2 Measure Theory

Measure Theory is the part of mathematical analysis that deals with the development of a precise way to measure large classes of sets and how to integrate functions. It started at the end of the 19th century with the works of Jordan, Borel, Young, and Lebesgue. By that time it was evident that the Riemann integral had serious limitations and had to be replaced by a new integral that was more general (that is, more functions could be integrated) and more flexible (that is, it led to more efficient calculus rules and in particular convergence theorems). The construction of Lebesgue turned out to be extremely fruitful and launched "Measure Theory." The idea of Lebesgue to partition the $f(x)$-axis (instead of the x-axis as is done in the Riemann integral) was a remarkable conceptual insight, which allowed the full power of measure theory to reveal itself. In this chapter we present some basic aspects of this theory, which are needed to deal with the topics that follow.

2.1 Basic Notions, Measures, and Outer Measures

We start by defining algebras and σ-algebras. These are families of subsets of a given set. On σ-algebras, the theory exhibits its full strength.

Definition 2.1.1. Let X be a set and $\mathcal{L} \subseteq 2^X$ a nonempty family of subsets.
(a) We say that $\mathcal{L}$ is an **algebra** (or a **field**) if $A, B \in \mathcal{L}$ implies $A \cup B \in \mathcal{L}$ and $A^c = X \setminus A \in \mathcal{L}$. That is, $\mathcal{L}$ is closed under finite unions and complementation.
(b) We say that $\mathcal{L}$ is a **σ-algebra** (or a **σ-field**) if $\mathcal{L}$ is an algebra and it is closed under countable unions, that is, if $\{A_n\}_{n\geq 1} \subseteq \mathcal{L}$, then $\bigcup_{n\geq 1} A_n \in \mathcal{L}$.

Remark 2.1.2. Note that if $\mathcal{L}$ is an algebra, then $\emptyset, X \in \mathcal{L}$. Indeed, let $A \in \mathcal{L}$. Then $A^c \in \mathcal{L}$ and so $X = A \cup A^c \in \mathcal{L}$. Hence $\emptyset = X^c \in \mathcal{L}$. Moreover, by de Morgan's law, every algebra (resp. σ-algebra) is closed under finite (resp. countable) intersections. If $E \subseteq X$, then the **restriction** (or **trace**) of $\mathcal{L}$ on E is defined by $\mathcal{L}_E = \{E \cap A : A \in \mathcal{L}\}$.

Example 2.1.3. (a) There are two extreme cases: $\mathcal{L}_1 = \{\emptyset, X\}$ and $\mathcal{L}_2 = 2^X$. Both are σ-algebras with $\mathcal{L}_1$ being the smallest with respect to inclusion and $\mathcal{L}_2$ being the greatest one.
(b) Let $X = [0, 1)$ and let $\mathcal{L}$ be the finite union of intervals $[a, b) \subseteq [0, 1)$. Then $\mathcal{L}$ is an algebra but not an σ-algebra since $E = \bigcap_{n\geq 1}[0, 1/n) = \{0\} \notin \mathcal{L}$.

Evidently the intersection of σ-algebras is again a σ-algebra. This leads to the following definitions.

https://doi.org/10.1515/9783110532982-002

Definition 2.1.4. (a) Let X be a set and let $\mathcal{F} \subseteq 2^X$ be nonempty. The σ-algebra generated by $\mathcal{F}$, denoted by $\sigma(\mathcal{F})$, is defined by

$$\sigma(\mathcal{F}) = \bigcap \left\{\mathcal{L} \subseteq 2^X : \mathcal{F} \subseteq \mathcal{L}, \mathcal{L} \text{ is a } \sigma\text{-algebra}\right\} .$$

(b) Let (X, τ) be a Hausdorff topological space. The **Borel σ-algebra** is defined by $\mathcal{B}(X) = \sigma(\tau)$.

As we will see later in our discussion of measures it is often more convenient to start with families that have less structure than σ-algebras and eventually pass to the σ-algebra they generate.

Definition 2.1.5. Let X be a set and let $\mathcal{L} \subseteq 2^X$ be a nonempty family of subsets.
(a) We say that $\mathcal{L}$ is a **ring** if $A, B \in \mathcal{L}$ implies $A \cup B \in \mathcal{L}$ and $A \setminus B \in \mathcal{L}$. That is, $\mathcal{L}$ is closed under finite unions and relative complementation.
(b) We say that $\mathcal{L}$ is a **σ-ring** if $\mathcal{L}$ is a ring and it is closed under countable unions, that is, if $\{A_n\}_{n\ge1} \subseteq \mathcal{L}$, then $\bigcup_{n\ge1} A_n \in \mathcal{L}$.
(c) We say that $\mathcal{L}$ is a **semiring** if the following hold:
 (i) $\emptyset \in \mathcal{L}$;
 (ii) $A, B \in \mathcal{L}$ implies $A \cap B \in \mathcal{L}$;
 (iii) $A, B \in \mathcal{L}$ implies $A \setminus B = \bigcup_{k=1}^n C_k$ for some $n \in \mathbb{N}$ and disjoint $\{C_k\}_{k=1}^n \subseteq \mathcal{L}$.

Remark 2.1.6. Note that if $\mathcal{L}$ is a ring and $A \in \mathcal{L}$, then $\emptyset = A \setminus A \in \mathcal{L}$. So, the empty set is always an element of a ring. Hence if $\mathcal{L}$ is a ring and $X \in \mathcal{L}$, then $\mathcal{L}$ is an algebra. Thus we see that the collection of all finite subsets of X is a ring but not an algebra unless X is a finite set. On the other hand the collection of all finite subsets of X and of their complements is an algebra but not a σ-algebra unless X is a finite set. If $\mathcal{L}$ is a ring and $A, B \in \mathcal{L}$, then $A \cap B = A \setminus (A \setminus B) \in \mathcal{L}$. So, a ring is also closed under finite intersections. Similarly $A \Delta B = (A \setminus B) \cup (B \setminus A) \in \mathcal{L}$ and so a ring is also closed under symmetric differences.

We have the following relations among the notions introduced thus far:

$$\begin{tikzcd}
\boxed{\sigma\text{-algebra}} \arrow[r] \arrow[dr] & \boxed{\sigma\text{-ring}} \arrow[r] & \boxed{\text{ring}} \arrow[r] & \boxed{\text{semiring}} \\
 & \boxed{\text{algebra}} \arrow[ur] & &
\end{tikzcd}$$

Apart from trivial cases, $\sigma(\mathcal{L})$ (see Definition 2.1.4(a)) cannot be constructively obtained from $\mathcal{L}$. In order to overcome this difficulty, we introduce the following notions.

Definition 2.1.7. Let X be a set and $\mathcal{D} \subseteq 2^X$. We say that $\mathcal{D}$ is a **Dynkin system** (or a **λ-system**) if the following conditions hold:

(i) $X \in \mathcal{D}$;
(ii) $A, B \in \mathcal{D}$ with $B \subseteq A$ implies $A \setminus B \in \mathcal{D}$;
(iii) $\{A_n\}_{n\geq 1} \subseteq \mathcal{D}$ increasing implies $A = \bigcup_{n\geq 1} A_n \in \mathcal{D}$.

Remark 2.1.8. Evidently (ii) implies that $\emptyset$ is in every Dynkin system and $\{\emptyset, X\}$ as well as 2^X are both Dynkin systems. Consider also the following conditions on the family $\mathcal{D} \subseteq 2^X$:
(iv) $A \in \mathcal{D}$ implies $A^c \in \mathcal{D}$;
(v) for every disjoint sequence $\{A_n\}_{n\geq 1} \subseteq \mathcal{D}$ we have $\bigcup_{n\geq 1} A_n \in \mathcal{D}$.
It is easy to show that $\mathcal{D}$ is a Dynkin system if and only if (i), (iv), and (v) hold if and only if (i), (ii), and (v) hold.

Definition 2.1.9. Let X be a set and $\mathcal{L} \subseteq 2^X$ a nonempty family of subsets of X. We say that $\mathcal{L}$ is a **monotone class** if $\{A_n\}_{n\geq 1} \subseteq \mathcal{L}$ is increasing or decreasing, then

$$A = \bigcup_{n\geq 1} A_n \in \mathcal{L} \quad \text{or} \quad A = \bigcap_{n\geq 1} A_n \in \mathcal{L} \,.$$

Remark 2.1.10. Any σ-algebra is a monotone class but a topology is not in general. Of course 2^X is always a monotone class and the intersection of a family of monotone classes is a monotone class. So, there is a smallest monotone class containing a nonempty family $\mathcal{L} \subseteq 2^X$. A monotone class that is also an algebra is also a σ-algebra.

The next result is known as the "Dynkin System Theorem." The name "Dynkin's π–λ Theorem" can be also found in the literature.

Theorem 2.1.11 (Dynkin System Theorem). *If X is a set, $\mathcal{L} \subseteq 2^X$ is a nonempty family of subsets that is closed under finite intersections, and $\mathcal{D}$ is a Dynkin system such that $\mathcal{D} \supseteq \mathcal{L}$, then $\mathcal{D} \supseteq \sigma(\mathcal{L})$.*

Proof. Let $\mathcal{D}_0$ be the smallest Dynkin system containing $\mathcal{L}$. Evidently $\mathcal{D}_0 \subseteq \mathcal{D}$. Moreover, $\sigma(\mathcal{L})$ is a Dynkin system. So, we also have $\mathcal{D}_0 \subseteq \sigma(\mathcal{L})$. Let

$$\mathcal{R} = \{A \in \mathcal{D}_0 \colon A \cap B \in \mathcal{D}_0 \text{ for every } B \in \mathcal{L}\} \,.$$

Since $\mathcal{L}$ is closed under finite intersections we have $\mathcal{L} \subseteq \mathcal{R}$ and since $\mathcal{D}_0$ is a Dynkin system, we have that $\mathcal{R}$ is a Dynkin system as well. Therefore

$$\mathcal{D}_0 = \mathcal{R} \,. \tag{2.1.1}$$

Let $\mathcal{R}' = \{E \in \mathcal{D}_0 \colon E \cap D \in \mathcal{D}_0 \text{ for all } D \in \mathcal{D}_0\}$. Because of (2.1.1), it holds that $\mathcal{D}_0 = \mathcal{R}$ and so we have that $\mathcal{L} \subseteq \mathcal{R}'$, and $\mathcal{R}'$ is a Dynkin system. Hence, $\mathcal{D}_0 = \mathcal{R}'$, which means that $\mathcal{D}_0$ is closed under finite intersections. Thus, $\mathcal{D}_0$ is a σ-algebra; see Remark 2.1.8. Hence,

$$\sigma(\mathcal{L}) = \mathcal{D}_0 \subseteq \mathcal{D} \,. \qquad \square$$

Monotone classes are closely related to σ-algebras and by Theorem 2.1.11 are also related to Dynkin systems. The next result illustrates this and is known as the "Monotone Class Theorem."

Theorem 2.1.12 (Monotone Class Theorem). *If X is a set, $\mathcal{L} \subseteq 2^X$ is an algebra and $\mathcal{M} \subseteq 2^X$ is a nonempty, monotone class such that $\mathcal{M} \supseteq \mathcal{L}$, then $\mathcal{M} \supseteq \sigma(\mathcal{L})$.*

Proof. Let $\Sigma = \sigma(\mathcal{L})$ and let $\mathcal{M}_0$ be the smallest monotone class containing $\mathcal{L}$. Evidently $\mathcal{M}_0 \subseteq \mathcal{M}$. If we show that $\Sigma = \mathcal{M}_0$, then we are done.

To this end, we fix $A \in \mathcal{M}_0$ and let

$$\mathcal{M}_0^A = \{B \in \mathcal{M}_0 : A \cap B, B \setminus A \in \mathcal{M}_0\} .$$

Then $\mathcal{M}_0^A$ is a monotone class. If $A \in \mathcal{L}$, then since $\mathcal{L}$ is an algebra, we have $\mathcal{M}_0 \subseteq \mathcal{M}_0^A$, hence $\mathcal{M}_0 = \mathcal{M}_0^A$. So, for any $B \in \mathcal{M}_0$ we have

$$A \cap B, A \setminus B, B \setminus A \in \mathcal{M}_0 \quad \text{for any } A \in \mathcal{L} .$$

Thus, $\mathcal{L} \subseteq \mathcal{M}_0^B$, which implies $\mathcal{M}_0 = \mathcal{M}_0^B$.

Then we see that $\mathcal{M}_0$ is an algebra and so it follows that $\mathcal{M}_0$ is a σ-algebra; see Remark 2.1.10. It follows that $\Sigma \subseteq \mathcal{M}_0$ and because Σ is also a monotone class containing $\mathcal{L}$ we conclude that $\Sigma = \mathcal{M}_0 \subseteq \mathcal{M}$. ☐

Remark 2.1.13. From the proof above we see that if $\mathcal{L} \subseteq 2^X$ is an algebra, then $\sigma(\mathcal{L})$ coincides with the smallest monotone class generated by $\mathcal{L}$. Therefore, the algebra $\mathcal{L}$ is a monotone class if and only if $\mathcal{L}$ is a σ-algebra.

Since the Borel σ-algebra (see Definition 2.1.4(b)) is an important σ-algebra, we state some easy but useful facts concerning its generation. The first result is an immediate consequence of Theorem 2.1.11.

Proposition 2.1.14. *If X is a Hausdorff topological space, then the Borel σ-Algebra is the smallest Dynkin system containing the open sets or the closed sets.*

In the context of metric spaces we can state a little different characterization of the Borel sets.

Proposition 2.1.15. *If X is a metrizable space, then the Borel σ-Algebra $\mathcal{B}(X)$ is the smallest family of subsets of X that includes the open sets and it is closed under countable intersections and under countable disjoint unions.*

Proof. From Proposition 1.5.8 we know that every closed set is G_δ. Hence, every family of sets that contains the open sets and is also closed under countable intersections, must contain the closed sets. Then the result follows from Problem 2.1. ☐

For a similar result for families containing the closed sets, we need to require that we have closure under arbitrary unions, not just disjoint ones.

Proposition 2.1.16. *If X is a metrizable space, then the Borel σ-Algebra $\mathcal{B}(X)$ is the smallest family of subsets of X that includes the closed sets and it is closed under countable intersections and under countable unions.*

Proof. Recall again from Proposition 1.5.8 that every open set is F_σ. Hence every family of sets that contains the closed sets and is closed under countable unions, must contain the open sets as well. Again an appeal to Problem 2.1 concludes the proof. $\square$

Remark 2.1.17. In a Hausdorff topological space the closure of any set belongs to the Borel σ-algebra being closed. Similarly for the interior of any set being open and the boundary of any set being closed. Recalling that singletons are closed sets, we infer that countable sets are Borel. Finally, compact sets are also Borel being closed.

For the real line $\mathbb{R}$ we can choose among many different generators of the Borel σ-algebra. So let

$$\begin{aligned} &\mathcal{L}_1 = \{(a,b) : a < b\}, && \mathcal{L}_2 = \{[a,b) : a < b\}, && \mathcal{L}_3 = \{(a,b] : a < b\}, \\ &\mathcal{L}_4 = \{[a,b] : a < b\}, && \mathcal{L}_5 = \{(a,\infty) : a \in \mathbb{R}\}, && \mathcal{L}_6 = \{(-\infty,b) : b \in \mathbb{R}\}, \\ &\mathcal{L}_7 = \{[a,+\infty) : a \in \mathbb{R}\}, && \mathcal{L}_8 = \{(-\infty,b] : b \in \mathbb{R}\}, && \mathcal{L}_9 = \text{ open sets of } \mathbb{R}, \\ &\mathcal{L}_{10} = \text{ closed sets of } \mathbb{R}. \end{aligned}$$

Moreover, by $\mathcal{L}_k^r$, $k \in \{1,\dots,8\}$ we denote the collection of intervals in $\mathcal{L}_k$ with rational endpoints.

The next result is straightforward.

Proposition 2.1.18. $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{L}_k)$ *for all* $k \in \{1,\dots,10\}$ *and* $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{L}_k^r)$ *for all* $k \in \{1,\dots,8\}$.

In many cases we will deal with the extended real line $\mathbb{R}^* = \mathbb{R} \cup \{\pm\infty\}$. In this case we have the following.

Definition 2.1.19. It holds that $\mathcal{B}(\mathbb{R}^*) = \sigma\Big(\mathcal{B}(\mathbb{R}) \cup \{\{+\infty\},\{-\infty\}\}\Big)$.

Remark 2.1.20. Evidently $\mathcal{B}(\mathbb{R}^*) = \{$the $\mathcal{B}(\mathbb{R})$-sets or the $\mathcal{B}(\mathbb{R})$-sets with $+\infty$ or $-\infty$ or both attached to them$\}$.

From Proposition 2.1.18 and Definition 2.1.19 we obtain the following.

Proposition 2.1.21. *It holds that* $\operatorname{card}(\mathcal{B}(\mathbb{R})) = \operatorname{card}(\mathcal{B}(\mathbb{R}^*)) = \mathfrak{c}$ *being the cardinality of the continuum.*

Now we pass to set functions.

Definition 2.1.22. Let X be a set, $\emptyset \in \mathcal{L} \subseteq 2^X$ and $\mu : \mathcal{L} \to \mathbb{R}^*$ is a set function.

(a) We say that μ is **monotone** if

$$A \subseteq B \text{ with } A, B \in \mathcal{L} \text{ implies } \mu(A) \le \mu(B).$$

(b) We say that μ is **additive** (or **finitely additive**) if $\{A_k\}_{k=1}^n \subseteq \mathcal{L}$ are pairwise disjoint and $\bigcup_{k=1}^n A_k \in \mathcal{L}$ implies $\mu(\bigcup_{k=1}^n A_k) = \sum_{k=1}^n \mu(A_k)$.

(c) We say that μ is σ-**additive** (or **countably additive**) if $\{A_k\}_{k\ge1} \subseteq \mathcal{L}$ are pairwise disjoint and $\bigcup_{k\ge1} A_k \in \mathcal{L}$ implies $\mu(\bigcup_{k\ge1} A_k) = \sum_{k\ge1} \mu(A_k)$.

(d) We say that μ is **subadditive** if $\{A_k\}_{k=1}^n \subseteq \mathcal{L}$ and $\bigcup_{k=1}^n A_k \in \mathcal{L}$ imply $\mu(\bigcup_{k=1}^n A_k) \leq \sum_{k=1}^n \mu(A_k)$.
(e) We say that μ is **σ-subadditive** if $\{A_k\}_{k\geq 1} \subseteq \mathcal{L}$ and $\bigcup_{k\geq 1} A_k \in \mathcal{L}$ imply $\mu(\bigcup_{k\geq 1} A_k) \leq \sum_{k\geq 1} \mu(A_k)$.
(f) When $\mathcal{L} = \Sigma$ is a σ-algebra, then we say that the set function $\mu\colon \Sigma \to \mathbb{R}^* = \mathbb{R}\cup\{\pm\infty\}$ is a **signed-measure** if it takes only one of the values $+\infty$ and $-\infty$, $\mu(\emptyset) = 0$, and it is σ-additive. If μ takes only nonnegative values, then we say that μ is a **measure**.
(g) A pair (X, Σ) with X being a set and $\Sigma \subseteq 2^X$ being a σ-algebra is said to be a **measurable space**. If μ is a measure on (X, Σ), then (X, Σ, μ) is said to be a **measure space**. We say that μ is **finite** (or that the measure space (X, Σ, μ) is finite) if $\mu(X) < \infty$. We say that μ is **σ-finite** if $X = \bigcup_{n\geq 1} X_n$ with $X_n \in \Sigma$ and $\mu(X_n) < +\infty$ for all $n \in \mathbb{N}$.

Example 2.1.23. (a) Let X be a nonempty set and $\Sigma = 2^X$. The set function $\mu\colon \Sigma \to [0, +\infty]$ defined by

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

is a measure known as the **counting measure**. If X is finite (resp. countable), then $\mu\colon \Sigma \to [0, +\infty]$ is finite (resp. σ-finite). More generally, let $f\colon X \to [0, +\infty)$ be a function and define $\mu\colon 2^X \to [0, +\infty]$ by setting

$$\mu(A) = \sum_{x\in A} f(x) = \sup \left[\sum_{x\in F} f(x) \colon F \subseteq A \text{ is finite} \right] .$$

Then $\mu\colon 2^X \to [0, +\infty]$ is a measure that is σ-finite if $\{x \in X\colon f(x) > 0\}$ is countable. Evidently, if $f(x) = 1$ for all $x \in X$, then we have the counting measure. If $f(x_0) = 1$ and $f(x) = 0$ if $x \neq x_0$, then $\mu\colon 2^X \to [0, +\infty]$ is called the **Dirac measure** at x_0 and is denoted by δ_{x_0}.
(b) Let X be an uncountable set and let

$$\Sigma = \left\{A \subseteq X\colon A \text{ is countable or } A^c \text{ is countable}\right\} .$$

Then Σ is a σ-algebra being the σ-algebra of countable or co-countable sets. The set function $\mu\colon \Sigma \to [0, 1]$ defined by

$$\mu(A) = \begin{cases} 0 & \text{if } A \text{ is countable} , \\ 1 & \text{if } A^c \text{ is countable, that is, } A \text{ is co-countable} \end{cases}$$

is a finite measure.

The next proposition summarizes the main properties of measures.

Proposition 2.1.24. *Let (X, Σ, μ) be a measure space. Then the following hold:*
(a) $\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$ *for all* $A, B \in \Sigma$.
(b) $\mu(A) = \mu(B) + \mu(A \setminus B)$ *for all* $A, B \in \Sigma$ *with* $B \subseteq A$.
(c) $\mu(B) \leq \mu(A)$ *for all* $A, B \in \Sigma$ *with* $B \subseteq A$ *(monotonicity).*

(d) $\mu(\bigcup_{k\geq 1} A_k) \leq \sum_{k\geq 1} \mu(A_k)$ *for all* $\{A_k\}_{k\geq 1} \subseteq \Sigma$ *(σ-subadditivity).*
(e) *If* $\{A_k\}_{k\geq 1} \subseteq \Sigma$ *is increasing, then* $\mu(\bigcup_{k\geq 1} A_k) = \lim_{k\to\infty} \mu(A_k)$ *(continuity from below).*
(f) *If* $\{A_k\}_{k\geq 1} \subseteq \Sigma$ *is decreasing and* $\mu(A_1) < +\infty$*, then* $\mu(\bigcap_{k\geq 1} A_k) = \lim_{k\to\infty} \mu(A_k)$ *(continuity from above).*

Proof. (a) By additivity we have

$$\mu(A) = \mu(A \cap B) + \mu(A \setminus B) \quad \text{and} \quad \mu(B) = \mu(A \cap B) + \mu(B \setminus A)\,.$$

Adding these two equations gives

$$\begin{aligned}\mu(A) + \mu(B) &= \mu(A \cap B) + [\mu(A \cap B) + \mu(A \setminus B) + \mu(B \setminus A)] \\ &= \mu(A \cap B) + \mu(A \cup B)\end{aligned}$$

again by the additivity.

(b) Let $A = B \cup (A \setminus B)$ and use the additivity we obtain $\mu(A) = \mu(B) + \mu(A \setminus B)$.

(c) Since μ is nonnegative, the assertion follows from (b).

(d) Let $B_1 = A_1$ and $B_k = A_k \setminus \bigcup_{i=1}^{k-1} A_i$ for $k \geq 2$. Then the sets $\{B_k\}_{k\geq 1}$ are disjoint and $\bigcup_{k\geq 1} B_k = \bigcup_{k\geq 1} A_k$. Then, taking the σ-additivity and part (c) into account it follows

$$\mu\left(\bigcup_{k\geq 1} A_k\right) = \mu\left(\bigcup_{k\geq 1} B_k\right) = \sum_{k\geq 1} \mu(B_k) \leq \sum_{k\geq 1} \mu(A_k)\,.$$

(e) Let $A_0 = \emptyset$. Then

$$\mu\left(\bigcup_{k\geq 1} A_k\right) = \sum_{k\geq 1} \mu(A_k \setminus A_{k-1}) = \lim_{n\to\infty} \sum_{k=1}^{n} \mu(A_k \setminus A_{k-1}) = \lim_{n\to\infty} \mu(A_n)\,.$$

(f) Let $B_k = A_1 \setminus A_k$. Then $\{B_k\}_{k\geq 1} \subseteq \Sigma$ is increasing, $\mu(A_1) = \mu(A_k) + \mu(B_k)$ for all $k \in \mathbb{N}$, see part (b), and $\bigcup_{k\geq 1} B_k = A_1 \setminus \bigcap_{k\geq 1} A_k$. By parts (e) and (b) there holds

$$\mu(A_1) = \mu\left(\bigcap_{k\geq 1} A_k\right) + \lim_{k\to\infty} \mu(B_k) = \mu\left(\bigcap_{k\geq 1} A_k\right) + \lim_{k\to\infty} [\mu(A_1) - \mu(A_k)]\,.$$

Hence, subtracting $\mu(A_1) < \infty$ from both sides gives $\mu\left(\bigcap_{k\geq 1} A_k\right) = \lim_{k\to\infty} \mu(A_k)$. ☐

Remark 2.1.25. Clearly, the condition $\mu(A_1) < +\infty$ in Proposition 2.1.24(f) can be replaced by the hypothesis that $\mu(E_n) < +\infty$ for some $n \in \mathbb{N}$ since the first $(n-1)$ sets do not affect the intersection.

It turns out that continuity from below (see Proposition 2.1.24(e)) for an additive set function is equivalent to σ-additivity.

Proposition 2.1.26. *If X is a set, $\mathcal{L} \subseteq 2^X$ is an algebra of sets in X and $\mu\colon \mathcal{L} \to [0, +\infty]$ is an additive set function, then μ is σ-additive if and only if μ is continuous from below, that is, if $\{A_n\}_{n\geq 1} \subseteq \mathcal{L}$ is increasing, $\bigcup_{n\geq 1} A_n \in \mathcal{L}$, then $\mu(\bigcup_{n\geq 1} A_n) = \lim_{n\to\infty} \mu(A_n)$.*

Proof. $\Longrightarrow$: This follows from the proof of Proposition 2.1.24(e).

$\Longleftarrow$: Suppose we have continuity from below. Let $\{B_k\}_{k\geq 1} \subseteq \mathcal{L}$ be a sequence of pairwise disjoint sets such that $\bigcup_{k\geq 1} B_k \in \mathcal{L}$. We set $A_n = \bigcup_{k=1}^n B_k$. From the continuity from below hypothesis, it follows

$$\mu\left(\bigcup_{k\geq 1} B_k\right) = \mu\left(\bigcup_{k\geq 1} A_k\right) = \lim_{n\to\infty} \mu(A_n) = \lim_{n\to\infty} \sum_{k=1}^{n} \mu(B_k) = \sum_{k\geq 1} \mu(B_k)\,.$$

This shows that $\mu\colon \mathcal{L} \to [0, +\infty]$ is σ-additive. □

We get a similar result when we suppose continuity from above at the empty set.

Proposition 2.1.27. *If X is a set, $\mathcal{L} \subseteq 2^X$ is an algebra of sets in X and $\mu\colon \mathcal{L} \to [0, +\infty]$ is an additive set function with $\mu(X) < +\infty$, then μ is σ-additive if and only if μ is continuous from above at the empty set, that is, if $\{A_k\}_{k\geq 1} \subseteq \mathcal{L}$ is a decreasing sequence such that $\bigcap_{k\geq 1} A_k = \emptyset$, then $\lim_{k\to\infty} \mu(A_k) = 0$.*

Proof. $\Longrightarrow$: This implication follows again from the proof of Proposition 2.1.24(f).

$\Longleftarrow$: Let $\{A_k\}_{k\geq 1} \subseteq \mathcal{L}$ be an increasing sequence such that $\bigcup_{k\geq 1} A_k \in \mathcal{L}$. Let $B_n = (\bigcup_{k\geq 1} A_k)\setminus A_n$ for all $n \in \mathbb{N}$. Then $\{B_n\}_{n\geq 1} \subseteq \mathcal{L}$ is decreasing and $\bigcap_{n\geq 1} B_n = \emptyset$. Therefore, by hypothesis, we have

$$0 = \lim_{n\to\infty} \mu(B_n) = \mu\left(\bigcup_{k\geq 1} A_k\right) - \lim_{n\to\infty} \mu(A_n)\,.$$

Hence, $\mu\left(\bigcup_{k\geq 1} A_k\right) = \lim_{k\to\infty} \mu(A_k)$ and so μ is continuous from below. Then Proposition 2.1.26 implies that μ is σ-additive. □

The next result gives a necessary and sufficient condition for two finite measures to be equal. It suffices to know that they coincide on a generating family that is closed under finite intersections.

Proposition 2.1.28. *If (X, Σ) is a measurable space, $\Sigma = \sigma(\mathcal{L})$ with $\mathcal{L}$ closed under finite intersections, μ_1, μ_2 are two finite measures on Σ and $\mu_1(X) = \mu_2(X)$ as well as $\mu_1|_{\mathcal{L}} = \mu_2|_{\mathcal{L}}$, then $\mu_1 = \mu_2$.*

Proof. Let $\mathcal{D} = \{A \in \Sigma\colon \mu_1(A) = \mu_2(A)\}$. Applying Proposition 2.1.24(b) and (c), we see that $\mathcal{D}$ is a Dynkin system; see Definition 2.1.7. Moreover, by hypothesis, $\mathcal{L} \subseteq \mathcal{D}$. Then, invoking Theorem 2.1.11, we infer that $\Sigma = \sigma(\mathcal{L}) = \mathcal{D}$, which means that $\mu_1 = \mu_2$. □

Corollary 2.1.29. *If X is a Hausdorff topological space, $\mathcal{B}(X)$ is its Borel σ-field and μ_1, μ_2 are two finite measures on $\mathcal{B}(X)$, which coincide on the open or closed sets, then $\mu_1 = \mu_2$.*

In the next definition we introduce a notion that will lead us to a property reminiscent of the intermediate value property.

Definition 2.1.30. Let (X, Σ, μ) be a measure space.
(a) We say that the measure $\mu\colon \Sigma \to [0, +\infty]$ is **semifinite** if for every $A \in \Sigma$ with $\mu(A) > 0$, there exists $B \in \Sigma$ with $B \subseteq A$ such that $0 < \mu(B) < +\infty$.
(b) We say that $A \in \Sigma$ is an **atom** of μ if $0 < \mu(A) < +\infty$ and for every $B \subseteq A$ with $B \in \Sigma$ either $\mu(B) = 0$ or $\mu(B) = \mu(A)$. A measure without any atoms is called **nonatomic**.

Remark 2.1.31. The measure μ on Σ is nonatomic if for every set $A \in \Sigma$ with $\mu(A) > 0$, there exists $B \in \Sigma$ with $B \subseteq A$ such that $0 < \mu(B) < \mu(A)$. For the Dirac measure

$$\delta_{x_0}(A) = \begin{cases} 1 & \text{if } x_0 \in A\,, \\ 0 & \text{otherwise}\,, \end{cases} \qquad \text{with } x_0 \in X, A \in \Sigma\,,$$

we see that $\{x_0\}$ is an atom. The main examples of atoms are singletons $\{x\}$ with positive measure.

Here is the result that recalls the intermediate value property.

Proposition 2.1.32. *If (X, Σ, μ) is a nonatomic measure space, then the range of μ is the interval $[0, \mu(X)]$.*

Proof. We fix $\lambda \in (0, \mu(X))$ and define $\mathcal{L} = \{A \in \Sigma\colon 0 < \mu(A) \le \lambda\}$. First we show that $\mathcal{L} \neq \emptyset$. The nonatomicity of μ implies the existence of $B \in \Sigma$ such that $0 < \mu(B) < \mu(X)$. The same argument (nonatomicity of μ) implies that we can find $E_1, E_2 \in \Sigma$ such that $B = E_1 \cup E_2, E_1 \cap E_2 = \emptyset$ and $\mu(E_1), \mu(E_2) \in (0, \mu(B))$. It follows that at least one of the sets E_1, E_2 satisfies $\mu(E_1) \in (0, 1/2\mu(B)]$. Proceeding inductively, suppose that we produced $E_1, \ldots, E_n \in \Sigma$ such that

$$\mu(E_n) \in \left(0, \frac{1}{2^n}\mu(B)\right]\,. \tag{2.1.2}$$

Applying again the nonatomicity of μ there exists $E_{n+1} \in \Sigma$ with $E_{n+1} \subseteq E_n$ such that $\mu(E_{n+1}) \in (0, 1/2\mu(E_n)]$. Evidently, because of (2.1.2) we have $\mu(E_{n+1}) \le 1/2^{n+1}\mu(B)$. Therefore, (2.1.2) holds for all $n \in \mathbb{N}$. Moreover, for a large enough $n \in \mathbb{N}$, we have $\mu(E_n) \le \lambda$. Hence, $E_n \in \mathcal{L}$ for a large enough $n \in \mathbb{N}$, thus yielding $\mathcal{L} \neq \emptyset$.

Next we show that there exists a Σ-set with measure equal to λ. To this end, let $D_0 = \emptyset$ and suppose that $D_n \in \Sigma$ is given. Let

$$\lambda_n = \sup\left[\mu(C)\colon C \in \Sigma, D_n \subseteq C, \mu(C) \le \lambda\right]\,.$$

Choose $D_{n+1} \in \Sigma$ such that

$$D_n \subseteq C_{n+1} \quad \text{and} \quad \lambda_n - \frac{1}{n} \le \mu(D_{n+1}) \le \lambda_n\,. \tag{2.1.3}$$

It holds $0 < \lambda_{n+1} \le \lambda_n \le \lambda$ and so $\lim_{n\to\infty} \lambda_n = \hat{\lambda}$ exists and $\hat{\lambda} \le \lambda$. We define

$$\hat{D} = \bigcup_{n\ge 1} D_n\,. \tag{2.1.4}$$

This implies, due to (2.1.3) and Proposition 2.1.24(e), that

$$\mu(\hat{D}) = \lim_{n\to\infty} \mu(D_n) = \hat{\lambda} \,. \tag{2.1.5}$$

We need to show that $\hat{\lambda} = \lambda$. If $\hat{\lambda} < \lambda$, then $\mu(X \setminus \hat{D}) = \mu(X) - \mu(\hat{D}) > \lambda - \hat{\lambda} > 0$; see Proposition 2.1.24(b). Reasoning as in the first part of the proof with X replaced by $X \setminus \hat{D}$ and λ replaced by $\lambda - \hat{\lambda} > 0$, we produce

$$C \in \Sigma\,, \quad C \subseteq X \setminus \hat{D} \quad \text{and} \quad 0 < \mu(C) < \lambda - \hat{\lambda} \,. \tag{2.1.6}$$

Then, the subadditivity yields $\hat{\lambda} = \mu(\hat{D}) < \mu(C \cup \hat{D}) \le \lambda$, which gives, because of (2.1.5) and (2.1.6), that $\lambda_n < \mu(C \cup \hat{D})$ for all sufficiently large $n \in \mathbb{N}$. But $D_n \subseteq C \cup \hat{D}$ for all $n \in \mathbb{N}$; see (2.1.4). This contradicts the definition of λ_n for large enough $n \in \mathbb{N}$. We conclude that $\hat{\lambda} = \lambda$ and the proof is finished. □

The notion of outer measure is an abstract generalization of the "outer area" when we apply the exhaustion method of Archimedes to calculate the area of a bounded region in $\mathbb{R}^2$.

Definition 2.1.33. Let X be a nonempty set and $\mu^* : 2^X \to [0, +\infty]$ be a set function. We say that μ^* is an **outer measure** if it satisfies the following conditions:
(a) $\mu^*(\emptyset) = 0$;
(b) μ^* is monotone, that is, $A \subseteq B$ implies $\mu^*(A) \le \mu^*(B)$;
(c) μ^* is σ-subadditive, that is, $\mu^*(\bigcup_{n\ge1} A_n) \le \sum_{n\ge1} \mu^*(A_n)$.
We say that the outer measure μ^* is finite (resp. σ-finite) if $\mu^*(X) < +\infty$ (resp. $X = \bigcup_{n\ge1} X_n$ and $\mu^*(X_n) < +\infty$ for all $n \in \mathbb{N}$).

A way to produce an outer measure is to start with a family of elementary sets on which a measure is naturally defined (for example intervals in $\mathbb{R}$ and rectangles in $\mathbb{R}^2$) and approximate any set from above by countable unions of such elementary sets. This process is formalized in the proposition that follows.

Proposition 2.1.34. *If X is a nonempty set, $\mathcal{L} \subseteq 2^X$ is such that $\emptyset, X \in \mathcal{L}$, $\vartheta\colon \mathcal{L} \to [0, +\infty]$ satisfies $\vartheta(\emptyset) = 0$ and for any $A \in \mathcal{L}$ we set*

$$\mu^*(A) = \inf\left[\sum_{n\ge1} \vartheta(E_n) \colon E_n \in \mathcal{L}, A \subseteq \bigcup_{n\ge1} E_n\right], \tag{2.1.7}$$

then μ^* is an outer measure.

Proof. First note that in (2.1.7) the infimum is taken over by a nonempty set since $A \subseteq X$ and by hypothesis, $X \in \mathcal{L}$. Moreover, $\mu^*(\emptyset) = 0$ and it is clear from (2.1.7) that $A \subseteq B$ implies $\mu^*(A) \le \mu^*(B)$. Finally we show the σ-additivity of μ^*. So, let $\{A_k\} \subseteq 2^X$ and $\varepsilon > 0$. For each $k \in \mathbb{N}$ we can find $\{E_n^k\}_{n\ge1} \subseteq \mathcal{L}$ such that

$$A_k \subseteq \bigcup_{n\ge1} E_n^k \quad \text{and} \quad \sum_{n\ge1} \vartheta\left(E_n^k\right) \le \mu^*(A_k) + \frac{\varepsilon}{2^k} \,.$$

Let $A = \bigcup_{k\geq 1} A_k$. Then we have

$$A \subseteq \bigcup_{k,n\geq 1} E_n^k \quad \text{and} \quad \sum_{k,n\geq 1} \vartheta\left(E_n^k\right) \leq \sum_{k\geq 1} \mu^*(A_k) + \varepsilon\,.$$

This gives, due to (2.1.7), $\mu^*(A) \leq \sum_{k\geq 1} \mu^*(A_k) + \varepsilon$. Letting $\varepsilon \searrow 0$, we conclude that μ^* is σ-subadditive. Therefore μ^* is an outer measure. □

Example 2.1.35. Let $f\colon \mathbb{R} \to \mathbb{R}$ be an increasing function. Let $\mathcal{L}$ be the family of all intervals $(a, b]$ with $a, b \in \mathbb{R}$ and set $\vartheta((a, b]) = f(b) - f(a)$. Then the conditions in Proposition 2.1.34 are satisfied and by applying (2.1.7) we can define an outer measure μ^*. This outer measure is called the **Lebesgue–Stieltjes outer measure** and if $f(x) = x$ for all $x \in \mathbb{R}$ it is called the **Lebesgue outer measure**. Note that

$$\mu^*((a, b]) = f(b) - \lim_{x\to a^+} f(x) \leq f(b) - f(a) = \vartheta((a, b])\,.$$

Thus, the inequality is strict at those points where f is not continuous from the right.

Now we will pass from outer measures to measures. Outer measures, although defined on the entire power set 2^X have the disadvantage that they are not σ-additive. However, when restricted to a particular subset of 2^X, they become σ-additive. In this direction we need the following remarkable definition due to Carathéodory.

Definition 2.1.36. Let X be a nonempty set and μ^* is an outer measure on 2^X. We say that $A \subseteq X$ is μ^***-measurable**, if $\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^c)$ for all $B \subseteq X$, that is, A splits additively all sets in X.

Remark 2.1.37. From Definition 2.1.33 we know that it holds that

$$\mu^*(B) \leq \mu^*(B \cap A) + \mu^*(B \cap A^c) \quad \text{for all } B \subseteq X\,,$$

due to the subadditivity property of the outer measure. In order to check the μ^*-measurability of a set $A \subseteq X$ it suffices to show that

$$\mu^*(B) \geq \mu^*(B \cap A) + \mu^*(B \cap A^c) \quad \text{for all } B \subseteq X \text{ with } \mu^*(B) < +\infty\,.$$

This definition of Carathéodory essentially says that the outer measure $\mu^*(A)$ of A is equal to its **inner measure** $\mu^*(X) - \mu^*(A^c)$. For this reason Definition 2.1.36 is the right one and leads to a σ-algebra on which μ^* is σ-additive, hence a measure. This is shown in the next theorem known as the "Carathéodory Theorem."

Theorem 2.1.38 (Carathéodory Theorem). *If X is a nonempty set and $\mu^*\colon 2^X \to [0, +\infty]$ is an outer measure, then the family Σ^* of all μ^*-measurable sets is a σ-algebra and $\mu = \mu^*|_{\Sigma^*}$ is a measure.*

Proof. The symmetric character of Definition 2.1.36 implies that Σ^* is closed under complementation.

Next let $A, E \in \Sigma^*$ and let $B \subseteq X$. We have

$$\begin{aligned}\mu^*(B) &= \mu^*(B \cap A) + \mu^*(B \cap A^c)\\ &= \mu^*(B \cap A \cap E) + \mu^*(B \cap A \cap E^c) + \mu^*(B \cap A^c \cap E) + \mu^*(B \cap A^c \cap E^c)\,.\end{aligned}$$

Note that $A \cup E = (A \cap E) \cup (A \triangle E) = (A \cap E) \cup (A \cap E^c) \cup (A^c \cap E)$. Hence, by the subadditivity,

$$\mu^*(B \cap (A \cup E)) \le \mu^*(B \cap A \cap E) + \mu^*(B \cap A \cap E^c) + \mu^*(B \cap A^c \cap E)\,.$$

This implies

$$\mu^*(B \cap (A \cup E)) + \mu^*(B \cap (A \cup E)^c) \le \mu^*(B)\,.$$

Hence, see Remark 2.1.37, $A \cup E \in \Sigma^*$ and thus, Σ^* is an algebra.

In addition, if $A, E \in \Sigma^*$ and $A \cap E = \emptyset$, then

$$\mu^*(A \cup E) = \mu^*((A \cup E) \cap A) + \mu^*((A \cup E) \cap A^c) = \mu^*(A) + \mu^*(E)$$

where we recall that $\mu^*(A \cap E) = 0$. This means that μ^* is additive on Σ^*.

Now we show that Σ^* is a σ-algebra. Let $\{A_n\}_{n\ge1} \subseteq \Sigma^*$ and let $D = \bigcup_{n\ge1} A_n$. Since from the first part of the proof, we have

$$D_k = \bigcup_{n=1}^{k} A_n \in \Sigma^* \quad \text{and} \quad D_k \setminus \bigcup_{n=1}^{k-1} A_n \in \Sigma^* \quad \text{for all } k \in \mathbb{N}\,,$$

without any loss of generality we may assume that the sets $\{A_n\}_{n\ge1} \subseteq \Sigma^*$ are mutually disjoint. For any $B \subseteq X$, since $D_n, A_n \in \Sigma^*$, we have for all $n \in \mathbb{N}$

$$\begin{aligned}\mu^*(B) &= \mu^*(B \cap D_n) + \mu^*(B \cap D_n^c)\\ &= \mu^*(B \cap A_n) + \mu^*\left(B \cap \left(\bigcup_{i\le n-1} A_i\right)\right) + \mu^*(B \cap D_n^c)\,.\end{aligned}$$

Then, by induction on $n \in \mathbb{N}$, we show that

$$\mu^*(B) = \sum_{i=1}^{n} \mu^*(B \cap A_i) + \mu^*(B \cap D_n^c) \ge \sum_{i=1}^{n} \mu^*(B \cap A_i) + \mu^*(B \cap D^c)$$

since μ^* is additive and since $D_n \subseteq D$ for all $n \in \mathbb{N}$. We let $n \to \infty$ and obtain

$$\mu^*(B) \ge \sum_{i\ge1} \mu^*(B \cap A_i) + \mu^*(B \cap D^c) \ge \mu^*(B \cap D) + \mu^*(B \cap D^c)$$

by the σ-subadditivity; see Definition 2.1.36. This implies that $D \in \Sigma^*$ (see Remark 2.1.37) and $\mu^*(B) = \sum_{i\ge1} \mu(B \cap A_i) + \mu(B \cap D^c)$.

Let $B = D \subseteq X$. Then $\mu^*(D) = \sum_{i\ge1} \mu^*(A_i)$ and so we conclude that Σ^* is a σ-algebra and $\mu = \mu^*|_{\Sigma^*}$ is a measure. □

Definition 2.1.39. Let (X, Σ, μ) be a measure space.
(a) A set $A \in \Sigma$ is said to be μ**-null** (or simply **null** if μ is clearly understood) if $\mu(A) = 0$.
(b) We say that μ is **complete** if Σ contains all subsets of null sets.

Remark 2.1.40. If A is μ-null and $B \subseteq A$, then $\mu(B) = 0$, provided $B \in \Sigma$. But in general it need not be the case that $B \in \Sigma$. For example this is the case with the Borel σ-algebra $\mathcal{B}(\mathbb{R})$. However, completeness can always be achieved by simply extending the domain of the measure. This is done in the next proposition whose proof is straightforward and so it is omitted.

Proposition 2.1.41. *If* (X, Σ, μ) *is a measure space,* $\mathcal{N} = \{D \in \Sigma\colon \mu(D) = 0\}$, $\Sigma_\mu = \{A \cup E\colon A \in \Sigma, E \subseteq D \in \mathcal{N}\}$ *and* $\overline{\mu}(A \cup E) = \mu(A)$ *for all* $A \cup E \in \Sigma_\mu$*, then* Σ_μ *is a* σ*-algebra and* $\overline{\mu}$ *is a complete measure on* Σ_μ.

Let (X, Σ^*, μ) be the measure space produced in Theorem 2.1.38.

Proposition 2.1.42. (X, Σ^*, μ) *is a complete measure space.*

Proof. Assume that $\mu^*(A) = 0$. Then, by the subadditivity, the monotonicity, and since $\mu^*(A) = 0$, for any $B \subseteq X$, we have

$$\mu^*(B) \leq \mu^*(B \cap A) + \mu^*(B \cap A^c) \leq \mu^*(B \cap A^c) \leq \mu^*(B) .$$

This gives $A \in \Sigma^*$ and so $\mu = \mu^*|_{\Sigma^*}$ is complete. □

Now let X be a set and let $\mathcal{L} \subseteq 2^X$ be a semiring. We consider a σ-additive set function $\mu\colon \mathcal{L} \to [0, +\infty]$. Applying Proposition 2.1.34, we can define the outer measure $\mu^*\colon 2^X \to [0, +\infty]$ corresponding to μ. It holds that $\mu^*(A) = \mu(A)$ for all $A \in \mathcal{L}$. We have the following result.

Proposition 2.1.43. *If* $\mathcal{D}$ *is a semiring satisfying* $\mathcal{L} \subseteq \mathcal{D} \subseteq \Sigma^*$*, then* μ^* *is the unique extension of* μ *to a* σ*-additive set function on* $\mathcal{D}$.

Proof. Let $\lambda\colon \mathcal{D} \to [0, +\infty]$ be a σ-additive extension of μ on $\mathcal{D}$ and let λ^* be the corresponding outer measure; see Proposition 2.1.34. If $A \subseteq X$ and $\{E_n\}_{n\geq 1} \subseteq \mathcal{L}$ are such that $A \subseteq \bigcup_{n\geq 1} E_n$, then

$$\lambda^*(A) \leq \sum_{n\geq 1} \lambda^*(E_n) = \sum_{n\geq 1} \lambda(E_n) = \sum_{n\geq 1} \mu(E_n) .$$

This implies

$$\lambda^*(A) \leq \mu^*(A) \quad \text{for every } A \subseteq X . \tag{2.1.8}$$

In order to show that $\lambda = \mu^*$ on $\mathcal{D}$, it suffices to show that $\mu^*(A) \leq \lambda(A)$ for all $A \in \mathcal{D}$ with $\mu^*(A) < +\infty$. Recall that μ is σ-additive. Fix $A \in \mathcal{D}$ with $\mu^*(A) < +\infty$ and $\varepsilon > 0$. Consider $\{E_n\}_{n\geq 1} \subseteq \mathcal{L}$ such that

$$A \subseteq \bigcup_{n\geq 1} E_n \quad \text{and} \quad \sum_{n\geq 1} \mu(E_n) \leq \mu^*(A) + \varepsilon ; \tag{2.1.9}$$

see Proposition 2.1.34. Taking Problem 2.2 into account we find pairwise disjoint $\{C_n\}_{n\geq 1} \subseteq \mathcal{L}$ such that

$$\hat{E} = \bigcup_{n\geq 1} E_n = \bigcup_{n\geq 1} C_n \in \sigma(\mathcal{D}) .$$

We know that $\mu^*|_{\sigma(\mathcal{D})}$ and $\lambda^*|_{\sigma(\mathcal{D})}$ are both measures that coincide with μ on $\mathcal{L}$. Therefore

$$\mu^*(\hat{E}) = \sum_{n\geq 1} \mu^*(C_n) = \sum_{n\geq 1} \mu(C_n) = \sum_{n\geq 1} \lambda(C_n) = \lambda^*(\hat{E}) . \tag{2.1.10}$$

Moreover, because of (2.1.8) and (2.1.9) as well as the σ-subadditivity of μ^* and since $\mu^*|_{\mathcal{L}} = \mu$, we have

$$\lambda^*(\hat{E} \setminus A) \leq \mu^*(\hat{E} \setminus A) = \mu^*(\hat{E}) - \mu^*(A) \leq \sum_{n\geq 1} \mu(E_n) - \mu^*(A) \leq \varepsilon . \tag{2.1.11}$$

Hence $\mu^*(A) \leq \mu^*(\hat{E}) = \lambda^*(\hat{E}) = \lambda(A) + \lambda^*(\hat{E} \setminus A) \leq \lambda(A) + \varepsilon$; see (2.1.10) and (2.1.11). Letting $\varepsilon \searrow 0$, we obtain $\mu^*(A) \leq \lambda(A)$. Therefore, $\lambda(A) = \mu^*(A)$ for all $A \in \mathcal{D}$. □

The Lebesgue measure on $\mathbb{R}$ was the starting point of "Measure Theory." So, let us look in some detail at how we can produce it using the previous abstract theory. To this end, we introduce

$$\mathcal{L} = \{(a, b] : a \leq b,\ a, b \in \mathbb{R}\}$$

with $(a, a] = \emptyset$. This is a semiring of subsets of $\mathbb{R}$. Let $\lambda : \mathcal{L} \to [0, +\infty]$ be the set function defined by $\lambda((a, b]) = b - a$. This set function is σ-additive and σ-finite. Using Proposition 2.1.43, we know that λ has a unique extension to $\Sigma^* = \Sigma_\lambda$ being the σ-field of λ^*-measurable sets; see Definition 2.1.36. We continue to denote this extension by λ. Then

- λ is the Lebesgue measure on $\mathbb{R}$.
- $\Sigma^* = \Sigma_\lambda$ is the σ-algebra of the Lebesgue measurable subsets of $\mathbb{R}$.

Note that λ is translation invariant, that is $\lambda(A) = \lambda(A + x)$ for all $A \in \Sigma_\lambda$ and for all $x \in \mathbb{R}$. Moreover, we have $\lambda(\theta A) = |\theta|\lambda(A)$ for all $A \in \Sigma_\lambda$ and for all $\theta \in \mathbb{R}$.

From the previous discussion it is not clear if $\Sigma_\lambda = 2^{\mathbb{R}}$. In fact the next theorem shows that this is not the case. Indeed there are subsets of $\mathbb{R}$ that are not Lebesgue measurable.

Theorem 2.1.44. *There is no translation invariant measure defined on all of $2^{\mathbb{R}}$, which assigns to every interval its length.*

Proof. We will define a subset of $\mathbb{R}$, which is not Lebesgue measurable. On $\mathbb{R}$ we consider the following equivalence relation

$$x \sim u \quad \text{if and only if} \quad x - u \in \mathbb{Q} .$$

Choose a single element $x \in [0, 1]$ from every equivalence class formed by $\sim$. Here we assume that the Axiom of Choice holds. Let $A \subseteq [0, 1]$ be the set formed by these representatives. Suppose that $A \in \Sigma_\lambda$. Then by translation invariance we have that $\{A + r\}_{r\in\mathbb{Q}}$ is a countable, Lebesgue measurable partition of $\mathbb{R}$ with $\lambda(A + r) = \eta$ independent of $r \in \mathbb{Q}$. If $\eta = 0$, then we have a contradiction to the fact that $\lambda(\mathbb{R}) = +\infty$. If $\eta > 0$, then, with $D = \mathbb{Q} \cap [0, 1]$, we obtain $2 = \lambda([0, 2]) = \sum_{r\in D} \lambda(A + r) = +\infty$, again a contradiction. Hence, $A \notin \Sigma_\lambda$. □

In general the measure theoretic and topological properties of sets in $\mathbb{R}$ differ.

Example 2.1.45. Singletons have a Lebesgue measure of zero. Hence, $\lambda(\mathbb{Q}) = 0$. Let $\{r_n\}_{n\geq1} \subseteq [0, 1]$ be an enumeration of the rationals in $[0, 1]$. Let $I_n = (r_n-\varepsilon/2^n, r_n+\varepsilon/2^n)$ and let $U = (0, 1) \cap (\bigcup_{n\geq1} I_n)$. Evidently, $U \subseteq [0, 1]$ is open and dense, so topologically "large." On the other hand we have $\lambda(U) \leq \sum_{n\geq1} \varepsilon/2^n = \varepsilon$. Hence, U is measure theoretically "small." Similarly, $C = [0, 1] \setminus U$ is nowhere dense and closed, thus topologically small, but $\lambda(C) \geq 1 - \varepsilon$, thus it is measure theoretically "large."

The Cantor set will help us to get an idea on what the relation is between $\mathcal{B}(\mathbb{R})$ and Σ_λ.

Example 2.1.46. The **Cantor set** is constructed as follows. Let $C_0 = [0, 1]$. We trisect $[0, 1]$ and remove the open middle third $(1/3, 2/3)$. We set $C_1 = [0, 1/3] \cup [2/3, 1]$. Then we trisect each of the two intervals of C_1 and remove the open middle thirds. We obtain $C_2 = [0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1]$. We proceed inductively. So, suppose we have C_n. This consists of 2^n closed intervals. We trisect each one of them and remove the open middle thirds. The remaining part of C_n is the set C_{n+1}, which is the union of 2^{n+1} disjoint closed intervals. Evidently $\{C_n\}_{n\geq1}$ is decreasing. Then the Cantor set C of $[0, 1]$ is defined by $C = \bigcap_{n\geq1} C_n$. This set consists of those points $x \in [0, 1]$, which in base -3 have an expansion $x = \sum_{k\geq1} a_k 1/3^k$ with $a_k \neq 1$ for all $k \in \mathbb{N}$.

Proposition 2.1.47. *The Cantor set C has the following properties:*
(a) *C is compact and nowhere dense.*
(b) $\lambda(C) = 0$.
(c) $\operatorname{card}(C) = \mathfrak{c}$ = *the cardinality of the continuum.*

Proof. (a) Clearly C is closed since it is the intersection of closed sets. Hence C is compact. Moreover, $\operatorname{int} C = \emptyset$ as it contains no interval since at each stage, each interval has length $1/3^n$. Therefore, C is nowhere dense.

(b) At each stage we remove 2^{n-1} open intervals each one of length $1/3^n$. Therefore the total measure of the removed set at the nth step is $2^{n-1}/3^n$. Hence, we have

$$\lambda([0, 1] \setminus C) = \sum_{n\geq1} \frac{2^{n-1}}{3^n} = \frac{1}{2} \sum_{n\geq1} \left(\frac{2}{3}\right)^n = 1 .$$

Thus, $\lambda(C) = 0$.

(c) Let $x \in C$. Then $x = \sum_{k\geq1} a_k/3^k$ with $a_k = 0$ or $a_k = 2$ for all $k \in \mathbb{N}$. Let $f(x) = \sum_{k\geq1} c_k/2^k$ with $c_k = a_k/2$ for all $k \in \mathbb{N}$, the base -2 expansion of $x \in C$. Hence, $f : C \to [0, 1]$ is onto, thus $\operatorname{card}(C) = \mathfrak{c}$. □

Remark 2.1.48. The Cantor set is interesting because it is "large" from the cardinality point of view but negligible from the measure theoretic point of view. We can generalize the above construction and have "Cantor-like sets" that still satisfy (a) and (c) from Proposition 2.1.47. So, let I be a bounded interval and $\vartheta \in (0, 1)$. We call the open interval with the same midpoint as I and length $\vartheta\lambda(I)$ the **open middle** ϑ. Now let $\{\vartheta_k\}_{k\geq 1} \subseteq (0, 1)$ and produce a decreasing sequence $\{\hat{C}_k\}_{k\geq 1}$ of closed sets in $[0, 1]$ as follows: $\hat{C}_0 = [0, 1]$ and $\hat{C}_k$ is produced by removing the open middle ϑ_k from each component interval of $\hat{C}_{k-1}$. We set $\hat{C} = \bigcap_{k\geq 1} \hat{C}_k$. We still have that $\hat{C}$ is compact and nowhere dense and $\operatorname{card}(\hat{C}) = \mathfrak{c}$. Concerning the Lebesgue measure, note that $\lambda(\hat{C}_k) = (1 - \vartheta_k)\lambda(\hat{C}_{k-1})$ for all $k \geq 2$. So, $\lambda(\hat{C}) = \prod_{k\geq 1}(1 - \vartheta_k) = \lim_{n\to\infty} \prod_{k=1}^{n}(1 - \vartheta_k)$. If $\vartheta_k = \vartheta \in (0, 1)$ for all $k \in \mathbb{N}$, then $\lambda(\hat{C}) = 0$. Note that the Cantor set corresponds to the particular case of $\vartheta = 1/3$. If $\vartheta_k \to 0$ sufficiently fast as $k \to \infty$, then $\lambda(\hat{C}) > 0$. In particular, $\prod_{k\geq 1}(1 - \vartheta_k) > 0$ if and only if $\sum_{k\geq 1} \vartheta_k < +\infty$. We point out that part (c) of the proposition above implies that there are $2^{\mathfrak{c}}$ Lebesgue measurable subsets of $\mathbb{R}$. On the other hand $\operatorname{card}(\mathcal{B}(\mathbb{R})) = \mathfrak{c}$. So, there are many more Lebesgue measurable sets than Borel sets in $\mathbb{R}$ although it is not easy to produce a set that is Lebesgue measurable but not a Borel set. For such a concrete set we refer to Federer [109, p. 68].

2.2 Measurable Functions – Integration

The Lebesgue integral is defined for measurable functions. For this reason we start this section with a discussion of measurable functions.

Definition 2.2.1. Let (X, Σ) and $(Y, \mathcal{L})$ be two measurable spaces and $f: X \to Y$ be a map. We say that f is $(\Sigma, \mathcal{L})$**-measurable** if $f^{-1}(A) \in \Sigma$ for all $A \in \mathcal{L}$. If X, Y are Hausdorff topological spaces, then they become measurable spaces by considering their Borel σ-algebras $\mathcal{B}(X)$, $\mathcal{B}(Y)$ and then f is said to be **Borel measurable** (or simply a **Borel function**). When $Y = \mathbb{R}$ or $Y = \mathbb{R}^*$ we always use the Borel σ-field of Y.

Remark 2.2.2. The reason that we use the Borel σ-algebra on $\mathbb{R}$ as range space is that the Lebesgue σ-algebra Σ_λ, as the completion of $\mathcal{B}(\mathbb{R})$, is in general too large for the Lebesgue measure; see Remark 2.1.48. In particular, there exists a continuous, nondecreasing function $h: [0, 1] \to [0, 1]$ and a Lebesgue measurable set $C \subseteq [0, 1]$ such that $h^{-1}(C)$ is not Lebesgue measurable (assuming the Axiom of Choice). In fact $h(x) = 1/2[\hat{f}(x) + x]$ with $\hat{f}$ being the function from the proof of Proposition 2.1.47(c) extended to all of $[0, 1]$ by declaring it to be constant on each interval missing from C. Then $\hat{f}$ is nondecreasing and continuous and is known as the **Cantor function**.

Proposition 2.2.3. *If (X, Σ) and $(Y, \mathcal{L})$ are measurable spaces, $\mathcal{L} = \sigma(\mathfrak{a})$ and $f: X \to Y$, then f is $(\Sigma, \mathcal{L})$-measurable if and only if $f^{-1}(A) \in \Sigma$ for all $A \in \mathfrak{a}$.*

Proof. $\Longrightarrow$: This is immediate from Definition 2.2.1.

$\Longrightarrow$: Let $\mathcal{D} = \{A \subseteq Y : f^{-1}(A) \in \Sigma\}$. Evidently $\mathcal{D} \supseteq \mathfrak{a}$ and $\mathcal{D}$ is a σ-algebra. Therefore, $\mathcal{D} \supseteq \sigma(\mathfrak{a}) = \mathcal{L}$ and this proves the $(\Sigma, \mathcal{L})$-measurability of f. $\square$

Combining Propositions 2.1.18 and 2.2.3 we have the following result.

Proposition 2.2.4. *If (X, Σ) is a measurable space and $f : X \to \mathbb{R}$, then the following statements are equivalent:*
(a) *f is Σ-measurable;*
(b) *$f^{-1}((a, +\infty)) \in \Sigma$ for all $a \in \mathbb{R}$;*
(c) *$f^{-1}([a, +\infty)) \in \Sigma$ for all $a \in \mathbb{R}$;*
(d) *$f^{-1}((-\infty, a]) \in \Sigma$ for all $a \in \mathbb{R}$;*
(e) *$f^{-1}((-\infty, a)) \in \Sigma$ for all $a \in \mathbb{R}$.*

Remark 2.2.5. In case f is $\mathbb{R}^*$-valued, we need to add the requirement that $f^{-1}(\pm\infty) \in \Sigma$ in the statements (b)–(e). Evidently we can take $a \in \mathbb{Q}$ in (b)–(e).

Immediately from Definition 2.2.1, we have that the composition preserves measurability.

Proposition 2.2.6. *If $(X, \Sigma), (Y, \mathcal{L}), (Z, \mathcal{D})$ are measurable spaces and $f : X \to Y$, $g : Y \to Z$ are measurable maps, then $h = g \circ f : X \to Z$ is measurable as well.*

Moreover, we have the following as a consequence of Proposition 2.2.3.

Proposition 2.2.7. *If X, Y are Hausdorff topological spaces and $f : X \to Y$ is continuous, then f is Borel measurable.*

Proposition 2.2.8. *If (X, Σ) is a measurable space and $f, g : X \to \mathbb{R}$ are Σ-measurable functions, then $f \pm g$ and fg are both Σ-measurable.*

Proof. If $f(x) + g(x) < a$, then $f(x) < a - g(x)$. Let $c \in \mathbb{Q}$ be such that $f(x) < c < a - g(x)$. So, we have that

$$\begin{aligned}&\{x \in X : f(x) + g(x) < a\} \\ &= \bigcup_{c \in \mathbb{Q}} \left[\{x \in X : f(x) < c\} \bigcap \{x \in X : g(x) < a - c\}\right] \in \Sigma .\end{aligned}$$

Hence $f + g$ is Σ-measurable.

Since $-g$ is Σ-measurable, if g is, it follows that $f - g$ is Σ-measurable as well.

For any $h : X \to \mathbb{R}$ being Σ-measurable and $a \geq 0$, we have

$$\left\{x \in X : h(x)^2 > a\right\} = \left\{x \in X : h(x) > a^{\frac{1}{2}}\right\} \bigcup \left\{x \in X : h(x) < -a^{\frac{1}{2}}\right\} \in \Sigma .$$

Therefore h^2 is Σ-measurable.

Since $fg = 1/2\left[(f + g)^2 - f^2 - g^2\right]$ using the fact above and the Σ-measurability of $f + g$, we conclude that fg is Σ-measurable. $\square$

Remark 2.2.9. The result above is also valid for R^*-valued functions, provided we always take the same value for $f \pm g$ at the points where it is undefined, that is, of the form $\infty - \infty$. In addition, recalling that we always define $0(\pm\infty) = 0$, the function fg is Σ-measurable for R^*-valued f and g.

Proposition 2.2.10. *If (X, Σ) is a measurable space and $f_n : \Sigma \to \mathbb{R}^*$ with $n \in \mathbb{N}$ are Σ-measurable, then*

$$\sup\{f_n\}_{n=1}^m\,, \quad \inf\{f_n\}_{n=1}^m\,, \quad \sup_{n\geq 1} f_n\,, \quad \inf_{n\geq 1} f_n\,, \quad \liminf_{n\to\infty} f_n\,, \quad \limsup_{n\to\infty} f_n$$

are all Σ-measurable.

Proof. Let $g(x) = \sup_{1\leq n\leq m} f_n(x)$. Then for all $a \in \mathbb{R}$, we have

$$\{x \in X : g(x) > a\} = \bigcup_{n=1}^{m} \{x \in X : f_n(x) > a\} \in \Sigma\,.$$

Thus g is Σ-measurable. Similarly, if $\hat{g}(x) = \sup_{n\geq 1} f_n(x)$, then for all $a \in \mathbb{R}$, we have

$$\{x \in X : \hat{g}(x) > a\} = \bigcup_{n\geq 1} \{x \in X : f_n(x) > a\} \in \Sigma\,.$$

In a similar fashion we also show that $\inf_{1\leq n\leq m} f_n$ and $\inf_{n\geq 1} f_n$ are both Σ-measurable.

Finally, recall that $\liminf_{n\to\infty} f_n = \sup_{k\geq 1} \inf_{n\geq k} f_n$ and $\limsup_{n\to\infty} f_n = \inf_{k\geq 1} \sup_{n\geq k} f_n$, to conclude that both are Σ-measurable. $\square$

When a sequence of measurable functions does not converge pointwise, we can still have the measurability of the set of points where pointwise convergence occurs.

Proposition 2.2.11. *If (X, Σ) is a measurable space and $f_n : X \to \mathbb{R}$ with $n \geq 1$ is a sequence of Σ-measurable functions, then the set $C = \{x \in X : \lim_{n\to\infty} f_n(x)$ exists$\} \in \Sigma$.*

Proof. Given $x \in C$, we have that $\{f_n(x)\}_{n\geq 1} \subseteq \mathbb{R}$ is a Cauchy sequence. So, for $\varepsilon = 1/n$ with $n \in \mathbb{N}$ we can find $m = m(\varepsilon) \in \mathbb{N}$ such that

$$|f_{m+k}(x) - f_m(x)| < \frac{1}{n} \quad \text{for all } k \in \mathbb{N}\,.$$

Therefore it follows

$$\begin{aligned} C &= \left\{x \in X : \forall n \in \mathbb{N}\, \exists m \in \mathbb{N} \text{ such that } |f_{m+k}(x) - f_m(x)| < \frac{1}{n}\ \forall k \in \mathbb{N}\right\} \\ &= \bigcap_{n\geq 1} \bigcup_{m\geq 1} \bigcap_{k\geq 1} \left\{x \in X : |f_{m+k}(x) - f_m(x)| < \frac{1}{n}\right\} \in \Sigma\,. \end{aligned}$$

$\square$

In Proposition 2.2.10 we saw that the pointwise limit of Σ-measurable, $\mathbb{R}^*$-valued functions is Σ-measurable as well. This result can be extended to maps with values in a metric space.

Proposition 2.2.12. *If (X, Σ) is a measurable space, Y is a metrizable space and $f_n : X \to Y$ with $n \in \mathbb{N}$ is a sequence of Σ-measurable functions such that $f_n(x) \to f(x)$ in Y for all $x \in X$, then f is Σ-measurable as well.*

Proof. Let $C \subseteq Y$ be a closed set. According to Proposition 2.2.3 it suffices to show that $f^{-1}(C) \in \Sigma$. Let d be a compatible metric on Y. Let $U_n = \{y \in Y : d(y, C) < 1/n\}$ with $n \in \mathbb{N}$. These sets are open and $C = \bigcap_{n\geq 1} U_n$; see Proposition 1.5.8. Let $x \in f^{-1}(C)$. Then $f(x) \in C$ and $f_n(x) \to f(x)$ in Y. Since for each $n \in \mathbb{N}$, U_n is a neighborhood of $f(x)$ there exists $m \in \mathbb{N}$ such that $f_k(x) \in U_n$ for all $k \geq m$, which implies

$$x \in \bigcap_{n\geq 1} \bigcup_{m\geq 1} \bigcap_{k\geq m} f_k^{-1}(U_n) .$$

This yields

$$f^{-1}(C) \subseteq \bigcap_{n\geq 1} \bigcup_{m\geq 1} \bigcap_{k\geq m} f_k^{-1}(U_n) . \tag{2.2.1}$$

Next suppose that $x \in \bigcap_{n\geq 1} \bigcup_{m\geq 1} \bigcap_{k\geq m} f_k^{-1}(U_n)$. So for every $n \in \mathbb{N}$, $f_k(x)$ is eventually in U_n, hence $f(x) = \lim_{k\to\infty} f_k(x) \in \overline{U}_n$. Therefore $f(x) \in \bigcap_{n\geq 1} \overline{U}_n$. But $\overline{U}_{n+1} \subseteq U_n$. Hence $f(x) \in \bigcap_{n\geq 1} U_n = C$, which gives $x \in f^{-1}(C)$. Hence

$$\bigcap_{n\geq 1} \bigcup_{m\geq 1} \bigcap_{k\geq m} f_k^{-1}(U_n) \subseteq f^{-1}(C) . \tag{2.2.2}$$

From (2.2.1) and (2.2.2) it follows that

$$f^{-1}(C) = \bigcap_{n\geq 1} \bigcup_{m\geq 1} \bigcap_{k\geq m} f_k^{-1}(U_n) \in \Sigma .$$

Thus, f is Σ-measurable. ☐

Remark 2.2.13. The result above fails if Y is not metrizable. To see this let $Y = I^I$ with $I = [0, 1]$ furnished with the product topology. Then Y is compact by Tychonoff's Theorem (see Theorem 1.4.56), but it is not metrizable. Let $f_n : I \to Y$ with $n \in \mathbb{N}$ be the sequence of maps defined by

$$f_n(x)(t) = [1 - n|x - t|]^+ \quad \text{for all } x, t \in I .$$

Note that each $f_n : I \to Y$ is continuous, thus Borel measurable. In addition, $f_n(x)(t) \to \chi_{\{x\}}(t)$ for all $t \in I$. Here

$$\chi_{\{x\}}(t) = \begin{cases} 1 & \text{if } t = x , \\ 0 & \text{if } t \neq x \end{cases}$$

is the indicator function of the singleton $\{x\}$.

For each $x \in I$ there exists an open set $U_x \subseteq Y$ such that $f^{-1}(U_x) = \{x\}$ (for example, let $U_x = \{f \in Y = I^I : f(x) > 0\}$). Let $D \subseteq I$ be a non-Borel set and let $V = \bigcup_{x\in D} U_x$. Evidently $V \subseteq I^I$ is open and $f^{-1}(V) = D$. This shows that f is not measurable.

Definition 2.2.14. Let (X, Σ, μ) be a measure space. A statement about $x \in X$ is said to hold **almost everywhere** or **a.e.** (for **almost all** x or **a.a.** $x \in X$) if it holds for all $x \notin D$ with $\mu(D) = 0$. Note that the set of all $x \in X$ for which the statement holds will be in Σ_μ but not necessarily in Σ.

Measurability is not affected by changing the function on a μ-null set.

Proposition 2.2.15. *If (X, Σ, μ) is a complete measure space, $(Y, \mathcal{L})$ is a measurable space, $f\colon X \to Y$ is $(\Sigma, \mathcal{L})$-measurable and $g\colon X \to Y$ satisfies $f(x) = g(x)$ for μ-a.a. $x \in X$, then g is $(\Sigma, \mathcal{L})$-measurable as well.*

Next we will introduce the functions, which are the building blocks for the theory of integration.

Definition 2.2.16. Let (X, Σ) be a measurable space.

(a) Given $A \subseteq X$, the **characteristic function** χ_A of A is defined by

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

(b) A **simple function** is a measurable function $s\colon X \to \mathbb{R}$, which has finite range. So, if $a_1, \dots, a_n$ are the distinct values of s, then we can write $s(x) = \sum_{k=1}^n a_k \chi_{A_k}(x)$ with $A_k = \{x \in X\colon s(x) = a_k\} \in \Sigma$. We call this the **standard representation** of s.

Remark 2.2.17. Since in probability theory a characteristic function is a Fourier transform, probabilists use the name indicator function and denote it by i_A. On the other hand, in nonsmooth analysis and optimization, this name and symbol are reserved for another function, namely

$$i_A(x) = \begin{cases} 0 & \text{if } x \in A\,, \\ +\infty & \text{if } x \notin A\,. \end{cases}$$

A simple function is a linear combination with distinct coefficients of characteristic functions of disjoint sets whose union is X. One of the coefficients a_k may well be zero, but still the term $a_k \chi_{A_k}$ is implicitly understood in the standard representation so as to have $X = \bigcup_{k=1}^n A_k$. If s and τ are simple functions, then so are $s + \tau$ and $s\tau$.

Simple functions approximate measurable functions.

Proposition 2.2.18. *If (X, Σ) is a measurable space and $f\colon \ \to [0, +\infty]$ is a Σ-measurable function, then there exists a sequence $\{s_n\}_{n\geq 1}$ of simple functions on X such that*

$$0 \leq s_1(x) \leq s_2(x) \leq \dots \leq s_n(x) \to f(x) \quad \textit{for all } x \in X \textit{ as } n \to \infty\,.$$

Moreover the convergence is uniform on any set on which f is bounded from above.

Proof. Given $n \in \mathbb{N}$ we partition the interval $[0, n)$ into $n2^n$ half-open intervals of length $1/2^n$. Then for each $1 \leq k \leq n2^n$ with $k \in \mathbb{N}$ we define

$$D_{n,k} = \left\{x \in X\colon \frac{k-1}{2^n} \leq f(x) < \frac{k}{2^n}\right\}\,, \qquad D_n = \{x \in X\colon f(x) \geq n\}\,.$$

The Σ-measurability of f implies that $D_{n,k}, D_n \in \Sigma$. We set

$$s_n = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \chi_{D_{n,k}} + n\chi_{D_n}\,.$$

Evidently this is a simple function for every $n \in \mathbb{N}$. Let $x \in D_{n,k}$. Then

$$\frac{2k-2}{2^{n+1}} \le f(x) < \frac{2k}{2^{n+1}} ,$$

which implies that $s_{n+1}(x) = (2k-2)/2^{n+1}$ or $s_{n+1}(x) = (2k-1)/2^{n+1}$. Hence $s_n(x) \le s_{n+1}(x)$.

Now let $x \in D_n$. Then $f(x) \ge n$ and we have $f(x) \ge n+1$ or $n \le f(x) < n+1$. If the first case holds, then $s_{n+1}(x) \ge n+1 > n = s_n(x)$. In the second case, let $k \in \{1, \dots, (n+1)2^{n+1}\}$ such that $(k-1)/2^{n+1} \le f(x) < k/2^{n+1}$. Since $f(x) > n$ it follows that $k/2^{n+1} > n$, hence $k = (n+1)2^{n+1}$. Therefore, $s_{n+1}(x) = n+1-1/2^{n+1} > n = s_n(x)$. This proves that $s_n \le s_{n+1}$.

Now we prove the pointwise convergence. So, fix $x \in X$ such that $f(x) \in [0, +\infty)$ and let $n > f(x)$. Then

$$0 \le f(x) - f_n(x) < \frac{1}{2^n} , \tag{2.2.3}$$

which gives $f_n(x) \to f(x)$ as $n \to \infty$.

On the other hand, if $f(x) = +\infty$, then $f_n(x) = n \to +\infty$. Finally if $0 \le f(x) \le M$ for some $M > 0$ and for all $x \in X$, then (2.2.3) holds for every $x \in X$ provided $n > M$. Therefore $f_n \to f$ uniformly. ☐

If $f^+ = \max\{f, 0\}$ and $f^- = \{-f, 0\}$, then $f = f^+ - f^-$ as well as $|f| = f^+ + f^-$ and if $f\colon X \to \mathbb{R}$ is Σ-measurable, then so are f^+ and f^-; see Proposition 2.2.10. So using Proposition 2.2.18 on each of the functions f^+ and f^- we have the following.

Corollary 2.2.19. *If (X, Σ) is a measurable space and $f\colon X \to \mathbb{R}$ is Σ-measurable, then there exists a sequence $\{s_n\}_{n\ge1}$ of simple functions on X such that*

$$|s_1| \le |s_2| \le \dots \le |s_n| \le \dots |f| \dots , \quad s_n(x) \to f(x) \quad \textit{for all } x \in X .$$

Moreover if f is bounded, then the convergence is uniform.

We can extend these results to maps with values in a separable metric space. This is useful when studying integration of Banach space-valued maps; see the Lebesgue–Bochner integral in Section 4.2.

Proposition 2.2.20. *If (X, Σ) is a measurable space, (Y, d) is a separable metric space and $f\colon X \to Y$, then the following hold:*

(a) *If (Y, d) is in addition totally bounded, then f is Σ-measurable if and only if it is the d-uniform limit of a sequence of simple functions with values in Y.*

(b) *f is Σ-measurable if and only if f is the d-pointwise limit of a sequence of simple functions with values in Y.*

Proof. (a) $\Longrightarrow$: Suppose that $f\colon X \to Y$ is Σ-measurable and let $\varepsilon > 0$. Since Y is by hypothesis totally bounded, there exists $y_1, \dots, y_m \in Y$ such that $Y = \bigcup_{k=1}^m B_\varepsilon(y_k)$ with $B_\varepsilon(y_k) = \{y \in Y\colon d(y, y_k) < \varepsilon\}$. We set $A_1 = B_\varepsilon(y_1)$ and $A_{k+1} = B_\varepsilon(y_{k+1}) \setminus \bigcup_{i=1}^k B_\varepsilon(y_i)$

for all $k \in \{1, \dots, m-1\}$. Then $\{A_k\}_{k=1}^m$ are mutually disjoint Borel sets in Y whose union is Y. We have

$$X = \bigcup_{k=1}^{m} f^{-1}(A_k) \quad \text{and} \quad f^{-1}(A_k) \cap f^{-1}(A_n) = \emptyset \quad \text{if } k \neq n\,.$$

We define $s\colon X \to Y$ by $s(x) = y_k$ if $x \in f^{-1}(A_k)$. Evidently s is a simple function and $d(s(x), f(x)) < \varepsilon$ for all $x \in X$. Therefore f is the d-uniform limit of a sequence of simple functions with values in Y.

$\Longleftarrow$: This is a consequence of Proposition 2.2.12.

(b) By Theorem 1.5.21 there is a homeomorphism (embedding) $\xi\colon Y \to \mathbb{H}$ onto a subset of the Hilbert cube $\mathbb{H} = [0,1]^{\mathbb{N}}$. Let $e(u, y) = d_{\mathbb{H}}(\xi(u), \xi(y))$ for all $u, y \in Y$. Then e is a metric on Y, compatible with d and (Y, e) is totally bounded. By part (a) we know that f is the e-uniform limit of a sequence of simple functions. Since e and d are topologically equivalent, we have that the sequence of simple functions is d-pointwise convergent to f. □

Definition 2.2.21. Let $\{(Y_\alpha, \mathcal{L}_\alpha)\}_{\alpha \in I}$ be a family of measurable spaces and $f_\alpha\colon X \to Y_\alpha$ be a map for each $\alpha \in I$. There is a unique σ-algebra on X with respect to which the f_α's are all measurable and this is the σ-algebra generated by the sets $f_\alpha^{-1}(A_\alpha)$ for all $A_\alpha \in \mathcal{L}_\alpha$ and all $\alpha \in I$. It is called the σ-algebra generated by $\{f_\alpha\}_{\alpha \in I}$ and is denoted by $\sigma(\{f_\alpha\})$.

Proposition 2.2.22. *If $(Y, \mathcal{L})$ is a measurable space, $f\colon X \to Y$ and $g\colon X \to \mathbb{R}$ are given maps, then g is $\sigma(f)$-measurable if and only if there exists a $\mathcal{L}$-measurable $h\colon Y \to \mathbb{R}$ such that $g = h \circ f$.*

Proof. $\Longrightarrow$: First we assume that g is a $\sigma(f)$-simple function. Then $g = \sum_{k=1}^n a_k \chi_{A_k}$ with $a_k \in \mathbb{R}$ and $A_k \in \sigma(f)$. For $k \in \{1, \dots, n\}$ let $C_k \in \mathcal{L}$ be such that $A_k = f^{-1}(C_k)$. We set $h = \sum_{k=1}^n a_k \chi_{C_k}$. Then h is a $\mathcal{L}$-simple function on Y and clearly $g = h \circ f$.

Now suppose that g is a general $\sigma(f)$-measurable function. Then by Corollary 2.2.19 there exists a sequence $\{s_n\}_{n \geq 1}$ of $\sigma(f)$-simple functions such that $s_n(x) \to g(x)$ for all $x \in X$. From the first part of the proof we can find $h_n\colon Y \to \mathbb{R}$ with $n \in \mathbb{N}$ being $\mathcal{L}$-measurable functions such that $s_n = h_n \circ f$ with $n \in \mathbb{N}$. Let $E = \{y \in Y\colon \lim_{n\to\infty} h_n(y) \text{ exists in } \mathbb{R}\}$. Since $h_n(f(x)) = s_n(x) \to g(x)$ it follows that $f(X) \subseteq E$. Define

$$h(y) = \lim_{n\to\infty} h_n(y) \quad \text{if } y \in E \quad \text{and} \quad h(y) = 0 \quad \text{if } y \notin E\,.$$

From the inclusion $f(X) \subseteq E$ it follows that $g = h \circ f$. Moreover, from Proposition 2.2.11 we know that $E \in \mathcal{L}$. Hence $h_n \chi_E$ is $\mathcal{L}$-measurable and since $h_n \chi_E \to h \chi_E$ it follows that h is $\mathcal{L}$-measurable.

$\Longleftarrow$: This follows from Proposition 2.2.6. □

Definition 2.2.23. Let $\{(X_\alpha, \Sigma_\alpha)\}_{\alpha \in I}$ be a family of measurable spaces. Set $X = \prod_{\alpha \in I} X_\alpha$ and let $p_\alpha\colon X \to X_\alpha$ with $\alpha \in I$ be the corresponding projection (coordinate) maps. Then the **product σ-algebra** on X denoted by $\bigotimes_{\alpha \in I} \Sigma_\alpha$ is defined by $\bigotimes_{\alpha \in I} \Sigma_\alpha = \sigma(\{p_\alpha\})$.

Remark 2.2.24. Let (X, Σ), $(Y, \mathcal{L})$ be two measurable spaces. A set of the form $A \times B$ with $A \in \Sigma$, $B \in \mathcal{L}$ is said to be a **measurable rectangle** . By $\mathcal{R}$ we denote the family of measurable rectangles in $X \times Y$. It is easy to see that $\mathcal{R}$ is an algebra. Then $\Sigma \bigotimes \mathcal{L} = \sigma(\mathcal{R})$. More generally if the index set I is countable, then

$$\bigotimes_{\alpha \in I} \Sigma_\alpha = \sigma\left(\prod_{\alpha \in I} A_\alpha : A_\alpha \in \Sigma_\alpha\right) .$$

Proposition 2.2.25. *If $\{(X_\alpha, \Sigma_\alpha)\}_{\alpha \in I}$ are measurable spaces and each Σ_α is generated by $\mathfrak{a}_\alpha$, then $\bigotimes_{\alpha \in I} \Sigma_\alpha$ is generated by $\hat{\mathfrak{a}} = \{p_\alpha^{-1}(B_\alpha) : B_\alpha \in \mathfrak{a}_\alpha, \alpha \in I\}$. Moreover, if the index set I is countable, then $\bigotimes_{\alpha \in I} \Sigma_\alpha$ is generated by $\tilde{\mathfrak{a}} = \{\prod_{\alpha \in I} B_\alpha : B_\alpha \in \mathfrak{a}_\alpha\}$.*

Proof. From Definition 2.2.23 it is clear that $\sigma(\hat{\mathfrak{a}}) \subseteq \bigotimes_{\alpha \in I} \Sigma_\alpha$. Let

$$\mathcal{D}_\alpha = \left\{B \subseteq X_\alpha : p_\alpha^{-1}(B) \in \sigma(\hat{\mathfrak{a}})\right\}, \alpha \in I .$$

It is easy to see that $\mathcal{D}_\alpha$ is a σ-algebra and $\mathfrak{a}_\alpha \subseteq \mathcal{D}_\alpha$. Therefore $\Sigma_\alpha \subseteq \mathcal{D}_\alpha$ for all $\alpha \in I$. Hence $\bigotimes_{\alpha \in I} \Sigma_\alpha \subseteq \sigma(\hat{\mathfrak{a}})$ and so equality holds.

The second assertion follows from Remark 2.2.24. ☐

Proposition 2.2.26. *If $\{X_k\}_{k=1}^n$ are Hausdorff topological spaces, then the following hold:*
(a) $\bigotimes_{k=1}^n \mathcal{B}(X_k) \subseteq \mathcal{B}(\prod_{k=1}^n X_k)$*;*
(b) *If $\{X_k\}_{k=1}^n$ are second countable, then $\bigotimes_{k=1}^n \mathcal{B}(X_k) = \mathcal{B}(\prod_{k=1}^n X_k)$.*

Proof. (a) By Proposition 2.2.25, $\bigotimes_{k=1}^n \mathcal{B}(X_k)$ is generated by the sets $p_k^{-1}(U_k)$ with open $U_k \subseteq X_k$ for all $k \in \{1, \dots, n\}$. These sets are open in $X = \prod_{k=1}^n X_k$ and so, we infer that $\bigotimes_{k=1}^n \mathcal{B}(X_k) \subseteq \mathcal{B}(X)$.

(b) Let $\mathcal{D}_k$ be a countable basis of X_k, $k \in \{1, \dots, n\}$. Recall that every open set in X_k is a countable union of elements in $\mathcal{D}_k$. Therefore $\mathcal{B}(X)$ is generated by $\mathcal{D}_k$ and $\mathcal{B}(X)$ is generated by $\hat{\mathcal{D}} = \{\prod_{k=1}^n B_k : B_k \in \mathcal{D}_k\}$. Hence, we conclude that $\bigotimes_{k=1}^n \mathcal{B}(X_k) = \mathcal{B}(X)$. ☐

Definition 2.2.27. Let X, Y be nonempty sets and $A \subseteq X \times Y$. For each $x \in X$ and each $y \in Y$, the x-**section** of A (resp. the y-**section** of A) are defined by

$$A_x = \{y \in Y : (x, y) \in A\} \quad (\text{resp. } A^y = \{x \in X : (x, y) \in A\}) .$$

Clearly for every $x \in X$ and every $y \in Y$ we have $\emptyset_x = \emptyset^y = \emptyset$ and $(X \times Y)_x = Y$ as well as $(X \times Y)^y = X$.

Remark 2.2.28. If $\{A_\alpha\}_{\alpha \in I} \subseteq X \times Y$, then for all $x \in X$ and for all $y \in Y$ we have

$$\left(\bigcup_{\alpha \in I} A_\alpha\right)_x = \bigcup_{\alpha \in I} (A_\alpha)_x , \quad \left(\bigcap_{\alpha \in I} A_\alpha\right)_x = \bigcap_{\alpha \in I} (A_\alpha)_x ,$$
$$\left(\bigcup_{\alpha \in I} A_\alpha\right)^y = \bigcup_{\alpha \in I} (A_\alpha)^y , \quad \left(\bigcap_{\alpha \in I} A_\alpha\right)^y = \bigcap_{\alpha \in I} (A_\alpha)^y .$$

So, it follows that if $\mathcal{L}$ is a σ-algebra on X and $\mathcal{D} = \{A \subseteq X \times Y : A^y \in \mathcal{L}$ for all $y \in Y\}$, then $\mathcal{D}$ is a σ-algebra on $X \times Y$. Similarly for $\mathcal{F}$ being a σ-algebra on Y. Finally, if (X, Σ) and $(Y, \mathcal{L})$ are measurable spaces and $A \subseteq X \times Y$, then we say that A has **measurable sections** if for all $x \in X$ and for all $y \in Y$, $A_x \in \mathcal{L}$ and $A^y \in \Sigma$.

Proposition 2.2.29. *If (X, Σ) and $(Y, \mathcal{L})$ are measurable spaces and $A \in \Sigma \bigotimes \mathcal{L}$, then A has measurable sections.*

Proof. Let

$$\hat{\mathcal{D}} = \{A \subseteq X \times Y : A_x \in \mathcal{L} \text{ and } A^y \in \Sigma \text{ for all } x \in X \text{ and for all } y \in Y\} .$$

Then $\hat{\mathcal{D}}$ is a σ-algebra that contains measurable rectangles. Note that

$$(A \times B)_x = \begin{cases} B & \text{if } x \in A \\ \emptyset & \text{if } x \notin A \end{cases} \quad \text{and} \quad (A \times B)^y = \begin{cases} A & \text{if } y \in B \\ \emptyset & \text{if } y \notin B . \end{cases}$$

Therefore, we have that $\sigma(\mathbb{R}) = \Sigma \bigotimes \mathcal{L} \subseteq \hat{\mathcal{D}}$, see Remark 2.2.24. □

Definition 2.2.30. Let (X, Σ) be a measurable space, Y and V are two Hausdorff topological spaces and $f : X \times Y \to V$. We say that f is a **Carathéodory function** if the following properties hold:
(a) $x \mapsto f(x, y)$ is Σ-measurable for every $y \in Y$;
(b) $y \mapsto f(x, y)$ is continuous for every $x \in X$.

Proposition 2.2.31. *If (X, Σ) is a measurable space, Y is a separable metrizable space, V is a metrizable space and $f : X \times Y \to V$ is a Carathéodory function, then f is jointly measurable, that is, f is $(\Sigma \bigotimes \mathcal{B}(Y), \mathcal{B}(V))$-measurable.*

Proof. Let d be a compatible metric for Y and e a compatible metric for V. Recall that Y is separable. So, let $D = \{y_k\}_{k \geq 1}$ be dense in Y. Moreover, let $C \subseteq V$ be a closed set. Then $f(x, u) \in C$ if and only if for every $n \in \mathbb{N}$ there exists $y_k \in D$ such that

$$d(u, y_k) < \frac{1}{n} \quad \text{and} \quad e(f(z, y_k), C) < \frac{1}{n} .$$

Therefore we have

$$f^{-1}(C) = \bigcap_{n \geq 1} \bigcup_{k \geq 1} \left\{x \in X : f(z, y_k) \in C_{\frac{1}{n}}\right\} \times B_{\frac{1}{n}}(y_k)$$

with $C_{1/n} = \{v \in V : e(v, C) < 1/n\}$. The measurability of $f(\cdot, y_k)$ and the openness of $C_{1/n}$ imply that $\{x \in X : f(z, y_k) \in C_{1/n}\} \in \Sigma$ for all $n, k \in \mathbb{N}$. Thus $f^{-1}(C) \in \Sigma \bigotimes \mathcal{B}(Y)$. □

The next theorem, known as "Egorov's Theorem," says that in a finite measure space, pointwise convergence of a sequence of measurable functions is in fact "almost" uniform.

Theorem 2.2.32 (Egorov's Theorem). *If (X, Σ, μ) is a finite measure space, (Y, d) is a metric space and $f_n \colon X \to Y$ with $n \in \mathbb{N}$ is a sequence of Σ-measurable functions such that $f_n(x) \xrightarrow{d} f(x)$ for μ-a.a. $x \in X$, then for any given $\varepsilon > 0$ there exists $A_\varepsilon \in \Sigma$ with $\mu(A_\varepsilon) < \varepsilon$ such that $f_n \xrightarrow{d} f$ uniformly on $X \setminus A_\varepsilon$. That is,* $\limsup_{n\to\infty}[d(f_n(x), f(x)) \colon x \in A_\varepsilon] = 0$.

Proof. From Proposition 2.2.12 we know that f is Σ-measurable. For $m, k \in \mathbb{N}$ let

$$A_{m,k} = \left\{x \in X \colon d(f_n(x), f(x)) \le \frac{1}{m} \text{ for all } n \ge k\right\} .$$

For every $m \in \mathbb{N}$ we have $\mu(X \setminus A_{m,k}) \searrow 0$ as $k \to +\infty$. We choose $k(m) \in \mathbb{N}$ such that $\mu(X \setminus A_{m,k(m)}) < \varepsilon/2^m$ and $D_\varepsilon = \bigcap_{m\ge1} A_{m,k(m)} \in \Sigma$. Then for $A_\varepsilon = X \setminus D_\varepsilon$ we have $\mu(A_\varepsilon) < \varepsilon$ and $f_n \xrightarrow{d} f$ uniformly on $D_\varepsilon = X \setminus A_\varepsilon$. ☐

From Chapter 1 we know that a continuous function for the subspace (relative) topology on $A \subseteq X$ cannot always be extended in a continuous fashion to all of X. Think of $f_1(x) = 1/x$ for $x \in (0, 1]$ and $f_2(x) = \sin(1/x)$ for $x \in (0, 1]$ (being bounded as well), which cannot be extended continuously to $[0, 1]$. In contrast, a measurable function from $A \subseteq X$ with the trace σ-algebra can be extended measurably to all of X. The point that we want to emphasize is that A need not be measurable, otherwise the result is obvious. We start with an easy observation that is useful in many circumstances.

Lemma 2.2.33. *If (X, Σ) and $(Y, \mathcal{L})$ are measurable spaces, $\{A_n\}_{n\ge1} \subseteq \Sigma$ are mutually disjoint sets such that $X = \bigcup_{n\ge1} A_n$ and $f_n \colon A_n \to Y$ with $n \in \mathbb{N}$ are $(\Sigma_{A_n}, \mathcal{L})$-measurable functions, then $f \colon X \to Y$ defined by $f|_{A_n} = f_n$ for all $n \in \mathbb{N}$ is $(\Sigma, \mathcal{L})$-measurable.*

Proof. For every $B \in \mathcal{L}$ we have $f_n^{-1}(B) \in \Sigma_{A_n} = \{A_n \cap D \colon D \in \Sigma\}$; see Remark 2.1.2. So, $f_n^{-1}(B) = A_n \cap D_n$ with $D_n \in \Sigma$. Note that $f^{-1}(B) = \bigcup_{n\ge1} f_n^{-1}(B) = \bigcup_{n\ge1}(A_n \cap D_n) \in \Sigma$. ☐

Theorem 2.2.34. *If (X, Σ) is a measurable space, $A \subseteq X$ (not necessarily in Σ), and $f \colon A \to \mathbb{R}$ is Σ_A-measurable (see Remark 2.1.2), then there exists a Σ-measurable function $\hat{f} \colon X \to \mathbb{R}$ such that $\hat{f}|_A = f$.*

Proof. Let V be the set of all functions $f \colon A \to \mathbb{R}$ that are Σ_A-measurable and admit a Σ-measurable extension on X. Evidently V is a vector space and it contains the simple functions. Recall that $f = f^+ - f^-$, so we may assume that $f \ge 0$. Proposition 2.2.18 implies that there exist Σ_A-simple functions $\{s_n\}_{n\ge1}$ such that $0 \le s_n \nearrow f$. Let $\hat{s}_n$ be the Σ-measurable extension of s_n and recall that $s_n \in V$ for all $n \in \mathbb{N}$. Let $\hat{f}(x) = \lim_{n\to\infty} \hat{s}_n(x)$ when this limit exists and it is finite. Otherwise we set $\hat{f}(x) = 0$. Evidently $\hat{f}|_A = f$. If C is the set of $x \in X$ where the sequence $\{\hat{s}_n(x)\}$ converges, then from Proposition 2.2.11 we have that $C \in \Sigma$. We define

$$\hat{h}_n = \hat{s}_n \quad \text{on } C \quad \text{and} \quad \hat{h}_n = 0 \quad \text{on } X \setminus C \quad \text{for all } n \in \mathbb{N} .$$

From Lemma 2.2.33 we know that for each $n \in \mathbb{N}$, $\hat{h}_n$ is Σ-measurable and $\hat{h}_n(x) \to \hat{f}(x)$ for all $x \in X$. Therefore by Proposition 2.2.11, $\hat{f}$ is Σ-measurable. ☐

Now we are ready to define the Lebesgue integral of a measurable function.

Definition 2.2.35. Let (X, Σ, μ) be a measure space.
(a) If $s\colon X \to [0, +\infty]$ is a simple function with standard representation $s = \sum_{k=1}^{n} a_k \chi_{A_k}$, then the integral of s with respect to the measure μ is defined by

$$\int_X s d\mu = \sum_{k=1}^{n} a_k \mu(A_k) .$$

(b) If $f\colon X \to [0, +\infty]$ is Σ-measurable, then the integral of f with respect to the measure μ is defined by

$$\int_X f d\mu = \sup \left[\int_X s d\mu\colon 0 \le s \le f \text{ and } s \text{ is simple} \right] .$$

(c) If $f\colon X \to \mathbb{R}^*$ is Σ-measurable and at least one of $\int_X f^+ d\mu$ and $\int_X f^- d\mu$ is finite, then the integral of f with respect to the measure μ is defined by

$$\int_X f d\mu = \int_X f^+ d\mu - \int_X f^- d\mu .$$

If both $\int_X f^+ d\mu$ and $\int_X f^- d\mu$ are finite, then we say that f is (μ)-**integrable**.

Remark 2.2.36. Since $|f| = f^+ + f^-$ we see that f is integrable if and only if $\int_X |f| d\mu < \infty$. Moreover, we have $\left|\int_X f d\mu\right| \le \int_X |f| d\mu$.

Definition 2.2.37. Let (X, Σ, μ) be a measure space and $f\colon X \to \mathbb{R}^*$ a μ-integrable function. The **integral of** f **over** A with respect to the measure μ is defined by

$$\int_A f d\mu = \int_X f \chi_A d\mu .$$

Remark 2.2.38. Recalling that any set $A \in \Sigma$ defines in a natural way a measure space with the trace σ-algebra $\Sigma_A = \{A \cap D\colon D \in \Sigma\}$ (see Remark 2.1.2), we see that it suffices to define the integral over the whole space X and we have it automatically defined over $A \in \Sigma$.

Some straightforward observations concerning the integral are listed below.

Proposition 2.2.39. *If (X, Σ, μ) is a measure space and V is the set of all μ-integrable functions, then V is a vector space, the integral is a linear functional on V and $f \le g$ μ-a.e. implies $\int_X f d\mu \le \int_X g d\mu$.*

Proposition 2.2.40. *If (X, Σ, μ) is a measure space and $f, g\colon X \to R^*$ are μ-integrable functions, then the following hold:*
(a) *$f \ge 0$ and $\int_X f d\mu = 0$ imply $f = 0$ μ-a.e.;*
(b) *the set $A = \{x \in X\colon f(x) \ne 0\}$ is σ-finite;*
(c) *$\int_C f d\mu = \int_C g d\mu$ for all $C \in \Sigma$ if and only if $f = g$ μ-a.e. if and only if $\int_X |f - g| d\mu = 0$.*

Proof. (a) Let $A = \{x \in X : f(x) > 0\}$ and $A_n = \{x \in X : f(x) \geq 1/n\}$ with $n \in \mathbb{N}$. Then $A_n \nearrow A$ and so $\mu(A_n) \nearrow \mu(A)$; see Proposition 2.1.26. If $\mu(A) > 0$, then there exists $n \in \mathbb{N}$ such that $\mu(A_n) > 0$. We have

$$0 < \frac{1}{n}\mu(A_n) \leq \int_{A_n} f d\mu \leq \int_X f d\mu = 0 ,$$

which is a contradiction. Therefore $\mu(A) = 0$ and so $f(x) = 0$ for μ-a.a. $x \in X$.

(b) As above, let $A_n = \{x \in X : |f(x)| \geq 1/n\}$ with $n \in \mathbb{N}$. Then $A_n \in \Sigma$ and $A = \bigcup_{n\geq 1} A_n$. Moreover

$$\frac{1}{n}\mu(A_n) \leq \int_{A_n} |f| d\mu \leq \int_X |f| d\mu < +\infty ,$$

which gives $\mu(A_n) \leq cn$ for all $n \in \mathbb{N}$ and for some $c > 0$. Hence A is σ-finite.

(c) The second equivalence is obvious. Moreover, if $f = g$ μ-a.e., then $\int_C f d\mu = \int_C g d\mu$ for all $C \in \Sigma$. So, it remains to show that $\int_C f d\mu = \int_C g d\mu$ for all $C \in \Sigma$ implies that $f = g$ μ-a.e. To this end let $C = \{x \in X : (f - g)(x) \neq 0\} \in \Sigma$. Suppose that $\mu(C) > 0$. Setting $C_n = \{x \in X : |(f - g)(x)| \geq 1/n\} \in \Sigma$. As above there exists $n \in \mathbb{N}$ such that $\mu(C_n) > 0$. We have $C_n = C_n^+ \cup C_n^-$ with

$$C_n^+ = \left\{x \in X : (f - g)(x) \geq \frac{1}{n}\right\} \in \Sigma$$

and

$$C_n^- = \left\{x \in X : (f - g)(x) \leq -\frac{1}{n}\right\} \in \Sigma .$$

So, at least one of C_n^+, C_n^- has positive μ-measure. To fix things, suppose that $\mu(C_n^+) > 0$. Then

$$0 = \int_{C_n^+} (f - g) d\mu \geq \frac{1}{n}\mu(C_n^+) > 0 ,$$

a contradiction. Therefore $\mu(C) = 0$ and so $f = g$ μ-a.e. as in the assertion. □

The next result is known as "Markov inequality."

Proposition 2.2.41 (Markov inequality). *If (X, Σ, μ) is a measure space and $f : X \to \mathbb{R}^*$ is μ-integrable, then for any $\lambda \in (0, +\infty)$ we have*

$$\mu(\{x \in X : |f(x)| \geq \lambda\}) \leq \frac{1}{\lambda}\int_X |f| d\mu .$$

Proof. Let $A_\lambda = \{x \in X : |f(x)| \geq \lambda\} \in \Sigma$. Then

$$\infty > \int_X |f| d\mu \geq \int_{A_\lambda} |f| d\mu \geq \lambda\mu(A_\lambda) \quad \text{implies} \quad \mu(A_\lambda) \leq \frac{1}{\lambda}\int_X |f| d\mu . \qquad \square$$

Proposition 2.2.42. *If (X, Σ, μ) is a measure space and $f\colon X \to \mathbb{R}^*$ is μ-integrable, then the following hold:*
(a) $\mu(\{x \in X\colon |f(x)| = +\infty\}) = 0$*, that is, f is μ-a.e. $\mathbb{R}$-valued;*
(b) *if $B \in \Sigma$ and $\mu(B) = 0$, then $\int_B f d\mu = 0$.*

Proof. (a) From Proposition 2.2.41 we see that for all $\lambda > 0$, $\mu(\{x \in X\colon |f(x)| \geq \lambda\}) < +\infty$ and $\lim_{\lambda\to+\infty} \mu(\{x \in X\colon |f(x)| \geq \lambda\}) = 0$. Note that

$$\{x \in X\colon |f(x)| \geq n\} \searrow \{x \in X\colon |f(x)| = +\infty\} \quad \text{as } n \to \infty\,.$$

This gives, due to Proposition 2.1.24(f),

$$\mu(\{x \in X\colon |f(x)| = +\infty\}) = \lim_{n\to\infty} \mu(\{x \in X\colon |f(x)| \geq n\}) = 0\,.$$

(b) We may assume that $f \geq 0$ since $f = f^+ - f^-$. If f is a simple function, then clearly from Definitions 2.2.35(a) and 2.2.37 we have $\int_B f d\mu = 0$. Then Definition 2.2.35(b) implies that $\int_B f d\mu = 0$. □

2.3 Convergence Theorems and L^p-Spaces

We start with certain convergence theorems that reveal the continuity properties of the Lebesgue integral.

The first such result is the so-called "Beppo Levi Theorem."

Theorem 2.3.1 (Beppo Levi Theorem). *If (X, Σ, μ) is a measure space and $f_n\colon X \to \mathbb{R}^*_+$ with $n \in \mathbb{N}$ is an increasing sequence of Σ-measurable functions such that $f_n \nearrow f$, then $\lim_{n\to\infty} \int_X f_n d\mu = \int_X f d\mu$.*

Proof. From Proposition 2.2.10 we have that f is Σ-measurable. The monotonicity of the integral function implies that

$$\lim_{n\to\infty} \int_X f_n d\mu \leq \int_X f d\mu\,. \tag{2.3.1}$$

Claim: If s is a simple function and $s \leq f$, then $\int_X s d\mu \leq \lim_{n\to\infty} \int_X f_n d\mu$.

For every $x \in X$ and every $\eta \in (0,1)$ there exists $n_0 = n_0(x, \eta) \in \mathbb{N}$ such that $\eta s(x) \leq f_n(x)$ for all $n \geq n_0$.

If we set $B_n = \{x \in X\colon \eta s(x) \leq f_n(x)\}$, then $\{B_n\}_{n\geq1} \subseteq \Sigma$ and $B_n \nearrow X$. We have $\eta\chi_{B_n} s \leq \chi_{B_n} f_n \leq f_n$.

Let $s = \sum_{k=1}^m a_k\chi_{A_k}$ be the standard representation of the simple function s. Then one gets

$$\begin{aligned}\eta\sum_{k=1}^m a_k\mu(A_k \cap B_n) &= \eta\int_X \chi_{B_n} s d\mu \leq \int_X f_n d\mu \leq \sup_{n\geq1}\int_X f_n d\mu \\ &= \lim_{n\to\infty}\int_X f_n d\mu\,.\end{aligned} \tag{2.3.2}$$

Note that for every $k \in \{1, \dots, m\}$, due to Proposition 2.1.24(e), it holds that $\mu(A_k \cap B_n) \nearrow \mu(A_k)$ as $n \to \infty$. This implies, because of (2.3.2), that

$$\eta \sum_{k=1}^{m} a_k \mu(A_k) = \eta \int_X s d\mu \le \lim_{n\to\infty} \int_X f_n d\mu .$$

Recall that $\eta \in (0, 1)$ is arbitrary. So, let $\eta \to 1^-$. Then $\int_X s d\mu \le \lim_{n\to\infty} \int_X f_n d\mu$. This proves the claim.

From the claim and Definition 2.2.35(b), we derive

$$\int_X f d\mu \le \lim_{n\to\infty} \int_X f_n d\mu . \tag{2.3.3}$$

From (2.3.1) and (2.3.3) we conclude that $\int_X f_n d\mu \nearrow \int_X f d\mu$. □

Corollary 2.3.2. *If (X, Σ, μ) is a measure space and $f : X \to \mathbb{R}_+^*$ is Σ-measurable, then $\int_X f d\mu = \lim_{n\to\infty} \int_X s_n d\mu$ for every increasing sequence of simple functions $s_n \nearrow f$.*

Now we can prove the famous “Monotone Convergence Theorem.”

Theorem 2.3.3 (Monotone Convergence Theorem). *If (X, Σ, μ) is a measure space and $f_n : X \to \mathbb{R}^*$ with $n \in \mathbb{N}$ is a sequence of Σ-measurable functions such that $f_n \nearrow f$ and $\int_X f_1 d\mu > -\infty$, then $\int_X f_n d\mu \nearrow \int_X f d\mu$ as $n \to \infty$.*

Proof. Just let $g_n = f_n - f_1 \ge 0$ for all $n \in \mathbb{N}$ and apply Theorem 2.3.1 to this sequence. □

Remark 2.3.4. The hypothesis that $\int_X f_1 d\mu > -\infty$ cannot be removed. To see this, consider the sequence $f_n = -\chi_{[n,\infty)}$ with $n \in \mathbb{N}$. Then $f_n \nearrow 0$ but $\int_X f_n d\mu = -\infty$ for all $n \in \mathbb{N}$. Moreover, there is a “decreasing” version of the theorem, namely $f_n \searrow f$ and $\int_X f_1 d\mu < +\infty$ imply that $\int_X f_n d\mu \searrow \int_X f d\mu$.

We can also formulate Theorem 2.3.3 in a series form.

Theorem 2.3.5. *If (X, Σ, μ) is a measure space and $f_n : X \to \mathbb{R}_+^*$ with $n \in \mathbb{N}$ is a sequence of Σ-measurable functions, then*

$$\int_X \left(\sum_{n\ge1} f_n \right) d\mu = \sum_{n\ge1} \int_X f_n d\mu .$$

The next convergence theorem is known as “Fatou’s Lemma.”

Theorem 2.3.6 (Fatou’s Lemma). *If (X, Σ, μ) is a measure space and $f_n, h : X \to \mathbb{R}^*$ with $n \in \mathbb{N}$ are Σ-measurable functions, then the following hold:*

(a) *If $h \le f_n$ μ-a.e. for all $n \in \mathbb{N}$ and $-\infty < \int_X h d\mu$, then*

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

(b) *If* $f_n \leq h$ *μ-a.e. for all* $n \in \mathbb{N}$ *and* $\int_X h d\mu < +\infty$, *then*

$$\limsup_{n\to\infty} \int_X f_n d\mu \leq \int_X \limsup_{n\to\infty} f_n d\mu .$$

Proof. (a) Let $g_n = \inf_{k\geq n} f_k$ with $n \in \mathbb{N}$. Then $g_n \geq h$ for all $n \in \mathbb{N}$ and $g_n \nearrow \liminf_{n\to\infty} f_n$. Invoking the Monotone Convergence Theorem (see Theorem 2.3.3) we have

$$\int_X g_n d\mu \nearrow \int_X \liminf_{n\to\infty} f_n d\mu .$$

It follows $\int_X g_n d\mu \leq \int_X f_n d\mu$ for all $n \in \mathbb{N}$ which implies

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

(b) Just apply (a) to the sequence $\{-f_n\}_{n\geq 1}$. □

Remark 2.3.7. The bound by h cannot be removed. To see this, consider $X = \mathbb{R}$ and $\mu = \lambda$ being the Lebesgue measure. Let $f_n = -1/n\chi_{[0,n]}$ for all $n \in \mathbb{N}$. Then $\liminf_{n\to\infty} \int_{\mathbb{R}} f_n d\lambda = -1 < 0 = \int_X \liminf_{n\to\infty} f_n d\mu$ and so Fatou's Lemma fails.

Now we will present the main convergence theorem for the Lebesgue integral known as the "Lebesgue Dominated Convergence Theorem." It allows us to interchange limits and integrals under general conditions and is the main reason why the Lebesgue integral is more powerful than the Riemann integral.

Theorem 2.3.8 (Lebesgue Dominated Convergence Theorem). *If* (X, Σ, μ) *is a measure space and* $f_n \colon X \to \mathbb{R}^*$ *with* $n \in \mathbb{N}$ *is a sequence of* Σ*-measurable functions such that*
- $f_n(x) \to f(x)$ *for μ-a.a.* $x \in X$;
- $|f_n(x)| \leq h(x)$ *for μ-a.a.* $x \in X$ *and for all* $n \in \mathbb{N}$

with h *being a μ-integrable function, then* f *is μ-integrable and* $\int_X |f_n - f| d\mu \to 0$. *In particular there holds*

$$\int_X f_n d\mu \to \int_X f d\mu \quad \text{as } n \to \infty .$$

Proof. From Proposition 2.2.12 we know that f is Σ-measurable. Moreover, $|f(x)| \leq h(x)$ for μ-a.a. $x \in X$. Therefore, f is μ-integrable.

Note that $0 \leq |f_n - f| \leq 2h$ μ-a.e. for all $n \in \mathbb{N}$. Applying Fatou's Lemma, Theorem 2.3.6, gives

$$0 \leq \liminf_{n\to\infty} \int_X |f_n - f| d\mu \leq \limsup_{n\to\infty} \int_X |f_n - f| d\mu \leq 0 ,$$

which implies $\int_X |f_n - f| d\mu \to 0$ as $n \to \infty$. Hence,

$$\left| \int_X (f_n - f) d\mu \right| \to 0 \quad \text{and so} \quad \int_X f_n d\mu \to \int_X f d\mu \quad \text{as } n \to \infty . \qquad \square$$

Remark 2.3.9. If the dominating function h is not μ-integrable, then the theorem fails in general. To see this, consider $X = [0, 1]$ and $\mu = \lambda$ being the Lebesgue measure. Let $f_n = n\chi_{[0,1/n]}$ with $n \in \mathbb{N}$. Then $\lim_{n\to\infty} \int_0^1 f_n d\lambda = 1 \neq 0 = \int_0^1 \lim_{n\to\infty} f_n d\lambda$.

We have already seen in Proposition 2.2.42(b) that integration is insensitive to changes on null sets. Hence, we can integrate functions f that are only defined on a measurable set A with a null complement by simply setting $f|_{A^c} = 0$. This also implies that if f is $\mathbb{R}^*$-valued and it is a.e. $\mathbb{R}$-valued, then for the purposes of integration we can treat f as $\mathbb{R}$-valued. With this in mind we are led to the introduction of the following spaces of integrable functions.

Definition 2.3.10. Let (X, Σ, μ) be a measure space and let $1 \leq p < \infty$. For any Σ-measurable function $f: X \to \mathbb{R}^*$ we define

$$\|f\|_p = \left(\int_X |f|^p d\mu \right)^{\frac{1}{p}} .$$

Let

$$\mathscr{L}^p(X) = \{f: X \to \mathbb{R}^* : f \text{ is } \Sigma\text{-measurable}, \|f\|_p < +\infty\} .$$

Evidently $\mathscr{L}^p(X)$ is a vector space. However in order to have a vector space on which $\|\cdot\|_p$ is a norm, we need to take care of functions that differ only on a μ-null set. So, we consider the following equivalence relation on $\mathscr{L}^p(X)$

$$f \sim h \quad \text{if and only if} \quad f(x) = h(x) \quad \text{for } \mu\text{-a.a. } x \in X .$$

Then we define $L^p(X) = \mathscr{L}^p(X)/\sim$.

Next let $f: X \to \mathbb{R}^*$ be Σ-measurable and define the **essential supremum** $\|f\|_\infty$ by

$$\|f\|_\infty = \inf\{\vartheta \geq 0: \mu(\{x \in X: |f(x)| \geq \vartheta\}) = 0\}$$

with the convention that $\inf \emptyset = +\infty$. We define

$$\mathscr{L}^\infty(X) = \{f: X \to \mathbb{R}^* : f \text{ is } \Sigma\text{-measurable}, \|f\|_\infty < +\infty\}$$

and $L^\infty(X) = \mathscr{L}^\infty(X)/\sim$.

Given $1 \leq p < \infty$ we say that $1 < p' \leq \infty$ is the **conjugate** of p if $1/p + 1/p' = 1$. Note that $p' = p/(p-1)$.

Recall the following elementary inequality known as "Young's inequality." It is a very special case of the so-called "Young–Fenchel inequality," which we discuss in Section 5.3.

Lemma 2.3.11 (Young's inequality). *If $p, p' \in (1, \infty)$ are conjugate exponents and $a, b \geq 0$, then $ab \leq 1/pa^p + 1/p'b^{p'}$ with equality if and only $b = a^{p-1}$.*

Next we will present three inequalities that are very basic in the theory of L^P-spaces. The first inequality is known as "Hölder's inequality."

Theorem 2.3.12 (Hölder's inequality). *If (X, Σ, μ) is a measure space, $1 \le p < \infty$, $1 < p' \le \infty$ are conjugate exponents and $f \in L^p(X)$, $h \in L^{p'}(X)$, then $fh \in L^1(X)$ and $\|fh\|_1 \le \|f\|_p \|h\|_{p'}$.*

Moreover, for $1 < p < \infty$, equality holds if and only if

$$\frac{|f(x)|^p}{\|f\|_p^p} = \frac{|h(x)|^{p'}}{\|h\|_{p'}^{p'}} \quad \text{for } \mu\text{-a.a. } x \in X .$$

Proof. First assume that $p \in (1, \infty)$, hence $p' \in (1, \infty)$. Let $a = |f(x)|/\|f\|_p$ and $b = |h(x)|/\|h\|_{p'}$. Then by applying Young's inequality (see Lemma 2.3.11) it follows

$$\frac{|f(x)h(x)|}{\|f\|_p \|h\|_{p'}} \le \frac{1}{p} \frac{|f(x)|^p}{\|f\|_p^p} + \frac{1}{p'} \frac{|h(x)|^{p'}}{\|h\|_{p'}^{p'}} \tag{2.3.4}$$

with equality if and only if $|f(x)|^p/\|f\|_p^p = |h(x)|^{p'}/\|h\|_{p'}^{p'}$ for μ-a.a. $x \in X$.

Integrating (2.3.4) it follows

$$\frac{1}{\|f\|_p \|h\|_{p'}} \int_X |fh| d\mu \le \frac{1}{p} + \frac{1}{p'} = 1 ,$$

which implies $\|fh\|_1 \le \|f\|_p \|h\|_{p'}$.

If $p = 1$, then $p' = +\infty$ and from the definition of the L^∞-norm, we have

$$\|fh\|_1 = \int_X |fh| d\mu \le \|h\|_\infty \int_X |f| d\mu = \|f\|_1 \|h\|_\infty . \qquad \square$$

When $p = p' = 2$, the inequality is usually called the "Cauchy–Bunyakowsky–Schwarz inequality."

Corollary 2.3.13 (Cauchy–Bunyakowsky–Schwarz inequality). *If (X, Σ, μ) is a measure space and $f, h \in L^2(X)$, then $fh \in L^1(X)$ and $\|fh\|_1 \le \|f\|_2 \|h\|_2$. Moreover, equality holds if and only if $f(x)^2/\|f\|_2^2 = h(x)^2/\|h\|_2^2$ for μ-a.a. $x \in X$.*

The second inequality is known as the "Minkowski inequality." In fact it is a consequence of Hölder's inequality.

Theorem 2.3.14 (Minkowski inequality). *If (X, Σ, μ) is a measure space and $f, h \in L^p(X)$ with $1 \le p \le \infty$, then $\|f + h\|_p \le \|f\|_p + \|h\|_p$.*

Proof. Via the triangle inequality the result is clear if $p = 1$ or $p = +\infty$.

So, assume that $1 < p < \infty$ and that $f + h \neq 0$, otherwise the result is clear. We estimate

$$|f(x) + h(x)|^p \le (|f(x)| + |h(x)|)\, |f(x) + h(x)|^{p-1},$$

which gives

$$\|f + h\|_p^p \le \int_X |f(x)||f(x) + h(x)|^{p-1} d\mu + \int_X |h(x)||f(x) + h(x)|^{p-1} d\mu .$$

Recall that $p - 1 = p/p'$. So, let $|f + h|^{p-1} \in L^{p'}(X)$ and apply Hölder's inequality (see Theorem 2.3.12) to get

$$\|f + h\|_p^p \le \left(\|f\|_p + \|h\|_p\right) \|f + h\|_p^{p-1} .$$

This implies $\|f + h\|_p \le \|f\|_p + \|h\|_p$. ☐

The third inequality is the so-called "Jensen inequality."

Theorem 2.3.15 (Jensen inequality). *If (X, Σ, μ) is a finite measure space, $f \in L^1(X)$ and $\varphi \colon \mathbb{R} \to \mathbb{R}$ is a convex function, then*

$$\varphi\left(\frac{1}{\mu(X)} \int_X f d\mu\right) \le \frac{1}{\mu(X)} \int_X (\varphi \circ f) d\mu .$$

Moreover, if φ is strictly convex, then equality holds if and only if f is a constant function.

Proof. It is well-known that φ is continuous. See Section 5.1 for more general continuity results for convex functions. In what follows for notational economy we set

$$(f)_X = \frac{1}{\mu(X)} \int_X f d\mu \tag{2.3.5}$$

being the average of f over X.

The convexity of φ implies that there exists $\eta \in \mathbb{R}$ such that

$$\eta(t - (f)_X) \le \varphi(t) - \varphi((f)_X) \quad \text{for all } t \in \mathbb{R} . \tag{2.3.6}$$

So, if $t = f(x)$, then, due to (2.3.5),

$$\eta\left(\int_X f d\mu - (f)_X \mu(X)\right) = 0 \le \int_X (\varphi \circ f) d\mu - \varphi((f)_X)\mu(X) .$$

This yields

$$\varphi\left(\frac{1}{\mu(X)} \int_X f d\mu\right) \le \frac{1}{\mu(X)} \int_X (\varphi \circ f) d\mu .$$

Finally, if φ is strictly convex, then (2.3.6) is a strict inequality for all $t \ne (f)_X$. If f is not constant, then $f(x) - (f)_X$ takes on both positive and negative values on sets of positive measure. Therefore, we cannot have equality. ☐

Now let us state some consequences of theses inequalities. The first is a consequence of Hölder's inequality; see Theorem 2.3.12.

Proposition 2.3.16. *If (X, Σ, μ) is a measure space, $1 \le p_k \le \infty$ for all $k = 1, \dots, n$, $\sum_{k=1}^n 1/p_k = 1/r \le 1$ and $f_k \in L^{p_k}(X)$ for all $k = 1, \dots, n$, then $\prod_{k=1}^n f_k \in L^r(X)$ and $\left\|\prod_{k=1}^n f_k\right\|_r \le \prod_{k=1}^n \|f_k\|_{p_k}$.*

Proof. Let $F = \{k \in \{1, \dots, n\} : p_k < \infty\}$ and assume that $F \neq \emptyset$ or otherwise the result is clear. Then

$$\left\|\prod_{k=1}^{n} f_k\right\|_r \leq \left\|\prod_{k\in F} f_k\right\|_r \prod_{k\notin F} \|f_k\|_\infty \quad \text{and} \quad \sum_{k\in F} \frac{1}{p_k} = \frac{1}{r}\,.$$

So we may assume that $F = \{1, \dots, n\}$. First consider the case $n = 2$. By hypothesis one obtains

$$\frac{r}{p_1} + \frac{r}{p_2} = 1\,.$$

Applying Hölder's inequality for $p = p_1/r$ and $p' = p_2/r$ to the functions $|f_1|^r, |f_2|^r$ leads to

$$\|f_1 f_2\|_r^r \leq \|f_1\|_{p_1}^r \|f_2\|_{p_2}^r\,.$$

That shows the proof for $n = 2$. When $n > 2$, we argue by induction. So let $1/\vartheta = \sum_{k=2}^{n} 1/p_k$. Hence $1/r = 1/p_1 + 1/\vartheta$. Assuming that the result holds for $n-1$, we have, by the induction assumptions and the validity of the case $n = 2$, that

$$\left\|\prod_{k=1}^{n} f_k\right\|_r \leq \|f_1\|_{p_1} \left\|\prod_{k=2}^{n} f_k\right\|_\vartheta \leq \|f_1\|_{p_1} \prod_{k=2}^{n} \|f_k\|_{p_k} = \prod_{k=1}^{n} \|f_k\|_{p_k}\,. \qquad \square$$

Another useful consequence of Hölder's inequality (see Theorem 2.3.12) is the so-called "Interpolation inequality."

Proposition 2.3.17 (Interpolation inequality). *If (X, Σ, μ) is a measure space, $1 \leq p \leq q \leq \infty$ and $f \in L^p(X) \cap L^q(X)$, then $f \in L^r(X)$ for all $p \leq r \leq q$ and $\|f\|_r \leq \|f\|_p^t \|f\|_q^{1-t}$ with*

$$\frac{1}{r} = \frac{t}{p} + \frac{1-t}{q} \quad \textit{with } t \in [0, 1]\,. \tag{2.3.7}$$

Proof. If $q = \infty$, then $t = p/r$ and $|f|^r \leq \|f\|_\infty^{r-p} |f|^p$. Hence

$$\|f\|_r \leq \|f\|_\infty^{1-\frac{p}{r}} \|f\|_p^{\frac{p}{r}} = \|f\|_p^t \|f\|_\infty^{1-t}\,.$$

So, suppose now that $q < \infty$. Consider the conjugate exponents $p/(tr)$, $q/((1-t)r)$; see (2.3.7). Then by applying Hölder's inequality (see Theorem 2.3.12), it follows

$$\|f\|_r^r = \int_X |f|^r d\mu = \int_X |f|^{tr} |f|^{(1-t)r} d\mu \leq \|f\|_p^{tr} \|f\|_q^{(1-t)r}\,,$$

which gives $\|f\|_r \leq \|f\|_p^t \|f\|_q^{1-t}$. $\square$

In finite measure spaces, by using Hölder's inequality, we can show that the L^p-spaces decrease as p increases.

Proposition 2.3.18. *If (X, Σ, μ) is a finite measure space and $1 \leq p \leq q \leq \infty$, then $L^q(X) \subseteq L^p(X)$ and $\|f\|_p \leq \|f\|_q \mu(X)^{1/p-1/q}$.*

Proof. First assume that $q = \infty$. Then for $f \in L^\infty(X)$ we have

$$\|f\|_p^p = \int_X |f|^p \, d\mu \le \|f\|_\infty^p \mu(X) .$$

Next assume that $q < \infty$. Consider the conjugate exponents q/p and $q/(q-p)$ and apply Hölder's inequality for them and $f \in L^p(X)$ as well as 1. This gives

$$\|f\|_p^p = \int_X |f|^p \, d\mu \le \| |f|^p \|_{\frac{q}{p}} \|1\|_{\frac{q}{p-q}} = \|f\|_q^p \mu(X)^{\frac{1}{p}-\frac{1}{q}} < +\infty . \qquad \square$$

Now we turn our attention to the Minkowski inequality; see Theorem 2.3.14. Evidently this inequality implies that $(L^p(X), \|\cdot\|_p)$ with $1 \le p \le \infty$ is a normed space. In fact, it is a complete normed space, that is, a Banach space.

Theorem 2.3.19. *If (X, Σ, μ) is a measure space and $1 \le p \le \infty$, then $(L^p(X), \|\cdot\|_p)$ is a Banach space.*

Proof. First assume that $p = \infty$. Let $\{f_n\}_{n\ge1} \subseteq L^\infty(X)$ be a Cauchy sequence. From Definition 2.3.10 we obtain

$$|f_n(x) - f_m(x)| \le \|f_n - f_m\|_\infty \quad \text{for } \mu\text{-a.a. } x \in X \text{ and for all } n, m \in \mathbb{N} .$$

This gives $\{f_n(x)\}_{n\ge1} \subseteq \mathbb{R}$ is a Cauchy sequence for all $x \in X \setminus A$ with $\mu(A) = 0$. Then, for all $x \in X \setminus A$, $f_n(x) \to f(x)$. Let $f(x) = 0$ for $x \in A$. From Proposition 2.2.12 we know that f is Σ-measurable and

$$|f(x) - f_m(x)| \le \sup_{n\ge m} \|f_n - f_m\|_\infty \le 1$$

for $m \in \mathbb{N}$ large enough and for all $x \in X \setminus A$. This yields $\|f\|_\infty \le \|f_m\|_\infty + 1$ for $m \in \mathbb{N}$ large enough. Hence, $f \in L^\infty(X)$ and so $L^\infty(X)$ is a Banach space.

Next assume that $1 \le p < \infty$. Let $\{f_n\}_{n\ge1} \subseteq L^p(X)$ be a Cauchy sequence. Recall that a Cauchy sequence is convergent if it has a convergent subsequence. So we may assume that

$$\|f_m - f_n\|_p < \frac{1}{2^n} \quad \text{for all } n \in \mathbb{N} \text{ and for all } m > n \text{ with } m \in \mathbb{N} . \tag{2.3.8}$$

Let $A(n) = \{x \in X : |f_n(x) - f_{n+1}(x)| \ge 1/n^2\}$. Then $\chi_{A(n)} 1/n^2 \le |f_n - f_{n+1}|$ for all $n \in \mathbb{N}$. Thus, because of (2.3.8),

$$\mu(A(n)) \frac{1}{n^{2p}} \le \int_X |f_n - f_{n+1}|^p \, d\mu < 2^{-np} \quad \text{for all } n \in \mathbb{N} .$$

Therefore

$$\sum_{n\ge1} \mu(A(n)) \le \sum_{n\ge1} \frac{n^{2p}}{2^{np}} < +\infty .$$

Let $C(n) = \bigcup_{m \geq n} A(m)$. Then $\{C(n)\}_{n \geq 1}$ is decreasing and $\mu(C(n)) \to 0$ as $n \to \infty$. Hence, if $C = \bigcap_{n \geq 1} C(n)$, then $\mu(C) = 0$ and for $x \in X \setminus C$ we have

$$|f_n(x) - f_m(x)| \leq \frac{1}{n^2} \quad \text{for all } n \in \mathbb{N} \text{ large enough}\,.$$

Then for any $m > n$ it holds that $|f_m(x) - f_n(x)| \leq \sum_{k \geq n} 1/k^2 \to 0$ as $n \to \infty$. So it follows that, for μ-a.a. $x \in X$, $\{f_n\}_{n \geq 1}$ is a Cauchy sequence and so it converges to some $f(x)$. On the exceptional μ-null set, we put $f(x) = 0$. Clearly f is measurable and by Fatou's Lemma (see Theorem 2.3.6), one gets

$$\int_X |f|^p \, d\mu \leq \liminf_{n \to \infty} \int_X |f_n|^p \, d\mu < \infty$$

since a Cauchy sequence is bounded. Hence, $f \in L^p(X)$.

Similarly, we obtain

$$\int_X |f - f_n|^p \, d\mu \leq \liminf_{m \to \infty} \int_X |f_m - f_n|^p \, d\mu\,,$$

which implies that $f_n \to f$ in $L^p(X)$. $\square$

A useful consequence of the result above is the following corollary.

Corollary 2.3.20. *If (X, Σ, μ) is a measure space, $\{f_n\}_{n \geq 1} \subseteq L^p(X)$ with $1 \leq p \leq \infty$, and $f_n \to f$ in $L^p(X)$, then there is a subsequence $\{f_{n_k}\}_{k \geq 1}$ of $\{f_n\}_{n \geq 1}$ such that $f_{n_k}(x) \to f(x)$ μ-a.e.*

Example 2.3.21. We have to pass to a subsequence to get pointwise convergence. To see this, consider the sequence $f_k = \chi_{[(i-1)/n, i/n]}$ for $k = i + (n(n-1))/2$ with $n \in \mathbb{N}$ and $i = 1, \ldots, n$. Then $\int_0^1 f_k^p \, d\lambda = 1/n \to 0$, that is, $f_n \to 0$ in $L^p[0, 1]$. However, $\liminf_{k \to \infty} f_k(x) = 0 < 1 = \limsup_{k \to \infty} f_k(x)$ for all $x \in [0, 1]$ and so we do not have pointwise convergence.

The next result provides a useful dense subset of the Banach space $L^p(X)$. It is a straightforward consequence of Proposition 2.2.18.

Proposition 2.3.22. *If (X, Σ, μ) is a measure space, then the set of simple functions in $L^p(X)$ is dense in $L^p(X)$ for $1 \leq p \leq \infty$.*

We continue with the examination of the Banach spaces $L^p(X)$ for $1 \leq p \leq \infty$. Next we examine under what conditions we can have separability of $L^p(X)$. We start with a definition.

Definition 2.3.23. Let (X, Σ, μ) be a measure space. On Σ we define the semimetric

$$d_\mu(A, B) = \mu(A \triangle B) \quad \text{for all } A, B \in \Sigma\,.$$

According to Remark 1.5.2 if we introduce on Σ the equivalence relation $\sim$ defined by $A \sim B$ if and only if $\mu(A \triangle B) = 0$, then, on $\Sigma(\mu) = \Sigma/\sim$, d_μ is a metric. Clearly we have

$$d_\mu(A, B) = \|\chi_A - \chi_B\|_1 \quad \text{for all } A, B \in \Sigma(\mu) \,.$$

Proposition 2.3.24. *If (X, Σ, μ) is a measure space, then $(\Sigma(\mu), d_\mu)$ is a separable metric space if and only if the Banach space $L^1(X)$ is separable.*

Proof. $\Longrightarrow$: Let $\{A_k\}_{k\geq 1} \subseteq \Sigma(\mu)$ be a countable d_μ-dense subset. Then the set of all functions that are finite linear combinations of $\{\chi_{A_k}\}_{k\geq 1}$ with rational coefficients is a countable dense subset of $L^1(X)$. Hence $L^1(X)$ is separable.

$\Longleftarrow$: By identifying an element of Σ with its characteristic function, we see that $\Sigma(\mu)$ can be viewed as a subset of $L^1(X)$. Then the separability of $L^1(X)$ implies the separability of $\Sigma(\mu)$. ☐

The next proposition provides a condition for the separability of $(\Sigma(\mu), d_\mu)$.

Proposition 2.3.25. *If (X, Σ, μ) is a finite measure space and $\Sigma = \sigma(\mathcal{L})$ with $\mathcal{L}$ being countable, then $(\Sigma(\mu), d_\mu)$ is separable.*

Proof. Note that the ring generated by $\mathcal{L}$ is still countable. So we may assume that $\mathcal{L}$ is a ring. Then, using Problem 2.3, for every $A \in \Sigma(\mu)$ we can find $B \in \mathcal{L}$ such that $d_\mu(A, B) = \mu(A \triangle B) \leq \varepsilon$. Hence $\mathcal{L}$ is d_μ-dense in $\Sigma(\mu)$ and so $(\Sigma(\mu), d_\mu)$ is separable. ☐

Corollary 2.3.26. *If X is a separable metric space, $\Sigma = \mathcal{B}(X)$ and μ is a finite measure on Σ, then $(\Sigma(\mu), d_\mu)$ is separable.*

In fact combining Propositions 2.3.18, 2.3.24, and 2.3.25, we can state the following result.

Proposition 2.3.27. *If (X, Σ, μ) is a σ-finite measure space, $\Sigma = \sigma(\mathcal{L})$ with $\mathcal{L}$ countable and $\mathfrak{a}$ is the smallest algebra containing $\mathcal{L}$, then the simple functions of the form $s = \sum_{k=1}^n a_k \chi_{A_k}$ with $n \in \mathbb{N}, a_k \in \mathbb{Q}, A_k \in \mathfrak{a}, \mu(A_k) < \infty, k = 1, \ldots, n$ form a countable dense subset of $L^p(X)$ for $1 \leq p < \infty$. In particular, $L^p(X)$ is separable for $1 \leq p < \infty$.*

For the space $L^\infty(X)$ we show that it is not separable. In order to show this first we mention the following decomposition result, which can be found in Dudley [90, p. 82].

Proposition 2.3.28. *If (X, Σ, μ) is a σ-finite measure space, then $\mu = \mu_a + \mu_d$ with μ_a purely atomic and μ_d nonatomic. Moreover the atoms on which μ_a is defined are at most countable.*

We can use this result to establish the nonseparability of $L^\infty(X)$.

Proposition 2.3.29. *If (X, Σ, μ) is a σ-finite measure space, then the Banach space $L^\infty(X)$ is not separable.*

Proof. Applying Proposition 2.3.28, we split X into its atomic part X_a and its nonatomic (diffuse) part X_d. We consider two distinct cases: (a) X_d is not μ-null. (b) X_d is μ-null.

Suppose that (a) holds. Then for each $\eta \in (0, \mu(X_d))$ there exists $A_\eta \in \Sigma$ such that $\mu(A_\eta) = \eta$; see Proposition 2.1.32. Then $\{A_\eta\}_{\eta\in(0,\mu(X_d))}$ is an uncountable set of distinct Σ-sets, that is, $\mu(A_\eta \triangle A_{\eta'}) > 0$ if $\eta \neq \eta'$. Let

$$U_\eta = \left\{f \in L^\infty(X) \colon \|f - \chi_{A_\eta}\|_\infty < \frac{1}{2}\right\}, \quad \eta \in (0, \mu(X_d)) = I .$$

Then $\{U_\eta\}_{\eta\in I}$ is an uncountable family of nonempty, open, and mutually disjoint sets in $L^\infty(X)$. This means that $L^\infty(X)$ is not separable. Indeed, if $L^\infty(X)$ were separable, then there would be a countable dense set $\{f_n\}_{n\geq1} \subseteq L^\infty(X)$. For each $\eta \in I$ we have $U_\eta \cap \{f_n\}_{n\geq1} \neq \emptyset$. So we can choose $n(\eta) \in \mathbb{N}$ such that $f_{n(\eta)} \in U_\eta$. The map $\eta \to n(\eta)$ is injective; recall that the sets are mutually disjoint. Therefore I is countable, a contradiction. The case (b) follows from Proposition 2.3.28. ☐

The main convergence theorem in the theory of Lebesgue integration is the "Lebesgue Dominated Convergence Theorem"; see Theorem 2.3.8. Two of the main ingredients in that result are:

- $f_n(x) \to f(x)$ μ-a.e. as $n \to \infty$ (the pointwise convergence of the sequence);
- $|f_n(x)| \leq h(x)$ for μ-a.a. $x \in X$ and for all $n \in \mathbb{N}$ with $h \in L^1(X)$ (existence of a dominating integrable function).

Both can be weakened. To weaken the pointwise convergence requirement we introduce the following convergence concept.

Definition 2.3.30. Let (X, Σ, μ) be a measure space. A sequence $f_n \colon X \to \mathbb{R}^*$ with $n \in \mathbb{N}$ of Σ-measurable functions **converges in measure** to a Σ-measurable function f if for every $\varepsilon > 0$

$$\mu(\{x \in X \colon |f_n(x) - f(x)| \geq \varepsilon\}) \to 0 \quad \text{as } n \to \infty .$$

We denote the convergence in measure by $f_n \xrightarrow{\mu} f$.

If μ is a probability measure, that is, $\mu(X) = 1$, then we say that the sequence $\{f_n\}_{n\geq1}$ **converges in probability to** f.

We say that the sequence $\{f_n\}_{n\geq1}$ is a **Cauchy sequence in measure** if for every $\varepsilon > 0$,

$$\lim_{n,m\to\infty} \mu(\{x \in X \colon |f_n(x) - f_m(x)| \geq \varepsilon\}) = 0 .$$

The following proposition is a straightforward consequence of the definition above.

Proposition 2.3.31. *If (X, Σ, μ) is a measure space, then the following hold:*

(a) $f_n \xrightarrow{\mu} f$ *and* $h_n \xrightarrow{\mu} h$ *imply* $\eta f_n + \vartheta h_n \xrightarrow{\mu} \eta f + \vartheta h$ *for all* $\eta, \vartheta \in \mathbb{R}$;

(b) $f_n \xrightarrow{\mu} f$ *implies* $f_n^\pm \xrightarrow{\mu} f^\pm$ *and* $|f_n| \xrightarrow{\mu} |f|$;

(b) $f_n \xrightarrow{\mu} f$ *and* $f_n \xrightarrow{\mu} g$ *imply* $f = g$ μ*-a.e.*

Proposition 2.3.32. *If (X, Σ, μ) is a finite measure space and $f_n \to f$ μ-a.e., then $f_n \xrightarrow{\mu} f$.*

Proof. For every $n \in \mathbb{N}$, let

$$A_n = \{x \in X : |f_n(x) - f(x)| \geq \varepsilon\} = \left\{x \in X : \frac{|f_n(x) - f(x)|}{1 + |f_n(x) - f(x)|} \geq \frac{\varepsilon}{1 + \varepsilon}\right\} . \tag{2.3.9}$$

This gives $\mu(A_n) \leq (1 + \varepsilon)/\varepsilon \int_X (|f_n - f|)/(1 + |f_n - f|) d\mu$ by the Markov inequality; see Proposition 2.2.41. But from the Lebesgue Dominated Convergence Theorem (see Theorem 2.3.8), it follows

$$\frac{1 + \varepsilon}{\varepsilon} \int_X \frac{|f_n - f|}{1 + |f_n - f|} d\mu \to 0 \quad \text{as } n \to \infty .$$

Hence $\mu(A_n) \to 0$ and so $f_n \xrightarrow{\mu} f$; see (2.3.9). ☐

In fact in finite measure spaces convergence in measure is strictly weaker than pointwise convergence.

Example 2.3.33. Let $X = [0, 1]$, $\Sigma = \mathcal{B}([0, 1])$, $\mu = \lambda|_{[0,1]}$ with λ being the Lebesgue measure on $\mathbb{R}$. Consider the sequence of Σ-measurable functions

$$f_n(x) = \chi_{\left[\frac{i}{2^k}, \frac{i+1}{2^k}\right]}(x) \quad \text{for all } i \in \{0, 1, \ldots, 2^k - 1\}, n = i + 2^k .$$

It follows that

$$\lambda(\{x \in [0, 1] : |f_n(x)| \geq \varepsilon\}) = \frac{1}{2^k} \to 0 \quad \text{as } n = n(k) \to +\infty .$$

Hence, $f_n \xrightarrow{\mu} 0$. But the pointwise limit of the f_n's does not exist at any $x \in [0, 1]$.

The following is a variant of the Markov inequality (see Proposition 2.2.41) and is known as the "Chebyshev inequality."

Proposition 2.3.34 (Chebyshev inequality). *If (X, Σ, μ) is a measure space, $f \in L^p(X)$, $1 \leq p < \infty$, and $\lambda > 0$, then*

$$\mu(\{x \in X : |f(x)| \geq \lambda\}) \leq \frac{1}{\lambda^p} \|f\|_p^p .$$

Proof. Let $A_\lambda = \{x \in X : |f(x)| \geq \lambda\}$. Then $\|f\|_p^p \geq \int_{A_\lambda} |f|^p d\mu \geq \lambda^p \mu(A_\lambda)$. ☐

Using the Chebyshev inequality we can compare convergence in $L^p(X)$ for $1 \leq p < \infty$ with convergence in measure.

Proposition 2.3.35. *If (X, Σ, μ) is a measure space, $\{f_n\}_{n\geq 1} \subseteq L^p(X)$ with $1 \leq p < \infty$, and $\|f_n - f\|_p \to 0$, then $f_n \xrightarrow{\mu} f$.*

Proof. Applying the Chebyshev inequality (see Proposition 2.3.34) yields the assertion of the proposition. ☐

Although convergence in measure is strictly weaker than pointwise convergence, we can always extract from any convergent sequence in measure a pointwise convergent subsequence.

Proposition 2.3.36. *If (X, Σ, μ) is a measure space and $f_n \xrightarrow{\mu} f$, then there exists a subsequence $\{f_{n_k}\}_{k\geq 1} \subseteq \{f_n\}_{n\geq 1}$ such that $f_{n_k} \to f$ μ-a.e.*

Proof. Since $f_n \xrightarrow{\mu} f$ there is a strictly increasing sequence $\{k_n\}_{n\geq 1} \subseteq \mathbb{N}$ such that

$$\mu\left(\left\{x \in X\colon |f_k(x) - f(x)| \geq \frac{1}{n}\right\}\right) < \frac{1}{2^n} \quad \text{for all } k \geq k_n \,.$$

For each $n \in \mathbb{N}$, let $A_n = \{x \in X\colon |f_{k_n}(x) - f(x)| \geq 1/n\} \in \Sigma$. We set $A = \bigcap_{k\geq 1} \bigcup_{n\geq k} A_n \in \Sigma$. Then we have

$$\mu(A) \leq \mu\left(\bigcup_{n\geq k} A_n\right) \leq \sum_{n\geq k} \mu(A_n) \leq \frac{1}{2^{k+1}} \quad \text{for every } k \in \mathbb{N} \,.$$

Hence, $\mu(A) = 0$.

If $x \notin A$, then there exists $k_0 \in \mathbb{N}$ such that $x \notin \bigcup_{n\geq k_0} A_n$ and so $|f_{k_n}(x) - f(x)| < 1/n$ for all $n \geq k_0$. Thus $f_{k_n}(x) \to f(x)$ for all $x \notin A$ with $\mu(A) = 0$. □

Definition 2.3.37. Let (X, Σ, μ) be a measure space and let $\mathcal{M}(X) = \{f\colon X \to \mathbb{R}^* : f$ is Σ-measurable$\}$. As before, we define $f \sim h$ if and only if $f = h$ μ-a.e. Then we set $L^0(X) = \mathcal{M}(X)/\sim$. When $\mu(X) < \infty$ on $L^0(X)$ we introduce the translation invariant metric

$$d_\mu(f, h) = \int_X \frac{|f - h|}{1 + |f - h|} d\mu \quad \text{for all } f, h \in L^0(X) \,. \tag{2.3.10}$$

Remark 2.3.38. It is easy to check that d_μ is a metric on $L^0(X)$. For the triangle inequality, use the elementary inequality that says that

$$a, b, c \in \mathbb{R}_+, \; a \leq b + c \quad \text{implies} \quad \frac{a}{1+a} \leq \frac{b}{1+b} + \frac{c}{1+c} \,.$$

In the next proposition we show that in finite measure spaces, convergence in measure is in fact a metric convergence.

Proposition 2.3.39. *If (X, Σ, μ) is a finite measure space and $\{f_n\}_{n\geq 1} \subseteq L^0(X)$, $f \in L^0(X)$, then $f_n \xrightarrow{\mu} f$ if and only if $f_n \xrightarrow{d_\mu} f$ in $L^0(X)$; see* (2.3.10).

Proof. In what follows for a given $\varepsilon > 0$ let

$$\begin{aligned} A_n &= \{x \in X\colon |f_n(x) - f(x)| \geq \varepsilon\} \\ &= \left\{x \in X\colon \frac{|f_n(x) - f(x)|}{1 + |f_n(x) - f(x)|} \geq \frac{\varepsilon}{1+\varepsilon}\right\}, n \in \mathbb{N} \,. \end{aligned} \tag{2.3.11}$$

Suppose that $f_n \xrightarrow{\mu} f$. Then we can find $n_0 \in \mathbb{N}$ such that

$$\mu(A_n) \leq \varepsilon \quad \text{for all } n \geq n_0 \,. \tag{2.3.12}$$

Then, because of (2.3.11) and (2.3.12), it follows

$$d_\mu(f_n, f) = \int_{A_n} \frac{|f_n - f|}{1 + |f_n - f|} d\mu + \int_{X \setminus A_n} \frac{|f_n - f|}{1 + |f_n - f|} d\mu$$
$$\le \mu(A_n) + \frac{\varepsilon}{1+\varepsilon}\mu(X \setminus A_n) \le (1 + \mu(X))\varepsilon$$

for all $n \ge n_0$. This gives $d_\mu(f_n, f) \to 0$ as $n \to \infty$.

Now assume that $f_n \xrightarrow{d_\mu} f$. Then $\varepsilon/(1+\varepsilon)\chi_{A_n} \le (f_n - f)/(1 + |f_n - f|)$ for all $n \in \mathbb{N}$; see (2.3.11). This implies $\mu(A_n) \le (1+\varepsilon)/(\varepsilon) d_\mu(f_n, f) \to 0$ as $n \to \infty$. Hence $f_n \xrightarrow{\mu} f$. ☐

The next notion will allow us to relax the dominating function requirement in the Lebesgue Dominated Convergence Theorem; see Theorem 2.3.8.

Definition 2.3.40. Let (X, Σ, μ) be a measure space and $\mathcal{F} \subseteq L^0(X)$. We say that $\mathcal{F}$ is **uniformly integrable** if for every $\varepsilon > 0$ there exists $D_\varepsilon \in \Sigma$ with $\mu(D_\varepsilon) < \infty$ and $\sup_{f\in\mathcal{F}} \int_{X\setminus D_\varepsilon} |f| d\mu \le \varepsilon$ as well as $\lim_{c\to\infty} \sup_{f\in\mathcal{F}} \int_{\{|f|\ge c\}} |f| d\mu = 0$.

Remark 2.3.41. In the literature one can find other definitions of uniform integrability that are equivalent to the definition above when $\mu(X) < \infty$. Some of these alternative definitions are examined in the exercises. In particular we mention the following equivalent definition for a set $\mathcal{F} \subseteq L^1(X)$ to be uniformly integrable:

(UI)'(a) $\mathcal{F} \subseteq L^1(X)$ is bounded, that is $\sup_{f\in\mathcal{F}} \|f\|_1 < \infty$;

(b) for every $\varepsilon > 0$ there exists $D_\varepsilon \in \Sigma$ with $\mu(D_\varepsilon) < \infty$ such that $\sup_{f\in\mathcal{F}} \int_{X\setminus D_\varepsilon} |f| d\mu \le \varepsilon$;

(c) for every $\varepsilon > 0$ there exists $\delta > 0$ such that $\mu(A) \le \delta$ implies $\sup_{f\in\mathcal{F}} \int_A |f| d\mu \le \varepsilon$.

The next result is a key property of the Lebesgue integral and will help us identify uniformly integrable subsets of $L^1(X)$. The result is referred to as the **absolute continuity** property of the integral.

Proposition 2.3.42. *If (X, Σ, μ) is a measure space and $f \in L^1(X)$, then for any given $\varepsilon > 0$ there exists $\delta = \delta(\varepsilon) > 0$ such that*

$$A \in \Sigma, \mu(A) \le \delta \quad \textit{implies} \quad \int_A |f| d\mu \le \varepsilon .$$

Proof. Since $f = f^+ - f^-$, without any loss of generality, we may assume that $f \ge 0$. Let $f_n = \min\{f, n\}$ with $n \in \mathbb{N}$. Then $f_n \nearrow f$ and so by the Monotone Convergence Theorem (Theorem 2.3.3), we have $\int_X f_n d\mu \nearrow \int_X f d\mu$. So, given $\varepsilon > 0$ there exists $n_0 = n_0(\varepsilon) \in \mathbb{N}$ such that

$$0 \le \int_X (f - f_n) d\mu \le \frac{\varepsilon}{2} \quad \text{for all } n \ge n_0 . \tag{2.3.13}$$

If $\delta = \varepsilon/(2n_0)$ and $A \in \Sigma$ satisfies $\mu(A) \le \delta$, then, due to (2.3.13),

$$\int_A f d\mu \le \int_A f_{n_0} d\mu + \int_X (f - f_{n_0}) d\mu \le \varepsilon .$$ ☐

Corollary 2.3.43. *If (X, Σ, μ) is a measure space and $\mathcal{F} \subseteq L^0(X)$ satisfies*

$$|f(x)| \leq h(x) \quad \text{for } \mu\text{-a.a. } x \in X \text{ and for all } f \in \mathcal{F} \text{ with } h \in L^1(X)\,,$$

then $\mathcal{F}$ is uniformly integrable. In particular, every finite set $\mathcal{F} \subseteq L^1(X)$ is uniformly integrable.

Now we can state the generalization of the Lebesgue Dominated Convergence Theorem; see Theorem 2.3.8. The result is known as the “Vitali Convergence Theorem” or “Extended Dominated Convergence Theorem.”

Theorem 2.3.44 (Vitali Convergence Theorem). *If (X, Σ, μ) is a measure space, $\{f_n\}_{n\geq 1} \subseteq L^1(X)$ is uniformly integrable and $f_n \xrightarrow{\mu} f$ as $n \to \infty$, then $f \in L^1(X)$ and $\|f_n - f\|_1 \to 0$. In particular, we have $\int_X f_n d\mu \to \int_X f d\mu$.*

Proof. On account of Proposition 2.3.36, we may assume that $f_n \to f$ μ-a.e. Given $\varepsilon > 0$, let $\delta > 0$ and $D_\varepsilon \in \Sigma$ be as postulated by (UI)$'$; see Remark 2.3.41. Moreover, thanks to Egorov’s Theorem, Theorem 2.2.32, we know that there exists $A_\varepsilon \in \Sigma$ with $A_\varepsilon \subseteq D_\varepsilon$ and $\mu(A_\varepsilon) \leq \delta$ such that

$$f_n \to f \quad \text{uniformly on } D_\varepsilon \setminus A_\varepsilon\,. \tag{2.3.14}$$

We have

$$\begin{aligned}\int_{D_\varepsilon} |f_n - f| d\mu &= \int_{A_\varepsilon} |f_n - f| d\mu + \int_{D_\varepsilon \setminus A_\varepsilon} |f_n - f| d\mu \\ &\leq \int_{A_\varepsilon} |f_n| d\mu + \int_{A_\varepsilon} |f| d\mu + \|f_n - f\|_{L^\infty(D_\varepsilon \setminus A_\varepsilon)} \mu(D_\varepsilon)\,.\end{aligned} \tag{2.3.15}$$

Note that according to (UI)$'$ (see also Definition 2.3.40), it holds that

$$\int_{A_\varepsilon} |f_n| d\mu \leq \varepsilon\,, \quad \int_{X\setminus D_\varepsilon} |f_n| d\mu \leq \varepsilon \quad \text{for all } n \in \mathbb{N}\,. \tag{2.3.16}$$

Moreover, by Fatou’s Lemma, one gets

$$\int_{A_\varepsilon} |f| d\mu \leq \varepsilon\,, \quad \int_{X\setminus D_\varepsilon} |f| d\mu \leq \varepsilon\,. \tag{2.3.17}$$

Taking (2.3.15), (2.3.16) and (2.3.17) into account it follows that

$$\begin{aligned}\int_X |f_n - f| d\mu &\leq \int_{X\setminus D_\varepsilon} |f_n| d\mu + \int_{X\setminus D_\varepsilon} |f| d\mu + \int_{D_\varepsilon} |f_n - f| d\mu \\ &\leq 4\varepsilon + \|f_n - f\|_{L^\infty(D_\varepsilon \setminus A_\varepsilon)} \mu(D_\varepsilon) \quad \text{for all } n \in \mathbb{N}\,.\end{aligned}$$

Hence, because of (2.3.14) and since $\mu(D_\varepsilon)$ is finite and $\varepsilon > 0$ is arbitrary, it follows that $f_n \to f$ in $L^1(X)$. ☐

Now that once we have the convergence theorems for the Lebesgue integral, we can establish the existence and uniqueness of the product measure.

So, let (X, Σ, μ) and $(Y, \mathcal{L}, \nu)$ be two measure spaces. Suppose that $\Sigma = \sigma(\mathfrak{a})$ and $\mathcal{L} = \sigma(\mathfrak{b})$. We want to define a measure ξ on rectangles of the form $A \times B$ with $A \in \mathfrak{a}$ and $B \in \mathfrak{b}$ such that

$$\xi(A \times B) = \mu(A)\nu(B) \quad \text{for all } A \in \mathfrak{a}, B \in \mathfrak{b} . \tag{2.3.18}$$

If the generators $\mathfrak{a}$ and $\mathfrak{b}$ are rich enough, we can have the uniqueness of the measure ξ satisfying (2.3.18).

Proposition 2.3.45. *If (X, Σ, μ) and $(Y, \mathcal{L}, \nu)$ are two measure spaces, $\Sigma = \sigma(\mathfrak{a})$, $\mathcal{L} = \sigma(\mathfrak{b})$ and*
(i) *$\mathfrak{a}$ and $\mathfrak{b}$ are closed under finite intersections;*
(ii) *there exists sequences $\{A_n\}_{n\geq1} \subseteq \mathfrak{a}$, $\{B_n\}_{n\geq1} \subseteq \mathfrak{b}$ with $A_n \nearrow X$, $B_n \nearrow Y$ and $\mu(A_n) < \infty$, $\nu(B_n) < \infty$ for all $n \in \mathbb{N}$,*

then there is at most on measure ξ on $\Sigma \bigotimes \mathcal{L}$ satisfying (2.3.18).

Proof. From Proposition 2.2.25 we know that $\Sigma \bigotimes \mathcal{L} = \sigma(\mathfrak{a} \times \mathfrak{b})$. Moreover we have

$$A_n \times B_n \nearrow X \times Y \quad \text{and} \quad \xi(A_n \times B_n) = \mu(A_n)\nu(B_n) < \infty \quad \text{for all } n \in \mathbb{N} .$$

Proposition 2.1.28 implies the uniqueness of ξ. ☐

Now we examine the issue of the existence of the product measure.

Theorem 2.3.46. *If (X, Σ, μ) and $(X, \mathcal{L}, \nu)$ are two σ-finite measure spaces, then the set function $\xi\colon \Sigma \times \mathcal{L} \to [0, +\infty]$ defined by $\xi(A \times B) = \mu(A)\nu(B)$ for all $A \in \Sigma, B \in \mathcal{L}$, extends uniquely to a σ-finite measure on $\Sigma \bigotimes \mathcal{L}$ such that*

$$\xi(C) = \int_Y \int_X \chi_C(x, y) d\mu d\nu = \int_X \int_Y \chi_C(x, y) d\nu d\mu \quad \text{for all } C \in \Sigma \bigotimes \mathcal{L}$$

and $x \to \chi_C(x, y)$, $y \to \chi_C(x, y)$, $x \to \int_Y \chi_C(x, y)d\nu$ and $y \to \int_X \chi_C(x, y)d\mu$ are measurable.

Proof. Uniqueness follows from Proposition 2.3.45. Consider sequences $\{A_n\}_{n\geq1} \subseteq \Sigma$ and $\{B_n\}_{n\geq1} \subseteq \mathcal{L}$ such that

$$A_n \nearrow X\,, \quad B_n \nearrow Y \quad \text{and} \quad \mu(A_n) < \infty\,, \quad \nu(B_n) < \infty \quad \text{for all } n \in \mathbb{N} .$$

Note that $C_n = A_n \times B_n \nearrow X \times Y$. For every $n \in \mathbb{N}$, let D_n be the family of all subsets $E \subseteq X \times Y$ such that

- $x \to \chi_{E\cap C_n}(x, y)$ and $y \to \chi_{E\cap C_n}(x, y)$ are measurable.
- $x \to \int_Y \chi_{E\cap C_n}(x, y)d\nu$ and $y \to \int_X \chi_{E\cap C_n}(x, y)d\mu$ are measurable.
- $\int_Y \int_X \chi_{E\cap C_n}(x, y)d\mu d\nu = \int_X \int_Y \chi_{E\cap C_n}(x, y)d\nu d\mu$.

It is a straightforward procedure to check that D_n is a Dynkin system; see Definition 2.1.7, which contains $\Sigma \times \mathcal{L}$. So, applying the Dynkin System Theorem (see Theorem 2.1.11) yields that $\Sigma \bigotimes \mathcal{L} \subseteq D_n$ for all $n \in \mathbb{N}$. Since $C_n \nearrow X \times Y$, Proposition 2.2.10 implies the measurability of $x \to \chi_C(x, y)$ and $y \to \chi_C(x, y)$ and then the Monotone Convergence Theorem (see Theorem 2.3.3) gives the measurability of $x \to \int_Y \chi_C(x, y)dv$ and of $y \to \int_X \chi_C(x, y)d\mu$.

Finally, if $E = X \times Y$, then we have that

$$C \to \xi(C) = \int_Y \int_X \chi_C(x, y)d\mu dv = \int_X \int_Y \chi_C(x, y)dv d\mu$$

is indeed a measure on $\Sigma \bigotimes \mathcal{L}$ and $\xi(A \times B) = \mu(A)v(B)$ for all $A \in \Sigma$ and for all $B \in \mathcal{L}$. □

Definition 2.3.47. Let (X, Σ, μ) and $(X, \mathcal{L}, v)$ be two σ-finite measure spaces. The unique measure ξ on $\Sigma \bigotimes \mathcal{L}$ produced in Theorem 2.3.46 is called the **product measure** of μ and v and is denoted by $\mu \times v$. The measure space $(X \times Y, \Sigma \bigotimes \mathcal{L}, \mu \times v)$ is called the **product measure space**.

Remark 2.3.48. Now we can define the Lebesgue measure λ^n on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ such that

$$\lambda^n(R) = \prod_{k=1}^{n}(b_k - a_k) \quad \text{for all rectangles } R = \prod_{k=1}^{n}[a_k, b_k) .$$

The next two theorems enable us to interchange the order of integration and to calculate integrals with respect to product measures using iteration. Their proofs are straightforward. Indeed, the results are true for characteristic functions, hence for simple functions. Then exploit the density of the simple functions to pass to the general case.

The first result is known as "Tonelli's Theorem."

Theorem 2.3.49 (Tonelli's Theorem). *If (X, Σ, μ) and $(X, \mathcal{L}, v)$ are two σ-finite measure spaces and if $f: X \times Y \to [0, \infty]$ is $\Sigma \bigotimes \mathcal{L}$-measurable, then the following hold:*

(a) *for all $y \in Y$, $x \to f(x, y)$ is Σ-measurable and for all $x \in X$, $y \to f(x, y)$ is $\mathcal{L}$-measurable;*

(b) *$x \to \int_X f(x, y)dv$ is Σ-measurable and $y \to \int_X f(x, y)d\mu$ is $\mathcal{L}$-measurable;*

(c) *$\int_{X\times Y} fd(\mu \times v) = \int_Y \int_X f(x, y)d\mu dv = \int_X \int_Y f(x, y)dv d\mu$.*

The second is known as "Fubini's Theorem."

Theorem 2.3.50 (Fubini's Theorem). *If (X, Σ, μ) and $(X, \mathcal{L}, v)$ are two σ-finite measure spaces, $f: X \times Y \to \mathbb{R}^*$ is $\Sigma \bigotimes \mathcal{L}$-measurable and at least one of the following three integrals is finite*

$$\int_{X\times Y} |f| d(\mu \times v) , \quad \int_Y \int_X |f| d\mu dv , \quad \int_X \int_Y |f| dv d\mu ,$$

then all three integrals are finite, $f \in L^1(X \times Y)$ and

(a) *$x \to f(x, y) \in L^1(X)$ for v-a.a. $y \in Y$;*

(b) *$y \to f(x, y) \in L^1(Y)$ for μ-a.a. $x \in X$;*

(c) $y \to \int_X f(x,y)d\mu \in L^1(Y)$;
(d) $x \to \int_Y f(x,y)d\nu \in L^1(X)$;
(e) $\int_{X\times Y} fd(\mu\times\nu) = \int_Y \int_X f(x,y)d\mu d\nu = \int_X \int_Y f(x,y)d\nu d\mu$.

2.4 Signed Measures and Radon–Nikodym Theorem

In this section we examine the notion of differentiating a measure ν with respect to another measure μ defined on the same σ-algebra. This differentiation theory can be developed more precisely if we extend the notion of measure and allow also negative values. This leads us to the concept of signed measure already introduced in Definition 2.1.22(f). For convenience, let us recall the definition here.

Definition 2.4.1. Let (X,Σ) be a measurable space and $\mu\colon \Sigma \to \mathbb{R}^*$ is a set function. We say that μ is a **signed measure** if the following hold:
(a) $\mu(\emptyset) = 0$;
(b) μ takes at most one of the values $+\infty$ and $-\infty$, that is, either $\mu\colon \Sigma \to (-\infty,+\infty]$ or $\mu\colon \Sigma \to [-\infty,+\infty)$;
(c) for every sequence $\{A_n\}_{n\ge1} \subseteq \Sigma$ of pairwise disjoint sets, we have

$$\mu\left(\bigcup_{n\ge1} A_n\right) = \sum_{n\ge1} \mu(A_n)\,. \tag{2.4.1}$$

Remark 2.4.2. If $\mu\left(\bigcup_{n\ge1} A_n\right)$ is finite in (2.4.1), then the sum on the right-hand side must converge independently of any rearrangement since the left-hand side is independent of the order of the terms. So the sum in (2.4.1) converges absolutely. Note that if μ_1, μ_2 are two measures on Σ and at least one of them is finite, then $\mu = \mu_1 - \mu_2$ is a signed measure.

Straightforward modifications in the proofs of Propositions 2.1.26 and 2.1.27 lead to the following characterization of signed measures.

Proposition 2.4.3. *If (X,Σ) is a measurable space and $\mu\colon \Sigma \to \mathbb{R}$ is an additive set function such that $\mu(\emptyset) = 0$, then μ is a signed measure if and only if one of the following equivalent properties holds:*
(a) $\{A_n\}_{n\ge1} \subseteq \Sigma$ *and* $A_n \nearrow A$ *imply* $\mu(A_n) \to \mu(A)$;
(b) $\{A_n\}_{n\ge1} \subseteq \Sigma$ *and* $A_n \searrow A$ *imply* $\mu(A_n) \to \mu(A)$;
(c) $\{A_n\}_{n\ge1} \subseteq \Sigma$ *and* $A_n \searrow \emptyset$ *imply* $\mu(A_n) \to 0$.

As we will see in the sequel, in order to study signed measures it is convenient to write them as differences of measures. For this reason we state the following definition.

Definition 2.4.4. Let (X,Σ) be a measurable space and $\mu\colon \Sigma \to \mathbb{R}^*$ is a signed measure. A set $A \in \Sigma$ is said to be a **positive** (resp. **negative**) set for μ, if $\mu(B) \ge 0$ (resp. $\mu(B) \le 0$) for all $B \in \Sigma$, $B \subseteq A$.

Example 2.4.5. Suppose that (X, Σ, μ) is a measure space and let $f: X \to \mathbb{R}^*$ be a Σ-measurable function such that at least one of $\int_X f^+ d\mu$ and $\int_X f^- d\mu$ is finite. Then the set function $\nu: \Sigma \to \mathbb{R}^*$ defined by $\nu(A) = \int_A f d\mu = \int_X f\chi_A d\mu$ is a signed measure and a set $A \in \Sigma$ is positive (resp. negative, null) for ν if $f \geq 0$ (resp. $f \leq 0$, $f = 0$) μ-a.e. on A.

It can happen that a set has positive μ-measure with μ being a signed measure but the set is not positive for μ.

Example 2.4.6. Let $X = \mathbb{R}$ and $\Sigma = \mathcal{B}(X)$. Consider $f: \mathbb{R} \to \mathbb{R}$ to be an odd function that is λ-integrable where λ denotes the Lebesgue measure. Assume that $f(x) > 0$ for all $x > 0$. Then $\nu(A) = \int_A f d\lambda$ is a signed measure (see Example 2.4.5), and any set of the form $[-a, b]$ with $0 < a < b$ has positive ν-measure without being a positive set for ν.

Next we will describe the structure of signed measures. We will show that X is the union of two disjoint sets, one positive and the other one negative. We start with a proposition for positive sets.

Proposition 2.4.7. *If (X, Σ) is a measurable space, $\mu: \Sigma \to \mathbb{R}^*$ is a signed measure and $A \in \Sigma$ is a positive set for μ, then any $B \in \Sigma$, $B \subseteq A$ is also a positive set for μ. Moreover, the union of any countable family of positive sets for μ is a positive set for μ.*

Proof. The first part of the conclusion is an immediate consequence from Definition 2.4.4.

Suppose that $\{A_n\}_{n\geq 1} \subseteq \Sigma$ are positive sets for μ. Let $C_n = A_n \setminus \bigcup_{k=1}^{n-1} A_k$. Then $C_n \in \Sigma$, $C_n \subseteq A_n$ and so from the first part C_n is positive for μ. Note that $\bigcup_{n\geq 1} A_n = \bigcup_{n\geq 1} C_n$ and the C_n's are mutually disjoint. So, if $B \in \Sigma$, $B \subseteq \bigcup_{n\geq 1} A_n$, then, by the σ-additivity of μ, $\mu(B) = \sum_{n\geq 1} \mu(B \cap C_n)$. Hence, $\mu(B) \geq 0$. So, we conclude that $\bigcup_{n\geq 1} A_n \in \Sigma$ is a positive set for μ. □

Now we can state the following important theorem for signed measures. The result is known as the “Hahn Decomposition Theorem.”

Theorem 2.4.8 (Hahn Decomposition Theorem). *If (X, Σ) is a measurable space and $\mu: \Sigma \to \mathbb{R}^*$ is a signed measure, then there exists a positive set $P \in \Sigma$ and a negative set $N \in \Sigma$ such that $X = P \cup N$ and $P \cap N = \emptyset$. Moreover, if P', N' is another such positive-negative decomposition of X, then $P \triangle P' = N \triangle N'$ is μ-null.*

Proof. Without any loss of generality we may assume that μ has values in $[-\infty, +\infty)$; see Definition 2.4.1. We define

$$\eta = \sup\left[\mu(A) : A \in \Sigma, A \text{ is a positive set for } \mu\right] \geq 0 . \tag{2.4.2}$$

Let $\{A_n\}_{n\geq 1} \subseteq \Sigma$ be a sequence of positive sets such that $\mu(A_n) \to \eta$. Let $P = \bigcup_{n\geq 1} A_n$. Then Propositions 2.4.7 and 2.4.3 imply that

$$P \text{ is positive for } \mu \text{ and } \mu(P) = \eta < +\infty . \tag{2.4.3}$$

Let $N = X \setminus P$. We claim that N is a negative set for μ. Arguing by contradiction, suppose that N is not negative for μ.

First we show that N cannot contain a positive set that is not μ-null. Indeed, if $A \subseteq N$ is positive and $\mu(A) > 0$, then $A \cup P$ is positive (see Proposition 2.4.7), and $\mu(A \cup P) = \mu(A) + \mu(P) \geq \eta$ (see (2.4.3)), a contradiction to the definition of $\eta \geq 0$ (see (2.4.2)).

Second, if $A \subseteq N$ and $\mu(A) > 0$, then there exists $B \in \Sigma$, $B \subseteq A$ with $\mu(B) > \mu(A)$. Indeed, since A is not positive, we can find $C \in \Sigma$, $C \subseteq A$ with $\mu(C) < 0$. Then if $B = A \setminus C$, we have $\mu(B) = \mu(A) - \mu(C) > \mu(A)$.

Since we have assumed that N is not a negative set for μ, we can produce a sequence $\{A_n\}_{n\geq1} \subseteq \Sigma$ with $A_n \subseteq N$ for all $n \in \mathbb{N}$ and a sequence $\{k_n\}_{n\geq1} \subseteq \mathbb{N}$ as follows:

k_1 is the smallest natural number for which we can find $B \in \Sigma, B \subseteq N$ with $\mu(B) > 1/k_1$. We set $A_1 = B$. Continuing inductively, let k_n be the smallest natural number for which we can find $B \in \Sigma, B \subseteq A_{n-1}$ with $\mu(B) \geq \mu(A_{n-1}) + 1/k_n$. We set $A_n = B$. Let $A = \bigcap_{n\geq1} A_n$. Then by Proposition 2.4.3, it follows that $\infty > \mu(A) = \lim_{n\to\infty} \mu(A_n) \geq \sum_{n\geq1} 1/k_n$, which gives $k_n \to \infty$. But as before, there exists $B \in \Sigma, B \subseteq A$ with $\mu(B) \geq \mu(A) + 1/k$ for some $k \in \mathbb{N}$. Then for large enough $n \in \mathbb{N}$, we have $k < k_n$ and $B \subseteq A_{n-1}$, a contradiction to the construction of the sequences $\{A_n\}_{n\geq1} \subseteq \Sigma$ and $\{k_n\}_{n\geq1} \subseteq \mathbb{N}$. It follows that N is negative for μ.

Finally suppose that P', N' is another such positive-negative pair. We have $P \setminus P' \subseteq P$ and $P \setminus P' \subseteq N'$, which yields that $P \setminus P'$ is both positive and negative for μ; see Proposition 2.4.7. This gives $\mu(P \setminus P') = 0$. Similarly we can show this for the set $P' \setminus P$. This completes the proof of the theorem. □

Remark 2.4.9. The pair (P, N) is called a **Hahn decomposition** for the signed measure μ.

The Hahn decomposition will lead us to a canonical decomposition of a signed measure. First we state a definition that is central in our considerations in this section.

Definition 2.4.10. Let (X, Σ) be a measurable space and $\mu, \nu \colon \Sigma \to [0, +\infty]$ be two measures.

(a) We say that μ and ν are **mutually singular** denoted by $\mu \perp \nu$ if there exists two disjoint sets $X_\mu, X_\nu \in \Sigma$ such that $X = X_\mu \cup X_\nu$ and for every $A \in \Sigma$, it holds that

$$\mu(A) = \mu(A \cap X_\mu) \quad \text{and} \quad \nu(A) = \nu(A \cap X_\nu)\,.$$

(b) We say that ν is **absolutely continuous** with respect to μ denoted by $\nu \ll \mu$ if for every $A \in \Sigma$ with $\mu(A) = 0$ it holds that $\nu(A) = 0$.

Proposition 2.4.11. *If (X, Σ) is a measurable space and $\mu, \nu \colon \Sigma \to [0, +\infty]$ are two measures with ν being finite, then $\nu \ll \mu$ if and only if for every $\varepsilon > 0$ there exists $\delta > 0$ such that*

$$A \in \Sigma \text{ and } \mu(A) \leq \delta \text{ imply } \nu(A) \leq \varepsilon\,. \tag{2.4.4}$$

Proof. $\Longrightarrow$: Arguing by contradiction suppose that the implication is not true. Then there exist $\varepsilon > 0$ and a sequence $\{A_n\}_{n\geq 1} \subseteq \Sigma$ such that

$$\mu(A_n) \leq \frac{1}{2^n} \text{ and } \nu(A_n) \geq \varepsilon \text{ for all } n \in \mathbb{N}\,. \tag{2.4.5}$$

Set $B_k = \bigcup_{n\geq k} A_n \in \Sigma$ and $B = \bigcap_{k\geq 1} B_k \in \Sigma$. Then

$$\mu(B) \leq \mu(B_k) \leq \sum_{n\geq k} \frac{1}{2^n} = \frac{1}{2^{k+1}} \to 0 \quad \text{as } k \to +\infty\,.$$

Hence,

$$\mu(B) = 0\,. \tag{2.4.6}$$

On the other hand since ν is finite, Proposition 2.1.24(f) gives

$$\nu(B) = \lim_{n\to\infty} \nu(B_n) \geq \lim_{n\to\infty} \nu(A_n) \geq \varepsilon\,;$$

see (2.4.5). This contradicts the hypothesis that $\nu \ll \mu$; see (2.4.6).

$\Longleftarrow$: If $A \in \Sigma$ with $\mu(A) = 0$, then $\nu(A) \leq \varepsilon$ for all $\varepsilon > 0$ and so $\nu(A) = 0$. Therefore $\nu \ll \mu$. ☐

Remark 2.4.12. From the proposition above, we infer that if ν is finite, then $\nu \ll \mu$ if and only if $\lim_{\mu(A)\to 0} \nu(A) = 0$.

If ν is not finite, then only the implication "$\Longleftarrow$" is valid in Proposition 2.4.11.

Example 2.4.13. Let $X = (0, 1)$, $\Sigma = \mathcal{B}((0, 1))$ and $\mu = \lambda$ be the Lebesgue measure on $(0, 1)$. Define $\nu(A) = \int_A 1/x d\lambda(x)$ for all $A \in \mathcal{B}((0, 1))$. Then $\nu \ll \mu$, but (2.4.4) fails.

Now we will use the Hahn decomposition of X to produce a canonical representation of a signed measure as the difference of two measures. The result is known as the "Jordan Decomposition Theorem."

Theorem 2.4.14 (Jordan Decomposition Theorem). *If (X, Σ) is a measurable space and $\mu\colon \Sigma \to \mathbb{R}^*$ is a signed measure, then there exist unique positive measures $\mu_+, \mu_-\colon \Sigma \to [0, +\infty]$ with at least one of them finite such that $\mu = \mu_+ - \mu_-$ and $\mu_+ \perp \mu_-$.*

Proof. Let (P, N) be a Hahn decomposition for μ; see Theorem 2.4.8. We define

$$\mu_+(A) = \mu(A \cap P) \quad \text{and} \quad \mu_-(A) = -\mu(A \cap N) \quad \text{for all } A \in \Sigma\,.$$

Then we have $\mu = \mu_+ - \mu_-$ and $\mu_+ \perp \mu_-$.

Suppose that (ξ_+, ξ_-) is another pair of measures such that $\mu = \xi_+ - \xi_-$ and $\xi_+ \perp \xi_-$. Let $A, B \in \Sigma$ such that $A \cap B = \emptyset$, $A \cup B = X$ and $\xi_+(B) = \xi_-(A) = 0$. Then $X = A \cup B$ is another Hahn decomposition for μ and so $\mu(P \triangle A) = 0$; see Theorem 2.4.8. Therefore for any $D \in \Sigma$ it follows that

$$\xi_+(D) = \xi_+(D \cap A) = \mu(D \cap A) = \mu(D \cap P) = \mu_+(D)\,,$$

which gives $\xi_+ = \mu_+$.

Similarly we show that $\xi_- = \mu_-$ and this proves the uniqueness of the difference decomposition. ☐

Definition 2.4.15. The measures μ_+ and μ_- from the proposition above are called the **positive** and **negative variations** of μ and $\mu = \mu_+ - \mu_-$ is called the **Jordan decomposition** of μ. The **total variation** of μ is the measure $|\mu|$ defined by $|\mu| = \mu_+ + \mu_-$.

Remark 2.4.16. For every $A \in \Sigma$ we have

$$\mu_+(A) = \sup\left[\mu(C)\colon C \in \Sigma, C \subseteq A, C \text{ is positive}\right] = \sup[\mu(C)\colon C \in \Sigma, C \subseteq A]\,,$$
$$\mu_-(A) = -\inf\left[\mu(C)\colon C \in \Sigma, C \subseteq A, C \text{ is negative}\right] = -\inf[\mu(C)\colon C \in \Sigma, C \subseteq A]\,,$$
$$|\mu|(A) = \sup\left[\sum_{k=1}^{n} |\mu(A_k)|\colon n \in \mathbb{N}, \{A_k\}_{k=1}^{n} \subseteq \Sigma \text{ are disjoint and } A = \bigcup_{k=1}^{n} A_k\right]\,.$$

Moreover, using the Jordan decomposition, we can define the Lebesgue integral with respect to a signed measure. So, let (X, Σ) be a measurable space and let $\mu\colon \Sigma \to \mathbb{R}^*$ be a signed measure. Consider $f\colon X \to \mathbb{R}^*$ a Σ-measurable function and $A \in \Sigma$. Suppose that at least one of the integrals $\int_A df\mu_+$ and $\int_A fd\mu_-$ is finite. Then the Lebesgue integral of f over A is defined as

$$\int_A fd\mu = \int_A fd\mu_+ - \int_A fd\mu_-\,.$$

If both integrals $\int_A fd\mu_+$, $\int_A fd\mu_-$ are finite, then we say that f is **Lebesgue integrable** with respect to μ over the set $A \in \Sigma$.

The Jordan decomposition established in Theorem 2.4.14 is minimal in the following sense.

Proposition 2.4.17. *If (X, Σ) is a measurable space, $\mu\colon \Sigma \to \mathbb{R}^*$ is a signed measure and $\mu = \xi_1 - \xi_2$ with $\xi_1, \xi_2\colon \Sigma \to [0, +\infty]$ measures, then $\xi_1 \geq \mu_+$ and $\xi_2 \geq \mu_-$.*

Proof. We have $\mu \leq \xi_1$. Hence, for all $A \in \Sigma$,

$$\mu_+(A) = \mu(A \cap P) \leq \xi_1(A \cap P) \leq \xi_1(A)\,.$$

Therefore $\mu_+ \leq \xi_1$. Similarly we show that $\mu_- \leq \xi_2$. $\square$

We extend the notions introduced in Definition 2.4.10 to signed measures.

Definition 2.4.18. Let (X, Σ) be a measurable space and $\mu, \nu\colon \Sigma \to \mathbb{R}^*$ be two signed measures.
(a) We say that μ and ν are **mutually singular** denoted by $\mu \perp \nu$ if $|\mu| \perp |\nu|$; see Definition 2.4.10(a).
(b) We say that ν is **absolutely continuous** with respect to μ denoted by $\nu \ll \mu$ if $|\nu| \ll |\mu|$; see Definition 2.4.10(b).

Remark 2.4.19. If μ is a signed measure, then $\mu_+ \perp \mu_-$.

The notion of mutual singularity is the antithesis of the notion of absolutely continuity.

Proposition 2.4.20. *If (X, Σ) is a measurable space and $\mu, \nu \colon \Sigma \to \mathbb{R}^*$ are signed measures, then $\mu \perp \nu$ and $\nu \ll \mu$ imply $\nu = 0$.*

Proof. Since by hypothesis $\mu \perp \nu$, there exist $A, B \in \Sigma$ with $A \cap B = \emptyset$, $X = A \cup B$, and $|\mu|(A) = |\nu|(B) = 0$; see Definition 2.4.18(a). By hypothesis we also have that $\nu \ll \mu$ and so $|\nu|(A) = 0$; see Definition 2.4.18(b). For every $C \in \Sigma$, it holds that

$$\begin{aligned} |\nu|(C) &= |\nu|(C \cap A) + |\nu|(C \cap B) \geq |\nu(C \cap A)| + |\nu(C \cap B)| \\ &\geq |\nu(C \cap A) + \nu(C \cap B)| = |\nu(C)| , \end{aligned}$$

by the additivity of ν. Hence, $|\nu(C)| = 0$ for all $C \in \Sigma$ and so $\nu \equiv 0$. □

Proposition 2.4.21. *If (X, Σ) is a measurable space and $\mu, \nu \colon \Sigma \to \mathbb{R}^*$ are signed measures, then $\nu \ll \mu$ if and only if $\nu_+ \ll \mu$ and $\nu_- \ll \mu$.*

Proof. $\Longrightarrow$: Suppose that $A \in \Sigma$ satisfies $|\mu|(A) = 0$. Then for $B \in \Sigma$, $B \subseteq A$ it follows $|\mu|(B) = 0$ and so $|\nu(B)| \leq |\nu|(B) = 0$. From Remark 2.4.16 we have

$$\nu_+(A) = \sup[\nu(B) \colon B \in \Sigma, B \subseteq A] = 0 .$$

Hence $\nu_+ \ll \mu$. Similarly we show that $\nu_- \ll \mu$.

$\Longleftarrow$: Suppose that $A \in \Sigma$ satisfies $|\mu|(A) = 0$. By hypothesis one gets $\nu_+(A) = \nu_-(A) = 0$. Recall that $|\nu| = \nu_+ + \nu_-$; see Definition 2.4.15. Therefore $|\nu|(A) = 0$ and we have proved that $\nu \ll \mu$. □

Remark 2.4.22. Evidently $\nu \ll \mu$ if and only if $A \in \Sigma$ with $|\nu|(A) = 0$ imply $\nu(A) = 0$.

In a similar fashion we also show the following facts about singular and absolutely continuous signed measures.

Proposition 2.4.23. *If (X, Σ) is a measurable space and $\mu, \nu, \xi \colon \Sigma \to \mathbb{R}^*$ are signed measures, then the following hold:*
(a) *$\mu \ll \xi$ and $\nu \ll \xi$ imply $|\mu| + |\nu| \ll \xi$;*
(b) *$\mu \perp \xi$ and $\nu \perp \xi$ imply $|\mu| + |\nu| \perp \xi$;*
(c) *$\mu \ll \xi$ and $\nu \ll \mu$ imply $\nu \ll \xi$;*
(d) *$\mu \perp \xi$ and $\nu \ll \mu$ imply $\nu \perp \xi$.*

Definition 2.4.24. Let (X, Σ) be a measurable space and $\mu \colon \Sigma \to \mathbb{R}^*$ is a signed measure.
(a) We say that μ is **finite** if $\mu(A) \in \mathbb{R}$ for every $A \in \Sigma$.
(b) We say that μ is σ**-finite** if there exists a sequence $\{A_n\}_{n \geq 1} \subseteq \Sigma$ such that $X = \bigcup_{n \geq 1} A_n$ and $\mu(A_n) \in \mathbb{R}$ for all $n \in \mathbb{N}$.

Remark 2.4.25. A signed measure μ is finite if and only if $|\mu(X)| < +\infty$. Moreover, we can assume in Definition 2.4.24(b) that the A_n's are mutually disjoint.

Proposition 2.4.26. *If (X, Σ) is a measurable space, $\nu \colon \Sigma \to \mathbb{R}^*$ is a finite signed measure and $\mu \colon \Sigma \to [0, +\infty]$ is a measure, then $\nu \ll \mu$ if and only if for every $\varepsilon > 0$ there exists $\delta > 0$ such that $A \in \Sigma$, $\mu(A) \leq \delta$ imply $|\nu(A)| \leq \varepsilon$.*

Proof. According to Definition 2.4.18(b), $\nu \ll \mu$ if and only if $|\nu| \ll \mu$ and recall that $|\nu(A)| \le |\nu|(A)$ for all $A \in \Sigma$. Then the conclusion follows from Proposition 2.4.11. □

Corollary 2.4.27. *If (X, Σ, μ) is a measure space and $f \in L^1(X)$, then for a given $\varepsilon > 0$ there exists $\delta = \delta(\varepsilon) > 0$ such that $A \in \Sigma$ with $\mu(A) \le \delta$ imply $\left|\int_A f d\mu\right| \le \varepsilon$.*

The technical result, which we prove next, will be used in the proof of the main structural result concerning signed measures, the so-called "Radon–Nikodym Theorem."

Lemma 2.4.28. *If (X, Σ) is a measurable space, μ, ν are measures on Σ with μ being σ-finite, $\nu \not\equiv 0$ and $\nu \ll \mu$, then there exist $\varepsilon > 0$ and $B \in \Sigma$ with $0 < \mu(B) < +\infty$ such that $\varepsilon\mu(C) \le \nu(C)$ for all $C \in \Sigma$, $C \subseteq B$, that is, B is a positive set for $\mu - \varepsilon\nu$.*

Proof. Let $\{A_n\}_{n\ge1} \subseteq \Sigma$ be disjoint sets such that $X = \bigcup_{n\ge1} A_n$ and $\mu(A_n) < +\infty$ for all $n \in \mathbb{N}$. Since $\nu \not\equiv 0$ we can find $m \in \mathbb{N}$ such that $\nu(A_m) > 0$. We choose $\varepsilon > 0$ small such that

$$\nu(A_m) - \varepsilon\mu(A_m) = (\nu - \varepsilon\mu)(A_m) > 0\,.$$

From Problem 2.53 we know that there exists $B \in \Sigma$, $B \subseteq A_m$ such that

$$(\nu - \varepsilon\mu)(B) > 0 \quad \text{and} \quad B \text{ is a positive set for } \nu - \varepsilon\mu\,. \tag{2.4.7}$$

Evidently $(\nu - \varepsilon\mu)(B) < +\infty$. Moreover, if $\mu(B) = 0$, then from (2.4.7) we have $\nu(B) > 0$, which contradicts the hypothesis that $\nu \ll \mu$. Therefore $\mu(B) > 0$. In addition, (2.4.7) implies that $\varepsilon\mu(C) \le \nu(C)$ for all $C \in \Sigma$, $C \subseteq B$. □

We saw in Example 2.4.5 that for a given measure space (X, Σ, μ) and $f \in L^1(X)$, the set function $\Sigma \ni A \xrightarrow{\nu} \int_A f d\mu$ is a signed measure. It is natural to ask whether the converse is true as well. Namely, if $\nu \ll \mu$, then can we find $f \in L^1(X, \mu)$ such that $d\nu = f d\mu$? The answer to this fundamental question is given by the so-called "Radon–Nikodym Theorem."

Theorem 2.4.29 (Radon–Nikodym Theorem). *If (X, Σ) is a measurable space, $\mu\colon \Sigma \to [0, +\infty]$ is a σ-finite measure, $\nu\colon \Sigma \to \mathbb{R}$ is a σ-finite signed measure and $\nu \ll \mu$, then there exists a unique up to equality μ-a.e. Σ-measurable function $f\colon X \to \mathbb{R}^*$ such that $\nu(A) = \int_A f d\mu$ for all $A \in \Sigma$.*

Proof. We know that ν_+, ν_- are finite measures on Σ and from Proposition 2.4.21, we know that $\nu_+ \ll \mu$ and $\nu_- \ll \mu$. Moreover, one has $\nu = \nu_+ - \nu_-$. Therefore without any loss of generality we may assume that ν is a σ-finite measure. It holds that $\Sigma \subseteq \Sigma_\mu \subseteq \Sigma_\nu$.

First assume that ν is finite. We introduce the set

$$\mathfrak{L} = \left\{ h \in L^1(X)\colon\, h \ge 0\ \mu\text{-a.e. and} \int_A h d\mu \le \nu(A) \text{ for all } A \in \Sigma_\mu \right\}\,. \tag{2.4.8}$$

We have $0 \in \mathfrak{L}$ and so $\mathfrak{L} \neq \emptyset$. Let $h_1, h_2 \in \mathfrak{L}$ and $A \in \Sigma_\mu$ and let

$$B = \{x \in A\colon\, h_1(x) \ge h_2(x)\}\,, \quad C = A \setminus B = \{x \in A\colon\, h_2(x) > h_1(x)\}\,.$$

Evidently $B, C \in \Sigma_\mu$, $A = B \cup C$ and $B \cap C = \emptyset$. Hence

$$\int_A \max\{h_1, h_2\} d\mu = \int_B \max\{h_1, h_2\} d\mu + \int_C \max\{h_1, h_2\} d\mu$$
$$= \int_B h_1 d\mu + \int_C h_2 d\mu \le \nu(B) + \nu(C) = \nu(A) .$$

Thus, $\max\{h_1, h_2\} \in \mathfrak{L}$. We define

$$\eta = \sup\left[\int_X h d\mu : h \in \mathfrak{L}\right] \le \nu(X) < +\infty ;$$

see (2.4.8). Let $\{h_n\}_{n\ge1} \subseteq \mathfrak{L}$ be such that $\lim_{n\to\infty} \int_X h_n d\mu = \eta$. We set $g_n = \max\{h_k\}_{k=1}^n$. Then from the previous part of the proof we have $\{g_n\}_{n\ge1} \subseteq \mathfrak{L}$ is increasing and $\int_X g_n d\mu \nearrow \eta$. From the Monotone Convergence Theorem (see Theorem 2.3.3) we know that there exists $g \in L^1(X, \mu)$ such that $g_n \nearrow g$ and $\int_X g d\mu = \eta$. We have

$$0 \le g_n \chi_A \nearrow g \chi_A \quad \text{and} \quad \int_X g_n \chi_A d\mu = \int_A g_n d\mu \le \nu(A) \quad \text{for all } n \in \mathbb{N} ,$$

which implies $\int_A g d\mu \le \nu(A)$ for all $A \in \Sigma_\mu$ and so $g \in \mathfrak{L}$.

Finally we show that $\nu(A) = \int_A g d\mu$ for all $A \in \Sigma_\mu$. Let

$$\xi(A) = \nu(A) - \int_A g d\mu \quad \text{for all } A \in \Sigma_\mu . \tag{2.4.9}$$

Then ξ is a measure on Σ_μ and $\xi \ll \mu$. Suppose that $\xi \not\equiv 0$. Then Lemma 2.4.28 implies that there exist $\varepsilon > 0$ and $B \in \Sigma_\mu$ such that

$$0 < \mu(B) < \infty \quad \text{and} \quad \varepsilon\mu(C) \le \xi(C) \quad \text{for all } C \in \Sigma_\mu, C \subseteq B . \tag{2.4.10}$$

Let $h = g + \varepsilon\chi_B$. Then $h \ge 0$ μ-a.e. and $h \in L^1(X, \mu)$. We have $\eta = \int_X g d\mu < \int_X h d\mu$, which gives

$$h \notin \mathfrak{L} . \tag{2.4.11}$$

On the other hand, for every $A \in \Sigma_\mu$, we derive, combining (2.4.8), (2.4.9), (2.4.10),

$$\int_A h d\mu = \int_A [g + \varepsilon\chi_B] d\mu = \int_A g d\mu + \varepsilon\mu(B \cap A) \le \int_A g d\mu + \xi(B \cap A)$$
$$\le \int_A g d\mu + \nu(B \cap A) - \int_{B\cap A} g d\mu = \int_{A\setminus B} g d\mu + \nu(B \cap A)$$
$$\le \nu(A \setminus B) + \nu(B \cap A) = \nu(A) .$$

This yields

$$h \in \mathfrak{L} . \tag{2.4.12}$$

Comparing (2.4.11) and (2.4.12), we reach a contradiction. Therefore

$$\nu(A) = \int_A g d\mu \quad \text{for all } A \in \Sigma \, .$$

Proposition 2.2.40(c) implies that $g \in L^1(X, \mu)$ is unique.

Now suppose that ν is σ-finite. Then we find $\{A_n\}_{n\geq 1} \subseteq \Sigma$ of disjoint sets such that $X = \bigcup_{n\geq 1} A_n$ with $\nu(A_n) < +\infty$ for all $n \in \mathbb{N}$. Let $\nu_n = \nu|_{A_n}$ for every $n \in \mathbb{N}$, that is, $\nu_n(B) = \nu(B \cap A_n)$ for all $n \in \mathbb{N}$. Evidently, ν_n is a finite measure on Σ and $\nu_n \ll \mu$. So, from the first part of the proof there exists a unique $g_n \in L^1(X, \mu)$ such that $\nu_n(B) = \int_B g_n d\mu$ for all $B \in \Sigma$. Recall that the A_n's are disjoint. We define $g = \sum_{n\geq 1} g_n \chi_{A_n}$ and we have that $g \colon X \to \mathbb{R}$ is Σ-measurable as well as

$$\nu(B) = \sum_{n\geq 1} \nu(B \cap A_n) = \sum_{n\geq 1} \int_B g_n \chi_{A_n} d\mu = \int_B g d\mu \, ,$$

see Theorem 2.3.5. □

Definition 2.4.30. The unique (up to equality μ-a.e.) function $g \colon X \to \mathbb{R}^*$ postulated by Theorem 2.4.29 is called the **Radon–Nikodym derivative** of ν with respect to μ and is denoted by $d\nu/d\mu = g$ or by $d\nu = g d\mu$. If ν is finite, then $g \in L^1(X, \mu)$ and if ν is a measure then $g \geq 0$ μ-a.e.

Theorem 2.4.29 leads to an interesting decomposition of ν. This result is known as the "Lebesgue Decomposition Theorem."

Theorem 2.4.31 (Lebesgue Decomposition Theorem). *If (X, Σ) is a measurable space, $\mu \colon \Sigma \to [0, +\infty]$ a σ-finite measure and $\nu \colon \Sigma \to \mathbb{R}^*$ is a σ-finite signed measure, then $\nu = \nu_a + \nu_s$ with $\nu_a \ll \mu$, $\nu_s \perp \mu$ and this decomposition is unique.*

Proof. Let $\xi = \mu + \nu$. Then ξ is a σ-finite measure on Σ and $\mu \ll \xi, \nu \ll \xi$. Applying Theorem 2.4.29, we can find Σ-measurable functions $g, h \colon X \to [0, +\infty]$ such that

$$\mu(A) = \int_A g d\xi \quad \text{and} \quad \nu(A) = \int_A h d\xi \quad \text{for all } A \in \Sigma \, . \tag{2.4.13}$$

Let $B = \{x \in X \colon g(x) > 0\}$ and $C = \{x \in X \colon g(x) = 0\}$. Then $B, C \in \Sigma$, $B \cap C = \emptyset$, $X = B \cup C$ and $\mu(C) = 0$; see (2.4.13). Let $\hat{\nu} = \nu|_C$, that is, $\hat{\nu}(E) = \nu(E \cap C)$ for all $E \in \Sigma$. Then $\hat{\nu}(B) = 0$ and so it follows that $\hat{\nu} \perp \mu$. Let $\tilde{\nu} = \nu|_B$, that is, $\tilde{\nu}(E) = \nu(E \cap B)$ for all $E \in \Sigma$. We obtain $\tilde{\nu}(E) = \nu(E \cap B) = \int_{E\cap B} h d\xi$; see (2.4.13) and $\nu = \tilde{\nu} + \hat{\nu}$.

We need to show that $\tilde{\nu} \ll \mu$. To this end, let $E \in \Sigma$ be such that $\mu(E) = 0$. Then $0 = \mu(E) = \int_E g d\xi$ (see (2.4.13)) and so, since $g \geq 0$ ξ-a.e., $g(x) = 0$ for ξ-a.a. $x \in E$. As $g|_{E\cap B} > 0$, we must have $\xi(E \cap B) = 0$, hence $\nu(E \cap B) = 0$ since $\nu \ll \xi$. Therefore $\tilde{\nu}(E) = \nu(E \cap B)$ and this shows that $\tilde{\nu} \ll \mu$.

Finally we show the uniqueness of this decomposition. So, suppose that (ν_a, ν_s) and (ν_a', ν_s') are two such decompositions. Then

$$\nu_a - \nu_a' = \nu_s' - \nu_s \, . \tag{2.4.14}$$

From Proposition 2.4.23 we have

$$\nu_a - \nu_a' \ll \mu \quad \text{and} \quad (\nu_s' - \nu_s) \perp \mu . \tag{2.4.15}$$

From (2.4.14), (2.4.15) and Proposition 2.4.20, we conclude that $\nu_a = \nu_a'$ and $\nu_s = \nu_s'$. Hence, the decomposition is unique. □

Definition 2.4.32. The decomposition $\nu = \nu_a + \nu_s$ provided by the previous theorem with $\nu_a \ll \mu$ as well as $\nu_s \perp \mu$ is called the **Lebesgue decomposition** of ν with respect to μ.

We conclude this section with two useful results concerning setwise limits of sequences of finite measures.

The first result is known as the "Vitali–Hahn–Saks Theorem."

Theorem 2.4.33 (Vitali–Hahn–Saks Theorem). *If (X, Σ) is a measurable space, $\{\nu_n\}_{n\geq 1}$ are finite signed measures, μ is a finite measure, $\nu_n \ll \mu$ for all $n \in \mathbb{N}$ and for all $A \in \Sigma$, the limit $\nu(A) = \lim_{n\to\infty} \nu_n(A)$ exists, then $\nu\colon \Sigma \to \mathbb{R}$ is a signed measure such that $\nu \ll \mu$.*

Proof. On account of the Jordan Decomposition Theorem (see Theorem 2.4.14) we may assume that the ν_n's are measures. First we show that $\{\nu_n\}_{n\geq 1}$ is in fact uniformly absolutely continuous with respect to μ, that is, for given $\varepsilon > 0$ there exists $\delta = \delta(\varepsilon) > 0$ such that $\mu(A) \leq \delta$ implies $\nu_n(A) \leq \varepsilon$ for all $n \in \mathbb{N}$; see Proposition 2.4.11.

Let $\Sigma(\mu)$ and d_μ be as in Definition 2.3.23. We claim that $(\Sigma(\mu), d_\mu)$ is a complete metric space. Indeed, let $S = \{\chi_A : A \in \Sigma_\mu\} \subseteq L^1(X, \mu)$. Let $\{\chi_{A_n}\}_{n\geq 1} \subseteq S$ and assume that $\chi_{A_n} \to f$ in $L^1(X, \mu)$. Then according to Corollary 2.3.20, there exists a subsequence $\{\chi_{A_{n_k}}\}_{k\geq 1}$ of $\{\chi_{A_n}\}_{n\geq 1}$ such that $\chi_{A_{n_k}}(x) \to f(x)$ for μ-a.a. $x \in X$. Therefore, range(f) = $\{0, 1\}$ and since f is measurable, there exists $A \in \Sigma_\mu$ such that $f = \chi_A$. This implies that S is a closed subset of $L^1(X, \mu)$, hence a complete metric space in its own right. But S is isometrically isomorphic to $(\Sigma(\mu), d_\mu)$. Therefore the latter is a complete metric space.

Note that for every $n \in \mathbb{N}$

$$|\nu_n(A) - \nu_n(B)| \leq \nu_n(A \triangle B) \quad \text{for all } A, B \in \Sigma \text{ and } \nu_n \ll \mu .$$

So, the map $\nu_n\colon \Sigma \to [0, +\infty)$ with $n \in \mathbb{N}$ is well-defined and continuous. We introduce the sets

$$D_k = \{A \in \Sigma\colon |\nu_n(A) - \nu_m(A)| \leq \varepsilon \text{ for all } n, m \geq k\} , k \in \mathbb{N} .$$

These sets are closed and $\Sigma = \bigcup_{k\in\mathbb{N}} D_k$. So, according to Theorem 1.5.68(b), we can find $k \in \mathbb{N}$ such that int $D_k \neq \emptyset$. This means that there exists $\tilde{A} \in D_k$ and $\delta_1 > 0$ such that $A \in \Sigma$ and $\mu(A \triangle \tilde{A}) \leq \delta_1$ imply $A \in D_k$. By hypothesis, $\nu_i \ll \mu$ for all $i \in \{1, \ldots, k\}$. So using Proposition 2.4.11 there is a $\delta \in (0, \delta_1]$ such that $A \in \Sigma$ with $\mu(A) \leq \delta$ imply $\nu_i(A) \leq \varepsilon$ for all $i \in \{1, \ldots, k\}$.

If $A \in \Sigma$ and $\mu(A) \le \delta$, then $\mu((A \cup \tilde{A}) \triangle \tilde{A}) \le \mu(A) \le \delta \le \delta_1$ and so

$$\begin{aligned} |\nu_n(A) - \nu_k(A)| &= \left|(\nu_n - \nu_k)(A \cup \tilde{A}) - (\nu_n - \nu_k)(\tilde{A} \setminus A)\right| \\ &\le \left|(\nu_n - \nu_k)(A \cup \tilde{A})\right| + \left|(\nu_n - \nu_k)(\tilde{A} \setminus A)\right| \le 2\varepsilon \end{aligned}$$

for all $n \ge k$. Therefore it follows that $A \in \Sigma$ with $\mu(A) \le \delta$ imply $\nu_n(A) \le 2\varepsilon + \nu_k(A) \le 3\varepsilon$ for all $n \in \mathbb{N}$, which is the uniform absolute continuity of $\{\nu_n\}_{n\ge1}$ with respect to μ.

Now let $\{A_n\}_{n\ge1} \subseteq \Sigma$ be mutually disjoint sets and $\varepsilon > 0$. We set $A = \bigcup_{n\ge1} A_n \in \Sigma$. Let $\delta > 0$ be as postulated by the uniform absolute continuity with respect to μ established in the first part of the proof. We choose $k \in \mathbb{N}$ such that $\mu(A \setminus \bigcup_{i=1}^k A_i) \le \delta$; see Proposition 2.1.24(e). This implies

$$\left|\nu_n(A) - \sum_{i=1}^m \nu_n(A_i)\right| = \left|\nu_n\left(A \setminus \bigcup_{i=1}^m A_i\right)\right| \le \varepsilon \quad \text{for all } n, m \ge k\,.$$

Hence

$$\left|\nu(A) - \sum_{i=1}^m \nu(A_i)\right| \le \varepsilon \quad \text{for all } m \ge k\,.$$

Since $\varepsilon > 0$ is arbitrary, it follows that $\nu(A) = \sum_{i\in\mathbb{N}} \nu(A_i)$ and so ν is a measure. Moreover, from the first part of the proof and Proposition 2.4.11 we have $\nu \ll \mu$. □

The next theorem, known as "Nikodym's Theorem", is an easy consequence of the theorem above.

Theorem 2.4.34 (Nikodym's Theorem). *If (X, Σ) is a measurable space and let $\{\nu_n\}_{n\ge1}$ be a sequence of nonzero finite measures defined on Σ such that the limit $\lim_{n\to\infty} \nu_n(A)$ exists for all $A \in \Sigma$, then $\nu(A) = \lim_{n\to\infty} \nu_n(A)$ with $A \in \Sigma$ is a finite measure.*

Proof. Consider the set function $\mu\colon \Sigma \to [0, +\infty)$ defined by

$$\mu(A) = \sum_{n\in\mathbb{N}} \frac{1}{2^n} \frac{\nu_n(A)}{\nu_n(X)} \quad \text{for all } A \in \Sigma\,.$$

Evidently μ is a finite measure on Σ and $\nu_n \ll \mu$ for all $n \in \mathbb{N}$. So, invoking Theorem 2.4.33, we conclude that ν is a finite measure on Σ. □

2.5 Regular and Radon Measures

In this section we investigate the connections between measure theory and topology. When we combine the measure theoretic and topological structures, we obtain stronger and more interesting results.

Throughout this section (X, τ) is a Hausdorff topological space. Additional conditions on X will be introduced as needed. By $C_c(X)$ we denote the space of all continuous functions $f: X \to \mathbb{R}$ with compact support. Recall that the support of f, denoted by $\operatorname{supp} f$, is defined to be the closure of the set $\{x \in X: f(x) \neq 0\}$.

Definition 2.5.1. The **Baire σ-algebra** of X, denoted by $\mathrm{Ba}(X)$, is defined to be the smallest σ-algebra on X, which makes all functions in $C_c(X)$ measurable. So, $\mathrm{Ba}(X)$ has as generators the sets $\{x \in X: f(x) \geq \eta\}$ with $f \in C_c(X)$ and $\eta \in \mathbb{R}$. These sets are known as **Baire sets**.

This new σ-algebra is most useful within the framework of locally compact spaces.

Lemma 2.5.2. *If X is locally compact, $K \subseteq X$ is compact and $W \subseteq X$ is open such that $K \subseteq W$, then we can find $U \in \tau \cap \mathrm{Ba}(X)$ and a compact G_δ-set C such that $K \subseteq U \subseteq C \subseteq W$.*

Proof. Proposition 1.4.66(c) says that there exists $D \in \tau$ being relatively compact such that $K \subseteq D \subseteq \overline{D} \subseteq W$. Then Proposition 1.4.68 implies that there is $f \in C_c(X)$ such that $f|_K = 1$ and $f|_{D^c} = 0$. Let $C = \{x \in X: f(x) \geq 1/2\}$. Then $C \subseteq X$ is compact, G_δ, $U = \{x \in X: f(x) > 1/2\} \in \tau$ and we have $K \subseteq U \subseteq C \subseteq W$. □

Corollary 2.5.3. *If X is locally compact, then $\tau \cap \mathrm{Ba}(X)$ is a basis for τ.*

Proof. Let $x \in X$ and $U \in \mathcal{N}(x)$. Then Lemma 2.5.2 implies that there exists $f \in C_c(X)$ such that $f(x) = 1$ and $f|_{U^c} = 0$. Consider the set $V = \{x \in X: f(x) > 1/2\}$. Then $V \in \tau \cap \mathrm{Ba}(X)$ and $V \subseteq U$. □

Now we can give an alternative characterization of $\mathrm{Ba}(X)$ when X is locally compact.

Theorem 2.5.4. *If X is locally compact, then*

$$\mathrm{Ba}(X) = \sigma(\{C \subseteq X: C \text{ is compact and a } G_\delta\text{-set}\}) .$$

Proof. Let $\mathcal{L} = \sigma(\{C \subseteq X: C \text{ is compact and a } G_\delta\text{-set}\})$. For every $f \in C_c(X)$ and $\eta > 0$, the set $\{x \in X: f(x) \geq \eta\}$ is compact and G_δ. Note that $\{f \geq \eta\} = \bigcap_{n \geq 1} \{f > \eta - 1/n\}$. Therefore $\{x \in X: f(x) \geq \eta\} \in \mathcal{L}$ for all $f \in C_c(X)$ and for all $\eta > 0$. For $\eta < 0$, we have $0 < -\eta + \eta/(2n) < -\eta$ and

$$\{f \geq \eta\} = \{f < \eta\}^c = \{-f > -\eta\}^c = \left(\bigcap_{n \geq 1} \left\{ -f \geq -\eta + \frac{\eta}{2n} \right\} \right)^c \in \mathcal{L} .$$

Moreover, note that $\{f \geq 0\} = \bigcap_{n \geq 1} \{f \geq -1/n\} \in \mathcal{L}$. So, every set $\{x \in X: f(x) \geq \eta\}$ for $f \in C_c(X)$ and $\eta \in \mathbb{R}$, belongs to $\mathcal{L}$ and we have

$$\mathrm{Ba}(X) \subseteq \mathcal{L} ; \tag{2.5.1}$$

see Definition 2.5.1. Now suppose that $K = \bigcap_{n \geq 1} W_n$ with $W_n \in \tau$ being compact. Lemma 2.5.2 implies that we can find $U_n \in \tau \cap \mathrm{Ba}(X)$ such that $K \subseteq U_n \subseteq W_n$ for all $n \in \mathbb{N}$. Then $K = \bigcap_{n \geq 1} U_n \in \mathrm{Ba}(X)$, which gives

$$\mathcal{L} \subseteq \mathrm{Ba}(X) . \tag{2.5.2}$$

From (2.5.1) and (2.5.2) we conclude that $\mathcal{L} = \mathrm{Ba}(X)$. □

Next we compare the Baire and Borel σ-algebras.

Theorem 2.5.5. (a) $\mathrm{Ba}(X) \subseteq \mathcal{B}(X)$
(b) *If X is locally compact, separable and metrizable, then* $\mathrm{Ba}(X) = \mathcal{B}(X)$.

Proof. (a) Just recall that every continuous function $f\colon X \to \mathbb{R}$ is Borel measurable.

(b) From Proposition 1.4.78 (see also Proposition 1.5.40), we know that X is σ-compact. Therefore, every closed subset of X is likewise σ-compact. It follows that it suffices to show that every compact set belongs to $\mathrm{Ba}(X)$. But Proposition 1.5.8 says that every compact set in X is G_δ. So, according to Theorem 2.5.4, it belongs to $\mathrm{Ba}(X)$ and we conclude that $\mathrm{Ba}(X) = \mathcal{B}(X)$. □

Using Proposition 1.4.66(d) we have at once the following result.

Proposition 2.5.6. *If X is locally compact and $\hat{\mathcal{B}}$ is a basis for τ, then* $\mathrm{Ba}(X) \subseteq \sigma(\hat{\mathcal{B}}) \subseteq \mathcal{B}(X)$.

The next theorem is the Baire counterpart of Proposition 2.2.26(b).

Theorem 2.5.7. *If X and Y are second countable, locally compact spaces, then* $\mathrm{Ba}(X\times Y) = \mathrm{Ba}(X) \bigotimes \mathrm{Ba}(Y)$.

Proof. Note that $X \times Y$ is locally compact. We define

$$\mathcal{M}(A) = \{B \subseteq Y\colon A \times B \in \mathrm{Ba}(X \times Y)\} .$$

It is routine to check that $\mathcal{M}(A)$ is a σ-ring for any A. Suppose that $C \subseteq X$ is compact and a G_δ-set. Then if $E \subseteq Y$ is compact and G_δ, then so is $C \times E \subseteq X \times Y$ and we infer that $\mathcal{M}(C)$ contains every compact G_δ-set in Y. Moreover, we have $Y \in \mathcal{M}(C)$; see Proposition 1.4.78 and Theorem 1.2.27. It follows that $\mathcal{M}(C)$ is a σ-algebra containing $\mathrm{Ba}(Y)$.

Let $\mathcal{L} = \{A \subseteq X\colon \mathrm{Ba}(Y) \subseteq \mathcal{M}(A)\}$. This family is closed under countable intersections and under complementation and we have seen above it contains every compact G_δ. Therefore

$$\mathrm{Ba}(X) \bigotimes \mathrm{Ba}(Y) \subseteq \mathrm{Ba}(X \times Y) . \tag{2.5.3}$$

On the other hand, from Corollary 2.5.3, we know that the family

$$\mathcal{B} = \{U \times V\colon U \subseteq X \text{ Baire open}, V \subseteq Y \text{ Baire open}\}$$

is a basis for $X \times Y$. Since $U \times V \in \mathrm{Ba}(X) \bigotimes \mathrm{Ba}(Y)$ it follows that $\sigma(\mathcal{B}) \subseteq \mathrm{Ba}(X) \bigotimes \mathrm{Ba}(Y)$. Then Proposition 2.5.6 gives

$$\mathrm{Ba}(X \times Y) \subseteq \mathrm{Ba}(X) \bigotimes \mathrm{Ba}(Y) . \tag{2.5.4}$$

From (2.5.3) and (2.5.4), we conclude that $\mathrm{Ba}(X \times Y) = \mathrm{Ba}(X) \bigotimes \mathrm{Ba}(Y)$. □

Definition 2.5.8. (a) A **(signed) Borel measure** is a (signed) measure defined on $\mathcal{B}(X)$.

(b) We say that a Borel measure μ is **regular** if for every $A \in \mathcal{B}(X)$

$$\begin{aligned}\mu(A) &= \inf\left[\mu(U)\colon U \subseteq X \text{ is open}, A \subseteq U\right] \quad \text{(outer regularity)}\\ &= \sup\left[\mu(C)\colon C \subseteq X \text{ is closed}, C \subseteq A\right] \quad \text{(inner regularity)}\,.\end{aligned}$$

(c) We say that a Borel measure μ is **compact regular** if for every $A \in \mathcal{B}(X)$

$$\mu(A) = \sup\left[\mu(K)\colon K \subseteq X \text{ is compact}, K \subseteq A\right]\,.$$

(d) We say that a Borel measure is a **Radon measure** if the following hold:
- $\mu(K) < +\infty$ for every compact $K \subseteq X$;
- $\mu(A) = \inf[\mu(U)\colon U \subseteq X$ is open, $A \subseteq U]$ for all $A \in \mathcal{B}(X)$;
- $\mu(A) = \sup[\mu(K)\colon K \subseteq X$ is compact, $K \subseteq A]$ for all $A \in \mathcal{B}(X)$.

For a signed Borel measure μ we say that μ is regular (resp. compact regular, Radon) if $|\mu|$ is such a measure or equivalently if μ_+ and μ_- have the corresponding properties.

Remark 2.5.9. Evidently two regular Borel measures are equal if and only if they coincide on the open or closed subsets. Similarly two compact regular measures are equal if and only if they coincide on the compact sets.

Proposition 2.5.10. *For finite Borel measures μ, outer and inner regularity are equivalent properties.*

Proof. Suppose that for all $A \in \mathcal{B}(X)$

$$\mu(A) = \inf[\mu(U)\colon U \subseteq X \text{ is open}, A \subseteq U]\,. \tag{2.5.5}$$

Taking Proposition 2.1.24(b) and (2.5.5) into account yields

$$\begin{aligned}\mu(X) - \mu(A) = \mu(A^c) &= \inf[\mu(U)\colon U \subseteq X \text{ is open}, A^c \subseteq U]\\ &= \mu(X) - \sup[\mu(C)\colon C \subseteq X \text{ is closed}, C \subseteq A]\,.\end{aligned}$$

Therefore, $\mu(A) = \sup[\mu(C)\colon C \subseteq X$ is closed, $C \subseteq A]$. Hence, outer regularity implies inner regularity.

In a similar way we show that the opposite implication holds as well. So, the two notions are equivalent. ☐

Theorem 2.5.11. *If $\mu\colon \mathcal{B}(X) \to [0, +\infty)$ is a finite, compact regular Borel measure, then μ is a Radon measure.*

Proof. Since every compact subset of X is closed, for every $A \in \mathcal{B}(X)$ we derive

$$\begin{aligned}\mu(A) &= \sup[\mu(K)\colon K \subseteq X \text{ is compact}, K \subseteq A]\\ &\le \sup[\mu(C)\colon C \subseteq X \text{ is closed}, C \subseteq A] \le \mu(A)\,.\end{aligned}$$

Hence,

$$\mu(A) = \sup[\mu(C)\colon\ C \subseteq X \text{ is closed, } C \subseteq A]\ . \tag{2.5.6}$$

From (2.5.6) and Proposition 2.5.10, we conclude that μ is a Radon measure. □

Theorem 2.5.12. *If X is metrizable and $\mu\colon \mathcal{B}(X) \to [0, +\infty)$ is a finite Borel measure, then μ is regular.*

Proof. Let $\mathcal{M} = \{A \in \mathcal{B}(X)\colon\ A \text{ is both outer and inner regular}\}$; see Definition 2.5.8(a). We are going to show that $\mathcal{M}$ is a σ-algebra containing all the open sets. Therefore $\mathcal{M} = \mathcal{B}(X)$.

Fact 1: $A \in \mathcal{M}$ implies $A^c \in \mathcal{M}$

This is immediate from the definition of $\mathcal{M}$. Recall that μ is finite and that $\mu(X) - \mu(A) = \mu(A^c)$; see Proposition 2.1.24(b).

Fact 2: $\{A_n\}_{n\geq 1} \subseteq \mathcal{M}$ implies $A = \bigcup_{n\geq 1} A_n \in \mathcal{M}$

For every $n \in \mathbb{N}$ there exist an open $U_n \subseteq X$ and a closed $C_n \subseteq X$ such that

$$C_n \subseteq A_n \subseteq U_n \quad \text{and} \quad \mu(U_n) \leq \mu(C_n) + \frac{\varepsilon}{2^n}\ . \tag{2.5.7}$$

Let $U = \bigcup_{n\geq 1} U_n$. Then $U \subseteq X$ is open and $A \subseteq U$. We know that $U \setminus A \subseteq \bigcup_{n\geq 1}(U_n \setminus A_n)$. Then, due to (2.5.7), this gives

$$\begin{aligned} 0 \leq \mu(U) - \mu(A) = \mu(U \setminus A) &\leq \sum_{n\geq 1} \mu(U_n \setminus A_n) \\ = \sum_{n\geq 1}(\mu(U_n) - \mu(A_n)) &\leq \sum_{n\geq 1} \frac{\varepsilon}{2^n} = \varepsilon\ . \end{aligned}$$

Hence,

$$\mu(A) = \inf[\mu(U)\colon\ U \subseteq X \text{ is open, } A \subseteq U] \quad (\text{outer regularity of } A)\ .$$

Let $C = \bigcup_{n\geq 1} C_n$. Arguing as above, we show that

$$\mu(A) \leq \mu(C) + \varepsilon\ . \tag{2.5.8}$$

For every $m \in \mathbb{N}$, let $\tilde{C}_m = \bigcup_{n=1}^m C_n$. Evidently $\tilde{C}_m$ is closed and $\tilde{C}_m \nearrow C$. Invoking Proposition 2.1.24(e), there exists $m \in \mathbb{N}$ such that $\mu(C) \leq \mu(\tilde{C}_m) + \varepsilon$ which gives, thanks to (2.5.8), that $\mu(A) \leq \mu(\tilde{C}_m) + 2\varepsilon$. This finally yields

$$\mu(A) = \sup[\mu(C)\colon\ C \subseteq X \text{ is closed, } C \subseteq A] \quad (\text{inner regularity of } A)\ .$$

Hence, $A \in \mathcal{M}$.

Fact 3: $\mathcal{M}$ contains all open sets

Let $U \subseteq X$ be open. Proposition 1.5.8 says that U is a F_σ-set. So, we can find closed subsets $\{C_n\}_{n\geq 1}$ of X such that $C_n \nearrow X$. Then $\mu(C_n) \nearrow \mu(X)$; see Proposition 2.1.24(e). Hence

$$\mu(U) = \sup[\mu(C)\colon\ C \subseteq X \text{ is closed, } C \subseteq U]\ ,$$

which gives $U \in \mathcal{M}$ since U is open.

Combining Facts 1–3 imply that $\mathcal{M} = \mathcal{B}(X)$. □

Proposition 2.5.13. *If X is metrizable and $\mu\colon \mathcal{B}(X) \to [0, +\infty)$ is a finite Borel measure, then μ is compact regular if and only if for every $\varepsilon > 0$ there exists a compact $K_\varepsilon \subseteq X$ such that $\mu(X) - \varepsilon \le \mu(K_\varepsilon)$.*

Proof. $\Longrightarrow$: This is immediate from Definition 2.5.8(c).

$\Longleftarrow$: From Theorem 2.5.12 we know that μ is regular. So, it suffices to show that for every closed $C \subseteq X$, we have

$$\mu(C) = \sup[\mu(K)\colon K \subseteq X \text{ is compact}, K \subseteq C]\,. \tag{2.5.9}$$

Arguing by contradiction suppose that there exists a closed $C \subseteq X$ such that (2.5.9) is not true. So we can find $\varepsilon > 0$ such that

$$\sup[\mu(K)\colon K \subseteq X \text{ is compact}, K \subseteq C] \le \mu(C) - \frac{\varepsilon}{2}\,. \tag{2.5.10}$$

For $K \subseteq X$ compact we have that $K \cap C \subseteq C$ is compact and, because of (2.5.10),

$$\mu(K) = \mu(K \cap C) + \mu(K \cap C^c) \le \mu(C) - \frac{\varepsilon}{2} + \mu(C^c) = \mu(X) - \frac{\varepsilon}{2}\,.$$

Since $K \subseteq X$ is arbitrary, we get a contradiction to our hypothesis. □

On Polish spaces all finite Borel measures are Radon measures.

Theorem 2.5.14. *If X is a Polish space and $\mu\colon \mathcal{B}(X) \to [0, +\infty)$ is a finite Borel measure, then μ is a Radon measure.*

Proof. On account of Theorem 2.5.11 we only need to show that μ is compact regular. Suppose that $D = \{x_k\}_{k\ge1} \subseteq X$ is dense. We consider the closed balls $\overline{B}_n(x_k) = \{x \in X\colon d(x, x_k) \le 1/n\}$ with $n, k \in \mathbb{N}$. Obviously $X = \bigcup_{k\ge1} \overline{B}_n(x_k)$ for every $n \in \mathbb{N}$. Given $\varepsilon > 0$, for every $n \in \mathbb{N}$, we can find $m_n \in \mathbb{N}$ such that

$$\mu\left(X \setminus \bigcup_{k=1}^{m_n} \overline{B}_n(x_k)\right) \le \frac{\varepsilon}{2^n}\,. \tag{2.5.11}$$

Let $K = \bigcap_{n\ge1} \bigcup_{k=1}^{m_n} \overline{B}_n(x_k)$. The set K is closed and totally bounded, hence K is compact; see Theorem 1.5.36. Taking (2.5.11) into account it follows

$$\begin{aligned}\mu(X) - \mu(K) = \mu(X \setminus K) &= \mu\left[\bigcup_{n\ge1}\left(X \setminus \bigcup_{k=1}^{m_n} \overline{B}_n(x_k)\right)\right]\\ &\le \sum_{n\ge1} \mu\left(X \setminus \bigcup_{k=1}^{m_n} \overline{B}_n(x_k)\right) \le \sum_{n\ge1} \frac{\varepsilon}{2^n} = \varepsilon\,.\end{aligned}$$

Hence, μ is compact regular (see Proposition 2.5.13), and so, μ is a Radon measure. □

In the next proposition we produce another useful dense subset of $L^p(X)$ for $1 \le p < \infty$.

Proposition 2.5.15. *If X is locally compact and $\mu\colon \mathcal{B}(X) \to [0, +\infty]$ is a Radon measure, then $C_c(X)$ is dense in $L^p(X)$ for $1 \le p < \infty$ where $C_c(X)$ is the space of all continuous functions $f\colon X \to \mathbb{R}$ that have a compact support.*

Proof. From Proposition 2.3.22, we know that simple functions are dense in $L^p(X)$. So, it suffices to show that for every $A \in \mathcal{B}(X)$ with $\mu(A) < +\infty$ we can approximate χ_A in the L^p-norm by $C_c(X)$-functions. Given $\varepsilon > 0$ there exist an open set $U \subseteq X$ and a compact set $K \subseteq X$ such that

$$K \subseteq A \subseteq U \quad \text{and} \quad \mu(U \setminus K) \le \varepsilon^p \,. \tag{2.5.12}$$

Since X is locally compact, combining Urysohn's Lemma (see Theorem 1.2.17) and Proposition 1.4.66(c), we can find $f \in C_c(X)$ such that $\chi_K \le f \le \chi_U$. Then, using (2.5.12), $\|\chi_A - f\|_p \le \mu(U \setminus K)^{1/p} \le \varepsilon$, which demonstrates that $C_c(X)$ is dense in $L^p(X)$ for $1 \le p < \infty$. □

Remark 2.5.16. Since $L^\infty(X)$ contains noncontinuous functions, the density result above fails for $p = +\infty$.

The next theorem is another remarkable result in the spirit of Egorov's Theorem; see Theorem 2.2.32. It asserts that a Borel measurable map between certain metric spaces is "almost" continuous. The result is known as "Lusin's Theorem."

Theorem 2.5.17 (Lusin's Theorem). *If X is a Polish space, Y is a separable metric space, $f\colon X \to Y$ is Borel measurable, and $\mu\colon \mathcal{B}(X) \to [0, +\infty)$ is a finite Borel measure, then given any $\varepsilon > 0$, there exists $K_\varepsilon \subseteq X$ being compact such that $\mu(X \setminus K_\varepsilon) \le \varepsilon$ and $f|_{K_\varepsilon}$ is continuous.*

Proof. We know that Y is second countable; see Proposition 1.5.5. So, let $\{V_n\}_{n\ge1}$ be a countable basis for the metric topology of Y. We have $f^{-1}(V_n) \in \mathcal{B}(X)$ for all $n \in \mathbb{N}$ and so using Theorem 2.5.12 there exists an open set $U_n \subseteq X$ such that

$$f^{-1}(V_n) \subseteq U_n \quad \text{and} \quad \mu\left(U_n \setminus f^{-1}(V_n)\right) \le \frac{\varepsilon}{2^{n+1}} \quad \text{for all } n \in \mathbb{N} \,. \tag{2.5.13}$$

The set $f^{-1}(V_n)$ is relatively open in $(X \setminus U_n) \cup f^{-1}(V_n)$. Note that $f^{-1}(V_n) = [(X \setminus U_n) \cup f^{-1}(V_n)] \cap U_n$, see (2.5.13). Let

$$A_\varepsilon = X \setminus \bigcup_{n\ge1} \left(U_n \setminus f^{-1}(V_n)\right) = \bigcap_{n\ge1} ((X \setminus U_n) \cup f^{-1}(V_n)) \,.$$

Thanks to (2.5.13), one gets

$$\mu(X \setminus A_\varepsilon) \le \frac{\varepsilon}{2} \,. \tag{2.5.14}$$

Using Theorem 2.5.14 there exists $K_\varepsilon \subseteq A_\varepsilon$ being compact such that $\mu(A_\varepsilon \setminus K_\varepsilon) \le \varepsilon/2$, which gives $\mu(X \setminus K_\varepsilon) \le \varepsilon$; see (2.5.14).

For every $n \in \mathbb{N}$, $f^{-1}(V_n)$ is relatively open in K_ε. Since $\{V_n\}_{n\ge1}$ is a basis for the metric topology of Y, it follows that for all open $V \subseteq Y$, $f^{-1}(V)$ is relatively open in K_ε. Hence $f|_{K_\varepsilon}$ is continuous. □

In addition there is also a second version of Lusin's Theorem.

Theorem 2.5.18 (Lusin's Theorem, Second Version). *If X is locally compact, μ is a Radon measure and $f: X \to \mathbb{R}$ is a Borel measurable function that vanishes outside a set of finite μ-measure, then for given $\varepsilon > 0$, there exist $A \in \mathcal{B}(X)$ and $h \in C_c(X)$ such that $\mu(A) \le \varepsilon$ and $f|_{X\setminus A} = h|_{X\setminus A}$. Moreover if f is bounded, then it holds that $\|h\|_\infty \le \|f\|_\infty$.*

Proof. First assume that f is bounded. Let $A = \{x \in X: f(x) \ne 0\} \in \mathcal{B}(X)$. By hypothesis, $\mu(A) < +\infty$. So, we can use Proposition 2.5.15 and find $\{h_n\}_{n\ge1} \subseteq C_c(X)$ such that $h_n \to f$ in $L^1(X)$. So, by passing to a suitable subsequence, if necessary we may assume that $h_n(x) \to f(x)$ for μ-a.a. $x \in X$; see Corollary 2.3.20. Invoking Egorov's Theorem (see Theorem 2.2.32), there exists $B \subseteq A$ such that

$$\mu(A \setminus B) \le \frac{\varepsilon}{3} \quad \text{and} \quad h_n \xrightarrow{\mu} f \text{ on } B\,. \tag{2.5.15}$$

Exploiting the fact that μ is a Radon measure, we find a compact set $K \subseteq B$ and an open set $U \supseteq B$ such that

$$\mu(B \setminus K) \le \frac{\varepsilon}{3} \quad \text{and} \quad \mu(U \setminus A) \le \frac{\varepsilon}{3}\,. \tag{2.5.16}$$

Since $h_n \xrightarrow{\mu} f$ on K, it follows that $f|_K$ is continuous. Invoking the locally compact version of the Tietze Extension Theorem (see Theorem 1.4.88), there exists $\hat{h} \in C_c(X)$ such that $\hat{h}|_K = f|_K$ and $\operatorname{supp} \hat{h} \subseteq U$. Hence, $D = \{x \in X: \hat{h}(x) \ne f(x)\} \subseteq U \setminus K$, which demonstrates, due to (2.5.15) and (2.5.16), that $\mu(D) \le \mu(U \setminus K) \le \varepsilon$.

Now let $\xi: \mathbb{R} \to \mathbb{R}$ be defined by

$$\xi(t) = \begin{cases} t & \text{if } |t| \le \|f\|_\infty\,, \\ \|f\|_\infty \operatorname{sgn} t & \text{if } |t| > \|f\|_\infty\,. \end{cases}$$

Evidently $\xi(0) = 0$, and so ξ is continuous. So, if we define $h = \xi \circ \hat{f}$, then $h \in C_c(X)$, $h = f$ on the set $\{\hat{h} = f\}$ and $\|h\|_\infty \le \|f\|_\infty$.

Finally we consider the general case in which f is unbounded. In this case we define $A_n = \{x \in X: 0 < |f(x)| \le n\} \in \mathcal{B}(X)$. Then $A_n \nearrow A$ and for large enough $n \ge 1$, we have that $\mu(A \setminus A_n) \le \varepsilon/2$. Then from the first part of the proof there exists $h \in C_c(X)$ such that $h = f\chi_{A_n}$ outside a set $D \in \mathcal{B}(X)$ with $\mu(D) \le \varepsilon/2$. Then finally we have $h = f$ outside a set $D_0 \in \mathcal{B}(X)$ with $\mu(D_0) \le \varepsilon$. ☐

There is a parametric variant of Lusin's Theorem concerning Carathéodory functions; see Definition 2.2.30. The result is known as "Scorza–Dragoni Theorem."

Theorem 2.5.19 (Scorza–Dragoni Theorem). *If T and X are Polish spaces, Y is a separable metric space, $\mu: \mathcal{B}(T) \to [0, +\infty)$ is a finite compact regular Borel measure, and $f: T \times X \to Y$ is a Carathéodory function, then for every $\varepsilon > 0$ there exists a compact set $K_\varepsilon \subseteq T$ with $\mu(T \setminus K_\varepsilon) \le \varepsilon$ such that $f|_{K_\varepsilon \times X}$ is continuous.*

Proof. From Theorem 1.5.21 we know that Y is homeomorphic to a subset of the Hilbert cube $\mathbb{H} = [0, 1]^{\mathbb{N}}$. Let $h = (h_n)_{n\in\mathbb{N}}: Y \to \mathbb{H}$ be this homeomorphism. Then f is a

Carathéodory function if and only if for every $n \in \mathbb{N}$, $h_n \circ f\colon T \times X \to [0,1]$ is a Carathéodory function. Therefore without any loss of generality we may assume that $Y = [0,1]$.

Let $\{U_n\}_{n\geq 1}$ be a basis for the topology of X and let $\{x_m\}_{m\geq 1} \subseteq X$ be dense. For every $q \in [0,1] \cap \mathbb{Q}$ let $\xi_{nq}\colon X \to [0,1]$ be defined by $\xi_{nq}(x) = q\chi_{U_n}(x)$. Since U_n is open, χ_{U_n} is lower semicontinuous (see Definition 1.7.1), and if $\varphi\colon X \to Y = [0,1]$ is lower semicontinuous, then $\varphi(x) = \sup[\xi_{nq}(x)\colon \xi_{nq} \leq \varphi]$ with $x \in X$. So, we define

$$A_{nqm} = \{t \in T\colon \xi_{nq}(x_m) \leq f(t, x_m)\} \in \mathcal{B}(T)\,.$$

Let $A_{nq} = \bigcap_{m\in\mathbb{N}} A_{nqm} \in \mathcal{B}(T)$. The density of $\{x_m\}_{m\geq 1}$ in X, the continuity of $f(t,\cdot)$, and the lower semicontinuity of ξ_{nq} imply that

$$A_{nq} = \{t \in T\colon \xi_{nq}(x) \leq f(t,x) \text{ for all } x \in X\}\,.$$

We set $\eta_{nq}(t,x) = \chi_{A_{nq}}(t)\xi_{nq}(x)$. Then $\eta_{nq} \leq f$ and for all $(t,x) \in T \times X$ we have $f(t,x) = \sup_{n,q} \eta_{nq}(t,x)$. Note that $\mathbb{N} \times ([0,1] \cap \mathbb{Q})$ is countable. So we can write that

$$f = \sup_{k\in\mathbb{N}} \chi_{B_k} h_k \text{ with } B_k \in \mathcal{B}(T)\,,\quad h_k \text{ is lower semicontinuous on } X\,.$$

Since by hypothesis μ is a finite, compact regular measure on T, there exist an open set $V_k \subseteq T$ and a compact set $K_k \subseteq T$ such that

$$K_k \subseteq B_k \subseteq V_k \quad \text{and} \quad \mu(V_k \setminus K_k) \leq \frac{\varepsilon}{2^{k+2}} \text{ for all } k \in \mathbb{N}\,. \tag{2.5.17}$$

Let $E_k = K_k \cup (X \setminus V_k)$ for all $k \in \mathbb{N}$. Then $\chi_{B_k}\big|_{E_k}$ is continuous (see (2.5.17)), and this implies that $\chi_{B_k} h_k$ is lower semicontinuous. Let $E = \bigcap_{k\in\mathbb{N}} E_k \subseteq T$ be compact. We see that $\mu(T \setminus E) \leq \varepsilon/2$ and $f\big|_{E\times X}$ is lower semicontinuous as the upper envelope of lower semicontinuous functions; see Proposition 1.7.4(a). The same argument applied to $1 - f$ produces another compact set $\tilde{E} \subseteq T$ with $\mu(T \setminus \tilde{E}) \leq \varepsilon/2$ and $(1-f)\big|_{\tilde{E}\times X}$ is lower semicontinuous. We set $T_\varepsilon = E \cap \tilde{E} \subseteq T$, which is compact. Then we see that $\mu(T \setminus T_\varepsilon) \leq \varepsilon$ and $f\big|_{T_\varepsilon\times X}$ continuous. □

Next we introduce an extension of the notion of a Carathéodory function (see Definition 2.2.30), which is important in calculus of variation, optimal control, and optimization.

Definition 2.5.20. Let (X, Σ) be a measurable space, Y a Hausdorff topological space, and $f\colon X \times Y \to \overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$. We say that f is a **normal integrand** if the following hold:

(a) f is $\Sigma \bigotimes \mathcal{B}(Y)$-measurable;

(b) $y \to f(x,y)$ is lower semicontinuous for all $x \in X$.

Proposition 2.5.21. *If (X, Σ, μ) is a complete measure space, Y is a Polish space, and $f\colon X \times Y \to \overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$ is a normal integrand such that there is a Carathéodory* *function $\xi\colon X\times Y\to\mathbb{R}$ satisfying $\xi(x,y)\le f(x,y)$ for all $(x,y)\in X\times Y$, then there is a sequence of Carathéodory functions $f_n\colon X\times Y\to\mathbb{R}$ such that $\xi(x,y)\le f_n(x,y)\le f(x,y)$ for all $(x,y)\in X\times Y$ and $f_n\nearrow f$ as $n\to\infty$.*

Proof. We reason as in the proof of Proposition 1.7.6. So, we define

$$f_n(x,y)=\inf[f(x,y)+nd(y,z)\colon z\in Y]\quad\text{for all } n\in\mathbb{N}$$

with d being the metric on Y. If $\{z_m\}_{m\ge1}\subseteq Y$ is dense in Y, then

$$f_n(x,y)=\inf_{m\in\mathbb{N}}[f(x,y)+nd(y,z_m)]\quad\text{for all } n\in\mathbb{N}\,.$$

This shows that f_n is $\Sigma\bigotimes\mathcal{B}(X)$-measurable; see Proposition 2.2.31. Clearly we have $\xi(x,y)\le f_n(x,y)$ for all $(x,y)\in X\times Y$, for all $n\in\mathbb{N}$ and as in the proof of Proposition 1.7.6, we show that $f_n\nearrow f$. ☐

Using this proposition we can have the following extension of the Scorza–Dragoni Theorem; see Theorem 2.5.19.

Theorem 2.5.22. *If T and Y are Polish spaces, μ is a finite, compact regular Borel measure on T and $f\colon T\times X\to\overline{\mathbb{R}}=\mathbb{R}\cup\{+\infty\}$ is a normal integrand bounded below by a Carathéodory function ξ, then for given $\varepsilon>0$ there is a compact set $T_\varepsilon\subseteq T$ such that $\mu(T\setminus T_\varepsilon)\le\varepsilon$ and $f|_{T_\varepsilon\times X}$ is lower semicontinuous.*

Proof. Using Proposition 2.5.21, there exist Carathéodory functions f_n such that $\xi\le f_n\le f$ for all $n\in\mathbb{N}$ and $f_n\nearrow f$. We apply the Scorza–Dragoni Theorem (see Theorem 2.5.19), and for each $n\in\mathbb{N}$ there is a compact set $T_n\subseteq T$ with $\mu(T\setminus T_n)\le\varepsilon/(2^n)$ and $f_n|_{T_n\times X}$ is continuous. Let $T_\varepsilon=\bigcap_{n\ge1}T_n\subseteq T$ being compact. Then, of course, $\mu(T\setminus T_\varepsilon)\le\varepsilon$ and $f|_{T_\varepsilon\times X}$ is lower semicontinuous. ☐

Definition 2.5.23. Let (X,Σ,μ) be a measure space, $(Y,\mathcal{L})$ a measurable space, and $f\colon X\to Y$ a $(\Sigma,\mathcal{L})$-measurable map. Then μ induces an **image measure** $\mu\circ f^{-1}$ on Y by $(\mu\circ f^{-1})(A)=\mu(f^{-1}(A))$ for all $A\in\mathcal{L}$.

Since f^{-1} preserves all the set theoretic operations, we see that indeed $\mu\circ f^{-1}$ is a measure on $(Y,\mathcal{L})$.

Proposition 2.5.24. *If (X,Σ,μ) is a measure space, $(Y,\mathcal{L})$ is a measurable space, $f\colon X\to Y$ is a $(\Sigma,\mathcal{L})$-measurable map, and $h\colon Y\to\mathbb{R}$ is a $\mathcal{L}$-measurable function, then*

$$\int_Y hd\left(\mu\circ f^{-1}\right)=\int_X(h\circ f)d\mu$$

whenever either side exists.

Proof. If $h=\chi_A$ with $A\in\mathcal{L}$, then the result follows from Definition 2.5.23. So, the result is also true for simple functions that are linear combinations of characteristic functions. Finally we use Proposition 2.2.18 to pass to the general case. ☐

Image measures via continuous maps preserve the property of being a Radon measure

Proposition 2.5.25. *If X, Y are Hausdorff topological spaces, X is compact, $f\colon X \to Y$ is continuous, and $\mu\colon \mathcal{B}(X) \to [0, +\infty]$ is a Radon measure, then $\mu \circ f^{-1}\colon \mathcal{B}(Y) \to [0, +\infty]$ is a Radon measure as well.*

Proof. According to Theorem 2.5.11, it suffices to show that $\mu \circ f^{-1}$ is compact regular. Since μ is a Radon measure, for every $A \in \mathcal{B}(Y)$ one gets

$$\left(\mu \circ f^{-1}\right)(A) = \sup[\mu(K)\colon K \subseteq X \text{ is compact}, K \subseteq f^{-1}(A)]\,; \tag{2.5.18}$$

see Definition 2.5.23. For a compact $K \subseteq f^{-1}(A)$ it follows $f(K) \subseteq A$ and so $K \subseteq f^{-1}(f(K)) \subseteq f^{-1}(A)$. Hence

$$\mu(K) \le \mu(f^{-1}(f(K))) \le \left(\mu \circ f^{-1}\right)(A)\,. \tag{2.5.19}$$

The continuity of f implies that $\tilde{K} = f(K) \subseteq Y$ is compact. Then from (2.5.18) and (2.5.19) it follows that

$$\left(\mu \circ f^{-1}\right)(A) = \sup\left[\left(\mu \circ f^{-1}\right)\left(\tilde{K}\right) : \tilde{K} \subseteq Y \text{ is compact}, \tilde{K} \subseteq A\right]\,,$$

which shows that $\mu \circ f^{-1}$ is compact regular, hence a Radon measure. □

2.6 Analytic (Souslin) Sets

In Definition 1.5.51 we introduced the notion of a Souslin space. Souslin spaces are of fundamental importance in measure theory since they give to the theory of Borel sets and Borel functions depth and power.

Let us start by recalling the definition of Souslin space.

Definition 2.6.1. A Hausdorff topological space X is said to be a **Souslin space** if it is the continuous image of a Polish space, that is, there exists a Polish space Y and a continuous surjection $f\colon Y \to X$. A subset of a Hausdorff topological space that is a Souslin space is called a **Souslin set**. A Souslin subset of a Polish space is called **analytic set** as well. The complement of a Souslin set is called **co-Souslin set** (or **coanalytic set**).

Remark 2.6.2. We have that a Souslin space is always separable but need not to be metrizable; see Remark 1.5.52. Moreover, using Remark 1.5.50, we see that a nonempty subset of a Hausdorff space is a Souslin set if it is the image of the Polish space $\mathbb{N}^\infty$ under a continuous map.

Given a set B, by B^f we denote the set of all finite sequences with terms in the set B. That is, $B^f = \bigcup_{n\ge 1} B_n^f$ with B_n^f being the set of n-sequences.

Of special interest to us is the set $\mathbb{N}^f$. Note that $\mathbb{N}^f$ is countable in contrast to $\mathbb{N}^\infty$, which is uncountable. Using $\mathbb{N}^f$ we introduce the following definition.

Definition 2.6.3. Let X be a nonempty set and $\mathcal{L} \subseteq 2^X$. An $\mathcal{L}$**-Souslin scheme** is a map $A\colon \mathbb{N}^f \to \mathcal{L}$. Let $\mathcal{D}$ be the family of all $\mathcal{L}$-Souslin schemes. The **Souslin operation** (or A**-operation**) over the class $\mathcal{L}$ is a map $\mathfrak{a}\colon \mathcal{D} \to \mathcal{L}$ such that

$$\mathfrak{a}(A) = \bigcup_{p \in \mathbb{N}^\infty} \bigcap_{k \in \mathbb{N}} A(p_1, \dots, p_k) \quad \text{for all } A \in \mathcal{D}\,. \tag{2.6.1}$$

The collection of all sets of this form is denoted by $S(\mathcal{L})$. The elements of $S(\mathcal{L})$ are called $\mathcal{L}$**-Souslin** (or $\mathcal{L}$**-analytic**) sets. A Souslin scheme A is said to be **regular** (or **monotone**) if $A(p_1, \dots, p_{k+1}) \subseteq A(p_1, \dots, p_k)$ with $p \in \mathbb{N}^\infty$.

Remark 2.6.4. If $\emptyset \in \mathcal{L}$ (or if $\mathcal{L}$ contains disjoint sets), then $\emptyset \in S(\mathcal{L})$. Note that in (2.6.1) the union is uncountable. So, if $\mathcal{L}$ is a σ-algebra and A is an $\mathcal{L}$-Souslin scheme, then $\mathfrak{a}(A)$ may be outside of $\mathcal{L}$. In what follows we will use the following notation. Given $s = (s_k)_{k=1}^n \in \mathbb{N}^f$ and $p \in \mathbb{N}^\infty$, we write $s \prec p$ if and only if $s_1 = p_1, \dots, s_n = p_n$.

In the next proposition we collect some basic properties of the operator S.

Proposition 2.6.5. *If X is a nonempty set and $\mathcal{L}, \mathcal{L}' \subseteq 2^X$, then the following hold:*
(a) *$S(\mathcal{L}) \subseteq S(\mathcal{L}')$ if $\mathcal{L} \subseteq \mathcal{L}'$, that is, S is monotone;*
(b) *$S(\mathcal{L})_\delta = S(\mathcal{L})$, that is, S is closed under countable intersections;*
(c) *$S(\mathcal{L})_\sigma = S(\mathcal{L})$, that is, S is closed under countable unions;*
(d) *$\mathcal{L} \subseteq S(\mathcal{L})$.*

Proof. (a) This is an immediate consequence of Definition 2.6.3.

(b) Clearly we have $S(\mathcal{L}) \subseteq S(\mathcal{L})_\delta$. Suppose that $\bigcap_{k\geq 1} \mathfrak{a}(A_k) \in S(\mathcal{L})_\delta$. We need to produce an $\mathcal{L}$-Souslin scheme $A\colon \mathbb{N}^f \to \mathcal{L}$ such that $\mathfrak{a}(A) = \bigcap_{k\geq 1} \mathfrak{a}(A_k)$. To this end for every $k \in \mathbb{N}$, let $T_k = \{(2m-1)2^{k-1}\colon m \in \mathbb{N}\}$. Then $\{T_k\}_{k\geq 1}$ is a partition of $\mathbb{N}$ into infinitely many infinite sets. For each $k \in \mathbb{N}$, let $\xi_k\colon \mathbb{N}^\infty \to \mathbb{N}^\infty$ be defined by

$$\xi_k((p_n)) = (p_{2^{k-1}}, p_{3\cdot 2^{k-1}}, p_{5\cdot 2^{k-1}}, \dots)\,,$$

that is, ξ picks from the sequence $(p_n)_{n\in\mathbb{N}}$ those elements with index in T_k. We will produce an $\mathcal{L}$-Souslin scheme A such that

$$\bigcap_{s \prec p} A(s) = \bigcap_{k\geq 1} \bigcap_{s \prec \xi_k(p)} A_k(s) \quad \text{for all } p \in \mathbb{N}^\infty\,. \tag{2.6.2}$$

We rewrite (2.6.1) as

$$\bigcap_{n\geq 1} A(p_1, \dots, p_n) = \bigcap_{k\geq 1} \bigcap_{m\geq 1} A_k(p_{2^{k-1}}, p_{3\cdot 2^{k-1}}, \dots, p_{(2m-1)\cdot 2^{k-1}}) \tag{2.6.3}$$

for all $p \in \mathbb{N}^\infty$. If $(p_1, \dots, p_n) \in \mathbb{N}^f$, then $n = (2m-1)2^{k-1}$ for exactly one pair $(m, k) \in \mathbb{N} \times \mathbb{N}$. Let

$$A(p_1, p_2, \dots, p_n) = A_k(p_{2^{k-1}}, p_{3\cdot 2^{k-1}}, \dots, p_{(2m-1)\cdot 2^{k-1}})\,. \tag{2.6.4}$$

Then (2.6.4) defines an $\mathcal{L}$-Souslin scheme, which satisfies (2.6.3) and consequently (2.6.2) as well.

Let $x \in \mathfrak{a}(A) = \bigcup_{p\in\mathbb{N}^\infty} \bigcap_{s<p} A(s)$; see (2.6.1). So, for some $p_0 \in \mathbb{N}^\infty$ we have

$$x \in \bigcap_{s<p_0} A(s) = \bigcap_{k\geq 1} \bigcap_{s<\xi_k(p_0)} A_k(s) ;$$

see (2.6.2). Hence

$$x \in \bigcap_{s<\xi_k(p_0)} A_k(s) \subseteq \bigcup_{p\in\mathbb{N}^\infty} \bigcap_{s<p} A_k(s) = \mathfrak{a}(A) \quad \text{for all } k \in \mathbb{N} ,$$

which implies that $x \in \bigcap_{k\geq 1} \mathfrak{a}(A_k)$. Hence

$$\mathfrak{a}(A) \subseteq \bigcap_{k\geq 1} \mathfrak{a}(A_k) . \tag{2.6.5}$$

Next suppose that $x \in \bigcap_{k\geq 1} \mathfrak{a}(A_k)$. Then, from (2.6.1), one gets $x \in \bigcup_{p\in\mathbb{N}^\infty} \bigcap_{s<p} A_k(s)$ for all $k \in \mathbb{N}$, which implies $x \in \bigcap_{s<p_k} A_k(s)$ for some $p_k \in \mathbb{N}^\infty$ and for all $k \in \mathbb{N}$.

Let $\hat{p} \in \mathbb{N}^\infty$ such that $\xi_k(\hat{p}) = p_k$ for all $k \in \mathbb{N}$. Then $x \in \bigcap_{k\geq 1} \bigcap_{s<\xi_k(\hat{p})} A_k(s)$, which implies, due to (2.6.2),

$$x \in \bigcap_{s<\hat{p}} A(s) \subseteq \bigcup_{p\in\mathbb{N}^\infty} \bigcap_{s<p} A(s) = \mathfrak{a}(A) .$$

Hence,

$$\bigcap_{k\geq 1} \mathfrak{a}(A_k) \subseteq \mathfrak{a}(A) . \tag{2.6.6}$$

From (2.6.5) and (2.6.6) we conclude that $\mathfrak{a}(A) = \bigcap_{k\geq 1} \mathfrak{a}(A_k)$.

(c) Clearly we have $S(\mathcal{L}) \subseteq S(\mathcal{L})_\sigma$. Consider $\bigcup_{k\geq 1} \mathfrak{a}(A_k) \in S(\mathcal{L})_\sigma$. We need to generate an $\mathcal{L}$-Souslin scheme A such that $\mathfrak{a}(A) = \bigcup_{k\geq 1} \mathfrak{a}(A_k)$.

If $s = (s_k)_{k=1}^n \in \mathbb{N}^f$, then $p_1 = (2m-1)2^{k-1}$ for exactly one pair $(m, k) \in \mathbb{N} \times \mathbb{N}$. We define

$$A(s_1, \ldots, s_n) = A((2m-1)2^{k-1}, s_2, \ldots, s_n) = A_k(m, s_2, \ldots, s_n) .$$

This is an $\mathcal{L}$-Souslin scheme for which we have

$$\bigcap_{n\geq 1} A\left((2m-1)2^{k-1}, s_2, \ldots, s_n\right) = \bigcap_{n\geq 1} A_k(m, s_2, \ldots, s_n) \tag{2.6.7}$$

for all $k \in \mathbb{N}$ and for all $(m, s_2, s_2, \ldots) \in \mathbb{N}^\infty$. Let $x \in \mathfrak{a}(A) = \bigcup_{p\in\mathbb{N}^\infty} \bigcap_{s<p} A(s)$; see (2.6.1). Then $x \in \bigcap_{n\geq 1} A(p_1, \ldots, p_n)$ for some $p \in \mathbb{N}^\infty$ which gives, choosing $(m, k) \in \mathbb{N} \times \mathbb{N}$ such that $p_1 = (2m-1)2^{k-1}$, $x \in \bigcap_{n\geq 1} A_k(m, p_2, \ldots, p_n) \subseteq \mathfrak{a}(A_k)$. Hence

$$\mathfrak{a}(A) \subseteq \bigcup_{k\geq 1} \mathfrak{a}(A_k) . \tag{2.6.8}$$

Next let $x \in \bigcup_{k\geq 1} \mathfrak{a}(A_k) = \bigcup_{k\geq 1} \bigcup_{p\in\mathbb{N}^\infty} \bigcap_{s<p} A_k(s)$. Then for some $k \in \mathbb{N}$ and some $(m, s_2, s_3, \dots) \in \mathbb{N}^\infty$, one gets $x \in \bigcap_{n\geq 1} A_k(m, s_2, \dots, s_n)$. Then, because of (2.6.7), it follows that

$$x \in \bigcap_{n\geq 1} A\left((2m-1)2^{k-1}, s_2, \dots, s_n\right) \subseteq \mathfrak{a}(A)\,.$$

This finally gives

$$\bigcup_{k\geq 1} \mathfrak{a}(A_k) \subseteq \mathfrak{a}(A)\,. \tag{2.6.9}$$

From (2.6.8) and (2.6.9) we conclude that $\mathfrak{a}(A) = \bigcup_{k\geq 1} \mathfrak{a}(A_k)$.

(d) For $B \in \mathcal{L}$ we set $A(s) = B$ for all $s \in \mathbb{N}^f$. Then $\mathfrak{a}(A) = B$. □

In fact S is an idempotent operator. For a proof of this result we refer to Klein–Thompson [178, Theorem 12.2.3, p. 143].

Proposition 2.6.6. *If X is a nonempty set and $\mathcal{L} \subseteq 2^X$, then $S(S(\mathcal{L})) = S(\mathcal{L})$.*

Concerning complementation, it is not true in general that $S(\mathcal{L})$ is closed under complementation. Hence, we cannot say in general that $S(\mathcal{L})$ is a σ-algebra. In order for $S(\mathcal{L})$ to contain $\sigma(\mathcal{L})$, we need additional hypotheses.

Proposition 2.6.7. *If X is a nonempty set, $\mathcal{L} \subseteq 2^X$ and for every $B \in \mathcal{L}$ we have that $X \setminus B \in S(\mathcal{L})$, then $\sigma(\mathcal{L}) \subseteq S(\mathcal{L})$.*

Proof. We know that the smallest algebra containing $\mathcal{L}$ is produced by taking finite intersections of finite unions of elements of $\mathcal{L}$ and of complements of elements of $\mathcal{L}$. Then Propositions 2.6.5 and 2.6.6 and the hypothesis imply that $S(S(S(\mathcal{L}))) = S(\mathcal{L})$. But $S(\mathcal{L})$ is a monotone class; see Proposition 2.6.5. So, using Theorem 2.1.12, we conclude that $\sigma(\mathcal{L}) \subseteq S(\mathcal{L})$. □

In Definition 2.6.1 we mentioned that a Souslin space that is a subset of a Polish space is called **analytic**. Next we give an alternative definition of analytic sets in terms of the Souslin operation and subsequently we show that the two notions of analyticity are in fact equivalent.

Definition 2.6.8. Let X be a Polish space and let $\mathcal{F}_X$ denote the family of closed subsets of X. The **analytic sets** of X are the elements of $S(\mathcal{F}_X)$.

Therefore we have two definitions of analytic sets; see Definition 2.6.1 and Definition 2.6.8. Next we show that they are equivalent and we also provide some other useful characterizations of analytic sets.

Proposition 2.6.9. *If X is a Polish space and $E \subseteq X$ is nonempty, then the following statements are equivalent:*

(a) *there exists a continuous function $f\colon \mathbb{N}^\infty \to X$ such that $E = f(\mathbb{N}^\infty)$;*

(b) *there exists a closed set $C \subseteq \mathbb{N}^\infty \times X$ such that $E = \mathrm{proj}_X\, C$;*

(c) *E is a Souslin space; see Definition* 1.5.51*;*
(d) *E is an analytic set and more precisely there is a regular Souslin scheme A consisting of closed subsets of X with a vanishing diameter such that $\mathfrak{a}(A) = E$.*

Proof. (a) $\Longrightarrow$ (b): Since $f\colon \mathbb{N}^\infty \to X$ is continuous, $\operatorname{Gr} f = C \subseteq \mathbb{N}^\infty \times X$ is closed and $\operatorname{proj}_X C = E$.

(b) $\Longrightarrow$ (c): We know that $\mathbb{N}^\infty \times X$ is Polish; see Remark 1.5.50 and Proposition 1.5.46. The set $C \subseteq \mathbb{N}^\infty \times X$ being closed is itself Polish; see Proposition 1.5.45. The projection map $\operatorname{proj}_X\colon C \to E$ is a continuous open surjection. Therefore, by Definition 1.5.51, we conclude that E is a Souslin space.

(c) $\Longrightarrow$ (a): According to Definition 1.5.51, there is a Polish space Y and a continuous surjection $h\colon Y \to E$. Moreover, from Remark 1.5.50 we know that there is a continuous surjection $g\colon \mathbb{N}^\infty \to Y$. Let $f = h \circ g\colon \mathbb{N}^\infty \to E$. Then f is a continuous surjection.

(a) $\Longrightarrow$ (d): By hypothesis there is a continuous surjection $f\colon \mathbb{N}^\infty \to E$. Consider the Souslin scheme defined by

$$A(p_1, \ldots, p_n) = \overline{f(U_{p_1,\ldots,p_n})} = f(\{p_1\} \times \ldots \times \{p_n\} \times \mathbb{N} \times \mathbb{N} \times \ldots)\,.$$

Clearly this Souslin scheme is regular (see Definition 2.6.3), and consists of closed sets. Moreover, the scheme $\{U_s\colon s \in \mathbb{N}^f\}$ has a vanishing diameter for the tree metric t; see Remark 1.5.50. Note that if $B \subseteq X$ is an F_σ-set and $\varepsilon > 0$, then we can write $B = \bigcup_{n\geq 1} B'_n$ with $\{B'_n\}$ pairwise disjoint F_σ-sets each having diameter less than $\varepsilon > 0$. Using this fact and an induction argument, we show that $E = \mathfrak{a}(A)$.

(d) $\Longrightarrow$ (a): By hypothesis we have $E = \bigcup_{p\in\mathbb{N}^\infty} \bigcap_{k\geq 1} A(p_1, \ldots, p_k)$. Since X is complete, in order for $\bigcap_{k\geq 1} A(p_1, \ldots, p_k)$ to be empty is that for some $k \in \mathbb{N}$, $A(p_1, \ldots, p_k) = \emptyset$. We define

$$\mathfrak{L} = \{p \in \mathbb{N}^\infty\colon A(p_1, \ldots, p_k) \neq \emptyset \text{ for all } k \in \mathbb{N}\}\,.$$

Using the definition of the tree metric (see Remark 1.5.50), we can easily see that $\mathfrak{L} \subseteq \mathbb{N}^\infty$ is closed. Hence Example 1.7.13(c) implies that $\mathfrak{L}$ is a retract of $\mathbb{N}^\infty$. We have

$$E = \bigcup_{p\in\mathfrak{L}} \bigcap_{k\geq 1} A(p_1, \ldots, p_k)\,.$$

For each $p \in \mathfrak{L}$ let $g(p)$ be the unique element of $\bigcap_{k\geq 1} A(p_1, \ldots, p_k)$. Recall that a Souslin scheme has a vanishing diameter, and apply Theorem 1.5.15. The map $g\colon \mathfrak{L} \to E$ is bijective and continuous. Let $r\colon \mathbb{N}^\infty \to \mathfrak{L}$ be a retraction map. Then $f = g \circ r\colon \mathbb{N}^\infty \to E$ is a continuous surjection. □

From Proposition 2.6.5, we have the following.

Proposition 2.6.10. *If X is a Polish space, then countable intersections and countable unions of analytic sets are analytic.*

Next we are going to show that the analytic sets contain the Borel sets.

Proposition 2.6.11. *If X is a Polish space and $B \in \mathcal{B}(X)$, then B is analytic.*

Proof. From Proposition 1.5.8, we know that every open set of X is F_σ. Hence, every open set is analytic; see Definition 2.6.8. Then Proposition 2.6.7 implies that $\mathcal{B}(X) \subseteq S(\mathcal{F}_X)$. Using Propositions 2.6.5 and 2.6.6 it follows that

$$S(\mathcal{F}_X) \subseteq S(\mathcal{B}(X)) \subseteq S(S(\mathcal{F}_X)) = S(\mathcal{F}_X)\ . \qquad \square$$

Remark 2.6.12. From the proof above we see that $S(\mathcal{F}_X) = S(\mathcal{B}(X))$. If X is countable, then $\mathcal{B}(X) = S(\mathcal{F}_X)$, that is, Borel and analytic sets coincide. If X is uncountable, then the class of analytic sets $S(\mathcal{F}_X)$ is strictly larger than the Borel σ-algebra $\mathcal{B}(X)$. In fact we can have an analytic set whose complement is not analytic.

We want to have a closer look at the relation between Borel and analytic sets. We start with a definition.

Definition 2.6.13. Let X be a Polish space and let $A_1, A_2 \subseteq X$ be nonempty. We say that A_1 and A_2 can be **separated by Borel sets** if there are disjoint Borel sets $B_1, B_2 \subseteq X$ such that $A_1 \subseteq B_1$ and $A_2 \subseteq B_2$.

Lemma 2.6.14. *Let X be a Polish space.*
(a) *If $\{A_n\}_{n\ge1}$ and C are nonempty subsets of X such that for every $n \in \mathbb{N}$ the sets A_n and C can be separated by Borel sets, then $\bigcup_{n\ge1} A_n$ and C can be separated by Borel sets.*
(b) *If $\{A_n\}_{n\ge1}$ and $\{C_n\}_{n\ge1}$ are nonempty subsets of X such that for each $(n, m) \in \mathbb{N} \times \mathbb{N}$ the sets A_n and C_m can be separated by Borel sets, then the sets $\bigcup_{n\ge1} A_n$ and $\bigcup_{n\ge1} C_n$ can be separated by Borel sets.*

Proof. (a) By hypothesis, for each $n \in \mathbb{N}$ there exist disjoint Borel sets B_n and D_n such that $A_n \subseteq B_n$ and $C \subseteq D_n$. Then $\bigcup_{n\ge1} B_n$ and $\bigcap_{n\ge1} D_n$ are disjoint Borel sets and $\bigcup_{n\ge1} A_n \subseteq \bigcup_{n\ge1} B_n$ and $C \subseteq \bigcap_{n\ge1} D_n$.

(b) From part (a) above for each $n \in \mathbb{N}$, the sets A_n and $\bigcup_{m\ge1} C_m$ can be separated by Borel sets. A second application of part (a) implies that $\bigcup_{n\ge1} A_n$ and $\bigcup_{m\ge1} C_m$ can be separated by Borel sets. $\square$

Now we show that disjoint analytical sets can be separated by Borel sets. The result is known as the "Separation Theorem" and has important consequences, some of which we explore here.

Theorem 2.6.15 (Separation Theorem). *If X is a Polish space and $A_1, A_2 \subseteq X$ are nonempty disjoint analytical sets, then A_1 and A_2 can separated by Borel sets.*

Proof. Invoking Proposition 2.6.9, there exist continuous surjections

$$f_1 : \mathbb{N}^\infty \to A_1 \quad \text{and} \quad f_2 : \mathbb{N}^\infty \to A_2\ .$$

For any $s \in \mathbb{N}^f$, we set $U_s = \{s_1\} \times \ldots \times \{s_k\} \times \mathbb{N} \times \mathbb{N} \times \ldots$ and then define $A_1^s = f_1(U_s)$ as well as $A_2^s = f_2(U_s)$.

Arguing indirectly, suppose that A_1 and A_2 cannot be separated by Borel sets. Since it holds that $A_1 = \bigcup_{n\geq 1} A_1^n$ and $A_2 = \bigcup_{n\geq 1} A_2^n$, using Lemma 2.6.14, there exist $n_1, m_1 \in \mathbb{N}$ such that the sets $A_1^{n_1}$ and $A_2^{m_1}$ cannot be separated by Borel sets. Note that

$$A_1^{n_1} = \bigcup_{n\geq 1} A_1^{n_1,n} \quad \text{and} \quad A_2^{m_1} = \bigcup_{n\geq 1} A_2^{m_1,n} .$$

Hence, a new application of Lemma 2.6.14 gives $n_2, m_2 \in \mathbb{N}$ such that $A_1^{n_1,n_2}$ and $A_2^{m_1,m_2}$ cannot be separated by Borel sets. Continuing this way, we produce $p(1) = (n_k)$ and $p(2) = (m_k) \in \mathbb{N}^\infty$ such that

$$A_1^{n_1,\dots,n_k} \quad \text{and} \quad A_2^{m_1,\dots,m_k} , \qquad k \in \mathbb{N}$$

cannot be separated by Borel sets. Let $x = f_1(p(1)) \in A_1$ and $u = f_2(p(2)) \in A_2$. We have $x \neq u$ since the sets A_1 and A_2 are disjoint. Let $U_1 \in \mathcal{N}(x)$ and $U_2 \in \mathcal{N}(u)$ such that $U_1 \cap U_2 = \emptyset$. The continuity of f_1 and f_2 implies that for $k \in \mathbb{N}$ large enough we have

$$A_1^{n_1,\dots,n_k} = f_1(U_{n_1,\dots,n_k}) \subseteq U_1 \quad \text{and} \quad A_2^{m_1,\dots,m_k} = f_2(U_{m_1,\dots,m_k}) \subseteq U_2 .$$

Therefore the open sets U_1 and U_2, which are Borel as well, separate $A_1^{n_1,\dots,n_k}$ and $A_2^{m_1,\dots,m_k}$, a contradiction. □

Corollary 2.6.16. *If X is a Polish space and $\{A_n\}_{n\geq 1}$ are pairwise disjoint analytic sets, then there exists a sequence $\{B_n\}_{n\geq 1}$ of pairwise disjoint Borel sets such that $A_n \subseteq B_n$ for every $n \in \mathbb{N}$.*

Corollary 2.6.17. *If X is a Polish space and $A \subseteq X$ is both analytic and coanalytic, that is, $X \setminus A$ is analytic as well, then $A \in \mathcal{B}(X)$.*

Proof. Using Theorem 2.6.15 there are disjoint Borel sets B_1, B_2 such that $A \subseteq B_1$ and $X \setminus A \subseteq B_2$. Evidently $A = B_1$ and $X \setminus A = B_2$. Therefore $A \in \mathcal{B}(X)$. □

Remark 2.6.18. Clearly the converse of the corollary above is true as well. Namely, every Borel set in X is both analytic and coanalytic.

Applying Corollary 2.6.17 we obtain the following characterizations of Borel measurable maps between Polish spaces.

Proposition 2.6.19. *If X, Y are Polish spaces and $f\colon X \to Y$, then the following statements are equivalent:*
(a) *f is Borel measurable;*
(b) $\operatorname{Gr} f \in \mathcal{B}(X \times Y) = \mathcal{B}(X) \bigotimes \mathcal{B}(Y)$;
(c) $\operatorname{Gr} f \subseteq X \times Y$ *is analytic.*

Proof. (a) $\Longrightarrow$ (b): Let $\varphi\colon X \times Y \to Y \times Y$ be defined by $\varphi(x, y) = (f(x), y)$. Since by hypothesis f is Borel measurable, for every $B, C \in \mathcal{B}(X)$ we have $\varphi^{-1}(B \times C) \in \mathcal{B}(X) \bigotimes \mathcal{B}(Y) = \mathcal{B}(X \times Y)$; see Proposition 2.2.26(b). Therefore φ is Borel measurable. Let $D = \{(y, z) \in Y \times Y\colon y = z\}$. Then $D \subseteq Y \times Y$ is closed and $\operatorname{Gr} f = \varphi^{-1}(D) \in \mathcal{B}(X \times Y) = \mathcal{B}(X) \bigotimes \mathcal{B}(Y)$.

(b) $\Longrightarrow$ (c): This implication is a consequence of Proposition 2.6.11.

(c) $\Longrightarrow$ (a): Let $B \in \mathcal{B}(Y)$. Then $X \times B \in \mathcal{B}(X \times Y)$ and so it is analytic. It follows that $\operatorname{Gr} f \cap (X \times B) \subseteq X \times Y$ is analytic. Note that

$$f^{-1}(B) = \operatorname{proj}_X(\operatorname{Gr} f \cap (X \times B)) \tag{2.6.10}$$

with $\operatorname{proj}_X : X \times Y \to X$ being the projection map defined by $\operatorname{proj}_X(x, y) = x$ for all $(x, y) \in X \times Y$. We know that proj_X is continuous. Since $\operatorname{Gr} f \cap (X \times B)$ is analytic, we find a continuous surjection $h : \mathbb{N}^\infty \to \operatorname{Gr} f \cap (X \times B)$; see Proposition 2.6.9. Then $\operatorname{proj}_X \circ h : \mathbb{N}^\infty \to f^{-1}(B)$ (see (2.6.10)) is a continuous surjection. Hence $f^{-1}(B) \subseteq X$ is analytic; see Proposition 2.6.9. In a similar way we show that $f^{-1}(Y \setminus B) \subseteq X$ is analytic. But $f^{-1}(Y \setminus B) = X \setminus f^{-1}(B)$. Therefore $f^{-1}(B) \subseteq X$ is coanalytic. Invoking Corollary 2.6.17, we conclude that $f^{-1}(B) \in \mathcal{B}(X)$ and so f is Borel measurable. ☐

Definition 2.6.20. Let (X, Σ) and $(Y, \mathcal{L})$ be two measurable spaces. A bijection $f : X \to Y$ is said to be an **isomorphism** if f is $(\Sigma, \mathcal{L})$-measurable and f^{-1} is $(\mathcal{L}, \Sigma)$-measurable. Then the measurable spaces (X, Σ) and $(Y, \mathcal{L})$ are said to be **isomorphic**. If X, Y are Hausdorff topological spaces and $\Sigma = \mathcal{B}(X)$, $\mathcal{L} = \mathcal{B}(Y)$, then we use the term **Borel isomorphism**.

Proposition 2.6.21. *If X, Y are Polish spaces and $f : X \to Y$ is a Borel isomorphism, then $E \subseteq X$ is analytic if and only if $f(E) \subseteq Y$ is analytic.*

Proof. $\Longrightarrow$: Since $E \subseteq X$ is analytic, we have $E = \mathfrak{a}(A)$ with A being a $\mathcal{F}_X$-Souslin scheme. Then $f(E) = S(f \circ A)$ with $f \circ A$ being the $\mathcal{B}(Y)$-Souslin scheme defined by $(f \circ A)(x) = f(A(x))$. Hence, $f(E)$ is analytic; see Remark 2.6.12.

$\Longleftarrow$: This is proven in a similar way. ☐

Corollary 2.6.22. *If X, Y are Polish spaces, $f : X \to Y$ is Borel measurable, $E \in \mathcal{B}(X)$ and $f|_E$ is one-to-one, then $f(E) \in \mathcal{B}(Y)$.*

Now we examine the measurability of analytic sets. Although analytic sets need not be Borel, it turns out that they will always be measurable for the completion of any probability measure defined on the Borel sets.

Definition 2.6.23. Let X be a Polish space and let $M_1^+(X)$ be the set of probability measures on X. Given $\mu \in M_1^+(X)$ let $\mathcal{B}(X)_\mu$ be the completion of the Borel σ-algebra $\mathcal{B}(X)$. Recall that $\mathcal{B}(X)_\mu$ can be described as the family of all sets of the form $B \cup N$ with $B \in \mathcal{B}(X)$ and N is a subset of a μ-null set. The **universal σ-algebra** $\hat{\Sigma}_X$ is defined by

$$\hat{\Sigma}_X = \bigcap_{\mu \in M_1^+(X)} \mathcal{B}(X)_\mu .$$

The elements of $\hat{\Sigma}_X$ are said to be **universally measurable sets**.

Next we will see that analytic sets are universally measurable.

Theorem 2.6.24. *If X is a Polish space and $E \subseteq X$ is analytic, then $E \in \hat{\Sigma}_X$, that is, E is universally measurable.*

Proof. According to Proposition 2.6.9 there exists $f\colon \mathbb{N}^\infty \to X$ being a continuous map such that $f(\mathbb{N}^\infty) = E$. Let $\mu \in M_1^+(X)$ and for any $k, m \in \mathbb{N}$ let

$$N(k, m) = \{p = (p_k) \in \mathbb{N}^\infty : p_k \le m\} \ .$$

We see that $f(N(k, m)) \nearrow f(\mathbb{N}^\infty) = E$ as $m \to +\infty$. So, for a given $\varepsilon > 0$ there exists $m_1 \in \mathbb{N}$ such that $\mu^*(f(N(1, m_1))) \ge \mu^*(E) - \varepsilon/2$ with μ^* being the outer measure corresponding to μ; see Proposition 2.1.34.

Similarly, for all $k \in \mathbb{N}$, we can find $m_k \in \mathbb{N}$ such that

$$\mu(\overline{f(C_k)}) \ge \mu^*(f(C_k)) \ge \mu^*(E) - \sum_{i=1}^{k} \frac{\varepsilon}{2^i} \ge \mu^*(E) - \varepsilon$$

with $C_k = \bigcap_{i=1}^k N(i, m_i)$. Letting $k \to \infty$ we see that $C_k \searrow C = \bigcap_{i\ge1} N(i, m_i)$. Note that each C_k is closed and C is compact. Let $U \supseteq C$ be open. Then U is a union of basic open sets and the compactness of C implies that this union is finite. Each basic open set depends on only finitely many coordinates. Let $j \in \mathbb{N}$ be the largest index of any coordinate in the definition of the sets of this finite subcover. We have $C_j \subseteq U$ and according to Problem 1.51 it holds that $\mu(f(C)) \ge \mu^*(E) - \varepsilon$. The set $f(X) \subseteq X$ is compact. Taking $\varepsilon = 1/n$ with $n \in \mathbb{N}$ we have a countable union of compact sets that is a Borel set $B \subseteq E$ with $\mu(B) = \mu^*(E)$. Therefore $\mu^*(E \setminus B) = 0$ and $E \in \mathcal{B}(X)_\mu$; see Proposition 2.1.41. We conclude that $E \in \hat{\Sigma}_X$. □

The following characterization of the universal σ-algebra $\hat{\Sigma}_X$ is immediate from Definition 2.6.23 and the proof of Theorem 2.6.24.

Proposition 2.6.25. *If X is a Polish space and $E \subseteq X$, then $E \in \hat{\Sigma}_X$ if and only if for any $\mu \in M_1^+(X)$ there exists $B \in \mathcal{B}(X)$ such that $\mu(E \triangle B) = 0$.*

There is a third σ-algebra that we can define for a Polish space X.

Definition 2.6.26. Let X be a Polish space. The **analytic σ-algebra** α_X is the smallest σ-algebra containing the analytic subsets of X, that is, $\alpha_X = \sigma(S(\mathcal{F}_X))$.

If $E \in \alpha_X$, then we say that E is **analytically measurable**. Therefore on any Polish space X we can define three important σ-algebras:

- $\mathcal{B}(X)$ = the Borel σ-algebra.
- α_X = the analytic σ-algebra.
- $\hat{\Sigma}_X$ = the universal σ-algebra.

These σ-algebras are related as follows

$$\mathcal{B}(X) \subseteq S(\mathcal{F}_X) \subseteq \alpha_X \subseteq \hat{\Sigma}_X \ . \tag{2.6.11}$$

If X is countable, then all classes in (2.6.11) are equal to 2^X. If X is uncountable, then all inclusions in (2.6.11) are strict.

Definition 2.6.27. Let X, Y be Polish spaces, $C \subseteq X$ be nonempty, and $f\colon C \to Y$. We say that f is **analytically** (resp. **universally**) **measurable** if $C \in \alpha_X$ (resp. $C \in \hat{\Sigma}_X$) and $f^{-1}(E) \in \alpha_X$ (resp. $f^{-1}(E) \in \hat{\Sigma}_X$) for all $E \in \mathcal{B}(Y)$.

The composition of functions preserves universal measurability.

Proposition 2.6.28. *If X, Y, Z are Polish spaces, $C \in \hat{\Sigma}_X$, $E \in \hat{\Sigma}_Y$, $f\colon C \to Y$, $g\colon E \to Y$, and $f(C) \subseteq E$, then $g \circ f\colon C \to Z$ is universally measurable.*

Proof. Let $B \in \mathcal{B}(Z)$. The universal measurability of g implies that $g^{-1}(B) \in \hat{\Sigma}_Y$. Since $(g \circ f)^{-1}(B) = f^{-1}(g^{-1}(B))$ we need to show that for every $D \in \hat{\Sigma}_Y$, $f^{-1}(D) \in \hat{\Sigma}_X$. Given $\mu \in M_1^+(X)$ we consider the image measure $\mu \circ f^{-1}$ on Y; see Definition 2.5.23. Let $F \in \mathcal{B}(Y)$ be such that $(\mu \circ f^{-1})(F \triangle D) = 0$. The universal measurability of f implies that $f^{-1}(F) \in \hat{\Sigma}_X$. Hence, by applying Proposition 2.6.25, there exists $G \in \mathcal{B}(X)$ such that $\mu(G \triangle f^{-1}(F)) = 0$. Therefore $\mu(G \triangle f^{-1}(D)) = 0$ and this implies, due to Proposition 2.6.25, that $f^{-1}(D) \in \hat{\Sigma}_X$. $\square$

From the proof above, we deduce the following corollary.

Corollary 2.6.29. *If X, Y are Polish spaces, $C \in \hat{\Sigma}_X$, and $f\colon C \to Y$ is universally measurable, then for every $E \in \hat{\Sigma}_Y$ we have $f^{-1}(E) \in \hat{\Sigma}_X$.*

Remark 2.6.30. Composition of functions does not preserve analytic measurability. The composition of two analytically measurable functions is universally measurable.

2.7 Selection and Projection Theorems

In this section we prove some results, which in addition to being interesting from a purely theoretical viewpoint, are used in many applied fields such as calculus of variations, optimization, optimal control, and mathematical economics.

The mathematical setting is the following: We are given a measurable space (Ω, Σ), a separable metric space (X, d), and a multifunction (so-called set-valued map) $F\colon \Omega \to 2^X$. The first basic question we want to study is whether we can find a single-valued, Σ-measurable map $f\colon \Omega \to X$ such that $f(w) \in F(w)$ for all $w \in \Omega$. Such a map is called a **measurable selection** of F. Its existence is not straightforward. First we need to introduce and discuss some notions of measurability for the multifunction F.

In what follows, (Ω, Σ) is a measurable space and (X, d) is a separable metric space. Additional hypotheses will be introduced as needed.

Definition 2.7.1. Let $F\colon \Omega \to 2^X$ be a multifunction.

(a) We say that F is **measurable** if for every open $U \subseteq X$,
$$F^-(U) = \{w \in \Omega\colon F(w) \cap U \neq \emptyset\} \in \Sigma\,.$$

(b) We say that F is **graph measurable** if
$$\operatorname{Gr} F = \{(w, x) \in \Omega \times X\colon x \in F(w)\} \in \Sigma \bigotimes \mathcal{B}(X)\,.$$

Remark 2.7.2. Note that in the definitions above we do not require that F be nonempty valued. By **domain of** F we mean the set $\operatorname{dom} F = \{w \in \Omega\colon F(w) \neq \emptyset\}$. If F is measurable, then clearly $\operatorname{dom} F \in \Sigma$ and so for measurable multifunctions, there is no loss of generality in assuming that $\operatorname{dom} F = X$. If F is single-valued, then measurability coincides with Σ-measurability. Evidently both notions make sense even if X is a general Hausdorff topological space. However, the most interesting properties and results can be established for X being a Polish space in the case of measurable multifunctions and for X being a Souslin space in the case of graph measurable multifunctions. Therefore, we see that the theory of measurable multifunctions requires separability of the ambient space. Without it we cannot go far. For economy in the presentation we have fixed X to be a separable metric space.

Proposition 2.7.3. *If $F\colon \Omega \to 2^X$ and for all closed $C \subseteq X$, $F^-(C) = \{w \in \Omega\colon F(w) \cap C \neq \emptyset\} \in \Sigma$, then F is measurable.*

Proof. From Proposition 1.5.8 we know that every open set $U \subseteq X$ is F_σ. So, $U = \bigcup_{n\ge1} C_n$ with closed $C_n \subseteq X$ for all $n \in \mathbb{N}$. Then, by hypothesis,

$$F^-(U) = F^-\left(\bigcup_{n\ge1} C_n\right) = \bigcup_{n\ge1} F^-(C_n) \in \Sigma\,.$$

Hence, F is measurable. $\square$

Remark 2.7.4. The converse of the proposition above is not true in general.

The measurability of F can be characterized functionally.

Proposition 2.7.5. *The multifunction $F\colon \Omega \to 2^X$ is measurable if and only if for all $x \in X$, the $\overline{\mathbb{R}}_+$-valued function $w \to d(x, F(w))$ is Σ-measurable.*

Proof. $\Longrightarrow$: Given $x \in X$ and $\eta > 0$, let $L_\eta(x) = \{w \in \Omega\colon d(x, F(w)) < \eta\}$. Then we see that $L_\eta(x) = F^-(B_\eta(x))$ with $B_\eta(x) = \{u \in X\colon d(u, x) < \eta\}$. Hence, $L_\eta(x) \in \Sigma$ and this implies the Σ-measurability of $w \to d(x, F(w))$.

$\Longleftarrow$: Given $x \in X$ and $\eta > 0$, by hypothesis, it holds that

$$F^-(B_\eta(x)) = L_\eta(x) \in \Sigma\,. \tag{2.7.1}$$

Let $U \subseteq X$ be open. The separability of X implies that $U = \bigcup_{n\ge1} B_{\eta_n}(x_n)$. Then

$$F^-(U) = \bigcup_{n\ge1} F^-(B_{\eta_n}(x_n)) \in \Sigma\,;$$

see (2.7.1). Thus, F is measurable. $\square$

Let us introduce some notation:

$$P_f(X) = \{A \subseteq X\colon A \text{ is nonempty and closed}\}\,, \qquad \hat{P}_f(X) = P_f(X) \cup \{\emptyset\}\,,$$
$$P_k = \{A \subseteq X\colon A \text{ is nonempty and compact}\}\,.$$

Proposition 2.7.6. *If $F\colon \Omega \to \hat{P}_f(X)$ is measurable, then F is graph measurable.*

Proof. Since F is closed valued, we have that

$$\operatorname{Gr} F = \{(w, x) \in \Omega \times X\colon d(x, F(w)) = 0\} . \tag{2.7.2}$$

But using Proposition 2.7.5 we see that $(w, x) \to d(x, F(w))$ is a Carathéodory function; see Definition 2.2.30. Then Proposition 2.2.3 implies that it is jointly measurable and so from (2.7.2) it follows that $\operatorname{Gr} F \in \Sigma \bigotimes \mathcal{B}(X)$, that is, F is graph measurable. $\square$

Recall that if $U \subseteq X$ is open, then $A \cap U \neq \emptyset$ if and only if $\overline{A} \cap U \neq \emptyset$. This straightforward observation leads to the following useful result.

Proposition 2.7.7. *The multifunction $F\colon \Omega \to 2^X$ is measurable if and only if $w \to \overline{F}(w) = \overline{F(w)}$ is measurable.*

For $P_k(X)$-valued multifunctions we obtain the converse of Proposition 2.7.3.

Proposition 2.7.8. *If $F\colon \Omega \to P_k(X)$ is measurable, then for all closed $C \subseteq X$, it holds that $F^-(C) = \{w \in \Omega\colon F(w) \cap C \neq \emptyset\} \in \Sigma$.*

Proof. In what follows for every $E \subseteq X$, we set

$$F^+(E) = \{w \in \Omega\colon F(w) \subseteq E\} . \tag{2.7.3}$$

Let $C \subseteq X$ be nonempty and closed and let $U_n = \{x \in X\colon d(x, C) > 1/n\}$ with $n \in \mathbb{N}$. Then U_n is open for each $n \in \mathbb{N}$ and $\{U_n\}_{n\geq 1}$ is increasing. We set $D_n = \overline{U}_n$ with $n \in \mathbb{N}$. Then

$$X \setminus C = \bigcup_{n\geq 1} U_n = \bigcup_{n\geq 1} D_n . \tag{2.7.4}$$

Let $w \in F^+(X \setminus C)$. Then $F(w) \subseteq X \setminus C$; see (2.7.3). Due to (2.7.4) and recalling that $\{U_n\}_{n\geq 1}$ is increasing as well as F is $P_k(X)$-valued, we see that there exists $n \in \mathbb{N}$ such that $F(w) \subseteq U_n \subseteq D_n$. Then, due to (2.7.3), it follows $F^+(X \setminus C) = \bigcup_{n\geq 1} F^+(D_n)$. Since F is measurable we derive

$$F^-(C) = X \setminus (F^+(X \setminus C)) = X \setminus \bigcup_{n\geq 1} F^+(D_n) = \bigcap_{n\geq 1} F^-(X \setminus D_n) \in \Sigma . \qquad \square$$

Proposition 2.7.9. *If $F\colon \Omega \to P_f(X)$ is measurable, then $F^-(K) \in \Sigma$ for all compact $K \subseteq X$.*

Proof. On account of Theorem 1.5.21 we may assume that X is dense in a compact metric space (Y, d_Y). Consider the multifunction $G\colon \Omega \to P_k(Y)$ defined by $G(w) = \overline{F(w)}^{d_Y}$. Proposition 2.7.7 guarantees the measurability of G. Now let $K \subseteq X$ compact. We have

$$F^-(K) = \{w \in \Omega\colon F(w) \cap K \neq \emptyset\} = \{w \in \Omega\colon G(w) \cap K \neq \emptyset\} = G^-(K) \in \Sigma$$

by Proposition 2.7.8. $\square$

When we introduce extra structure on the space, we can say more. To be more precise, we have the following result.

Proposition 2.7.10. *If X is σ-compact and $F\colon \Omega \to P_f(X)$, then the following statements are equivalent:*
(a) *$F^-(C) \in \Sigma$ for every closed $C \subseteq X$.*
(b) *F is measurable.*
(c) *$F^-(K) \in \Sigma$ for every compact $K \subseteq X$.*

Proof. (a) $\Longrightarrow$ (b): This implication follows from Proposition 2.7.3.

(b) $\Longrightarrow$ (c): This implication follows from Proposition 2.7.9.

(c) $\Longrightarrow$ (a): By hypothesis, $X = \bigcup_{n\geq 1} K_n$ with compact K_n. Then for closed $C \subseteq X$ it holds that

$$F^-(C) = \bigcup_{n\geq 1}(C \cap K_n) \in \Sigma$$

since $C \cap K_n \subseteq X$ is compact for every $n \in \mathbb{N}$. □

The next theorem summarizes the measurability properties of closed valued multifunctions.

Theorem 2.7.11. *Let (Ω, Σ) be a measurable space, (X, d) a separable metric space, and $F\colon \Omega \to P_k(X)$ a multifunction. Consider the following statements:*
(a) *$F^-(C) \in \Sigma$ for every closed $C \subseteq X$.*
(b) *F is measurable.*
(c) *For every $x \in X$, $w \to d(x, F(w))$ is Σ-measurable.*
(d) *F is graph measurable.*
Then (a) $\Longrightarrow$ (b) $\Longleftrightarrow$ (c) $\Longrightarrow$ (d) and if X is σ-compact, then (a) $\Longleftrightarrow$ (b) $\Longleftrightarrow$ (c) $\Longrightarrow$ (d).

Now we are ready for the first existence theorem for measurable selections. The result is known as the "Kuratowski–Ryll Nardzewski Selection Theorem."

Theorem 2.7.12 (Kuratowski–Ryll Nardzewski Selection Theorem). *If (Ω, Σ) is a measurable space, X is a Polish space, and $F\colon \Omega \to P_f(X)$ is a measurable multifunction, then F admits a measurable selection, that is, there exists a Σ-measurable function $f\colon \Omega \to X$ such that $f(w) \in F(w)$ for all $w \in \Omega$.*

Proof. Let d be a bounded compatible metric on X. We may assume that the d-diameter of X is strictly less than 1. Let $\{x_n\}_{n\geq 1}$ be dense in X. We produce inductively a sequence of Σ-measurable maps $f_n\colon \Omega \to X$ with $n \in \mathbb{N}_0$, which satisfy

$$d(f_n(w), F(w)) < \frac{1}{2^n} \quad \text{for all } n \in \mathbb{N}_0 \text{ and for all } w \in \Omega\,, \tag{2.7.5}$$

$$d(f_n(w), f_{n-1}(w)) < \frac{1}{2^{n-1}} \quad \text{for all } n \in \mathbb{N} \text{ and for all } w \in \Omega\,. \tag{2.7.6}$$

Let us start with f_0. We define $f_0\colon \Omega \to X$ by $f_0(w) = x_1$ for all $w \in \Omega$. Since by hypothesis $\operatorname{diam} X < 1$, inequality (2.7.5) holds for $n = 0$. For the induction hypothesis, we assume that we have already produced $f_0, f_1, \dots, f_{n-1}$, which satisfy (2.7.5) as well as (2.7.6). For every $k \in \mathbb{N}$, we define

$$A_k^n = \left\{w \in \Omega \colon d(x_k, F(w)) < \frac{1}{2^n}\right\}, \qquad C_k^n = \left\{w \in \Omega \colon d(x_k, f_{n-1}(w)) < \frac{1}{2^{n-1}}\right\}$$

and $E_k^n = A_k^n \cap C_k^n$. First we show that $\Omega = \bigcup_{k\geq 1} E_k^n$. So, let $w \in \Omega$. The induction hypothesis says that there exists $u \in F(w)$ such that $d(f_{n-1}(w), u) < 1/(2^{n-1})$; see (2.7.5). The density of $\{x_n\}_{n\geq 1}$ in X implies that there is $k \in \mathbb{N}$ such that $d(x_k, u) < 1/2^n$ and $d(x_k, u) + d(u, f_{n-1}(w)) < 1/2^{n-1}$. By the triangle inequality we have $d(x_k, f_{n-1}(w)) < 1/2^{n-1}$. Hence, we see that $w \in E_k^n$, thus $\Omega = \bigcup_{k\geq 1} E_k^n$. The measurability of F and Proposition 2.7.5 imply that $A_k^n \in \Sigma$. Taking the induction hypothesis into account, the Σ-measurability of f_{n-1} implies that $C_k^n \in \Sigma$. Therefore $E_k^n \in \Sigma$. We define a function $f_n \colon \Omega \to X$ by setting $f_n(w) = x_k$ for all $w \in E_k^n \setminus \bigcup_{i=1}^{k-1} E_i^n$. Hence f_n is Σ-measurable and satisfies (2.7.5) and (2.7.6). This completes the induction.

From (2.7.6) we infer that for every $w \in \Omega$, $\{f_n(w)\}_{n\geq 0} \subseteq X$ is a Cauchy sequence. Therefore

$$f_n(w) \xrightarrow{d} f(w) \quad \text{for all } w \in \Omega \text{ as } n \to \infty .$$

Proposition 2.2.12 implies that f is Σ-measurable and $d(f(w), F(w)) = 0$ for all $w \in \Omega$. Since $F(w) \in P_f(X)$ for all $w \in \Omega$, we conclude that $f(w) \in F(w)$ for all $w \in \Omega$. Therefore $f \colon \Omega \to X$ is a Σ-measurable selection of F. □

In fact we can produce a whole sequence of dense Σ-measurable selections of F.

Theorem 2.7.13. *If (Ω, Σ) is a measurable space, X is a Polish space and $F \colon \Omega \to P_f(X)$, then the following statements are equivalent:*
(a) *F is measurable;*
(b) *there exists a sequence of Σ-measurable selections $f_n \colon \Omega \to X$ of F such that $F(w) = \overline{\{f_n(w)\}_{n\geq 1}}$ for all $w \in \Omega$.*

Proof. (a) $\Longrightarrow$ (b): Let $\{U_n\}_{n\geq 1}$ be a countable basis for the metric topology of X. For every $n \in \mathbb{N}$, we define the multifunction

$$F_n(w) = \begin{cases} F(w) \cap U_n & \text{if } F(w) \cap U_n \neq \emptyset , \\ F(w) & \text{otherwise} , \end{cases}$$

for all $w \in \Omega$. Let $\Omega_n = F^-(U_n) \in \Sigma$ with $n \in \mathbb{N}$. Then for every open set $V \subseteq X$ we obtain

$$F_n^-(V) = \{w \in \Omega_n \colon F(w) \cap U_n \neq \emptyset\} \cup \{w \in (\Omega \setminus \Omega_n) \colon F(w) \cap V \neq \emptyset\} \in \Sigma ,$$

which implies that F_n is measurable for all $n \in \mathbb{N}$. Then, thanks to Proposition 2.7.7, it follows that $\overline{F}_n$ is measurable for all $n \in \mathbb{N}$.

Invoking Theorem 2.7.12 there exists a sequence $f_n \colon \Omega \to X$ with each f_n being a Σ-measurable selection of F_n. Note that $\overline{F}_n(w) \subseteq F(w)$ for all $n \in \mathbb{N}$ and for all $w \in \Omega$.

Hence, f_n is a Σ-measurable selection of F. Evidently, $F(w) = \overline{\{f_n(w)\}_{n\geq 1}}$ for all $w \in \Omega$.

(b) $\Longrightarrow$ (a): For every $x \in X$, it holds that

$$d(x, F(w)) = \inf_{n\geq 1} d(x, f_n(w)) \quad \text{for all } w \in \Omega ,$$

which demonstrates, because of Proposition 2.2.10, that $w \to d(x, F(w))$ is Σ-measurable. Hence, due to Proposition 2.7.5, we get that F is measurable. $\square$

We can state another measurable selection theorem for graph measurable multifunctions. First we start with a definition.

Definition 2.7.14. (a) A family $\mathcal{L}$ of subsets of a set X is said to **separate points** in X if for every two distinct points $x, u \in X$ there is $A \in \mathcal{L}$ such that $x \in A, u \notin A$ or $x \notin A, u \in A$.

(b) A family $\mathcal{D}$ of $\mathbb{R}$-valued functions on X is said to **separate points** in X if for every two distinct points $x, u \in X$ there is $f \in \mathcal{D}$ such that $f(x) \neq f(w)$.

(c) A σ-algebra $\mathcal{L}$ of subsets of a set X is said to be **countably generated** if there is a countable family $\{A_n\}_{n\geq 1} \subseteq \mathcal{L}$ such that $\mathcal{L} = \sigma(\{A_n\}_{n\geq 1})$.

(d) A σ-algebra $\mathcal{L}$ of subsets of a set X is said to be **countably separated** if there is a countable family $\{A_n\}_{n\geq 1} \subseteq \mathcal{L}$ that separates points in X, see (a).

Example 2.7.15. Suppose X is a separable metric space and $\mathcal{L} = \mathcal{B}(X)$ being the Borel σ-algebra. Then $\mathcal{B}(X)$ is countably generated and countably separated. To see this consider $\{U_n\}_{n\geq 1}$ being a countable basis for the metric topology. Then $\sigma(\{U_n\}_{n\geq 1}) = \mathcal{B}(X)$, that is, $\mathcal{B}(X)$ is countably generated and clearly, $\{U_n\}_{n\geq 1}$ separates points in X, that is, $\mathcal{B}(X)$ is countably separated.

Proposition 2.7.16. *If (Ω, Σ) is a measurable space, Y is a Hausdorff topological space and $D \in \Sigma \bigotimes \mathcal{B}(Y)$, then there exists $\Sigma_0 \subseteq \Sigma$ being a countably generated sub-σ-algebra of Σ such that $D \in \Sigma_0 \bigotimes \mathcal{B}(Y)$.*

Proof. Let $\mathcal{L} = \{C \in \Sigma \bigotimes \mathcal{B}(X):$ the conclusion of the proposition holds$\}$. Clearly $\mathcal{L}$ includes all measurable rectangles; see Remark 2.2.24. Moreover $\mathcal{L}$ is closed under complementation. Let $\{C_n\}_{n\geq 1} \subseteq \mathcal{L}$. Then $C_n \in \Sigma_{on} \bigotimes \mathcal{B}(X)$ with $\Sigma_{on} \subseteq \Sigma$ being a countably generated sub-σ-algebra. Then $\bigcup_{n\geq 1} C_n \in \sigma(\bigcup_{n\geq 1} \Sigma_{on}) \bigotimes \mathcal{B}(X)$ and $\sigma(\bigcup_{n\geq 1} \Sigma_{on})$ is countably generated. Therefore $\mathcal{L}$ is a σ-algebra and so we must have $\mathcal{L} = \Sigma \bigotimes \mathcal{B}(X)$. $\square$

Extending the notion of universal σ-algebra (see Definition 2.6.23) to arbitrary measurable spaces, we state the following definition.

Definition 2.7.17. Let (Ω, Σ) be a measurable space. The **universal σ-algebra** corresponding to Σ is defined by $\hat{\Sigma} = \bigcap_{\mu \in M_1^+(\Omega)} \Sigma_\mu$ where $M_1^+(\Omega)$ denotes the set of all probability measures on Ω and Σ_μ is the μ-completion of Σ. We say that the measurable space (Ω, Σ) is complete if $\Sigma = \hat{\Sigma}$.

Using this definition and Corollary 2.6.29 (see also the proof of Proposition 2.6.28) we have the following result.

Proposition 2.7.18. *If (Ω_1, Σ_1) and (Ω_2, Σ_2) are measurable spaces and $f\colon \Omega_1 \to \Omega_2$ is a (Σ_1, Σ_2)-measurable map, then f is $(\hat{\Sigma}_1, \hat{\Sigma}_2)$-measurable.*

The next result is the original version of the so-called “Yankov-von Neumann Selection Theorem.” For its proof we refer to Klein–Thompson [178, Theorem 14.3.2,p. 166].

Theorem 2.7.19 (Yankov-von Neumann Selection Theorem). *If X, Y are Polish spaces, $F\colon X \to 2^Y \setminus \{\emptyset\}$, and $\operatorname{Gr} F \in \alpha_{X\times Y}$, then there exists an analytically measurable function $f\colon X \to Y$ such that $f(x) \in F(x)$ for all $x \in X$.*

Recalling that a Souslin space is the continuous image of a Polish space (see Definition 1.5.51), from Theorem 2.7.19 we easily deduce the following result.

Theorem 2.7.20. *If X is a Borel subset of a Polish space, Y is a Souslin space, $F\colon X \to 2^Y \setminus \{\emptyset\}$, and $\operatorname{Gr} F \subseteq X \times Y$ is a Souslin subset, then there exists an analytically measurable map $f\colon X \to Y$ such that $f(x) \in F(x)$ for all $x \in X$.*

Remark 2.7.21. Note that Borel sets of Polish spaces are usually called **Borel spaces**.

Proposition 2.7.22. *If (Ω, Σ) is a measurable space such that Σ is countably generated and countably separated, then there is a subset E of $\{0, 1\}^{\mathbb{N}}$ such that (Ω, Σ) and $(E, \mathcal{B}(E))$ are isomorphic; see Definition* 2.6.20.

Proof. Let $\{A_n\}_{n\geq 1}$ be the generators of Σ. We are going to show that they separate points in Ω. Arguing by contradiction, suppose that for some $w, w' \in \Omega$, $w \neq w'$ it holds that $\chi_{A_n}(w) = \chi_{A_n}(w')$ for all $n \in \mathbb{N}$. Let $\Sigma_0 = \{A \subseteq \Omega\colon \chi_{A_n}(w) = \chi_{A_n}(w')\}$. Evidently Σ_0 is a σ-algebra and $A_n \in \Sigma_0$ for all $n \in \mathbb{N}$, thus $\Sigma \subseteq \Sigma_0$, which contradicts the fact that Σ is countably separated. Let $f\colon \Omega \to \{0, 1\}^{\mathbb{N}}$ be defined by $f(w) = \{\chi_{A_n}(w)\}_{n\geq 1}$. Clearly f is one-to-one and Σ-measurable. We need to show that $f^{-1}\colon E = f(\Omega) \to \Omega$ is measurable. So, we want to show that if $A \in \Sigma$, then $f(A) \in \mathcal{B}(E)$. Let $\Sigma_1 = \{A \subseteq \Omega\colon f(A) \in \mathcal{B}(E)\}$. This is a σ-algebra and $A_n \in \Sigma_1$ for all $n \in \mathbb{N}$ since $f(A_n) = \{(e_k) \in \{0, 1\}^{\mathbb{N}}\colon e_n = 1\} \cap E$. Therefore, $\Sigma \subseteq \Sigma_1$ and we have proven the measurability of f^{-1}. Hence, we have that (Ω, Σ) and $(E, \mathcal{B}(E))$ are isomorphic. ☐

Remark 2.7.23. Recall that $\{0, 1\}^{\mathbb{N}}$ and $\mathbb{N}^{\infty}$ are isometrically isomorphic. Hence, $\{0, 1\}^{\mathbb{N}}$ is Polish.

Proposition 2.7.24. *If (Ω, Σ) is a measurable space such that Σ is countably generated and countably separated, X is a Souslin space and $F\colon \Omega \to 2^X \setminus \{\emptyset\}$ is a graph measurable multifunction, then F admits a $\hat{\Sigma}$-measurable selection.*

Proof. Invoking Proposition 2.7.22 we know that there exists $E \subseteq \{0, 1\}^{\mathbb{N}}$ such that (Ω, Σ) and $(E, \mathcal{B}(E))$ are isomorphic. Let $h\colon \Omega \to E$ be this isomorphism. The measurable spaces $(\Omega \times X, \Sigma \bigotimes \mathcal{B}(X))$ and $(E \times X, \mathcal{B}(E) \bigotimes \mathcal{B}(X))$ are isomorphic. Moreover, from Proposition 2.2.26(b) we know that $\mathcal{B}(E) \bigotimes \mathcal{B}(X) = \mathcal{B}(E \times X)$.

We introduce the multifunction $F_1 : E \to 2^X \setminus \{\emptyset\}$ defined by $F_1 = F \circ h^{-1}$. We have $\operatorname{Gr} F_1 = (h, \mathrm{id}_X)(\operatorname{Gr} F)$ with id_X being the identity map on X. Therefore $\operatorname{Gr} F_1 \in \mathcal{B}(E) \bigotimes \mathcal{B}(X) = \mathcal{B}(E \times X)$.

Hence, there exists $D_1 \in \mathcal{B}(P \times X)$ with $P = \{0, 1\}^{\mathbb{N}}$ such that $\operatorname{Gr} F_1 = D_1 \cap (E \times X)$. Then $E = \operatorname{proj}_P \operatorname{Gr} F_1 \subseteq E_1 = \operatorname{proj}_P D_1$. Let $h' : \Omega \to E_1$ be defined by $h'(w) = h(w)$ for all $w \in \Omega$. Then h' is injective and Σ-measurable. Let $F_2 : E_1 \to 2^X \setminus \{\emptyset\}$ be the multifunction defined by $\operatorname{Gr} F_2 = D_1$. We claim that

$$F_2(h'(w)) = F_1(h(w)) \quad \text{for all } w \in \Omega \, . \tag{2.7.7}$$

To this end, note that for every $u \in E$ we have

$$F_1(u) = \operatorname{proj}_X[\operatorname{Gr} F_1 \cap (\{u\} \times X)] \quad \text{and} \quad F_2(u) = \operatorname{proj}_X[\operatorname{Gr} F_2 \cap (\{u\} \times X)] \, .$$

Recall that $\operatorname{Gr} F_1 = \operatorname{Gr} F_2 \cap (E \times X)$. So

$$\operatorname{Gr} F_1 \cap (\{u\} \times X) = \operatorname{Gr} F_2 \cap (\{u\} \times X) \, ,$$

which gives $F_1(u) = F_2(u)$ for all $u \in E$ and this proves (2.7.7).

Since $D_1 \in \mathcal{B}(E \times X)$, D_1 is a Souslin subset of $E \times X$. Hence, we can apply Theorem 2.7.20 and obtain $f_2 : E_1 \to X$ being an analytically measurable map such that $f_2(u) \in F_2(u)$ for all $u \in E_1$. Since h' is $(\Sigma, \mathcal{B}(E_1))$-measurable, using Proposition 2.7.18 we have that h' is $(\hat{\Sigma}, \hat{\mathcal{B}}(E_1))$-measurable. Let $f = f_2 \circ h'$. Then $f : \Omega \to X$ is $\hat{\Sigma}$-measurable and $f(w) \in F(w)$ for all $w \in \Omega$. ☐

Now we are ready for the second measurable selection theorem which is graph conditioned. The result is usually known as the "Yankov-von Neumann–Aumann Selection Theorem."

Theorem 2.7.25 (Yankov-von Neumann–Aumann Selection Theorem). *If (Ω, Σ) is a complete measurable space, X is a Souslin space, and $F : \Omega \to 2^X \setminus \{\emptyset\}$ is graph measurable, then F admits a Σ-measurable selection.*

Proof. Using Proposition 2.7.16 there is a countably generated sub-σ-algebra $\Sigma_0 \subseteq \Sigma$ such that $\operatorname{Gr} F \in \Sigma_0 \bigotimes \mathcal{B}(X)$. On Ω we define an equivalence relation $\sim$ by

$$w \sim w' \quad \text{if and only if} \quad \chi_A(w) = \chi_A(w') \quad \text{for all } A \in \Sigma_0 \, . \tag{2.7.8}$$

Let $\Omega_* = \Omega / \sim$ and let $p : \Omega \to \Omega_*$ be the canonical projection on the quotient space, that is, $p(w) = \dot{w}$ being the equivalence class of $w \in \Omega$. Let $\Sigma_* = p(\Sigma_0) = \{p(A) : A \in \Sigma_0\}$. It is easy to see that Σ_* is a σ-algebra and if $\{A_n\}_{n \ge 1}$ are the generators of Σ_0, that is, $\Sigma_0 = \sigma(\{A_n\}_{n\ge1})$, then $\Sigma_* = \sigma(\{p(A_n)\}_{n\ge1})$. Therefore Σ_* is countably generated.

Next suppose that $\dot{w} \neq \dot{w}'$. Then we can find $A \in \Sigma_0$ such that $\chi_A(w) \neq \chi_A(w')$; see (2.7.8). This is equivalent saying that $\chi_{p(A)}(\dot{w}) = \chi_{p(A)}(\dot{w}')$. It follows that Σ_* is also countably separated. Moreover, note that p is a one-to-one correspondence between Σ_0 and Σ_*. Let id_X be the identity map on X and let $\eta\colon \Omega \times X \to \Omega_* \times X$ be defined by $\eta = (p, \mathrm{id}_X)$. Then $\operatorname{Gr} F \in \Sigma \bigotimes \mathcal{B}(X)$ implies that $\eta(\operatorname{Gr} F) \in \Sigma_* \bigotimes \mathcal{B}(X)$. Let $F_1\colon \Omega_* \to 2^X \setminus \{\emptyset\}$ defined by $\operatorname{Gr} F_1 = \eta(\operatorname{Gr} F)$. We can now apply Proposition 2.7.24 and produce a $\hat{\Sigma}_*$-measurable selection $f_1\colon \Omega_* \to X$ of F_1, that is, $f_1(w) \in F_1(w)$ for all $w \in \Omega$. Let $f = f_1 \circ p$ and for $w \in \Omega$ we define $\mathcal{D}(w) = \{A \in \Sigma_0 \bigotimes \mathcal{B}(X)\colon A_{w'} = A_w \text{ for all } w' \in \dot{w}\}$. Recall that A_w is the w-section of A; see Definition 2.2.27. Note that $\mathcal{D}(w)$ is an algebra and a monotone class. Hence, Theorem 2.1.12 implies that $\mathcal{D}(w)$ is a σ-algebra. It follows that $\mathcal{D}(w) = \Sigma_0 \bigotimes \mathcal{B}(X)$. Since $\operatorname{Gr} F_w = F(w)$, we see that F is constant on $\dot{w}$ and we have $F(w') = F_1(\dot{w})$ for all $w' \in \dot{w}$. Because $f(w') = f(\dot{w})$ we obtain that $f(w) \in F(w)$ for all $w \in \Omega$. Proposition 2.7.18 implies that f is Σ-measurable. This finishes the proof. □

As for the Kuratowski–Ryll Nardzewski Selection Theorem (see Theorem 2.7.12), we can improve the result above and produce a whole dense sequence of measurable selections. To do this, we will need the following result due to Leese [194, p. 407].

Proposition 2.7.26. *If (Ω, Σ) is a complete measurable space, X is a Souslin space, and $F\colon \Omega \to 2^X \setminus \{\emptyset\}$ is graph measurable, then there exists a Polish space Y, a measurable multifunction $G\colon \Omega \to P_f(Y)$, and a continuous map $h\colon Y \to X$ such that $F(w) = h(G(w))$ for all $w \in \Omega$.*

Remark 2.7.27. Using this proposition and the Kuratowski–Ryll Nardzewski Selection Theorem (see Theorem 2.7.12), we have at once the Yankov-von Neumann–Aumann Selection Theorem; see Theorem 2.7.25. The conclusion of this proposition looks similar to the definition of Souslin spaces; see Definition 1.5.51. For this reason graph measurable multifunctions into a Souslin space are also called multifunctions of **Souslin-type**.

Theorem 2.7.28. *If (Ω, Σ) is a complete measurable space, X is a Souslin space, and $F\colon \Omega \to 2^X \setminus \{\emptyset\}$ is graph measurable, then there exists a sequence of Σ-measurable selections $f_n\colon \Omega \to X$ of F such that $F(w) = \overline{\{f_n(w)\}}_{n\geq 1}$ for all $w \in \Omega$.*

Proof. Applying Proposition 2.7.26 there is a Polish space Y, a measurable multifunction $G\colon \Omega \to P_f(Y)$, and a continuous map $h\colon Y \to$ such that

$$F(w) = h(G(w)) \quad \text{for all } w \in \Omega\,. \tag{2.7.9}$$

Invoking Theorem 2.7.13 there is a sequence of Σ-measurable selections $g_n\colon \Omega \to Y$ of G such that

$$G(w) = \overline{\{g_n(w)\}}_{n\geq 1} \quad \text{for all } w \in \Omega\,. \tag{2.7.10}$$

The continuity of h implies that $f_n = h \circ g_n\colon \Omega \to X$ with $n \in \mathbb{N}$ is a sequence of Σ-measurable selections of F (see (2.7.9)), and using Proposition 1.1.35(b) as well as

(2.7.10) we derive that

$$F(w) \subseteq \overline{\{f_n(w)\}}_{n\geq 1} \quad \text{for all } w \in \Omega\,. \qquad \square$$

Given a Borel subset in a Cartesian product it is natural to ask whether its projection on a factor is Borel as well. The next example shows that the answer to this question is negative. This fact was the starting point for Souslin to develop the theory of analytic sets; see Remarks 2.8.

Example 2.7.29. We show that the projection of a Borel set in $\mathbb{R}^2$ need not be Borel. So, let $X = [0, 1]$, $Y = [0, 1] \cap (\mathbb{R} \setminus \mathbb{Q})$ being the set of the irrationals in $[0, 1]$. From Corollary 1.5.49 we know that Y is a Polish space. Let $A \subseteq X$ be analytic but not Borel and let $f: Y \to A$ be a continuous function. Then $\operatorname{Gr} f \in \mathcal{B}(X \times Y) = \mathcal{B}(X) \bigotimes \mathcal{B}(Y)$ but $\operatorname{proj}_X \operatorname{Gr} f = A \notin \mathcal{B}(X)$.

Next we will show that the projection of a Borel set is universally measurable. We will need two auxiliary lemmata.

Lemma 2.7.30. *If*

$$K_n = \left\{ \sum_{k\geq 1} \frac{s_k}{4^k} : s \in \{0, 1\}^{\mathbb{N}}, s_n = 1 \right\} ,$$

then $K_n \subseteq \mathbb{R}$ is compact and for every $s \in \{0, 1\}^{\mathbb{N}}$, it holds that $\sum\limits_{k\geq 1} s_k/4^k \in K_n$ if and only if $s_n = 1$.

Proof. We know that $\{0, 1\}^{\mathbb{N}}$ is compact. Let $C_n = \{s \in \{0, 1\}^{\mathbb{N}} : s_n = 1\}$. This set is closed, hence compact. Consider the function $f: \{0, 1\}^{\mathbb{N}} \to \mathbb{R}$ defined by $f(s) = \sum_{k\geq 1} s_k/4^k$. Then f is the uniform limit of continuous functions, hence it is continuous. It follows that $f(C_n) = K_n$ is compact. Note that f is injective, hence it is a homeomorphism (see Theorem 1.4.54), and $f(s) \in K_n$ if and only if $s \in C_n$. $\square$

Lemma 2.7.31. *If (Ω, Σ) is a measurable space, Y is a Hausdorff topological space, and $D \in \Sigma \bigotimes \mathcal{B}(Y)$, then there exists $C \in \mathcal{B}(\mathbb{R} \times Y)$ and a Σ-measurable function $f: \Omega \to \mathbb{R}$ such that $D = \{(w, y) \in \Omega \times Y : (f(w), y) \in C\}$.*

Proof. Invoking Proposition 2.7.16 there exists a countably generated sub-σ-algebra $\Sigma_0 \subseteq \Sigma$ such that $D \in \Sigma_0 \bigotimes \mathcal{B}(Y)$. Suppose $\Sigma_0 = \sigma(\{A_n\}_{n\geq 1})$ and consider the function $f: \Omega \to \mathbb{R}$ defined by $f(w) = \sum_{k\geq 1} 1/4^k \chi_{A_k}(w)$. Lemma 2.7.31 says that for every $n \in \mathbb{N}$ and every $w \in \Omega$ we have $f(w) \in K_n$ if and only if $\chi_{A_n}(w) = 1$ if and only if $w \in A_n$. Hence

$$f^{-1}(K_n) = A_n\,. \tag{2.7.11}$$

Evidently f is Σ-measurable and we define $\xi(w, y) = (f(w), y)$ and $\mathcal{L} = \{\xi^{-1}(E) : E \in \mathcal{B}(\mathbb{R}\times Y)\}$. Clearly $\mathcal{L}$ is a σ-algebra and from (2.7.11) we see that $\xi^{-1}(K_n\times B) = f^{-1}(K_n)\times B = A_n \times B$ with $B \in \mathcal{B}(Y)$. This implies $A_n \times B \in \mathcal{L}$ for all $n \in \mathbb{N}$ and for all $B \in \mathcal{B}(Y)$. Therefore $D \in \Sigma_0 \times \mathcal{B}(Y) \subseteq \mathcal{L}$. So, there is a set $C \in \mathcal{B}(\mathbb{R} \times Y)$ such that $D = \xi^{-1}(C)$ and this proves the lemma. $\square$

Now we are ready for the measurable projection theorem known as the "Yankov-von Neumann–Aumann Projection Theorem."

Theorem 2.7.32 (Yankov-von Neumann–Aumann Projection Theorem). *If (Ω, Σ) is a complete measurable space, X is a Souslin space, and $D \in \Sigma \bigotimes \mathcal{B}(X)$, then $proj_\Omega D \in \Sigma$.*

Proof. Lemma 2.7.31 says that there exist $C \in \mathcal{B}(\mathbb{R} \times X)$ and a Σ-measurable function $f\colon \Omega \to \mathbb{R}$ such that $D = \{(w, x) \in \Omega \times X\colon (f(w), x) \in C\}$. Then $\operatorname{proj}_\Omega D = f^{-1}(\operatorname{proj}_\mathbb{R} C)$. The space $X \times \mathbb{R}$ is Souslin (see Proposition 1.5.54(b)), and since $C \in \mathcal{B}(\mathbb{R} \times X)$ it follows that C is Souslin; see Proposition 2.6.11. The set $\operatorname{proj}_\mathbb{R} C$ is the continuous image of a Souslin space, therefore it is a Souslin space as well. As f is Σ-measurable, invoking Proposition 2.7.18, we conclude that $D \in \hat{\Sigma}$. ☐

We mention two more measurable projection theorems. The first is due to Brown–Purves [59].

Theorem 2.7.33. *If X, Y are Polish spaces, $D \in \mathcal{B}(X \times Y) = \mathcal{B}(X) \bigotimes \mathcal{B}(Y)$ and for every $x \in D$, $D_x \subseteq Y$ is σ-compact, then* $\operatorname{proj}_X D \in \mathcal{B}(X)$.

For the second projection theorem, we need to introduce a special class of spaces.

Definition 2.7.34. Let Y be a Hausdorff topological space. We say that Y is of **class** σ MK, if $Y = \bigcup_{n\geq 1} K_n$ with each K_n with $n \in \mathbb{N}$ large enough, being metrizable compact.

Remark 2.7.35. Recall that every metrizable compact space is the continuous image of a Cantor set; see Kuratowski [183, p. 444]. Therefore X is σ MK if and only if X is the continuous image of a closed set in $\mathbb{R}$. A separable, metrizable, locally compact space belongs to the class σ MK. But the space need not be metrizable. Again anticipating some material from Chapter 3, let X be a separable Banach space and let X^* be its topological dual. We have $X^* = \bigcup_{n\geq 1} n\overline{B}_1^*$ with $\overline{B}_1^* = \{x^* \in X^*\colon \|x^*\|_* \leq 1\}$ being the closed unit ball in X^*. We know that $\overline{B}_1^*$ equipped with the relative w*-topology is metrizable compact; see Section 3.3. So, $X^*_{\mathrm{w}^*}$, that is, X^* furnished with the w*-topology, is a σ MK-space.

The next measurable projection theorem is due to Levin [199].

Theorem 2.7.36. *If X is a Borel subset of a Polish space, that is, a Borel space, Y is a σ* MK*-space and $D \in \mathcal{B}(X \times Y) = \mathcal{B}(X) \bigotimes \mathcal{B}(Y)$ with $D_x \in P_f(Y)$ for every $x \in X$, then* $\operatorname{proj}_X D \in \mathcal{B}(X)$.

Remark 2.7.37. Note that in this case the projection of a Borel set is Borel.

Comparable Souslin topologies on a set X generate the same Borel σ-algebras.

Proposition 2.7.38. *If τ_1 and τ_2 are two comparable Souslin topologies on X, then* $\mathcal{B}(X_{\tau_1}) = \mathcal{B}(X_{\tau_2})$.

Proof. To fix things we assume that $\tau_2 \subseteq \tau_1$. Then $\mathcal{B}(X_{\tau_2}) \subseteq \mathcal{B}(X_{\tau_1})$. Let $A \in \mathcal{B}(X_{\tau_1})$. Then A is τ_1-Souslin; see Propositions 2.6.11 and 2.6.9. Hence, there exist a Polish space Y and a continuous surjection $f: Y \to (A, \tau_1(A))$; see Definition 1.1.24. Then $f: Y \to (A, \tau_2(A))$ is continuous as well and so A is τ_2-Souslin. The same argument applied to $A^c = X \setminus A$ shows that A^c is τ_2-Souslin as well. Invoking Corollary 2.6.17 we conclude that $A \in \mathcal{B}(X_{\tau_2})$. Hence $\mathcal{B}(X_{\tau_1}) \subseteq \mathcal{B}(X_{\tau_2})$ and so finally we conclude that $\mathcal{B}(X_{\tau_1}) = \mathcal{B}(X_{\tau_2})$. □

Remark 2.7.39. More generally if τ_1 and τ_2 are two Souslin topologies on X and $\tau_1 \cap \tau_2$ is Hausdorff, then $\mathcal{B}(X_{\tau_1}) = \mathcal{B}(X_{\tau_2}) = \mathcal{B}(X_{\tau_1 \cap \tau_2})$.

Proposition 2.7.40. *If (Ω, Σ) is a complete measurable space, X is a Polish space, and $F: \Omega \to 2^X \setminus \{\emptyset\}$ is graph measurable, then $F^-(D) \in \Sigma$ for all $D \in \mathcal{B}(X)$.*

Proof. Note that $F^-(D) = \operatorname{proj}_\Omega[\operatorname{Gr} F \cap (\Omega \times B)] \in \Sigma$; see Theorem 2.7.32. □

Therefore, we can state the following theorem, which summarizes the measurability properties of closed valued multifunctions.

Theorem 2.7.41. *Let (Ω, Σ) be a measurable space, (X, d) is a separable metric space and $F: \Omega \to P_f(X)$. Consider the following statements:*
(a) *$F^-(D) \in \Sigma$ for all $D \in \mathcal{B}(X)$;*
(b) *$F^-(C) \in \Sigma$ for all closed $C \subseteq X$;*
(c) *F is measurable;*
(d) *for every $x \in X$, $w \to d(x, F(w))$ is Σ-measurable;*
(e) *there exists a sequence of Σ-measurable selections $f_n: \Omega \to X$ such that $F(w) = \overline{\{f_n(w)\}}_{n \geq 1}$ for all $w \in \Omega$;*
(f) *F is graph measurable.*
We have the following implications:
(1) *(a) $\Longrightarrow$ (b) $\Longrightarrow$ (c) $\Longleftrightarrow$ (d) $\Longrightarrow$ (f).*
(2) *If X is complete, that is, X is a Polish space, then (c) $\Longleftrightarrow$ (d) $\Longleftrightarrow$ (e).*
(3) *If X is σ-compact, then (b) $\Longleftrightarrow$ (c).*
(4) *If $\Sigma = \hat{\Sigma}$, that is, the measurable space is complete, and X is complete, then (a) to (f) are all equivalent.*

2.8 Remarks

(2.1) Cantor [61] was one of the first to give a general definition of the measure of a set. However, the definition he gave produced a nonadditive measure. Then came the French mathematician Jordan [168] who defined a set to be measurable if its topological boundary has zero measure. So, the set of rational numbers in an interval is not measurable. Moreover, there are open sets that are not measurable. Finally, the measure that Jordan defined is only finitely additive. Then came Borel [39] who showed that the length of intervals can be extended to a σ-additive set function on the σ-algebra generated by intervals, the Borel σ-algebra. The Borel measure is based on the fact that any open set $U \subseteq \mathbb{R}$ is the union of countably many disjoint intervals. However, we should mention that Borel did not use the terminology of open sets. At that time mathematicians focused on closed – even more specifically on perfect – sets. The notion, together with the name of open set, was introduced by Baire [20] in his thesis. Borel did not use his theory of measure to develop a corresponding theory of integration. Borel sets are produced by infinite applications of certain set-theoretic operations and so we cannot have a good insight concerning their structure. This led to an axiomatic definition of measurable sets. An important contribution to this came from Carathéodory [62] who introduced the notion of outer measure in the sense of Definition 2.1.33. Carathéodory worked on $\mathbb{R}^N$. Moreover, Definition 2.1.36 about μ^*-measurable sets is also due to Carathéodory [62]. It is a rather strange definition, not that intuitive. It singles out as measurable those sets which split all sets in X in two parts on which μ is additive. It is not clear how Carathéodory came up with this definition. Nevertheless, it turned out to be a very fruitful one. It gives a σ-algebra – in general not the largest possible – which contains the Borel sets and on which μ is a measure. Vitali [295] was the first to establish the existence of a nonmeasurable set in $\mathbb{R}$; see Theorem 2.1.44. A detailed account of the historical development of measurable sets can be found in Chapter 4 of Hawkins [141]. Concerning the atoms of a measure (see Definition 2.1.30(b)), we mention the following result known as "Saks Lemma," see Dunford–Schwartz [94, Lemma IV.9.7, p. 308].

Lemma 2.8.1 (Saks Lemma). *If (X, Σ, μ) is a finite measure space, then for every $\varepsilon > 0$ there exists a finite partition of X into pairwise disjoint sets $\{A_k\}_{k=1}^n \subseteq \Sigma$ such that either $\mu(A_k) \leq \varepsilon$ for all $k \in \{1, \dots, k\}$ or A_k is an atom with $\mu(A_k) > \varepsilon$ for all $k \in \{1, \dots, k\}$.*

Proposition 2.1.32 is a particular case of a more general result due to Lyapunov [209] known as the "Lyapunov Convexity Theorem." The result has important applications in many applied areas such as optimal control and mathematical economics; see Hermes-LaSalle [144] and Klein-Thompson [178].

Theorem 2.8.2 (Lyapunov Convexity Theorem). *If (X, Σ) is a measurable space and $\mu_1, \dots, \mu_n : X \to \mathbb{R}$ are nonatomic measures, then the set $R = \{(\mu_k(A))_{k=1}^n : A \in \Sigma\} \subseteq \mathbb{R}^n$ is compact and convex.*

The Cantor set (see Example 2.1.46) plays an important role in foundational work and it is also a useful tool in topology.

Further details on measure theory can be found in the books of Bogachev [36, 37], Dudley [90], Folland [114], Halmos [139], Hewitt-Stromberg [145], Royden [258], and Rudin [259].

(2.2) There is no doubt that Lebesgue's theory of integration is one of the major mathematical breakthroughs in the 20 $\underline{\text{th}}$-century. Lebesgue was influenced by the ideas of Borel, but his theory of measure is more general. His theory was first presented in his thesis [189]. Many of the questions left open in his thesis were resolved in his book [190] published two years later. It was based on lectures he gave to the College de France in the period 1902–1903. With his integral, Lebesgue was able to overcome a number of difficulties that were associated with Riemann's theory of integration. In particular the limit theorems for the new integral are substantially more general and helped in the dissemination of Lebesgue's theory. Proposition 2.2.12 goes back to Hausdorff [140] while the example produced in Remark 2.2.13 is due to Dudley [89]; see also Dudley [90, Proposition 4.2.3, p. 96]. Theorem 2.2.32 was proven by Egorov [98]. Egorov was the mathematical mentor of Lusin. We mention that Egorov's Theorem as well as Lusin's Theorem (see Theorem 2.5.17) were stated without proof in Lebesgue [190]. Theorem 2.2.34 is due to von Alexits [299] and Sierpinski [271]. The use of simple functions in the definition of the Lebesgue integral (see Definition 2.2.35) underlines the main difference with Riemann's method. More precisely, in contrast to Riemann, Lebesgue does not consider partitions of the domain $[a, b]$ of f. Instead he considers partitions of the range of f. A detailed discussion of the development of Lebesgue's method can be found in Hawkins [141].

We conclude our remarks on this subsection with two useful observations. The first concerns Egorov's Theorem (see Theorem 2.2.32) and indicates when we can drop the hypothesis that $\mu(X) < \infty$.

Proposition 2.8.3. *If (X, Σ, μ) is a measure space, $f_n\colon X \to \mathbb{R}$ with $n \in \mathbb{N}$ is a sequence of Σ-measurable functions such that*

$$f_n \to f \quad \mu\text{-a.e.} \qquad \text{and} \qquad |f_n(x)| \le h(x) \quad \mu\text{-a.e. with } h \in L^1(X)\,,$$

then given $\varepsilon > 0$ there exists $A_\varepsilon \in \Sigma$ with $\mu(A_\varepsilon) < \varepsilon$ such that $f_n \to f$ uniformly in $X \setminus A_\varepsilon$.

The second observation shows how the Lebesgue measure changes under nonsingular linear transformations.

Proposition 2.8.4. *If $L\colon \mathbb{R}^N \to \mathbb{R}^N$ is linear and nonsingular, then the following hold:*
(a) $L(A) \in \mathcal{B}(\mathbb{R}^N)$ *for all* $A \in \mathcal{B}(\mathbb{R}^N)$;
(b) $\lambda^N(L(A)) = |\det(L)|\lambda^N(A)$ *for all* $A \in \mathcal{B}(\mathbb{R}^N)$.

(2.3) Theorem 2.3.1 – and consequently Theorems 2.3.3 as well as 2.3.5 – are due to Beppo Levi [197]. Theorem 2.3.6 is due to Fatou [108]. Theorem 2.3.8 is the "crown jewel" of Lebesgue's theory and was proved by Lebesgue [192]. The L^p-spaces were defined by Riesz [243] when $p = 2$, [244] when $1 < p < 2$ and [245] when $2 < p < \infty$. Riesz [244, 245] proved the completeness of L^p, $p \neq 2$ while the completeness of L^2 was proved by Fischer [111]. The Cauchy–Bunyakowsky-Schwarz inequality (see Corollary 2.3.13) was first proven by Cauchy (1821) for finite sums, then by Bunyakowsky (1859) for Riemann integrals and finally by Schwarz (1885) for double integrals. Hölder's inequality (see Theorem 2.3.12) can be found in Rogers [254] and Hölder [154]. Of course the inequalities proven by Rogers and Hölder do not have the form of Theorem 2.3.12, but it can be shown that they imply Theorem 2.3.12. Note that Hölder acknowledges that he was inspired by the work of Rogers. For this reason Dudley [90] calls the result "Rogers-Hölder inequality." Theorem 2.3.14 was proven by Minkowski [217] for finite sums and by Riesz [245] for integrals. Jensen's inequality (see Theorem 2.3.15) was obtained by Jensen [166]. Convergence in measure, initially called also **asymptotic convergence**, can be found in early works of Borel and Lebesgue but a systematic study of it can be found in Riesz [244], who pointed out a gap in the book of Lebesgue concerning this mode of convergence and in Fréchet [119, 120]. In fact Fréchet [119] showed that convergence in measure is metrizable by the metric

$$d_F(f,h) = \inf_{\varepsilon>0}[\varepsilon + \mu\{x \in X\colon |f(x) - h(x)| > \varepsilon\}] .$$

Another metric was introduced by Fan [107] who defined

$$d_K(f,h) = \inf[\varepsilon \geq 0\colon \mu\{x \in X\colon |f(x) - h(x)| > \varepsilon\} < \varepsilon] .$$

The metric in (2.3.10) was first introduced by Nikodym [230].

The notion of uniform integrability and the main results concerning it go back to the works of Lebesgue, Vitali, and de la Vallee Poussin. Additional equivalent formulations of this notion can be found in Gasiński-Papageorgiou [125, see Problems 1.7, 1.15, 1.16, 1.17].

Lebesgue [190] was the first to establish for bounded measurable functions of two variables the reduction of multiple integrals to repeated ones. Later Fubini [122] proved Theorem 2.3.50 and the appearance of his result marked a real triumph for Lebesgue's method. As Fubini pointed out, the Lebesgue integral is necessary for this kind of study. Theorem 2.3.49 is due to Tonelli [286].

We conclude the remarks of this subsection with a result on the existence of the essential supremum for a family of functions. The result is useful in probability theory and elliptic partial differential equations.

Proposition 2.8.5. *If (X, Σ, μ) is a σ-finite measure space and $\mathcal{F}$ is a family of Σ-measurable, $\mathbb{R}$-valued functions, then there exists a unique (up to μ-a.e. equality) Σ-measurable function $h\colon X \to \mathbb{R}$ such that $f(x) \leq h(x)$ for μ-a.a. $x \in X$ and for all $f \in \mathcal{F}$.*

If h' is another Σ-measurable function such that $f(x) \leq h'(x)$ for μ-a.a. $x \in X$ and for all $f \in \mathcal{F}$, then $h(x) \leq h'(x)$ for μ-a.a. $x \in X$.

We call $h = \operatorname{ess\,sup} \mathcal{F}$. In addition there is a sequence $\{f_n\}_{n\geq 1} \subseteq \mathcal{F}$ such that $\operatorname{ess\,sup} \mathcal{F} = \sup_{n\geq 1} f_n$. Finally if $\mathcal{F}$ is upward directed, that is, if $f_1, f_2 \in \mathcal{F}$, then there exists $f \in \mathcal{F}$ such that $f_1 \leq f, f_2 \leq f$, then $\{f_n\}_{n\geq 1}$ can be chosen to be increasing.

(2.4) Signed measures were first considered by Lebesgue [192] who studied such measures of the form

$$\mu(A) = \int_A f(x)d\nu(x) \quad \text{with } f \in L^1(\nu) .$$

The Hahn Decomposition Theorem (see Theorem 2.4.8) was proven by Hahn [136]. Concerning the Jordan Decomposition Theorem (see Theorem 2.4.14), we mention that

Jordan (1881) introduced functions of bounded variation on an interval $[a, b]$ and proved that such a function can be written as the difference of two nondecreasing functions; see also Section 4.3.

The more general Theorem 2.4.14 was named after Jordan as a tribute of his important contributions on the subject. Note that if μ is a finite signed measure on $[a, b]$, then $f(x) = \mu([a, x])$ with $x \in [a, b]$ is a function of bounded variation and $f = g - h$ with $g(x) = \mu_+([a, x])$ and $h(x) = \mu_-([a, x])$ for all $x \in [a, b]$.

The Radon–Nikodym Theorem (see Theorem 2.4.29) started with Lebesgue who obtained the special case of absolute continuity with respect to the Lebesgue measure. The case of Borel measures on $\mathbb{R}^N$ was proven by Radon [238] and a little later by Daniell [72] as well. The general form of the theorem is due to Nikodym [230]. The Lebesgue decomposition in the general abstract setting (see Theorem 2.4.31) can be found in Saks [262]. There is a unifying short proof of Theorems 2.4.29 and 2.4.33 due to von Neumann [304]; see also Dudley [90, p. 134] and Rudin [259, p. 130]. Although Theorem 2.4.33 is called the Vitali–Hahn–Saks Theorem, others also contributed to its formulation, like Lebesgue and Nikodym. It appears the general form was proven by Saks [262]. Theorems 2.4.33 and 2.4.34 are very useful in general measure theory.

(2.5) The definition of the Baire σ-algebra (see Definition 2.5.1) is not the same in all authors. For example, Dudley [90, p. 174] defines the Baire σ-algebra to be the smallest σ-algebra for which all $f \in C_b(X)$ are measurable. Recall that $C_b(X)$ is the space of all $\mathbb{R}$-valued, continuous, and bounded functions. Other definitions of $\mathrm{Ba}(X)$ are provided by Bogachev [37, p. 12] and Halmos [139, p. 220]. Here we follow Royden [258, p. 301]. We should point out that for the Borel σ-algebra, there are some different definitions. More precisely, some of the older texts define the Borel σ-algebra to be the σ-algebra generated by the compact sets. This in in general smaller than the Borel σ-algebra of Definition 2.1.4(b).

Similarly the terminology introduced in Definition 2.5.8 is not uniform. People use other names for the same notions, see, for example Aliprantis–Border [6, pp. 434–435]. Topological measure theory started with the seminal paper of Radon [238] who worked on $\mathbb{R}^N$. A classical reference on Radon measures is the book of Schwartz [268].

The topological structure of the ambient space leads to the definition of the support of a measure.

Definition 2.8.6. Let X be a Hausdorff topological space and $\mu\colon \mathcal{B}(X) \to [0, \infty]$ a Borel measure. The **support of** μ is the set

$$\operatorname{supp}\mu = \{x \in X\colon \mu(U) > 0 \text{ for all } U \in \mathcal{N}(x)\}\,.$$

Remark 2.8.7. Evidently $\operatorname{supp}\mu$ is closed and if $A \in \mathcal{B}(X)$, $A \subseteq X \setminus \operatorname{supp}\mu$, then $\mu(A) = 0$. Every Radon measure has a unique support.

We have a regularity result for functions that are integrable with respect to a Radon measure. The result is known as the “Vitali–Carathéodory Theorem;” see Rudin [259, p. 57].

Theorem 2.8.8 (Vitali–Carathéodory Theorem). *If X is a locally compact topological space, $\mu\colon \mathcal{B}(X) \to [0,\infty]$ is a Radon measure, $f \in L^1(X,\mu)$ and $\varepsilon > 0$, then there exist $g\colon X \to \mathbb{R}$ being upper semicontinuous, bounded above and $h\colon X \to \mathbb{R}$ being lower semicontinuous, bounded below such that $g(x) \leq f(x) \leq h(x)$ for μ-a.a. $x \in X$ and $\int_X (h-g)d\mu \leq \varepsilon$.*

Remark 2.8.9. There is an alternative approach to Lebesgue integration due to Daniell [71] based on the extension of positive linear functionals. Within that theory, the Vitali–Carathéodory Theorem is essentially the definition of the measurability and integrability of f.

As was the case with Egorov’s Theorem (see Theorem 2.2.32), Lusin’s Theorem (see Theorem 2.5.17) was first stated without proof by Lebesgue [190]. Lusin [206] proved the result later. There is a category analog to Lusin’s Theorem.

Theorem 2.8.10. *If X is a separable metric space and $f\colon X \to \mathbb{R}$ is Borel measurable, then there is a set D of first category such that $f|_{X\setminus D}$ is continuous.*

Theorem 2.5.19 is due to Scorza Dragoni [269]. Normal integrands (see Definition 2.5.20) is a basic tool in many applied fields such as calculus of variations, optimization and optimal control; see Buttazzo [60], Ekeland–Temam [103], and Papageorgiou–Kyritsi [232].

Finally we mention an important class of measures that allows us to measure the size of lower dimensional sets in $\mathbb{R}^N$, for example, curves and surfaces in $\mathbb{R}^3$. So, let (X,d) be a metric space, $p \geq 0$, $\delta > 0$, and $A \subseteq X$. We set

$$H_{p,\delta}(A) = \inf\left(\sum_{k\geq 1} (\operatorname{diam} B_k)^p \colon A \subseteq \bigcup_{k\geq 1} B_k, \operatorname{diam} B_k \leq \delta\right). \tag{2.8.1}$$

As usual we set $\inf \emptyset := +\infty$. $H_{p,\delta}(A)$ increases as $\delta \to 0^+$. So, the following definition makes sense.

Definition 2.8.11. For every $A \in \mathcal{B}(X)$, the limit

$$\lim_{\delta\to 0^+} H_{p,\delta}(A) = H_p(A)$$

is the p-**dimensional Hausdorff measure** of A. The measure $H_p\colon \mathcal{B}(X) \to [0,\infty]$ is regular.

Remark 2.8.12. Note that in (2.8.1) there is no loss of generality if B_k is closed or open for all $k \in \mathbb{N}$.

For more on Hausdorff measures we refer to Evans–Gariepy [105].

(2.6) The theory of Souslin or analytic or A-sets started when Souslin, a student of Lusin, discovered an error in Lebesgue [191]. Lebesgue claimed that the projection of a Borel set in $\mathbb{R}^2$ onto the x-axis is again a Borel set. Souslin realized that this is not true and went on to introduce analytic sets and started their study. Souslin [275] also produced an analytic set in the real line whose complement is not analytic and so it is not Borel; see Proposition 2.6.11 and Remark 2.6.12. Lusin [207] proved that analytic sets in $\mathbb{R}$ are Lebesgue measurable. Unfortunately, Souslin died very young at the age of 25 in 1919. The work on analytic sets was continued initially by Lusin and subsequently by many other mathematicians. Theorem 2.6.15 is due to Lusin [208] and is one of the most important results in the theory of analytic sets with far-reaching consequences. In addition to the σ-algebras $\mathcal{B}(X)$, α_X, $\hat{\Sigma}_X$ there is a fourth σ-algebra known as the **limit σ-algebra** denoted by $\mathcal{L}_X$ and it is between α_X and $\hat{\Sigma}_X$. For a discussion of this σ-algebra see Bertsekas–Shreve [31, Appendix B4]. Analytic (Souslin) sets are discussed in the books of Aliprantis–Border [6], Bertsekas–Shreve [31], Bogachev [37], Cohn [69], Dudley [90], Klein–Thompson [178], and Srivastava [276].

(2.7) Measurable multifunctions are an important tool in many applied areas. Detailed studies of measurable multifunctions can be found in the books of Aliprantis-Border [6], Aubin-Frankowska [17], Castaing-Valadier [64], Denkowski–Migórski-Papageorgiou [77], Hu-Papageorgiou [157], and Klein-Thompson [178]. Theorem 2.7.12 was proven by Rohlin [255] and later by Kuratowski-Ryll Nardzewski [185]. There is a gap in the proof of Rohlin and for this reason the result is attributed to Kuratowski-Ryll Nardzewski. Theorem 2.7.25 as stated is due to Sainte-Beuve [261]. Earlier versions of it were proven by Yankov [310], von Neumann [305] and Aumann [18]. The same can be said for Theorem 2.7.32.

Problems

Problem 2.1. Let X be a set and let $\mathcal{L} \subseteq 2^X$ be nonempty. Show that $\sigma(\mathcal{L})$ is the smallest family $\mathfrak{L} \subseteq 2^X$, which contains $\mathcal{L}$ and satisfies the following assertions:
(a) $A \in \mathcal{L}$ implies $A^c \in \mathfrak{L}$;
(b) $\mathfrak{L}$ is closed under countable intersections;
(c) $\mathfrak{L}$ is closed under countable disjoint unions.

Problem 2.2. Let X be a set and let $\mathcal{L} \subseteq 2^X$ be a semiring. Show that:
(a) If $A, A_1, \ldots, A_n \in \mathcal{L}$, then there exist $\{B_i\}_{i=1}^m \subseteq \mathcal{L}$ pairwise disjoint such that $A \setminus \bigcup_{k=1}^n A_k = \bigcup_{i=1}^m B_i$.
(b) If $\{A_n\}_{n\geq 1} \subseteq \mathcal{L}$, then there exist $\{C_k\}_{k\geq 1} \subseteq \mathcal{L}$ pairwise disjoint such that $\bigcup_{n\geq 1} A_n = \bigcup_{k\geq 1} C_k$ and for each $k \geq 1$ there exists $n \geq 1$ such that $C_k \subseteq A_n$.

Problem 2.3. Let (X, Σ, μ) be a finite measure space and $\{A_i\}_{i\in I} \subseteq \Sigma$ are pairwise disjoint with an arbitrary index set I. Show that $\mu(A_i) = 0$ for all $i \in T \setminus I_0$ with I_0 is at most countable.

Problem 2.4. Let (X, Σ, μ) be a finite nonatomic measure space and let $\{\eta_n\}_{n\geq 1} \subseteq (0, +\infty)$ be such that $\sum_{n\geq 1} \eta_n \leq \mu(X)$. Show that there is $\{A_n\}_{n\geq 1} \subseteq \Sigma$ pairwise disjoint such that $\mu(A_n) = \eta_n$ for all $n \in \mathbb{N}$.

Problem 2.5. Let (X, Σ, μ) be a measure space with μ being semifinite (see Definition 2.1.30) and $A \in \Sigma$, $\mu(A) = +\infty$. Show that there exists $C \in \Sigma$, $C \subseteq A$ with $\mu(C) = +\infty$ and that C is σ-finite.

Problem 2.6. Let (X, Σ, μ) be a measure space. Show that μ is semifinite (see Definition 2.1.30) if and only if for all $A \in \Sigma$ with $\mu(A) > 0$ there holds

$$\mu(A) = \sup[\mu(C) \colon C \in \Sigma, C \subseteq A, 0 < \mu(C) < \infty] .$$

Problem 2.7. Let X be a σ-compact metric space, $\mathcal{B}(X)$ is the Borel σ-algebra of X, and μ_1, μ_2 are two finite measures on $\mathcal{B}(X)$, which are equal on compact sets. Show that $\mu_1 = \mu_2$.

Problem 2.8. Let (X, Σ, μ) be a measure space and μ^* the outer measure defined in (2.1.7) with $\mathcal{L} = \Sigma$ and $\vartheta = \mu$. Show that:
(a) $\mu^*(A) = \inf[\mu(B) \colon B \in \Sigma, A \subseteq B]$ for every $A \subseteq X$.
(b) For every $A \subseteq X$ there exists $B \in \Sigma_{\mu^*}$ such that $A \subseteq B$ and $\mu^*(A) = \mu(B)$.

Problem 2.9. Let (Ω, Σ, μ) be a measure space, $\{A_n\}_{n\geq 1} \subseteq \Sigma$ with $\sum_{n\geq 1} \mu(A_n) < \infty$, and $\liminf_{n\to\infty} \mu(A_n) \geq \vartheta \geq 0$. Let D_∞ be the set of elements in Ω that belong to an infinity of sets A_n. Show that $D_\infty \in \Sigma$ and $\mu(D_\infty) \geq \vartheta$.

Problem 2.10. Let X be a nonempty set, $\mathcal{L} \subseteq 2^X$ is an algebra, and $\mu \colon \mathcal{L} \to [0, \infty]$ is an additive set function. Let μ^* be the outer measure defined in (2.1.7) with $\mathcal{L} = \Sigma$ and $\vartheta = \mu$. Show that every element in $\mathcal{L}$ is μ^*-measurable; see Definition 2.1.36. Moreover, show that if μ is σ-additive, then $\mu^*|_{\mathcal{L}} = \mu$.

Problem 2.11. Let $\mathcal{L}$ be a σ-algebra of sets in $\mathbb{R}$. Show that $\mathcal{B}(\mathbb{R}) \subseteq \mathcal{L}$ if and only if any continuous function $f \colon \mathbb{R} \to \mathbb{R}$ is $\mathcal{L}$-measurable.

Problem 2.12. Let (X, Σ, μ) be a measure space, $f \colon X \to [0, \infty]$ a Borel function, and let $d_f(t) = \mu(\{x \in X \colon f(x) > t\})$. Show that:
(a) d_f is right continuous.
(b) If $\mu(X) < \infty$, then for every $t_0 > 0$ it holds that $\lim_{t\to t_0^-} d_f(t) = \mu(\{x \in X \colon f(x) \geq t_0\})$.

Problem 2.13. Given $\varepsilon > 0$, produce a dense open set $U \subseteq \mathbb{R}$ such that $\lambda(U) \leq \varepsilon$, where λ is the Lebesgue measure on $\mathbb{R}$.

Problem 2.14. Suppose that $1 \leq p < \infty$ and let $f \in L^p(\mathbb{R}^N)$ for the Lebesgue measure on $\mathbb{R}^N$. Show that

$$\lim_{h\to 0} \int\limits_{\mathbb{R}^N} |f(x+h) - f(x)| d\lambda = 0 .$$

Problem 2.15. (a) Suppose that $f: \mathbb{R}^N \to \mathbb{R}$ is integrable and $K \subseteq \mathbb{R}^N$ is nonempty and compact. Show that $\lim_{|y|\to\infty} \int_{K+y} |f(x)|dx = 0$.

(b) Suppose that $f: \mathbb{R}^N \to \mathbb{R}$ is uniformly continuous and $f \in L^p(\mathbb{R}^N)$ for some $1 \le p < \infty$. Show that $\lim_{|x|\to\infty} f(x) = 0$.

Problem 2.16. Let X be a nonempty set, Y is a metrizable space and $f: X \to Y$ is a map that is the pointwise limit of simple functions. Show that $f(X) \subseteq Y$ is separable.

Problem 2.17. Let (X, Σ) be a measurable space, Y a second countable Hausdorff topological space, and $f: X \to Y$ a Σ-measurable multifunction. Show that $\operatorname{Gr} f \in \Sigma \bigotimes \mathcal{B}(Y)$.

Problem 2.18. Let (Ω, Σ, μ) be a measure space and $\mathcal{L} \subseteq \Sigma$ a countable subset such that if $A \in \Sigma$, $\mu(A) < \infty$, then there exists $B \in \mathcal{L}$ with $\mu(A \triangle B) \le \varepsilon$. Show that $L^p(\Omega)$ is separable for all $1 \le p < \infty$.

Problem 2.19. Let (Ω, Σ, μ) be a σ-finite measure space and assume that $f \in L^p(\Omega)$ for all $p \ge p_0 \ge 1$. Show that $\lim_{p\to+\infty} \|f\|_p = \|f\|_\infty$.

Problem 2.20. Let (X, Σ), $(Y, \mathcal{L})$, and $(V, \mathcal{D})$ be measurable spaces, $f: X \to Y$, $g: X \to V$, and let $h: X \to Y \times V$ be defined by $h(x) = (f(x), g(x))$ for all $x \in X$. Show that h is $(\Sigma, \mathcal{L} \bigotimes \mathcal{D})$-measurable if and only if f is $(\Sigma, \mathcal{L})$-measurable and g is $(\Sigma, \mathcal{D})$-measurable.

Problem 2.21. Let (X, Σ) be a measurable space, Y, Y_1, Y_2 separable metrizable spaces, and V a Hausdorff topological space. Suppose that

$$f_k: X \times Y \to Y_k,\ k = 1, 2 \text{ are Carathéodory functions },$$
$$g: Y_1 \times Y_2 \to V \text{ is Borel measurable }.$$

Show that $h: X \times Y \to V$ defined by $h(x, y) = g(f_1(x, y), f_2(x, y))$ is $\Sigma \bigotimes \mathcal{B}(X)$-measurable.

Problem 2.22. Let $E \subseteq \mathbb{R}$ be Lebesgue measurable with $\lambda(E) > 0$. Show that there exists a nonmeasurable subset of E.

Problem 2.23. Let (X, Σ, μ) be a finite measure space and $f_{nm}: X \to \mathbb{R}$ with $n, m \in \mathbb{N}$ a family of Σ-measurable functions such that

$$f_{nm}(x) \to f_n(x) \quad \mu\text{-a.e. as } m \to \infty \quad \text{and} \quad f_n(x) \to f(x) \quad \mu\text{-a.e. as } n \to \infty .$$

Show that there exists an increasing sequence $m_n \in \mathbb{N}$ with $n \ge 1$ such that

$$f_{nm_n}(x) \to f(x) \quad \mu\text{-a.e. as } n \to \infty .$$

Problem 2.24. Let X be a compact metrizable space and Y be a separable metrizable space, and consider the function space $C(X, Y)$ with the τ_u-topology; see Remark 1.6.17. Let

$$\mathcal{L} = \left\{e_x^{-1}(C), C \subseteq Y \text{ is closed}\right\} ;$$

see Definition 1.6.7. Show that $\mathcal{B}(C(X, Y)) = \sigma(\mathcal{L})$.

Problem 2.25. Let (X, Σ) be a measurable space, V a compact metrizable space, Y a separable metrizable space, and consider the function space $C(V, Y)$ endowed with the τ_u-topology; see Remark 1.6.17.

(a) Given a Carathéodory function $f: X \times V \to Y$, show that $\hat{f}: X \to C(V, Y)$ defined by $\hat{f}(x)(\cdot) = f(x, \cdot)$ is Σ-measurable.

(b) If $h: X \to C(V, Y)$ is Σ-measurable, show that $\tilde{h}: X \times V \to Y$ defined by $\tilde{h}(x, \cdot) = h(x)(\cdot)$ is a Carathéodory function.

Problem 2.26. Let (X, Σ, μ) be a measure space and $f: X \to \mathbb{R}$ is a μ-integrable function. Show that the set $C = \{x \in X : f(x) \neq 0\}$ has σ-finite μ-measure.

Problem 2.27. Suppose that X and Y are Hausdorff topological spaces such that

$$D(Y) = \{(y, v) \in Y \times Y : y = v\} \in \mathcal{B}(Y) \bigotimes \mathcal{B}(Y) .$$

Show that the graph of any Borel function $f: X \to Y$ belongs to $\mathcal{B}(X) \otimes \mathcal{B}(Y)$.

Problem 2.28. Let (X, Σ, μ) be a finite measure space. Show that there exists an at most countable family $\{A_n\}_{n\geq 1} \subseteq \Sigma$ of atoms such that $X \setminus \bigcup_{n\geq 1} A_n$ is nonatomic.

Problem 2.29. Let (X, Σ, μ) be a measure space with μ being semifinite (see Definition 2.1.30(a)), and let $f, g: X \to [0, +\infty]$ be two Σ-measurable functions such that

$$\int_A f d\mu \leq \int_A g d\mu \quad \text{for all } A \in \Sigma \text{ with } \mu(A) < \infty .$$

Show that $f(x) \leq g(x)$ for μ-a.a. $x \in X$.

Problem 2.30. Let $A \subseteq \mathbb{R}$ be a set of finite Lebesgue measure and let $f: \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = \lambda(A \cap (-\infty, x])$ for all $x \in \mathbb{R}$. Here λ denotes the Lebesgue measure on $\mathbb{R}$. Show that f is continuous.

Problem 2.31. Let $A \subseteq \mathbb{R}$ be a Lebesgue measurable set with $\lambda(A) > 0$ with λ being the Lebesgue measure on $\mathbb{R}$. Show that $A - A$ contains an open set.

Problem 2.32. Let (X, Σ, μ) be a measure space and $f: X \to [0, \infty]$ is a Σ-measurable function. Show that $\int_X f d\mu = \int_0^\infty \mu(\{x \in X : f(x) > s\}) ds$.

Problem 2.33. Let (X, Σ, μ), $(Y, \mathcal{L}, \nu)$ be two σ-finite measure spaces. Show that $(X \times Y, \Sigma \bigotimes \mathcal{L}, \mu \times \nu)$ is σ-finite as well.

Problem 2.34. Let (X, Σ, μ) be a measure space, $f_n, f: X \to [0, +\infty)$ with $n \geq 1$ are Σ-measurable functions and suppose that $f_n \xrightarrow{\mu} f$. Show that for every $\vartheta > 0$, $f_n^\vartheta \xrightarrow{\mu} f^\vartheta$.

Problem 2.35. Let (X, Σ, μ) be a nonatomic measure space and $f: X \to [0, \infty]$ is a Σ-measurable function. Show that the measure $\Sigma \ni A \to \xi(A) = \int_A f d\mu$ is nonatomic if and only if $\mu(\{x \in X : f(x) = +\infty\}) = 0$.

Problem 2.36. Let X be a Hausdorff topological space, $\mu\colon \mathcal{B}(X) \to [0, +\infty)$ be a finite Borel measure, and $f\colon X \to \mathbb{R}$ be a continuous function. Show that there exists an at most countable set $D \subseteq \mathbb{R}$ such that $\mu(\{x \in X\colon f(x) = \eta\}) > 0$ for all $\eta \in D$.

Problem 2.37. Let X, Y be two metric spaces and $f\colon X \to Y$. Let $C_f = \{x \in X\colon f$ is continuous$\}$. Show that $C_f \in \mathcal{B}(X)$.

Problem 2.38. Does the Lebesgue Dominated Convergence Theorem (see Theorem 2.3.8) hold for nets? Justify your answer.

Problem 2.39. Let X be a Polish space and $A \subseteq X$. Show that A is analytic if and only if $A = \operatorname{proj}_X B$ with $B \in \mathcal{B}(X \times X) = \mathcal{B}(X) \bigotimes \mathcal{B}(X)$.

Problem 2.40. Let (X, Σ) be a measurable space and Y a metric space. Show that $f\colon X \to Y$ is Σ-measurable if and only if for all continuous $\varphi\colon Y \to \mathbb{R}$ we have that $\varphi \circ f$ is Σ-measurable.

Problem 2.41. Let (Ω, Σ) be a measurable space, X a separable metrizable space, Y a Hausdorff topological space, $f\colon \Omega \times X \to Y$ a Carathéodory map, and $U \subseteq Y$ be open. Show that the multifunction $w \to G(w) = \{x \in X\colon f(w, x) \in U\}$ is measurable.

Problem 2.42. Let (Ω, Σ) be a measurable space, X is a Polish space and $F_n\colon \Omega \to P_f(X)$ with $n \in \mathbb{N}$ are measurable multifunctions such that for every $w \in \Omega$, there exists $n \in \mathbb{N}$ such that $F_n(w) \in P_k(X)$. Show that $w \to \bigcap_{n\geq 1} F_n(w)$ is measurable.

Problem 2.43. Let $\{X_n\}_{n\geq 1}$ be a sequence of Polish spaces and for each $n \in \mathbb{N}$, $A_n \subseteq X_n$ is analytic. Show that $\prod_{n\geq 1} A_n$ is an analytic subset of $\prod_{n\geq 1} X_n$.

Problem 2.44. Let X, Y be a Polish spaces, $A \in \mathcal{B}(X)$, $f\colon A \to Y$ is a Borel measurable map, and $E = f(A)$. Assume that f is injective and $B \in \mathcal{B}(Y)$. Show that f^{-1} is Borel measurable.

Problem 2.45. Let X, Y be Polish spaces and $f\colon X \to Y$ be Borel measurable.
(a) Show that if $A \subseteq X$ is analytic, then $f(A) \subseteq Y$ is analytic.
(b) Show that if $B \subseteq Y$ is analytic, then $f^{-1}(B) \subseteq X$ is analytic.

Problem 2.46. Let X, Y be Hausdorff topological spaces and $f\colon X \to Y$ be a map that has a graph that is a Souslin subset of $X \times Y$. Show that f is Borel measurable.

Problem 2.47. Let (X, Σ, μ) be a finite measure space, $K \subseteq L^1(X)$ be uniformly integrable, and K^* be the sequential closure for the μ-almost everywhere convergence in K. Show that K^* is uniformly integrable as well.

Problem 2.48. Let (X, Σ, μ) be a measure space and $C \subseteq L^1(X)$ a uniformly integrable set. Show that for given $\varepsilon > 0$ there exist $\xi_\varepsilon \in L^1(X)_+$ and $\delta > 0$ such that $A \in \Sigma$, $\int_A \xi_\varepsilon d\mu \leq \delta$ implies $\sup_{f\in C} \int_A |f| d\mu \leq \varepsilon$.

Problem 2.49. Let (X, Σ, μ) be a measure space and $C \subseteq L^1(X)$ a uniformly integrable set. Show that for given $\varepsilon > 0$ there is $\xi_\varepsilon \in L^1(X)_+$ such that $\sup_{f \in C} \int_{\{|f| \geq \xi_\varepsilon\}} |f| d\mu \leq \varepsilon$.

Problem 2.50. Let (X, Σ, μ) be a measure space and $C \subseteq L^1(X)$. Assume that for every $\varepsilon > 0$ we can find $\xi_\varepsilon \in L^1(\Omega)_+$ such that

$$\sup_{f \in C} \int\limits_{\{|f| \geq \xi_\varepsilon\}} |f| d\mu \leq \varepsilon \,.$$

Show that C is uniformly integrable.

Problem 2.51. Let (X, Σ, μ) be a measure space and $C \subseteq L^1(X)$ be a bounded set, and suppose that for every $\varepsilon > 0$ we can find $\xi_\varepsilon \in L^1(\Omega)_+$ and $\delta > 0$ such that $A \in \Sigma$, $\int_A h_\varepsilon d\mu \leq \delta$ implies that $\sup_{f \in C} \int_A |f| d\mu \leq \varepsilon$. Show that C is uniformly integrable.

Problem 2.52. Let (Ω, Σ) be a measurable space, X a separable metrizable space, $f\colon \Omega \times X \to \mathbb{R}$ a Carathéodory function, and $F\colon \Omega \to P_k(X)$ a measurable multifunction. Let $m(w) = \min[f(w, x)\colon x \in F(w)]$ and $M(w) = \{x \in F(w)\colon m(w) = f(w, x)\}$. Show that m and M are both measurable.

Problem 2.53. Let (X, Σ) be a measurable space and μ, ν be finite measures on (X, Σ). Show that either $\mu \perp \nu$ or that there exist $\varepsilon > 0$ and $B \in \Sigma$ with $\mu(B) > 0$ and $\nu \geq \varepsilon\mu$ on B, that is, B is a positive set for $\nu - \varepsilon\mu$.