increases as $\varepsilon$ decreases for $A \subset X$, $\lim_{\varepsilon \to 0} \tau_\varepsilon(A)$ exists and we define

$$\tau^d(A) = \lim_{\varepsilon \to 0} \tau_\varepsilon(A), \quad A \subset X.$$

**Exercise 3.6.1**

(i) Show that $\tau^d$ 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 $G_\delta$-set $B \supset A$ such that $\tau^d(A) = \tau^d(B)$.

We shall call $\tau^d$ the measure **constructed** from premeasure $\tau$ 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 $\tau^d$ be the measure on $\mathbb{R}$ constructed from $\tau$ by Method II. Show that every measurable set in $\mathbb{R}$ is $\tau^d$-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 $\tau^s$ 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 $\tau^s$ 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 $0 \leq t < s$.

**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 $0 \leq s \leq l$. Then for $s_1, s_2$ in $[0, l]$,

$$\operatorname{diam} t[s_1, s_2] \leq |s_1 - s_2|.$$

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,

$$l = \sum_{j=1}^{k} |s_j - s_{j-1}| \geq \sum_{j=1}^{k} \operatorname{diam} t[s_{j-1}, s_j] \geq \tau_\varepsilon^1(\mathcal{C}),$$

hence $l \geq H^1(\mathcal{C})$.

To show $l \leq H^1(\mathcal{C})$, 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 $A \subset \mathbb{R}^2$, $H^1(PA) \leq H^1(A)$. 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 $\mu_g$ 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 $\tau$ is a pre-measure on $\mathbb{R}$. The measure $\tau^*$ on $\mathbb{R}$ constructed from $\tau$ by Method $I$ is called the **Lebesgue–Stieltjes** measure generated by $g$ and is denoted by $\mu_g$; when $g(x) = x$, $\mu_g$ is the Lebesgue measure on $\mathbb{R}$.

It turns out that $\mu_g$ is also the measure $\tau^d$ on $\mathbb{R}$ constructed from $\tau$ 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 $\tau_\varepsilon$ and $\tau^d$).

**Theorem 3.7.1** *The measure $\mu_g$ 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 $G_\delta$-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 $G_\delta$-set).*

***Proof*** Since, as we have just shown, $\mu_g$ is a measure on $\mathbb{R}$ constructed from the premeasure $\tau$ by Method II, $\mu_g$ 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 $G_\delta$-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 $n_1, \dots, n_l$ and integers $1 \leq j_1 < \cdots < j_l = k$ such that $[x_{j_m} - x_{j_{m+1}}] \subset I_{n_{m+1}}$ for $m = 0, 1, \dots, l-1$. Now,

$$g(b+) - g(a-) \leq g(b') - g(a') = \sum_{m=1}^{l} \{g(x_{j_m}) - g(x_{j_{m-1}})\}$$
$$\leq \sum_{m=1}^{l} \tau(I_{n_m}) \leq \sum_{n} \tau(I_n),$$

from which, since $\{I_n\}$ is any sequence of finite open intervals with $\bigcup_n I_n \supset [a, b]$, it follows that $g(b+) - g(a-) \leq \mu_g([a, b])$. ■

**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 $\mathbb{R}$, $\mu_g$ 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 $(a_n, b_n]$, $n = 1, 2, \dots$, 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 $\nu$ 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 $g(y) - g(x) \leq \sum_{n>n_0}\nu(t_n) < \varepsilon$, 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 $\nu$ 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

$$g_d(x) = \begin{cases} \sum_{t\in(0,x]\cap D}\nu(t), & x \geq 0 \\ -\sum_{t\in(x,0]\cap D}\nu(t), & x < 0. \end{cases}$$

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 $\mu_g$ on $[a, b]$ generated by $g$ is the restriction of $\mu_h$ 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 $\mu_g$ 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 $\mu$ on a set $\Omega$ is called *regular* if for any $A \subset \Omega$, there is a $\mu$-measurable set $B \supset A$ such that $\mu(B) = \mu(A)$. Such a regularity endows $\mu$ 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 $\mu$ on a metric space $X$ is called a **Borel** measure if every Borel set is $\mu$-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, $\lambda^n$ 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 $\mu$ is a Borel measure on a metric space $X$ and $f$ is a nonnegative $\Sigma^\mu$-measurable function on $X$. Let $\nu$ 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 $\nu$ the indefinite integral of $f$ with respect to $\mu$, or simply the $\mu$-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 $\mu$-integrable on every compact subset of $X$, then $\{f\mu\}^*$ is a Radon measure. Note that if $\mu$ is $\sigma$-finite and Borel regular, then for any $\Sigma^\mu$-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 $\mu$ 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 $\tau$ a premeasure defined on $\mathcal{G} \subset \mathcal{B}(X)$. Then the measure $\tau^d$ on $X$ constructed from $\tau$ 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 $\tau$ 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 $\mu$ is a measure on $\Omega$ and $A \subset \Omega$, then the restriction to $A$ of $\mu$, 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 $\mu$ be a Borel regular measure on a metric space $X$ and suppose that $A \subset X$ is $\mu$-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 $\mu$-measurable set is $\nu$-measurable, $\nu$ is a Borel measure. It remains to show that $\nu$ 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 $\overset{\circ}{I} \cap \overset{\circ}{J} = \emptyset$.

**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 $\sigma$-algebra generated by the family of all compact sets.

**Lemma 3.9.1** *Suppose that $\mu$ 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 $\mu$ by $\mu \lfloor B$ if necessary, we may assume that $\mu$ 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,

$$\mu\left(B^c\setminus\bigcap_j K''_j\right) = \mu\left(\bigcap_j B^c_j\setminus\bigcap_j K''_j\right) \le \mu\left(\bigcup_j (B^c_j\setminus K''_j)\right)$$
$$\le \sum_j \mu(B^c_j\setminus K''_j) < \varepsilon;$$

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 $\lambda$-system. Since $\mathcal{M}$ contains all compact sets, and since the family of all compact sets is a $\pi$-system, $\mathcal{M}$ contains $\mathcal{B}^n$ by the $(\pi$-$\lambda)$ theorem, because $\mathcal{B}^n$ is the $\sigma$-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 $\mu$ 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 $\mu$ 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 $\mu$-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 $\mu$-measurable with $\mu(A) < +\infty$ and denote $\mu\lfloor A$ by $\nu$; then by Proposition 3.8.2, $\nu$ 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 $\mu$-measurable and

$$\mu(A) = \sum_j \mu(A_j).$$

Since $\mu$ 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

$$\sup\{\mu(K) : K \subset A,\ K \text{ is compact}\} \geq \sup\left\{\mu\left(\bigcup_{j=1}^{l} K_j\right) : l = 1, 2, \ldots\right\} = +\infty. \qquad \blacksquare$$

**Remark** Because of Theorem 3.9.1 (i), a set $E \subset \mathbb{R}^n$ is $\mu$-measurable if and only if $\mu(G) = \mu(G \cap E) + \mu(G \cap E^c)$ for all open sets $G$, where $\mu$ is a Radon measure on $\mathbb{R}^n$.

**Corollary 3.9.1** *The Lebesgue measure $\lambda^n$ is also the measure on $\mathbb{R}^n$ constructed by Method I from the premeasure $\tau$ on the family of all oriented closed cubes I, defined by $\tau(I)$ = volume of I.*

***Proof*** Let $\tau^*$ be the measure on $\mathbb{R}^n$ constructed from $\tau$ 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 $F_\sigma$ set $M \subset A$ with $\lambda^n(A \backslash M) = 0$ (a $F_\sigma$-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 $\mu$ be a premeasure defined for open sets $G$ in $\mathbb{R}^n$ by

$$\mu(G) = \int_G w d\lambda^n.$$

Denote by $\mu^*$ the measure on $\mathbb{R}^n$ constructed from $\mu$ 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 $\mu^*$ 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 $\mu$ is a measure on a metric space $X$ with the property that compact sets are $\mu$-measurable. Let $E \subset A$ be subsets of $X$ of which $E$ is not $\mu$-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 $\mu$ 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 $\mu$. 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 $\mu$ is a finite Riesz measure on $X$, then for any $\mu$-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 $\mu$, 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 $\mu$. 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 $\ell$ 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 $\ell$ on $C_c(X)$, a related measure $\mu$ on $X$ is constructed as follows. Define first a premeasure $\tau$ 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 $\mu(\bigcup_{j=1}^n G_j) =$

$\tau(\bigcup_{j=1}^n G_j) \leq \sum_{j=1}^n \mu(G_j)$. Thus the first part of (2) is verified. Now if $G_1, \ldots, G_n$ are disjoint, we need to show that $\mu(\bigcup_{j=1}^n G_j) \geq \sum_{j=1}^n \mu(G_j)$. For this purpose, since $\mu(\bigcup_{j=1}^n G_j) \geq \mu(G_j)$ 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

$$\mu\left(\bigcup_{j=1}^n G_j\right) = \tau\left(\bigcup_{j=1}^n G_j\right) \geq \ell(u) = \sum_{j=1}^n \ell(u_j) \geq \sum_{j=1}^n \mu(G_j) - \varepsilon,$$

from which $\mu(\bigcup_{j=1}^n G_j) \geq \sum_{j=1}^n \mu(G_j)$ follows by letting $\varepsilon \to 0$. Thus (2) is verified.

We show next that $\mu$ is a Carathéodory measure on $X$. Let $\{A_n\}$ be a sequence of subsets of $X$; we claim that $\mu(\bigcup_n A_n) \leq \sum_n \mu(A_n)$. 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),

$$\ell(u) \leq \mu\left(\bigcup_{j=1}^{n_0} G_j\right) \leq \sum_{j=1}^{n_0} \mu(G_j) \leq \sum_n \mu(G_n) \leq \sum_n \mu(A_n) + \varepsilon;$$

consequently, $\ell(u) \leq \sum_n \mu(A_n) + \varepsilon$ for each $u \in U_c(\bigcup_n G_n)$ and hence $\mu(\bigcup_n G_n) \leq \sum_n \mu(A_n) + \varepsilon$. Thus, $\mu(\bigcup_n A_n) \leq \mu(\bigcup_n G_n) \leq \sum_n \mu(A_n) + \varepsilon$. Since $\varepsilon > 0$ is arbitrary, $\mu(\bigcup_n A_n) \leq \sum_n \mu(A_n)$. As $\mu(\emptyset) = 0$ and $\mu(A) \leq \mu(B)$ for $A \subset B$ are direct consequences of the definition of $\mu$, $\mu$ 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

$$\mu(G) \geq \mu(G \cap (H_1 \cup H_2)) = \mu(G_1) + \mu(G_2) \geq \mu(A) + \mu(B),$$

and consequently, $\mu(A \cup B) \geq \mu(A) + \mu(B)$, or $\mu(A \cup B) = \mu(A) + \mu(B)$. Thus, $\mu$ is a Carathéodory measure on $X$. The measure $\mu$ so constructed will be referred to as the **measure** for the positive linear functional $\ell$.

**Lemma 3.10.1** *Suppose that $\ell$ is a positive linear functional on $C_c(X)$ and let $\mu$ be the measure for $\ell$, then $\mu$ is a Radon measure on $X$.*

***Proof*** Since $\mu$ is a Carathéodory measure, it is a Borel measure. From the definition of $\mu$, 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 $\mu$ 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 $u \in U_c(\overset{\circ}{V})$, then $u \leq f$. Thus $\mu(K) \leq \mu(\overset{\circ}{V}) = \sup\{\ell(u) : u \in U_c(\overset{\circ}{V})\} \leq \ell(f) < \infty$. We have shown that $\mu$ is a Radon measure on $X$. ■

**Lemma 3.10.2** *Suppose that $\ell$ is a positive linear functional on $C_c(X)$ and $\mu$ is the measure for $\ell$. 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, $f = \sum_{|j|\leq k} f u_j$ 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 $\mu$, the Radon measure $\mu$ 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 $\mu$ for a positive linear functional $\ell$ 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 $\mu$ for $\ell$ is a Riesz measure on $X$ for which (3.6) holds, it remains to show that if $\nu$ 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 $\nu$ and $\mu$ 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 $\ell$ is a bounded positive linear functional on $C_c(X)$ as a n.v.s. with the norm previously defined, then the measure $\mu$ for $\ell$ 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 $\ell$ be a positive linear functional on $C[0, 1]$ and let $\mu$ be the measure for $\ell$. 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 $\mu_g$ is $\mu$.

## 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 $\alpha$. 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 $\alpha$ 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 $\alpha$ 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 $\alpha$ 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 $\alpha$, 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 $\tau$ 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.