MATH 50006

Measure and Integration

Lecture Notes

Pierre-François Rodriguez

Imperial College London

this version: May 13, 2024

Contents

Chapter 0

Introduction

The purpose of this course is to introduce and study the concept of a measure, which is a central notion in contemporary mathematics. We start by 'putting the needle in the wound' and highlight three seemingly unrelated, but serious issues, which will all be duly addressed by means of measure theory.

i) Limitations of the Riemann-integral. Recall the following:

Definition 0.1. A function $f : [a,b] \to \mathbb{R}$ is called **Riemann-integrable**, if for any partition $P = \{a = t_0 < t_1 < \cdots < t_{n-1} < t_n = b\}$, $n \geq 1$, of $[a,b]$, defining

$$L(f,P) = \sum_{i=1}^{n} (t_i - t_{i-1}) \inf_{t \in [t_{i-1}, t_i]} f(t) \tag{0.1}$$

and similarly

$$U(f,P) = \sum_{i=1}^{n} (t_i - t_{i-1}) \sup_{t \in [t_{i-1}, t_i]} f(t), \tag{0.2}$$

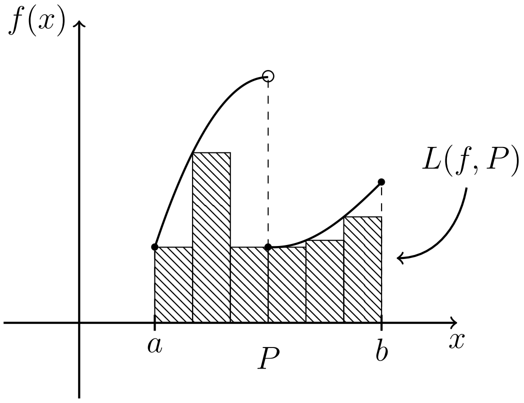
one has

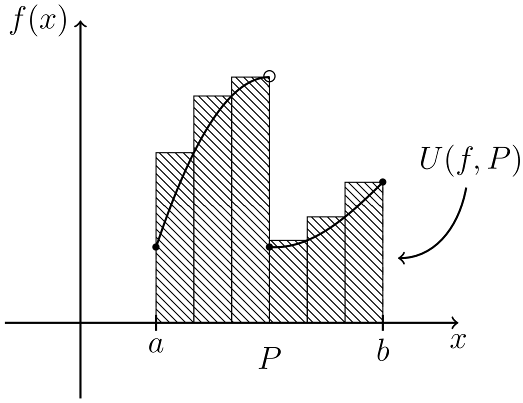
$$\sup_P L(f,P) = \inf_P U(f,P) \tag{0.3}$$

in which case the value in (0.3) is denoted by $\int_a^b f(x)\mathrm{d}x$ and called the Riemann(-Darboux) integral of f over $[a,b]$.

Remark 0.2. 1) One easily shows that all functions in $C^0_{pw}[a,b]$, the space of piecewise continuous functions on $[a,b]$ (i.e. the set of functions $f : [a,b] \to \mathbb{R}$ having at most finitely many points of discontinuity) is Riemann-integrable (exercise).

2) Intuitively, $\int_a^b f(x)\,\mathrm{d}x$, if existing, corresponds to a (signed) area under the graph of the function f.





Unfortunately, the class of Riemann-integrable functions is rather limited. Consider for instance the Dirichlet-function

$$1_{\mathbb{Q}}(x) \overset{\text{def.}}{=} \begin{cases} 1 & \text{if } x \in \mathbb{Q}, \\ 0 & \text{if } x \in \mathbb{R}\backslash\mathbb{Q} \end{cases} . \tag{0.4}$$

Then $1_{\mathbb{Q}}$ is <u>not</u> Riemann-integrable. This follows readily upon recalling that $\mathbb{Q}$ and $\mathbb{R}\backslash\mathbb{Q}$ are both dense in $\mathbb{R}$, whence, in view of (0.1) and (0.2), for any $a < b$ and any partition P of $[a,b]$, one obtains that $L(1_{\mathbb{Q}}, P) = 0$, $U(1_{\mathbb{Q}}, P) = 1$.

Note that, although $1_{\mathbb{Q}} \notin C^0_{pw}[a,b]$, there are 'many more' points in $\mathbb{R}\backslash\mathbb{Q}$ than in $\mathbb{Q}$ (uncountably vs. countably many) and one may expect that a more satisfactory theory of integration would assign $\int_a^b 1_{\mathbb{Q}}(x)\,\mathrm{d}x = 0$, as the function $1_{\mathbb{Q}}$ vanishes 'almost everywhere.' We will make this notion precise later on in the course. Moreover letting $\{q_0, q_1, q_2, \dots\}$ denote an arbitrary enumeration of $\mathbb{Q} \cap [a,b]$, and defining

$$f_n(x) = \begin{cases} 1 & \text{if } x \in \{q_0, q_1, \dots, q_n\}, \\ 0 & \text{else.} \end{cases} , \quad n \geq 0, \quad \text{and} \quad f = 1_{\mathbb{Q}} \tag{0.5}$$

one sees immediately by (0.4) that $f_n(x) \overset{n\to\infty}{\longrightarrow} f(x)$ for every $x \in \mathbb{R}$, i.e. $f_n \to f$ pointwise, and $f_n \in C^0_{pw}[a,b]$ for all $n \geq 0$. Hence f_n is Riemann-integrable by Remark 0.2,1) and in fact $\int_a^b f_n(x)\,\mathrm{d}x = 0$ for each n. But note that

$$0 = \lim_{n\to\infty} \int_a^b f_n(x)\,\mathrm{d}x \neq \int_a^b \lim_{n\to\infty} f_n(x)\,\mathrm{d}x \left(= \int_a^b f(x)\,\mathrm{d}x \right) \tag{0.6}$$

since the integral on the right-hand side (0.6) is not even defined (as a Riemann integral)!

The solution to this will come by means of the so-called *Lebesgue measure*, introduced in this course, and its associated *Lebesgue integral*, which represents an extension of the Riemann integral (in particular, everything you have learned regarding the integration of elementary functions will remain valid). The Lebesgue integral is more robust, it will allow for a much larger class of functions to be integrated (including $1_{\mathbb{Q}}$). It will also produce streamlined conditions concerning exchange of limits and integrals (making for instance (0.6) perfectly valid).

ii) (In-)completeness.

From the perspective of analysis, it is often convenient to work in the setting of complete normed vector spaces (cf. courses in Functional Analysis, PDE's,...), notably for applications of Banach's fixed point theorem.
For instance, one could endow the space $C^0_{pw}[a,b]$ with the (semi-)norm

$$\|f\|_{L^1} \stackrel{\text{def.}}{=} \int_a^b |f(x)|\,\mathrm{d}x \tag{0.7}$$

Then, referring to the sequence (f_n) from (0.5), one readily sees that (f_n) is a Cauchy-sequence with respect to $\|\cdot\|_{L^1}$ (in fact $\|f_n - f_m\|_{L^1} = 0$), but $f_n \stackrel{\text{ptw.}}{\to} f$. This will motivate the introduction of the *Banach space* $L^1[a,b]$ of (Lebesgue-)integrable functions (and more generally, L^p-space for $p \in [1,\infty]$) later in the course.

iii) Foundations of Probability Theory.

Measure theory also plays a fundamental role in supplying a rigorous framework for probability theory. In a nutshell, for a certain random experiment (e.g. repeated coin tosses) one defines a suitable measure space $(\Omega, \mathcal{A}, \mathbb{P})$. Here Ω is called the *sample space*, $\mathcal{A} \subset 2^\Omega$ is a set of *measurable* (cf. Chap.1) subsets of Ω, and $\mathbb{P}$ is a *probability measure*, which assigns a value $\mathbb{P}[A] \in [0,1]$ to each $A \in \mathcal{A}$. One interprets

$$\begin{aligned}
\Omega \quad &\longleftrightarrow \quad \text{possible outcomes of the random experiment} \\
&\qquad\quad \Big(\stackrel{\text{e.g.}}{=} \{\pm 1\}^{\mathbb{N}} \text{in case of repeated coin tosses}\Big) \\
\mathcal{A} \quad &\longleftrightarrow \quad \text{the set (σ-algebra, cf. below) of events} \\
\mathbb{P}[A] \quad &\longleftrightarrow \quad \text{the probability of the event } A \in \mathcal{A}
\end{aligned} \tag{0.8}$$

e.g. $A = \{\text{first toss is heads}\} = \{\omega \in \{\pm 1\}^{\mathbb{N}} : \omega_1 = 1\}$, $\mathbb{P}[A] = \frac{1}{2}$ assuming one models fair coin tosses. A *random variable* X (e.g. the number of heads among the first n tosses) is then a measurable (cf. Chap.2) function $X : \Omega \to \mathbb{R}$ and its expectation is given by the integral $\mathbb{E}[X] = \int_\Omega X(\omega)\,\mathrm{d}\mathbb{P}[\omega]$, assuming X is integrable w.r.t. $\mathbb{P}$. Here the integral refers to the one associated with the (probability)measure $\mathbb{P}$ (cf. Chap. 3).
For comparison, the Lebesgue measure referred to below (0.6) will also be introduced as a triplet

$$(\mathbb{R}, \mathcal{A}, \lambda) \tag{0.9}$$

where $\mathcal{A} \subset 2^{\mathbb{R}}$ is a collection (σ-algebra) of measurable subsets of $\mathbb{R}$, to which the measure $\lambda : \mathcal{A} \to [0,\infty]$ assigns size $\lambda(A) \in [0,\infty]$ for $A \in \mathcal{A}$. This indicates the benefits of studying so-called *measure spaces*, i.e. triplets as in (0.9) or above (0.8) in an abstract setting first, which is the approach we will take in this course.

Literature

— W. Rudin : Real and Complex Analysis, 3rd Ed. , McGraw-Hill, 1987. (mostly Chapters 1-8)

— H.L. Royden : Real Analysis, 3rd Ed., Macmillan, 1988.

— P.R. Halmos : Measure Theory, GTM, Springer, 1950.

Acknowledgements. I am grateful to Michael Struwe and his impeccable lectures, from which I learned most of what follows. I warmly thank Henri Elad Altman and Xue-Mei Li for sharing their material on a similar course they previously taught. Finally I am thankful to Lifan Xuan and Yuheng Wu for their help with the LaTeX-editing of these notes and to Yuriy Shulzhenko, Qiyu Xie and Yuxuan Du for their comments on successive versions of this mansucript.

Chapter 1

Measure Spaces

1.1 Abstract Measure Theory

Let X be an arbitrary set, $2^X \stackrel{\text{def.}}{=} \{A : A \subset X\}$ its power set. We denote by $\emptyset$ the empty set, $\emptyset \in 2^X$.

Definition 1.1. A family $\mathcal{A} \subset 2^X$ is called an **algebra** (over X) if

(1.1) $$X \in \mathcal{A}$$

(1.2) $$A \in \mathcal{A} \;\Rightarrow\; A^c \left(\stackrel{\text{def.}}{=} X \backslash A\right) \in \mathcal{A}$$

(1.3) $$A_1, \dots, A_m \in \mathcal{A} \;\Rightarrow\; \bigcup_{k=1}^{m} A_k \in \mathcal{A}$$

$\mathcal{A}$ is called a **σ-algebra** if (1.3) also holds for countable unions, i.e. if

(1.3') $$A_1, A_2, \dots \in \mathcal{A} \Rightarrow \bigcup_{k=1}^{\infty} A_k \in \mathcal{A}.$$

Remark 1.2. 1) The set $\{\emptyset, X\}$ is a σ-algebra, the coarsest (i.e. smallest, cf. 4) below) σ-algebra over X. More generally, $\{\emptyset, A, A^c, X\}$ is a σ-algebra, for every $A \subset X$.

2) Every σ-algebra is an algebra. The converse is false. Consider for instance $X = (0, 1]$ and

$$\mathcal{A} = \left\{ \emptyset, \text{ all sets of the form } \bigcup_{k=1}^{m} (a_k, b_k], \text{ for } 0 \le a_k < b_k \le 1,\ m \ge 1 \right\}.$$

Then $\mathcal{A}$ is an algebra (note that $(a, b]^c = (0, a] \cup (b, 1]$), but not a σ-algebra: indeed, choosing $A_k = (0, 1 - \frac{1}{k}]$, $k \ge 1$, one has that $\bigcup_{k=1}^{\infty} A_k = (0, 1) \notin \mathcal{A}$.

3) (Intersection of algebras/σ-algebras). Let $\mathcal{F}$ be an arbitrary collection of algebras/σ-algebras over X. Then the intersection

$$\bigcap_{\mathcal{A}\in\mathcal{F}} \mathcal{A} \overset{\text{def.}}{=} \{A : A \in \mathcal{A} \text{ for every } \mathcal{A} \in \mathcal{F}\} \tag{1.4}$$

is an algebra/σ-algebra. Indeed one checks that (1.4) satisfies (1.1)-(1.3)/(1.1)-(1.3').

4) (σ-algebra generated by a collection $\mathcal{C}$ of subsets). Let $\mathcal{C} \subset 2^X$. Then the set

$$\sigma(\mathcal{C}) \overset{\text{def.}}{=} \bigcap_{\substack{\mathcal{A}\subset 2^X \text{ a } \sigma\text{-algebra} \\ \text{with } \mathcal{C}\subset\mathcal{A}}} \mathcal{A} \qquad (\subset 2^X) \tag{1.5}$$

is well defined (cf. (1.4) regarding the meaning of the right-hand side) as the intersection is non-empty since 2^X is a σ-algebra (containing $\mathcal{C}$). Moreover $\sigma(\mathcal{C})$ defines a σ-algebra by 3), the smallest σ-algebra containing $\mathcal{C}$. Indeed if $\mathcal{A}$ is a σ-algebra with $\mathcal{C} \subset \mathcal{A}$, then $\sigma(\mathcal{C}) \subset \mathcal{A}$ by (1.5).

Referring to 1), one has for instance $\sigma(\emptyset) = \{\emptyset, X\}$, $\sigma(A) = \{\emptyset, A, A^c, X\}$. Moreover $\mathcal{C}$ is a σ-algebra if and only if $\sigma(\mathcal{C}) = \mathcal{C}$. A frequent instance of (1.5) is the following example. If (X, τ) is a topological space (with $\tau \subset 2^X$ the collection of open sets), then

$$\mathcal{B}(X) \overset{\text{def.}}{=} \sigma(\tau) \tag{1.6}$$

is called the **Borel σ-algebra** of (X, τ).

5) If $\mathcal{A}$ is a σ-algebra, then in view of (1.3') one also has:

$$A_k \in \mathcal{A},\ k \geq 1 \Rightarrow \bigcap_{k=1}^{\infty} A_k \in \mathcal{A},$$

which follows from (1.2), (1.3') and de Morgan's identity $\bigcap_{k=1}^{\infty} A_k = (\bigcup_{k=1}^{\infty} A_k^c)^c$.

A pair $(X, \mathcal{A})$, where $\mathcal{A}$ is a σ-algebra over X (cf. Definition 1.1) is called a **measurable space**. The elements of $\mathcal{A}$ are called **measurable sets**. A measure assigns a 'size' to each measurable set as follows:

Definition 1.3. Let $(X, \mathcal{A})$ be a measurable space. A **measure** on $(X, \mathcal{A})$ is a function $\mu : \mathcal{A} \to [0, \infty]$ such that:

$$\mu(\emptyset) = 0, \tag{1.7}$$

(1.8) If $A_k \in \mathcal{A},\ k = 1, 2, \dots,$ are pairwise disjoint (i.e. $A_l \cap A_k = \emptyset,\ \forall l \neq k$), then $\mu(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} \mu(A_k)$.

(note that $\bigcup_{k=1}^{\infty} A_k \in \mathcal{A}$ by (1.3')). The property (1.8) is called **σ-additivity**.

A **measure space** is a triplet $(X, \mathcal{A}, \mu)$ consisting of a set X, a σ-algebra $\mathcal{A} \subset 2^X$ and a measure $\mu : \mathcal{A} \to [0, \infty]$.

Note: there is no issue with the series appearing in (1.8), which is over non-negative real numbers. Its value is an element of the extended real line $[0,\infty] = [0,\infty) \cup \{\infty\}$ (i.e. the series may well diverge). By convention, in the sequel we set $x + \infty = \infty$, $\forall x \in [0,\infty]$ (in particular: $\infty + \infty = \infty$) and $x \times \infty = \infty$, $\forall x \in [0,\infty]$.

Example 1.4. 1) Let $(X, \mathcal{A})$ be a measurable space. The *counting measure* $\mu : \mathcal{A} \to [0,\infty]$ is defined by $\mu(A) = n$ if $A \in \mathcal{A}$ has exactly n elements and $\mu(A) = \infty$ otherwise. This is a measure (exercise).

2) Let $(X, \mathcal{A})$ be a measurable space and $x \in X$. Then $\delta_x : \mathcal{A} \to [0,\infty]$ defined by

$$\delta_x(A) \stackrel{\text{def.}}{=} \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{if } x \notin A \end{cases} \quad \textit{for } A \in \mathcal{A}.$$

defines a measure on $(X, \mathcal{A})$, the *Dirac measure at* x.

3) Let X be uncountable. Then $\mathcal{A} = \{A \subset X : A \text{ or } A^c \text{ is countable}\}$ defines a σ-algebra on X (check!). The function $\mu : \mathcal{A} \to [0,1]$ defined as $\mu(A) = 0$ if A is countable, $\mu(A) = 1$ if A^c is countable, is a measure on $(X, \mathcal{A})$.

We now collect a few basic properties of measures.

Proposition 1.5. *Let* $(X, \mathcal{A}, \mu)$ *be a measure space.*

(1.9) (Monotonicity) If $A, B \in \mathcal{A}$ are such that $A \subset B$, then $\mu(A) \leq \mu(B)$.

(1.10) (Finite additivity) If $n \geq 1, A_k \in \mathcal{A}, 1 \leq k \leq n$ and A_k are pairwise disjoint, then $\mu(\bigcup_{k=1}^n A_k) = \sum_{k=1}^n \mu(A_k)$.

(1.11) If $A_k \in \mathcal{A}$ are such that $A_k \subset A_{k+1}$, $\forall k \geq 1$, then $\mu\Big(\bigcup_{k=1}^\infty A_k\Big) = \lim_{k\to\infty} \mu(A_k)$.

(1.12) If $A_k \in \mathcal{A}, k \geq 1$ are such that $A_k \supset A_{k+1}$ for all $k \geq 1$, then $\mu(A_1) < \infty \to \mu\Big(\bigcap_{k=1}^\infty A_k\Big) = \lim_{k\to\infty} \mu(A_k)$.

(1.13) (σ-subadditivity) If $A, A_k \in \mathcal{A}, k \geq 1, A \subset \bigcup_{k=1}^\infty A_k$, then $\mu(A) \leq \sum_{k=1}^\infty \mu(A_k)$.

(Note: in view of (1.9)*, the limits on the RHS of* (1.11) *and* (1.12) *exist on* $[0,\infty]$ *since they are montone and in* (1.12)*,* $\bigcap_{k=1}^\infty A_k \in \mathcal{A}$ *by Remark 1.2, 5).)*

Proof. Property (1.10) follows from (1.8) by choosing $A_k = \emptyset$, $k \geq n+1$ and using (1.7). Property (1.9) follows from (1.10) by choosing $n = 2$, $A_1 = A$, $A_2 = B\backslash A (= B \cap A^c \in \mathcal{A})$ whence $\mu(B) = \mu(A) + \mu(B\backslash A) \geq \mu(A)$.

We now show (1.11). Partition $A = \bigcup_{k=1}^{\infty} A_k = \bigcup_{k=1}^{\infty} \widetilde{A}_k$ into disjoint sets $\widetilde{A}_k$ with $\widetilde{A}_1 = A_1$, $\widetilde{A}_k = A_k \backslash A_{k-1}$, to obtain

$$\mu(A) = \mu\Big(\bigcup_{k=1}^{\infty} \widetilde{A}_k\Big) \overset{(1.8)}{=} \sum_{k=1}^{\infty} \mu(\widetilde{A}_k) = \lim_{m\to\infty} \sum_{k=1}^{m} \mu(\widetilde{A}_k) \overset{(1.10)}{=} \lim_{m\to\infty} \mu(A_m).$$

To obtain (1.12), define $\widetilde{A}_k = A_1 \backslash A_k, k \geq 1$, so that $\emptyset = \widetilde{A}_1 \subset \widetilde{A}_2 \subset \cdots$ and note that by (1.10),

$$\mu(A_1) = \mu(\widetilde{A}_k) + \mu(A_k), \quad \text{for all } k \geq 1.$$

Taking limits as $k \to \infty$ (which exist because they are monotone) gives

$$\begin{aligned} \mu(A_1) - \lim_{k\to\infty} \mu(A_k) = \lim_{k\to\infty} \mu(\widetilde{A}_k) &\overset{(1.11)}{=} \mu\Big(\bigcup_{k=1}^{\infty} \widetilde{A}_k\Big) \\ &\overset{*}{=} \mu\Big(A_1 \backslash \Big(\bigcap_{k=1}^{\infty} A_k\Big)\Big) \overset{(1.10)}{=} \mu(A_1) - \mu\Big(\bigcap_{k=1}^{\infty} A_k\Big), \end{aligned}$$

from which (1.12) follows. In $*$ we used:

$$\bigcup_{k=1}^{\infty} \widetilde{A}_k = \bigcup_{k=1}^{\infty} (A_1 \cap A_k^c) = A_1 \cap \Big(\bigcup_{k=1}^{\infty} A_k^c\Big) \overset{\text{de Morgan}}{=} A_1 \cap \Big(\bigcap_{k=1}^{\infty} A_k\Big)^c = A_1 \backslash \Big(\bigcap_{k=1}^{\infty} A_k\Big).$$

Finally, for $A_k \in \mathcal{A}, k \geq 1$, let $\widetilde{A}_1 = A \cap A_1, \widetilde{A}_k = A \cap A_k \cap \big(\bigcap_{l=1}^{k-1} A_l\big)^c$, whence the sets $\widetilde{A}_k$ are disjoint, and

$$\mu(A) = \mu\Big(\bigcup_{k=1}^{\infty} \widetilde{A}_k\Big) \overset{(1.8)}{=} \sum_{k=1}^{\infty} \mu(\widetilde{A}_k) \overset{(1.9)}{\leq} \sum_{k=1}^{\infty} \mu(A_k).$$

□

Remark 1.6. 1) Let $X = \mathbb{N}$ and $\mu =$ counting measure on $(\mathbb{N}, 2^{\mathbb{N}})$, see Example 1.4, 1). Then defining $A_k = \{k, k+1, \ldots\}$ for $k \geq 1$, one has $A_k \supset A_{k+1}$ for all k, and $\bigcap_{k=1}^{\infty} A_k = \emptyset$, whence $\infty = \lim_{k\to\infty} \mu(A_k) \neq \mu\left(\bigcap_{k=1}^{\infty} A_k\right) = 0$. The condition $\mu(A) < \infty$ is thus needed in (1.12).

2) Probability theory studies measure spaces $(X, \mathcal{A}, \mu)$ such that $\mu(X) = 1$. The elements of $\mathcal{A}$ are usually called *events* (rather than measurable sets). It follows by (1.7) and (1.9) that $\mu(A) \in [0,1]$ for all $A \in \mathcal{A}$, and $\mu(A)$ is given the interpretation to be the *probability of* A. The contents of this course represent the starting point for probability theory. One of the first intrinsically probabilistic notions is that of *independence*, which is a certain property of measures. Many of the classical theorems in probability theory (law of large numbers, central limit theorem,...) derive from this notion. Independence will not be discussed in this course.

1.2 Construction of Measures

Let $X \neq \emptyset$ be arbitrary. We now provide a tool to construct measure a on X. This will roughly work as follows:

Given: $\widetilde{\mu}$ a **pre-measure** $\xrightarrow[\text{Step 1}]{\text{extend}}$ μ^*: an **outer measure** $\xrightarrow[\text{Step 2}]{\text{restrict}}$ μ: a **measure**.

The point of this is that $\widetilde{\mu}$ is typically "easy" to define in cases of interest. For instance, one may know which measure one wants a certain initial class of sets to be assigned. The theory then provides the rest (see Theorem 1.13 below).

Definition 1.7.

i) Let $\mathcal{A} \subset 2^X$ be an algebra. A function $\mu : \mathcal{A} \to [0, \infty]$ satisfying (1.7) and (1.8)*, i.e. (1.8) whenever $\bigcup_{k=1}^{\infty} A_k \in \mathcal{A}$, is called a **pre-measure** (on X).

ii) A function $\mu : 2^X \to [0, \infty]$ satisfying (1.7) and (1.13) with $\mathcal{A} = 2^X$ is called an **outer measure** (on X).

Step 1 in the above "construction" will be driven by the next proposition.

Definition 1.8. A family $\mathcal{K} \subset 2^X$ is called a **cover of** X if

$$\emptyset \in \mathcal{K}, \text{ and} \tag{1.14}$$

$$\exists (K_n)_{n\in\mathbb{N}} \subset \mathcal{K} \text{ such that } X = \bigcup_{n=1}^{\infty} K_n. \tag{1.15}$$

Example 1.9. 1) The open "intervals"

$$I = \prod_{k=1}^{n} (a_k, b_k) \overset{\text{def.}}{=} \{x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n : a_k < x_k < b_k \textit{ for all } 1 \le k \le n\},$$

with $a_k \le b_k \in \mathbb{R}$, form a cover of $X = \mathbb{R}^n$; similarly the closed intervals $\prod_{k=1}^{n}[a_k, b_k]$ (require $a_k \le x_k \le n_k$ instead) or the half-open intervals $\prod_{k=1}^{n}(a_k, b_k]$ or $\prod_{k=1}^{n}[a_k, b_k)$.

2) Every algebra $\mathcal{A} \subset 2^X$ is a cover of X, since $\emptyset, X \in \mathcal{A}$ by (1.1) and (1.2).

Proposition 1.10. *Let $\mathcal{K}$ be a cover of X, $\widetilde{\mu} : \mathcal{K} \to [0, \infty]$ be a map with $\widetilde{\mu}(\emptyset) = 0$. Then*

$$\mu^*(A) \overset{\text{def.}}{=} \inf \left\{ \sum_{j=1}^{\infty} \widetilde{\mu}(K_j) : K_j \in \mathcal{K},\ A \subset \bigcup_{j=1}^{\infty} K_j \right\}, \quad \textit{for } A \in 2^X, \tag{1.16}$$

defines an outer measure on X.

Proof. μ^* is well-defined, as the sequence $(K_n)_{n\in\mathbb{N}}$ supplied by (1.15) is always a valid choice on the right-hand side. Clearly $\mu^*(A) \in [0,\infty]$ for any $A \in 2^X$ and $\mu^*(\emptyset) = 0$ follows by choosing $K_j = \emptyset$, all j, and using $\widetilde{\mu}(\emptyset) = 0$. Thus, (1.7) holds for μ^*. It remains to show (1.13) (for μ^* in place of μ and with $\mathcal{A} = 2^X$). Let $A_k \in 2^X, k \in \mathbb{N}$. For each $k \in \mathbb{N}$ and $\varepsilon > 0$, by (1.16), we can find a sequence $(K_{k,j})_{j\in\mathbb{N}} \subset \mathcal{K}$ such that $A_k \subset \bigcup_{j=1}^{\infty} K_{k,j}$ and

$$\sum_{j=1}^{\infty} \widetilde{\mu}(K_{k,j}) < \mu^*(A_k) + 2^{-k}\varepsilon \tag{1.17}$$

Now assume $A \in 2^X$, $A \subset \bigcup_{k=1}^{\infty} A_k$ (as in (1.13)). Then $A \subset \bigcup_{k=1}^{\infty}\bigcup_{j=1}^{\infty} K_{k,j}$, thus

$$\mu^*(A) \overset{(1.16)}{\leq} \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} \widetilde{\mu}(K_{k,j}) \overset{(1.17)}{<} \sum_{k=1}^{\infty} \mu^*(A_k) + \varepsilon.$$

The claim follows by letting $\varepsilon \downarrow 0$. □

We now highlight one key property of outer measures.

Lemma 1.11. *If μ^* is an outer measure on X, then*

$$\Sigma \overset{\text{def.}}{=} \{A \subset X : \mu^*(B) = \mu^*(B \cap A) + \mu^*(B\backslash A) \text{ for all } B \subset X\} \tag{1.18}$$

is a σ-algebra on X.

Remark 1.12. By subadditivity of μ (cf. (1.13)), Σ is equivalently defined by requiring $\mu^*(B) \geq \mu^*(B \cap A) + \mu^*(B\backslash A)$ in (1.18).

We defer the proof of Lemma 1.11 for a few lines and proceed to the main result of this section. Its usefulness will be witnessed in the next section, which comprises Step 2.

Theorem 1.13 (Hahn-Carathéodory Extension Theorem)**.**
Let X be an arbitrary set, $\mathcal{A}$ an algebra over X and $\widetilde{\mu} : \mathcal{A} \to [0,\infty]$ a pre-measure on X. Then, defining μ^* by (1.16) with $\mathcal{K} = \mathcal{A}$, Σ by (1.18) and $\mu \overset{\text{def.}}{=} \mu^*|_\Sigma$ (i.e. $\mu : \Sigma \to [0,\infty]$, $\mu(A) = \mu^*(A)$ for all $A \in \Sigma$), one has:

$$(X, \Sigma, \mu) \text{ is a measure space.} \tag{1.19}$$

$$\mathcal{A} \subset \Sigma. \tag{1.20}$$

$$\mu^*(A) = \widetilde{\mu}(A) = \mu(A), \text{ for all } A \in \mathcal{A}. \tag{1.21}$$

Proof. We start with (1.19). μ^* is an outer measure by Proposition 1.10 and Example 1.9,2). Hence Σ is a σ-algebra by Lemma 1.11. Thus (X, Σ) is a measurable space. It remains to argue that μ is a measure on (X,Σ). The property (1.7) is inherited from μ^* (cf. Definition 1.7,ii)). To obtain (1.8), first note that for $A, B \in \Sigma$, $A \cap B = \emptyset$,

$$\begin{aligned}\mu(A\cup B) &= \mu^*(A\cup B) \overset{(1.18)}{=} \mu^*((A\cup B)\cap A) + \mu^*((A\cup B)\backslash A)\\ &\overset{A\cap B=\emptyset}{=} \mu^*(A)+\mu^*(B) = \mu(A)+\mu(B).\end{aligned}$$

By induction, this yields that μ is finitely additive (i.e. it satisfies (1.10)). In particular, it follows that μ satisfies (1.9), as can be seen by inspecting the proof of (1.9), which only relies on (1.10). Now assume $A_1, A_2, \dots$ are piecewise disjoint. Then for all $m \geq 1$,

$$\sum_{n=1}^m \mu(A_n) \overset{(1.10)}{=} \mu\left(\bigcup_{n=1}^m A_n\right) \overset{(1.9)}{\leq} \mu\left(\bigcup_{n=1}^\infty A_n\right). \tag{1.22}$$

Letting $m \to \infty$ in (1.22) yields $\sum_{n=1}^\infty \mu(A_n) \leq \mu\left(\bigcup_{n=1}^\infty A_n\right)$, the reverse inequality follows from the fact that μ^* satisfies (1.13) and that $\mu(A) = \mu^*(A)$ for all $A \in \mathcal{A}$.

We now show (1.20). Let $A \in \mathcal{A}$. By Remark 1.12, it suffices to show that

$$\mu^*(B) \geq \mu^*(B\cap A) + \mu^*(B\backslash A), \quad \text{for all } B \subset X. \tag{1.23}$$

Let $\varepsilon > 0$. By (1.16) we can find $K_j \in \mathcal{A}(=\mathcal{K})$, $j \geq 1$, such that

$$\sum_{j=1}^\infty \widetilde{\mu}(K_j) \leq \mu^*(B) + \varepsilon \quad \text{and} \quad B \subset \bigcup_{j=1}^\infty K_j. \tag{1.24}$$

Hence $B \cap A \subset \left(\bigcup_{j=1}^\infty (K_j \cap A)\right)$, $B\backslash A \subset \left(\bigcup_{j=1}^\infty (K_j\backslash A)\right)$ and therefore

$$\begin{aligned}\mu^*(A\cap B) + \mu^*(B\backslash A) &\overset{(1.13)}{\leq} \sum_{j=1}^\infty (\mu^*(K_j\cap A) + \mu^*(K_j\backslash A))\\ &\underset{(1.16)}{\overset{A,K_j\in\mathcal{A}}{\leq}} \sum_{j=1}^\infty (\widetilde{\mu}(K_j\cap A) + \widetilde{\mu}(K_j\backslash A))\\ &\underset{(1.8)}{\overset{A,K_j\in\mathcal{A}}{=}} \sum_{j=1}^\infty \widetilde{\mu}(K_j) = \sum_{j=1}^\infty \mu^*(K_j) \overset{(1.24)}{\leq} \mu^*(B)+\varepsilon.\end{aligned}$$

Letting $\varepsilon \downarrow 0$, (1.23) follows.

We now argue that (1.21) holds. Let $A \in \mathcal{A}(=\mathcal{K}$ in (1.16)). By choosing $K_1 = A(\in \mathcal{K})$, $K_j = \emptyset$, $j \geq 2$, we obtain that $\mu^*(A) \leq \widetilde{\mu}(A)$. For the reverse inequality, let $K_j \in \mathcal{A}$, $j \geq 1$, be arbitrary with $A \subset \bigcup_{j=1}^\infty K_j$. Define $\widetilde{K}_1 = K_1$, $\widetilde{K}_j = K_j \backslash \left(\bigcup_{l=1}^{j-1} K_l\right)$, $j \geq 2$. Since $\mathcal{A}$ is an algebra, $\widetilde{K}_j \in \mathcal{A}$ for all $j \geq 1$, the $\widetilde{K}_j$'s are disjoint by construction and $\bigcup_{j=1}^\infty \widetilde{K}_j = \bigcup_{j=1}^\infty K_j$ contains A. Hence the sets $\widetilde{\widetilde{K}}_j = \widetilde{K}_j \cap A$, $j \geq 1$, are disjoint sets in $\mathcal{A}$ whose union is A. Applying (1.8), we get

$$\widetilde{\mu}(A) = \sum_{j=1}^\infty \widetilde{\mu}(\widetilde{\widetilde{K}}_j) \leq \sum_{j=1}^\infty \widetilde{\mu}(\widetilde{K}_j) \leq \sum_{j=1}^\infty \widetilde{\mu}(K_j), \tag{1.25}$$

where in the first inequality, we used that

$$\widetilde{\mu}(\widetilde{K}_j) \overset{(1.8)}{=} \widetilde{\mu}(\widetilde{\widetilde{K}}_j) + \widetilde{\mu}(\widetilde{K}_j \cap A^c) \geq \widetilde{\mu}(\widetilde{\widetilde{K}}_j)$$

along with a similar reasoning for the second inequality in (1.25). In view of (1.25), (1.16), taking an infimum over collections $K_j \in \mathcal{A}$, $j \geq 1$, with $A \subset \bigcup_{j=1}^{\infty} K_j$ yields $\widetilde{\mu}(A) \leq \mu^*(A)$, whence (1.21). $\square$

We now supply the:

Proof of Lemma 1.11. We need to verify (1.1), (1.2), (1.3'). For ease of notation, we omit the superscript $*$ from μ^* throughout this proof.

(1.1): $\mu(B \cap X) + \mu(B \backslash X) = \mu(B) + \mu(\emptyset) \overset{(1.7)}{=} \mu(B)$, for all $B \subset X$, hence $X \in \Sigma$.

(1.2): Suppose that $A \in \Sigma$. Then, for all $B \subset X$, one has

$$\mu(B \cap A^c) + \mu(B \backslash A^c) = \mu(B \backslash A) + \mu(B \cap A) \overset{A \in \Sigma}{=} \mu(B),$$

hence $A^c \in \Sigma$.

(1.3'): We first show (1.3) by induction over $m \geq 1$. The case $m = 1$ is trivial. Assume (1.3) holds for $m-1$, for some integer $m \geq 2$. Let $A_1 \dots, A_m \in \Sigma$. Define $A = \bigcup_{k=1}^{m-1} A_k$. By induction assumption, $A \in \Sigma$. Now for arbitrary $B \subset X$, in view of (1.18), we have

$$\mu(B) = \mu(B \cap A) + \mu(B \backslash A) \qquad \text{(since } A \in \Sigma\text{)}, \tag{1.26}$$

$$\mu(B \backslash A) = \mu((B \backslash A) \cap A_m) + \mu((B \backslash A) \backslash A_m) \qquad \text{(since } A_m \in \Sigma\text{)}. \tag{1.27}$$

Hence,

$$\begin{aligned} \mu(B) &\overset{(1.26)(1.27)}{=} \mu(B \cap A) + \mu((B \backslash A) \cap A_m) + \mu((B \backslash A) \backslash A_m) \\ &\geq \quad \mu(B \cap (A \cup A_m)) + \mu(B \backslash (A \cup A_m)), \end{aligned}$$

(using subadditivity of the outer measure $\mu = \mu^*$ in the second line) i.e. $A \cup A_m = \bigcup_{k=1}^{m} A_k \in \Sigma$ in view of Remark 1.12.
We now show (1.3'). Let $A_1, A_2, \dots \in \Sigma$. We need to argue that $\bigcup_{k=1}^{\infty} A_k \in \Sigma$ and we may assume to that effect that $A_k \cap A_l = \emptyset$, $k \neq l$. (Indeed, otherwise consider $\widetilde{A}_1 = A_1$, $\widetilde{A}_k = A_k \backslash \bigcup_{l=1}^{k-1} A_l$, which satisfy $\widetilde{A}_k \in \Sigma$ (by what is already shown), for all $k \geq 1$, and observe that $\bigcup_{k=1}^{\infty} \widetilde{A}_k = \bigcup_{k=1}^{\infty} A_k$.) Now, for all $m \geq 1$, $B \subset X$, we have

$$\begin{aligned} \mu\Big(B \cap \Big(\bigcup_{k=1}^{m} A_k\Big)\Big) &\overset{A_m \in \Sigma}{=} \mu\Big(B \cap \Big(\bigcup_{k=1}^{m} A_k\Big) \cap A_m\Big) + \mu\Big(B \cap \Big(\bigcup_{k=1}^{m} A_k\Big) \backslash A_m\Big) \\ &\overset{A_k\text{'s disjt.}}{=} \mu(B \cap A_m) + \mu\Big(B \cap \Big(\bigcup_{k=1}^{m-1} A_k\Big)\Big) \qquad (1.28) \\ &\overset{\text{iterate}}{=} \dots = \sum_{k=1}^{m} \mu(B \cap A_k). \end{aligned}$$

Using (1.18) and the fact that $\mu(B) \geq \mu(A)$ for all $A, B \subset X$ with $A \subset B$ (this follows immediately by considering $A_1 = B$, $A_k = \emptyset$, $k \geq 2$ in (1.13) and using (1.7)) yields, for all $B \subset X$, $m \geq 1$,

$$\mu(B) \overset{\bigcup_{k=1}^m A_k \in \Sigma}{=} \mu\Big(B \cap \Big(\bigcup_{k=1}^m A_k\Big)\Big) + \mu\Big(B \backslash \Big(\bigcup_{k=1}^m A_k\Big)\Big) \geq \sum_{k=1}^m \mu(B \cap A_k) + \mu\Big(B \backslash \Big(\bigcup_{k=1}^\infty A_k\Big)\Big).$$

Thus, letting $m \to \infty$, we find that

$$\mu(B) \geq \sum_{k=1}^\infty \mu(B \cap A_k) + \mu\Big(B \backslash \Big(\bigcup_{k=1}^\infty A_k\Big)\Big) \overset{(1.13)}{\geq} \mu\Big(B \cap \Big(\bigcup_{k=1}^\infty A_k\Big)\Big) + \mu\Big(B \backslash \Big(\bigcup_{k=1}^\infty A_k\Big)\Big).$$

On account of Remark 1.12, this implies that $\bigcup_{k=1}^\infty A_k \in \Sigma$. □

Theorem 1.13 is an existence result. We conclude by discussing the uniqueness of the measure μ extending a given pre-measure $\widetilde{\mu} : \mathcal{A} \to [0, \infty]$ (with $\mathcal{A} \subset 2^X$ an algebra), which requires an additional assumption. A (pre-)measure $\widetilde{\mu}$ is called *σ-finite* if

$$\begin{aligned}&\text{there exist disjoint sets } S_k \in \mathcal{A},\ k \geq 1 \text{ such that}\\ &X = \bigcup_{k=1}^\infty S_k \text{ and } \widetilde{\mu}(S_k) < \infty \text{ for all } k \geq 1.\end{aligned} \tag{1.29}$$

Theorem 1.14 (Uniqueness of Hahn-Carathéodory Extension)**.**
Under the assumptions of Theorem 1.13, and if $\widetilde{\mu}$ is σ-finite, the following holds: let $\nu : 2^X \to [0, \infty]$ be an outer measure with $\nu|_{\mathcal{A}} = \widetilde{\mu}$. Then $\nu|_\Sigma = \mu$.

Proof. Let $A \in \Sigma$. We first show $\nu(A) \leq \mu(A)$. Let $A_k \in \mathcal{A}$, $k \geq 1$, be such that $A \subset \bigcup_{k=1}^\infty A_k$. Then by subadditivity of ν,

$$\nu(A) \leq \sum_{k=1}^\infty \nu(A_k) = \sum_{k=1}^\infty \widetilde{\mu}(A_k). \tag{1.30}$$

Taking infima over $A_k \in \mathcal{A}$, $k \geq 1$ s.t. $A \subset \bigcup_{k=1}^\infty A_k$ on either side of (1.30), it follows in view of (1.16) that $\nu(A) \leq \mu^*(A) \overset{(1.21)}{=} \mu(A)$.

We now show that $\mu(A) \leq \nu(A)$. First suppose that

$$\text{there exists } S \in \mathcal{A} \text{ such that } A \subset S \text{ and } \widetilde{\mu}(S) < \infty. \tag{1.31}$$

Then by the inequality just shown we know that

$$\begin{aligned}\nu(A) + \nu(S \backslash A) &\overset{\nu \leq \mu}{\leq} \mu(A) + \mu(S \backslash A) = \mu(S)\\ &\overset{S \in \mathcal{A}}{=} \widetilde{\mu}(S) \overset{\nu|_{\mathcal{A}} = \widetilde{\mu}}{=} \nu(S) \overset{(1.13)}{\leq} \nu(A) + \nu(S \backslash A).\end{aligned} \tag{1.32}$$

Thus (1.32) is a chain of equalities, and using that

$$\mu(S\backslash A) \overset{(1.9)}{\leq} \mu(S) \overset{(1.21)}{=} \widetilde{\mu}(S) < \infty,$$

it follows that

$$\mu(A) = \nu(A) + \nu(S\backslash A) - \mu(S\backslash A) \overset{\nu\leq\mu}{\leq} \nu(A),$$

for A satisfying (1.31).

We now remove this assumption using (1.29). Let $A_k = A \cap S_k$, $k \geq 1$. The sets A_k are disjoint and $A = \bigcup_{k=1}^{\infty} A_k$. Since $\mu(A) = \nu(A)$ on sets A satisfying (1.31), we have $\mu\left(\bigcup_{k=1}^{m} A_k\right) = \nu\left(\bigcup_{k=1}^{m} A_k\right)$ for all $m \geq 1$. Hence,

$$\nu(A) \overset{(1.9)}{\geq} \lim_{m\to\infty} \nu\Big(\bigcup_{k=1}^{m} A_k\Big) = \lim_{m\to\infty} \mu\Big(\bigcup_{k=1}^{m} A_k\Big) \overset{(1.8)}{=} \mu(A),$$

as desired. □

We return to the failure of the uniqueness property in absence of σ-finiteness below.

1.3 The Lebesgue Measure

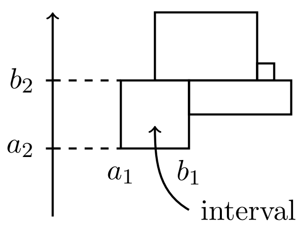
As a first application of Theorems 1.13 and 1.14, we construct the Lebesgue measure λ on $\mathbb{R}^n$, which extends the set function (pre-measure) assigning e.g. in case $n = 1$ "length $b - a$" to every interval $(a, b) \in \mathbb{R}$.

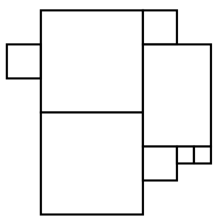
Definition 1.15. For $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$, $b = (b_1, \ldots, b_n) \in \mathbb{R}^n$, let

$$(a, b) \overset{\text{def.}}{=} \prod_{k=1}^{n} (a_k, b_k) = \{x = (x_1, \ldots, x_n) \in \mathbb{R}^n : a_k < x_k < b_k,\ 1 \leq k \leq n\}$$

if $a_k < b_k$ for all $1 \leq k \leq n$, and $(a, b) = \emptyset$ otherwise.

The sets $[a, b]$, $(a, b]$, $[a, b)$ are defined similarly. In case the endpoint is not included, the choice $a_k = -\infty$ or $b_k = +\infty$ is also permitted. We refer to any such set as **interval** (thus all intervals are subsets of $\mathbb{R}^n$). We call an **elementary figure** any union $\bigcup_{k=1}^{m} I_k$ of finitely many disjoint intervals $I_1, \ldots, I_m$.





: an elementary figure in $\mathbb{R}^2$.

We then introduce

$$\mathcal{A} = \{A \subset \mathbb{R}^n : A \textit{ is an elementary figure}\}. \tag{1.33}$$

One readily sees that $\mathcal{A}$ is an algebra on $X = \mathbb{R}^n$. Indeed, (1.1) is immediate, for (1.2) one note that if $A = \bigcup_{k=1}^m I_k \in \mathcal{A}$, then by de Morgan's law $A^c = \bigcap_{k=1}^m I_k^c$, the complement of an interval is a finite union of disjoint intervals (see also Remark 1.2, 2) in case $n = 1$), and the intersection of a finite number of intervals is an interval (use induction). The property (1.3), is obtained by similar considerations (for instance, the union of two intervals can always be expressed as the union of at most a bounded number of disjoint intervals).

We then define the function $\widetilde{\lambda} : \mathcal{A} \to [0, \infty]$ by declaring

$$\begin{aligned} \widetilde{\lambda}((a,b)) &= \widetilde{\lambda}((a,b]) = \widetilde{\lambda}([a,b)) = \widetilde{\lambda}([a,b]) \\ &\overset{\text{def.}}{=} \begin{cases} \prod_{k=1}^n (b_k - a_k), & \text{if } a_k < b_k,\ 1 \le k \le n. \\ 0, & \text{else.} \end{cases} \end{aligned} \tag{1.34}$$

$$\widetilde{\lambda}\Big(\bigcup_{k=1}^m I_k\Big) \overset{\text{def.}}{=} \sum_{k=1}^m \widetilde{\lambda}(I_k), \text{ if } I_1, \dots, I_m \text{ are disjoint intervals.}$$

Lemma 1.16. *The map $\widetilde{\lambda}$ defined by* (1.34) *is a pre-measure on $\mathcal{A}$.*

PROOF. In view of Definition 1.7, i) we only need to check (1.8)*, i.e. if $I \in \mathcal{A}$ is a disjoint union of countably many sets in $\mathcal{A}$, whence $I = \bigcup_{k=1}^\infty I_k$, for disjoint intervals I_k, $k \ge 1$, then

$$\widetilde{\lambda}(I) = \sum_{k=1}^\infty \widetilde{\lambda}(I_k). \tag{1.35}$$

Note that $\widetilde{\lambda}$ is finitely additive by definition (see (1.34)), hence monotone on $\mathcal{A}$, therefore

$$\widetilde{\lambda}(I) \overset{\text{monot.}}{\ge} \widetilde{\lambda}\left(\bigcup_{k=1}^m I_k\right) \overset{(1.34)}{=} \sum_{k=1}^m \widetilde{\lambda}(I_k),$$

for all m. Letting $m \to \infty$ gives "$\ge$" in (1.35).

For the reverse inequality, we use a compactness argument to reduce to finite additivity. We may assume that $\sum_{k=1}^\infty \widetilde{\lambda}(I_k) < \infty$ (else we already know (1.35) holds) and that I is an interval (since $I \in \mathcal{A}$). We will explain how to treat the general case at the end of the proof. Let $\bar{I}$ be the closure of I, $\bar{I}_L = \bar{I} \cap [-L, L]^n$, $L > 0$. By (1.34),

$$\widetilde{\lambda}(\bar{I}_L) \to \widetilde{\lambda}(\bar{I}) = \widetilde{\lambda}(I) \quad \text{as} \quad L \to \infty. \tag{1.36}$$

The set $\bar{I}_L$ is compact. Let $\varepsilon > 0$ and fix open intervals I_k^ε such that $I_k^\varepsilon \supset I_k$ for every $k \geq 1$ and satisfying

$$\widetilde{\lambda}(I_k^\varepsilon) < \widetilde{\lambda}(I_k) + \varepsilon 2^{-k-1},\ k \geq 1. \tag{$\widetilde{1.36}$}$$

Moreover let $K_\varepsilon \subset I_L$ be *closed* such that $\widetilde{\lambda}(K_\varepsilon) \geq \widetilde{\lambda}(I_L) - \frac{\varepsilon}{2}$. Now $K_\varepsilon \subset I_L \subset I \subset \bigcup_{k=1}^\infty I_k^\varepsilon$. Thus the open sets I_k^ε form an (open) cover of the compact set K_ε. By the Heine-Borel theorem, finitely many intervals I_k^ε, $1 \leq k \leq m(L)$, cover K_ε. It follows that

$$\widetilde{\lambda}(\bar{I}_L) \leq \widetilde{\lambda}(K_\varepsilon) + \frac{\varepsilon}{2} \overset{\text{monot.}}{\leq} \widetilde{\lambda}\left(\bigcup_{k=1}^{m(L)} I_k^\varepsilon\right) + \frac{\varepsilon}{2} \underset{\text{subadd.}}{\overset{\text{finite}}{\leq}} \sum_{k=1}^{m(L)} \widetilde{\lambda}(I_k^\varepsilon) + \frac{\varepsilon}{2} \overset{(\widetilde{1.36})}{\leq} \sum_{k=1}^{\infty} \widetilde{\lambda}(I_k) + \varepsilon.$$

Letting first $\varepsilon \downarrow 0$, then $L \to \infty$, and using (1.36), "$\leq$" follows in (1.35).

In the general case, i.e. without assuming I to be an interval, one argues as follows. Since $I \in \mathcal{A}$ one knows by (1.33) and the definition of an elementary figure that I can be expressed as $I = \bigcup_{k=1}^n J_k$ for disjoint intervals J_k and moreover that $\widetilde{\lambda}(I) = \sum_{k=1}^n \lambda(J_k)$ by (1.34). But since $I = \bigcup_{k=1}^\infty I_k$ and the intervals I_k are disjoint, this means that for every k, $J_k = \bigcup_\ell I_{k,\ell}$ for a certain subset $\{I_{k,1}, I_{k,2}, \dots\} \subset \{I_1, I_2, \dots\}$ and the collections $\{I_{k,1}, I_{k,2}, \dots\}$ are disjoint as k varies. Now one simply applies the above argument separately to the interval J_k (instead of I) and the collection $\{I_{k,1}, I_{k,2}, \dots\}$ (instead of $\{I_1, I_2, \dots\}$). This yields that

$$\widetilde{\lambda}(J_k) \leq \sum_\ell \widetilde{\lambda}(I_{k,\ell}), \text{ for every } k \in \{1, \dots, n\}$$

and "$\leq$" in (1.35) follows by summing over k on both sides. $\square$

With Lemma 1.16 at hand, and since (1.29) holds for $\widetilde{\mu} = \widetilde{\lambda}$ (for instance using the unit cubes $Q_z = \prod_{k=1}^n (z_k, z_k + 1]$, $z = (z_1, \dots, z_n) \in \mathbb{Z}^n$, which tile $\mathbb{R}^n$ and have $\widetilde{\lambda}(Q_z) = 1 < \infty$), Theorems 1.13 and 1.14 apply with $X = \mathbb{R}^n$, $\widetilde{\mu} = \widetilde{\lambda}$ and $\mathcal{A}$ given by (1.33), and yield the existence of a unique measure λ on $\mathbb{R}^n$ extending $\widetilde{\lambda}$ such that (1.19)-(1.21) holds. The σ-algebra Σ is possibly complicated to describe but one has the following

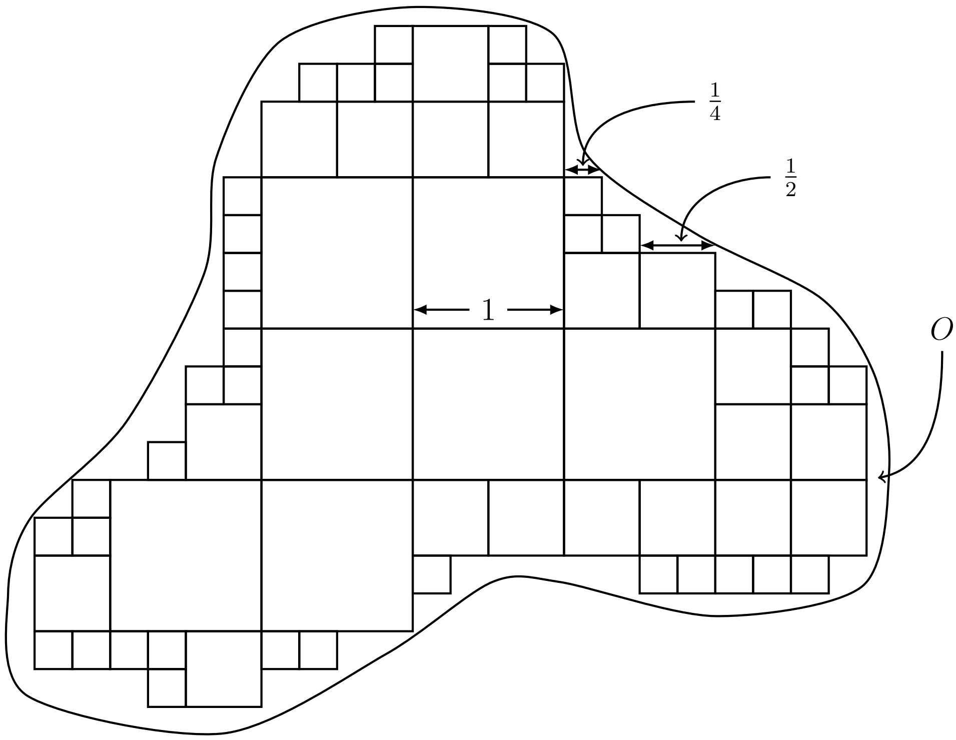
Lemma 1.17. *Let $\mathcal{B}(\mathbb{R}^n)$ denote the Borel σ-algebra on $\mathbb{R}^n$, see* (1.6)*. Then $\mathcal{B}(\mathbb{R}^n) \subset \Sigma$.*

Proof. Let $O \subset \mathbb{R}^n$ be an open set. We need to show $O \in \Sigma$, whence $\{O : O \subset \mathbb{R}^n \textit{ open}\} \subset \Sigma$ and therefore $\mathcal{B}(\mathbb{R}^n) = \sigma(\{O : O \subset \mathbb{R}^n \textit{ open}\}) \subset \sigma(\Sigma) = \Sigma$. In view of (1.20), $\sigma(\mathcal{A}) \subset \Sigma$, hence it suffices to argue that

$$O \in \sigma(\mathcal{A}). \tag{1.37}$$

We show that O can be written as countable union of disjoint half-open cubes (i.e. intervals of the form $[a, b)$), which implies (1.37) by (1.33). For $m \geq 0$, let $\mathcal{C}_m \subset \mathcal{A}$ consist of all sets of the form $[z, z + 2^{-m})$, $z \in 2^{-m}\mathbb{Z}^n$, where $z + 2^{-m}$ refers to the point

$(z_1 + 2^{-m}, \dots, z_n + 2^{-m})$ if $z = (z_1, \dots, z_n)$. For each $m \geq 0$, choose in the m-th step the cubes in $\mathcal{C}_m$ which are i) contained in O and ii) disjoint from the cubes selected in steps $0, \dots, m-1$ (this latter condition is absent for $m = 0$). The union of the cubes thereby obtained equals O and has the desired properties. $\square$



We then make the following

Definition 1.18. The **Lebesgue measure** $\lambda : \mathcal{B}(\mathbb{R}^n) \to [0, \infty]$ is the Hahn-Carathéodory extension of $\widetilde{\lambda}$ given by (1.34).

Remark 1.19. 1) (Restriction of a measure). Let $(X, \mathcal{F}, \mu)$ be a measure space, $A \in \mathcal{F}$. One checks that $\mathcal{F}|_A \overset{\text{def.}}{=} \{A \cap B : B \in \mathcal{F}\}$ is a σ-algebra on A, and that

$$\mu|_A(B) \overset{\text{def.}}{=} \mu(B), \text{ for } B \in \mathcal{F},\ B \subset A$$

is a measure on $(A, \mathcal{F}|_A)$. The measure $\mu|_A$ is called the **restriction of** μ **to** A.

2) Applying 1) to λ and $A = I$ an interval, one obtains the so-called **Lebesgue measure on** I. If $n = 1$, $I = [0, 1]$, the resulting measure $\lambda|_{[0,1]}$ is a probability measure (i.e. $\lambda|_{[0,1]}([0, 1]) = 1$) called the **uniform distribution on** $[0, 1]$. More generally, for $a < b$, the measure

$$\mathcal{B}(\mathbb{R})|_{[a,b]} = \mathcal{B}([a, b]) \to [0, 1], \quad A \mapsto \frac{1}{b-a}\lambda(A)$$

is called the **uniform distribution on** $[a, b]$.

We now investigate the question: which sets are measurable? To this effect we first collect useful approximation properties of measurable sets.

Proposition 1.20. *For all $A \in \mathcal{B}(\mathbb{R}^n)$,*

$$\lambda(A) = \inf_{G \supset A,\, G \text{ open}} \lambda(G) \tag{1.38}$$

Proof. The inequality "$\leq$" in (1.38) follows immediately from (1.9), which implies that $\lambda(A) \leq \lambda(G)$ for any $G \in \mathcal{B}(\mathbb{R}^n)$ with $G \supset A$.

We now argue that

$$\lambda(A) \geq \inf_{G \supset A,\, G \text{ open}} \lambda(G). \tag{1.39}$$

Let $\varepsilon > 0$. Recall that $\lambda(A) = \lambda^*(A)$ (see Theorem 1.13), where λ^* is the outer measure on $2^{\mathbb{R}^n}$ extending $\widetilde{\lambda}$ given by (1.34). In particular by (1.16) (with $\mathcal{K} = \mathcal{A}$ the algebra defined by (1.33)), there exist intervals I_j, $j \geq 1$ such that

$$A \subset \bigcup_{j=1}^{\infty} I_j \quad \text{and} \quad \sum_{j=1}^{\infty} \lambda(I_j) \overset{(1.21)}{=} \sum_{j=1}^{\infty} \widetilde{\lambda}(I_j) \overset{(1.16)}{\leq} \lambda^*(A) + \varepsilon = \lambda(A) + \varepsilon. \tag{1.40}$$

By (1.34) we further find open intervals $\widetilde{I}_j \supset I_j$ such that $\lambda(\widetilde{I}_j) \leq \lambda(I_j) + \varepsilon 2^{-j}$, for every $j \geq 1$. Set $G \overset{\text{def.}}{=} \bigcup_{j=1}^{\infty} \widetilde{I}_j$. Then G is open, $A \subset G$, and

$$\lambda(G) \overset{(1.13)}{\leq} \sum_{j=1}^{\infty} \lambda(\widetilde{I}_j) \leq \sum_{j=1}^{\infty} \lambda(I_j) + \varepsilon \overset{(1.40)}{\leq} \lambda(A) + 2\varepsilon.$$

Letting $\varepsilon \downarrow 0$, (1.39) follows. □

Inspection of the proof of Proposition 1.20 reveals that (1.38) is in fact true for any $A \subset \mathbb{R}^n$ upon replacing $\lambda(A)$ by $\lambda^*(A)$ on the left-hand-side of (1.38). The **regularity** of measurable sets is expressed in the following

Proposition 1.21. *If $A \in \mathcal{B}(\mathbb{R}^n)$, then for all $\varepsilon > 0$, there exists $G \supset A$ open s.t. $\lambda(G \backslash A) < \varepsilon$.*

In fact Proposition 1.21 is just an application (how?) of the following general approximation result for measures.

Proposition 1.22. *Let $\mathcal{A}$ be an algebra and μ a measure on $\sigma(\mathcal{A})$ which is σ-finite on $\mathcal{A}$ (i.e. it satisfies (1.29) with $\widetilde{\mu} = \mu$). Then for all $A \in \sigma(\mathcal{A})$ and $\varepsilon > 0$, there exist mutually disjoint sets $A_1, A_2, \dots \in \mathcal{A}$ such that $A \subset \bigcup_{n=1}^{\infty} A_n$ and $\mu(\bigcup_{n=1}^{\infty} A_n \backslash A) < \varepsilon$.*

Proof. See exercises. □

Remark 1.23. If $A \in \mathcal{B}(\mathbb{R}^n)$, then $A^c \in \mathcal{B}(\mathbb{R}^n)$ by (1.2). Applying Proposition 1.21 to A^c, we find an open set $\widetilde{G} \supset A^c$ such that $\lambda(\widetilde{G} \backslash A^c) < \varepsilon$. Set $F \overset{\text{def.}}{=} \widetilde{G}^c \subset A$. Then F is closed and

$$\lambda(A \backslash F) = \lambda(A \cap F^c) = \lambda(A \cap \widetilde{G}) = \lambda(\widetilde{G} \backslash A^c) < \varepsilon.$$

Overall we thus find that if $A \in \mathcal{B}(\mathbb{R}^n)$,

$$\forall \varepsilon > 0,\ \exists F, G,\ F \text{ closed},\ G \text{ open s.t. } F \subset A \subset G \text{ and } \lambda(G \backslash F) < \varepsilon. \tag{1.41}$$

Proposition 1.24 (translational invariance of λ)**.**

Let $\Phi_{x_0} : \mathbb{R}^n \to \mathbb{R}^n$, $\Phi_{x_0}(x) = x_0 + x$ *for* $x_0 \in \mathbb{R}^n$. *Then*

$$\lambda(\Phi_{x_0}(A)) = \lambda(A) \quad \text{for all } A \in \mathcal{B}(\mathbb{R}^n). \tag{1.42}$$

(Note $\Phi_{x_0}(A) \overset{\text{def.}}{=} \{\Phi_{x_0}(x) : x \in A\} \subset \mathbb{R}^n$*.)*

Proof. i) First suppose that $A = I$ is an interval (see Def. 1.15), say $I = (a, b)$ (the other cases are treated similarly). Then, abbreviating $\Phi = \Phi_{x_0}$, we have that $\Phi(A) = \Phi((a, b)) = (a + x_0, b + x_0)$ and (1.42) follows immediately from (1.34).

ii) Next, suppose that $A = G$ is open. Then, using the construction below (1.37), we find disjoint intervals $I_1, I_2, \ldots$ such that $G = \bigcup_{k=1}^\infty I_k$. It follows that $\Phi(G)$ is open and $\Phi(G) = \bigcup_{k=1}^\infty \Phi(I_k)$, with disjoint intervals $\Phi(I_k)$, $k \geq 1$. Hence

$$\lambda(\Phi(G)) \overset{(1.8)}{=} \sum_{k=1}^\infty \lambda(\Phi(I_k)) \overset{\text{case i)}}{=} \sum_{k=1}^\infty \lambda(I_k) \overset{(1.8)}{=} \lambda(G),$$

showing (1.42) in this case.

iii) Finally consider an arbitrary set $A \in \mathcal{B}(\mathbb{R}^n)$. Note that $A \subset G$, G open is equivalent to $\Phi(A) \subset \Phi(G)$, $\Phi(G)$ open and therefore

$$\lambda(\Phi(A)) \overset{(1.38)}{=} \inf_{\substack{A \subset G \\ G \text{ open}}} \lambda(\Phi(G)) \overset{\text{case ii)}}{=} \inf_{\substack{A \subset G \\ G \text{ open}}} \lambda(G) \overset{(1.38)}{=} \lambda(A).$$

□

We conclude this section by giving an example of a non-measurable set.

Proposition 1.25.

$$\mathcal{B}(\mathbb{R}) \neq 2^{\mathbb{R}}.$$

Proof. (Vitali). Define the following equivalence relation on $\mathbb{R}$: $x \sim y$ if and only if $x - y \in \mathbb{Q}$. We choose[1] a representative in $(0, 1]$ of each equivalent class $[x] = \{y : y \sim x\}$ for $\sim$. The (Vitali) set is defined as

$$V = \mathbb{R} \backslash \sim \qquad (\subset (0, 1]\ ) \tag{1.43}$$

[1] using the axiom of choice.

Let $A = \mathbb{Q} \cap (-1, 1]$. We then claim that

$$(0, 1] \subset \bigcup_{q \in A} (q + V) \subset [-1, 2], \tag{1.44}$$

which implies in particular that

$$\text{if } V \text{ is measurable then } 1 \le \lambda\Big(\bigcup_{q \in A} (q + V)\Big) \le 3. \tag{1.45}$$

The second inclusion in (1.44) is clear, for the first one simply note that, if $y \in (0, 1]$, then $y = q + [y]$ for some $q \in \mathbb{Q}$ (because $\sim$ is an equivalence relation) and in fact $q \in (-1, 1]$ since we chose $[y] \in (0, 1]$.

Now observe that

$$\text{if } q_1, q_2 \in \mathbb{Q},\ q_1 \neq q_2,\ \text{then } (q_1 + V) \cap (q_2 + V) = \emptyset. \tag{1.46}$$

Indeed, suppose not, and assume $[x] \in q_i + V$, $i = 1, 2$ (with $q_1 \neq q_2$). Then $[x] = [q_i + y_i]$, $i = 1, 2$ for some $y_1, y_2 \in (0, 1]$ i.e. $q_i + y_i - x = r_i \in \mathbb{Q}$ for $i = 1, 2$. It follows that $y_1 - y_2 \in \mathbb{Q}$, i.e. $y_1 \sim y_2$ whence $q_1 = q_2$. Thus (1.46) holds and implies together with (1.42) that

$$\begin{aligned} &\text{if } V \text{ is measurable, then} \\ &\lambda\Big(\bigcup_{q \in A} (q + V)\Big) \overset{(1.46),(1.8)}{=} \sum_{q \in A} \lambda(q + V) \overset{(1.42)}{=} \sum_{q \in A} \lambda(V) \in \{0, \infty\} \end{aligned} \tag{1.47}$$

(depending on whether $\lambda(V) = 0$ or $\lambda(V) > 0$). But (1.47) contradicts (1.45), hence V is not measurable. $\square$

Chapter 2

Integration

2.1 Measurable functions

Let $(X, \mathcal{A})$, $(Y, \mathcal{A}')$ be measurable spaces.

Definition 2.1. A function $f : X \to Y$ is called $\mathcal{A}$-$\mathcal{A}'$ **measurable** (or simply **measurable** when $\mathcal{A}$, $\mathcal{A}'$ are clear from the context) if

$$f^{-1}(A') \in \mathcal{A} \quad \text{for all } A' \in \mathcal{A}'. \tag{2.1}$$

(Notation: $f^{-1}(S) = \{x \in X : f(x) \in S\}$.)

Remark 2.2. 1) In practice, one checks measurability as follows. Let $\mathcal{E}' \subset \mathcal{A}'$ be a generating set, i.e. $\sigma(\mathcal{E}') = \mathcal{A}'$ (e.g. if $(Y, \mathcal{A}') = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, as is often the case, one could choose $\mathcal{E}' = \{U \subset \mathbb{R} : U \text{ open}\}$ or even $\mathcal{E}' = \{(a, b) : a < b,\ a, b \in \mathbb{R}\}$). Then

$$f \text{ is measurable} \quad \Longleftrightarrow \quad f^{-1}(A') \in \mathcal{A} \quad \text{for all } A' \in \mathcal{E}'. \tag{2.2}$$

Indeed, "$\Rightarrow$" is immediate. For "$\Leftarrow$", note that

$$\mathcal{G} \overset{\text{def.}}{=} \{A' \subset Y' : f^{-1}(A') \in \mathcal{A}\}.$$

is a σ-algebra (check this!). By (2.2) $\mathcal{G} \supset \mathcal{E}'$, therefore $\sigma(\mathcal{G}) \supset \sigma(\mathcal{E}')$, but $\sigma(\mathcal{G}) = \mathcal{G}$ since $\mathcal{G}$ is a σ-algebra while $\sigma(\mathcal{E}') = \mathcal{A}'$. All in all, $\mathcal{G} \supset \mathcal{A}'$ which gives (2.1).

2) If $f : X \to [-\infty, \infty]$, then we require for f to be measurable that (2.1) holds with $Y = \mathbb{R}$, $\mathcal{A}' = \mathcal{B}(\mathbb{R})$, and in addition that $f^{-1}(\{+\infty\}) \in \mathcal{A}$, $f^{-1}(\{-\infty\}) \in \mathcal{A}$.

Example 2.3. 1) Let $X = Y$, $\mathcal{A} = \mathcal{A}'$. Then the function $\mathrm{id} : X \to X$, $\mathrm{id}(x) = x$ is measurable.

2) Let $X = Y = \mathbb{R}$, $\mathcal{A} = \mathcal{A}' = \mathcal{B}(\mathbb{R})$. If $f : \mathbb{R} \to \mathbb{R}$ is continuous, then f is measurable. Indeed, for $U \subset \mathbb{R}$ open, $f^{-1}(U)$ is open by continuity, hence $f^{-1}(U) \in \mathcal{B}(\mathbb{R})$, and measurability follows using (2.2).

3) (Indicator function) Let $(X, \mathcal{A})$ be a measurable space, $A \subset X$. The indicator function of A is defined as

$$1_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases} \tag{2.3}$$

(another common notation is $\chi_A(x)$). Then 1_A is ($\mathcal{A}$-$\mathcal{B}(\mathbb{R})$) measurable if and only if $A \in \mathcal{A}$. Indeed

$$(\sigma(f) =)\{f^{-1}(S) : S \in \mathcal{B}\,(\mathbb{R})\} = \{\emptyset, A, A^c, X\}\ (= \sigma(\{A\})\,)$$

For instance, the function $1_{\mathbb{Q}}$ from (0.4) is $\mathcal{B}(\mathbb{R})$-$\mathcal{B}(\mathbb{R})$ measurable, since $\mathbb{Q} = \bigcup_{q \in \mathbb{Q}} \{q\} \in \mathcal{B}(\mathbb{R})$.

4) If $(X, \mathcal{A})$, $(Y, \mathcal{A}')$, $(Z, \mathcal{A}'')$ are measurable spaces, $f : X \to Y$ is $\mathcal{A}$-$\mathcal{A}'$ measurable, $g : Y \to Z$ is $\mathcal{A}'$-$\mathcal{A}''$ measurable, then $h = g \circ f$ is $\mathcal{A}$-$\mathcal{A}''$ measurable. For, if $A \in \mathcal{A}''$, then

$$h^{-1}(A) = f^{-1}(\underbrace{g^{-1}(A)}_{\in \mathcal{A}'}) \in \mathcal{A}.$$

5) As explained in Remark 1.6, 2), probability theory studies measure spaces, traditionally denoted by $(\Omega, \mathcal{A}, P)$ (rather than $(X, \mathcal{A}, \mu)$) with $P(\Omega) = 1$ (here P is a measure on the measurable space $(\Omega, \mathcal{A})$, cf. the discussion around (0.8)). In this context, an $\mathcal{A}$-$\mathcal{B}(\mathbb{R})$ measurable function $X : \Omega \to \mathbb{R}$ is referred to as a (real-valued) **random variable**.

Theorem 2.4. *Let $(X, \mathcal{A})$ be a measurable space and $f, g : X \to \mathbb{R}$ be $\mathcal{A}$-$\mathcal{B}(\mathbb{R})$ measurable functions. Then so are*

$$f + g,\ f \cdot g,\ |f|,\ f \wedge g = \min\{f, g\},\ f \vee g = \max\{f, g\},\ g^{-1} \tag{2.4}$$

(the latter if $g(x) \neq 0$).

If $f_k : X \to \mathbb{R}$, $k \geq 1$ are measurable, then so are

$$\inf_{k \geq 1} f_k,\ \sup_{k \geq 1} f_k,\ \limsup_{k \to \infty} f_k,\ \liminf_{k \to \infty} f_k \tag{2.5}$$

(Note: the first of these is the map $X \to \mathbb{R}$, $x \mapsto \inf_{k \geq 1} f_k(x)$; similarly for the others.)

Proof. (2.4): The measurability of $f + g$ follows from the representation

$$(f + g)^{-1}((-\infty, a)) = \bigcup_{\substack{r,s \in \mathbb{Q} \\ r+s<a}} f^{-1}((-\infty, r)) \cap g^{-1}((-\infty, s)). \tag{2.6}$$

The right-hand side of (2.6) is plainly in $\mathcal{A}$ since f, g are measurable and $\mathcal{A}$ is a σ-algebra. The measurability of $f+g$ then follows from (2.2) since $\mathcal{E}' \stackrel{\text{def.}}{=} \{(-\infty, a) : a \in \mathbb{R}\}$ generates $\mathcal{B}(\mathbb{R})$, i.e. $\sigma(\mathcal{E}') = \mathcal{B}(\mathbb{R})$.

The functions $s \mapsto s^2$, $s \mapsto -s$, $s \mapsto s/2$ are continuous on $\mathbb{R}$, hence measurable by Example 2.3, 2). The measurability of $f \cdot g$ then follows from the representation

$$f \cdot g = \frac{1}{2}[(f+g)^2 - f^2 - g^2],$$

Example 2.3, 4) and the measurability of sums already shown.
Let

$$s^+ = \max\{s, 0\}, \quad s^- = \max\{-s, 0\} \quad = (-s)^+. \tag{2.7}$$

Then by continuity of the maps $s \mapsto s^+$, $s \mapsto s^-$ on $\mathbb{R}$ and the representations

$$\begin{aligned} |f| &= f^+ + f^- \\ f \wedge g &= f - (g-f)^- \\ f \vee g &= f + (g-f)^+, \end{aligned} \tag{2.8}$$

the measurability of these functions follows. Finally, regarding g^{-1} observe that

$$\left(\frac{1}{g}\right)^{-1}((-\infty, a)) = \begin{cases} g^{-1}((\frac{1}{a}, 0)), & a < 0 \\ g^{-1}((-\infty, 0)), & a = 0 \\ g^{-1}((-\infty, 0) \cup (\frac{1}{a}, \infty)), & a > 0. \end{cases}$$

(2.5): One notes that

$$\left(\inf_{k \geq 1} f_k\right)^{-1}((-\infty, a)) = \bigcup_{k=1}^{\infty} f_k^{-1}((-\infty, a)) \in \mathcal{A}$$

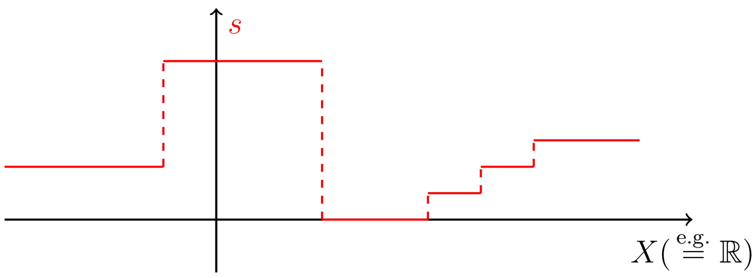
which yields measurability of $\inf_{k \geq 1} f_k$. For the suprema one uses that $\sup_{k \geq 1} f_k = -\inf_{k \geq 1}(-f_k)$. Finally observe that

$$\begin{aligned} \liminf_{k \to \infty} f_k &= \sup_{l \geq 1} \big(\inf_{k \geq l} f_k\big) \\ \limsup_{k \to \infty} f_k &= \inf_{l \geq 1} \big(\sup_{k \geq l} f_k\big), \end{aligned}$$

from which measurability of $\liminf / \limsup f_k$ follows. □

The following straightforward approximation result for measurable functions will be extremely useful.

Definition 2.5. Let X be a set. A function $s : X \to \mathbb{R}$ is called a **simple function** (or **step function**) if it takes on finitely many values, i.e. if $im(s) = \{y \in \mathbb{R} : y = s(x)$ *for some* $x \in X\}$ is a finite set.



Writing $im(s) = \{\alpha_1, \cdots, \alpha_l\}$, $\alpha_i \in \mathbb{R}$, $\alpha_i \neq \alpha_j$ for $i \neq j$, and defining $A_i \overset{\text{def.}}{=} s^{-1}(\{\alpha_i\})$, $1 \leq i \leq l$, one obtains that the sets $A_1, \ldots, A_l$ form a partition of X, i.e. $A_i \cap A_j = \emptyset$, and $X = \bigcup_{i=1}^l A_i$, and s admits the representation

$$(2.9) \qquad s(x) = \sum_{i=1}^{l} \alpha_i \, 1_{A_i}(x), \quad \text{for all } x \in X.$$

In particular, if $\mathcal{A}$ is a σ-algebra over X, using Example 2.3, 3) and (2.4) it follows from (2.9) that s is $(\mathcal{A}$-$\mathcal{B}(\mathbb{R}))$ measurable if and only if $A_i \in \mathcal{A}$ for all $i \leq i \leq l$.

Theorem 2.6 (Approximation by Simple Functions)**.**

Let $(X, \mathcal{A})$ be a measurable space and let $f : X \to [0, \infty]$ be a function. Then f is measurable if and only if there exists a sequence of measurable simple functions $s_n : X \to [0, \infty)$ such that

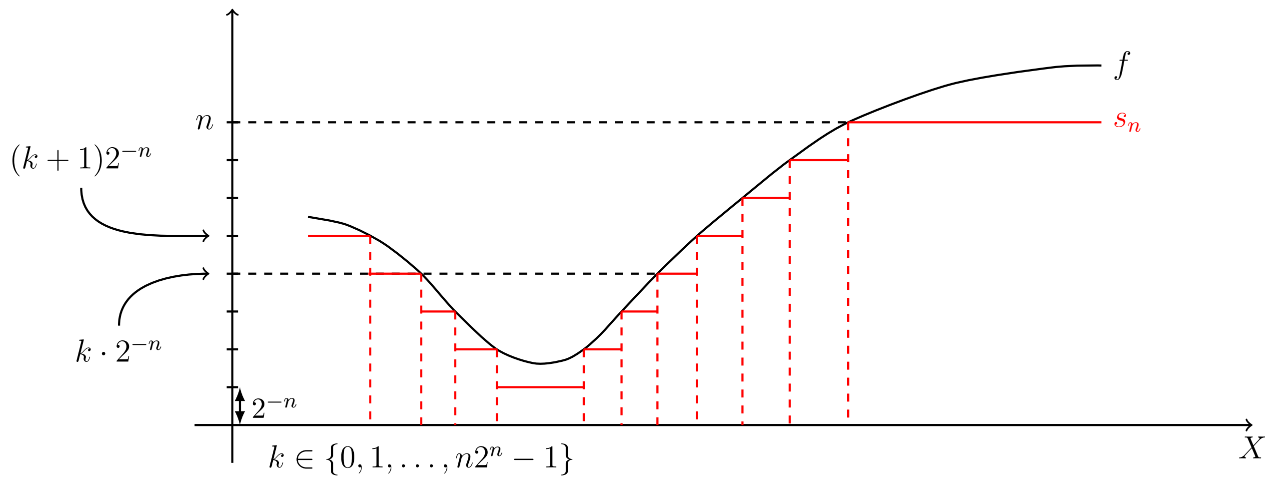
$$(2.10) \qquad 0 \leq s_1(x) \leq s_2(x) \leq \cdots \leq f(x), \quad f(x) = \lim_{n\to\infty} s_n(x), \quad \text{for all } x \in X.$$

Proof. If (s_n) satisfies (2.10), then $f = \lim_{n\to\infty} s_n$ is measurable by (2.5). Conversely, suppose that f is measurable. For $n \geq 1$ define $\varphi_n : [0, \infty] \to \mathbb{R}$ by

$$(2.11) \qquad \varphi_n(t) = \begin{cases} k \cdot 2^{-n}, & \text{if } k2^{-n} \leq t < (k+1)2^{-n},\ k = 0, 1, \ldots, n2^n - 1. \\ n, & \text{if } t \geq n. \end{cases}$$

The functions (φ_n) are measurable, increasing (i.e. $\varphi_n(t) \leq \varphi_{n+1}(t)$ for all t) and $t - 2^{-n} \leq \varphi_n(t) \leq t$ for all $t < n$. Hence $\lim_{n\to\infty} \varphi_n(t) = t$ for all $t \in [0, \infty]$. The functions $s_n \overset{\text{def.}}{=} \varphi_n \circ f$ thus satisfy (2.10). $\square$

The simple function s_n approximating f from below in the proof Theorem 2.6:



2.2 Measurability vs. Continuity

We now investigate the relationship between measurability and continuity.

Definition 2.7. Let $(X, \mathcal{A}, \mu)$ be measure space. The property $P(x)$, $x \in X$ is said to hold μ almost everywhere (abbreviated μ-a.e.) if

$$\mu(\{x \in X : P(x) \text{ does not hold}\}) = 0.$$

(assuming $\{x \in X : P(x) \textit{ holds}\} \in \mathcal{A}$).

For $\Omega \in \mathcal{A}$ we will also say that $P(x)$ holds μ-a.e. on Ω if Definition 2.7 holds on the measure space $(\Omega, \mathcal{A}|_\Omega, \mu|_\Omega)$, i.e. if $\mu(\{x \in \Omega : P(x) \text{ does not hold}\}) = 0$.

Example 2.8. 1) Let $f = 1_\mathbb{Q}$, λ be Lebesgue measure on $\mathbb{R}$. Then $f = 0$ λ-a.e. Indeed,

$$\lambda(\{x \in \mathbb{R} : f(x) \neq 0\}) = \lambda(\mathbb{Q}) = 0.$$

2) Let $f_n : X \to \mathbb{R}$, $n \geq 1$, be measurable functions. The sequence f_n **converges μ-a.e.** if $\mu(\{x \in X : \lim_{n\to\infty} f_n(x) \text{ does not exist}\}) = 0$. The set in question is indeed measurable (see exercises).

Throughout the remainder of this section consider $\Omega \subset \mathbb{R}^n$ measurable (i.e. $\Omega \in \mathcal{B}(\mathbb{R}^n)$) with

$$\lambda(\Omega) < \infty, \tag{2.12}$$

where λ denotes the Lebesgue measure on $\mathbb{R}^n$.

Theorem 2.9 (Egorov)**.**

Let $f_k : \Omega \to \mathbb{R}$, $k \geq 1$, $f : \Omega \to \mathbb{R}$ be measurable. Suppose that $f_k(x) \to f(x)$ as $k \to \infty$ λ-a.e. on Ω. Then for all $\delta > 0$, there exists $F \subset \Omega$ compact such that

$$\lambda(\Omega \backslash F) < \delta \quad \text{and} \quad \sup_{x \in F} |f_k(x) - f(x)| \to 0 \quad \text{as } k \to \infty, \tag{2.13}$$

(i.e. f_k converges to f uniformly on F).

Proof. Let $\delta > 0$. For $i, j \geq 1$ define

$$C_{i,j} = \bigcup_{k=j}^{\infty} \{x \in \Omega : |f_k(x) - f(x)| > 2^{-i}\}. \tag{2.14}$$

The set $C_{i,j}$ is measurable since f, f_k are and $C_{i,j+1} \subset C_{i,j}$. Using (2.12) and (1.12), it follows that

$$\lim_{j \to \infty} \lambda(C_{i,j}) = \lambda\Big(\bigcap_{j=1}^{\infty} C_{i,j}\Big) \overset{\text{monot.}}{\leq} \lambda(\{x \in \Omega : \lim_{k \to \infty} f_k(x) = f(x)\}^c) = 0$$

for all $i \geq 1$. Hence, there exists $J(i) \geq 1$ such that $\lambda(C_{i,J(i)}) < \delta \cdot 2^{-i-1}$, and thus, setting $A = \Omega \backslash (\bigcup_{i=1}^{\infty} C_{i,J(i)})$,

$$\lambda(\Omega \backslash A) \overset{(1.13)}{\leq} \sum_{i=1}^{\infty} \lambda(C_{i,J(i)}) < \frac{\delta}{2}. \tag{2.15}$$

We can now find $F \subset A$ compact such that $\lambda(A \backslash F) < \delta/2$: to see this, first use (1.13) and the assumption (2.12) to find $L = L(\delta)$ large enough such that $\lambda(A \cap [-L, L]^n) \geq \lambda(A) - \frac{\delta}{4}$. Then use (1.41) with $A \cap [-L, L]^n$ in place of A and choose $\varepsilon = \frac{\delta}{4}$. This yields the desired set F, which is closed by (1.41) and bounded since $F \subset [-L, L]^n$.

Together with (2.15), we then obtain that $\lambda(\Omega \backslash F) < \delta$. Finally, observe that, as $F \subset A \subset C_{i,J(i)}^c$ for all i, in view of (2.14), one has

$$|f_k(x) - f(x)| < 2^{-i}, \quad \text{for all } k \geq J(i),\ i \geq 1, \text{ and } x \in F,$$

from which the asserted uniform convergence in (2.13) follows. □

Example 2.10. One cannot choose $\delta = 0$ in general. Consider $\Omega = [0, 1]$, $f_k(x) = x^k$ then $f_k(x) \to 1_{\{1\}}(x)$ for all $x \in [0, 1]$ as $K \to \infty$ but f_k does not converge uniformly on $[0, 1]$. The conclusions of Theorem 2.9 hold e.g. with $F = [0, 1 - \delta]$.

Theorem 2.9 asserts that every convergent sequence of measurable functions (defined on a subset of $\mathbb{R}^n$ having finite Lebesgue measure) is "nearly" uniformly convergent on that set. The next result asserts that any measurable function is "nearly" continuous.

Theorem 2.11 (Lusin, $\Omega \subset \mathbb{R}^n$ satisfies (2.12)).

If $f : \Omega \to \mathbb{R}$ is measurable, then for all $\delta > 0$, there exists $F \subset \Omega$ compact such that

$$\lambda(\Omega \backslash F) < \delta \quad \text{and} \quad f|_F : F \to \mathbb{R} \text{ is continuous.} \tag{2.16}$$

Proof. 1) We first suppose that $f = s$ is a simple function. Thus, using the representation (2.9), where $A_1, \dots, A_l$ are disjoint sets with $\bigcup_{i=1}^l A_i = \Omega$, we find by (1.41) compact subsets $F_i \subset A_i$ such that

$$\lambda(A_i \backslash F_i) < \delta 2^{-i}, \quad \text{for all } 1 \leq i \leq l. \tag{2.17}$$

The sets F_i are disjoint, and f is continuous on $F \overset{\text{def.}}{=} \bigcup_{i=1}^l F_i$, since f is constant on each set F_i. Moreover, F is compact and

$$\lambda(\Omega \backslash F) = \lambda\Big(\bigcup_{i=1}^l (A_i \backslash F_i)\Big) = \sum_{i=1}^l \lambda(A_i \backslash F_i) \overset{(2.17)}{<} \delta$$

2) We now consider the general case for f. Write $f = f^+ - f^-$, see (2.7) for notation. The functions $f^\pm$ have values in $[0, \infty)$. Let $(s_n^\pm)$ be the sequences of simple functions approximating $f^\pm$ as in Theorem 2.6. The function $s_n \overset{\text{def.}}{=} s_n^+ - s_n^-$ is a simple function and by (2.10),

$$\lim_{n \to \infty} s_n(x) = f(x), \quad \text{for all } x \in \Omega. \tag{2.18}$$

Moreover, applying 1) individually for each s_n, we find compact sets F_n such that

$$\lambda(\Omega \backslash F_n) < \delta 2^{-n-1} \quad \text{and} \quad s_n|_{F_n} : F_n \to \mathbb{R} \text{ continuous, for all } n \geq 1. \tag{2.19}$$

In view of (2.13), Theorem 2.9 applies with $f_k \overset{\text{def.}}{=} s_k$ and yields a compact set $F_0 \subset \Omega$ such that

$$\lambda(\Omega \backslash F_0) \leq \frac{\delta}{2}, \quad \sup_{x \in F_0} |s_n(x) - f(x)| \overset{n \to \infty}{\longrightarrow} 0. \tag{2.20}$$

Now define $F = \bigcap_{n=0}^\infty F_n$. The set F is compact, and the bounds in (2.19), (2.20) readily imply that $\lambda(\Omega \backslash F) < \delta$. Moreover, by (2.19), one knows that each function s_n is continuous on $F (\subset F_n)$ and by (2.20) that s_n converges to f uniformly on $F (\subset F_0)$. It follows* that f is continuous on F, as desired. (*recall that the limit of a uniformly convergent sequence of continuous functions is continuous). $\square$

Remark 2.12. Let $\Omega = [0,1]$, $f = 1_{\mathbb{Q} \cap [0,1]}$. The function $f : [0,1] \to \mathbb{R}$ is nowhere continuous, but $f|_{F'}$ is where $F' = [0,1] \backslash \mathbb{Q}$ (indeed $f|_{F'} = 0$). To obtain for a given $\delta > 0$ a set F satisfying the conclusions of Theorem 2.11, let $\{q_0, q_1, \dots\}$ be an enumeration of $\mathbb{Q} \cap [0,1]$ and define $F \subset F'$ as

$$F = [0,1] \setminus \bigcup_{n \geq 0} (q_n - \delta 2^{-(n+1)}, q_n + \delta 2^{-(n+1)}),$$

which is closed. This example also shows that one cannot in general find F compact in Theorem 2.11 if one asks that $\delta = 0$.

2.3 The Integral

We now construct the integral on the measure space $(X, \mathcal{A}, \mu)$. In particular, as will turn out, in case $X = \mathbb{R}$, $\mathcal{A} = \mathcal{B}(\mathbb{R})$, $\mu = \lambda$, the resulting Lebesgue integral will extend the Riemann(-Darboux) integral, cf. (0.1)-(0.3).

We will introduce the integral of suitable **measurable** functions f in three steps:

Step 1: simple function

Step 2: non-negative function (by approximation)

Step 3: integrable real-valued function.

Step 1: We introduce the cone

$$S^+ = \{s : X \to [0, \infty) : s \text{ simple, measurable}\} \tag{2.21}$$

(see Definition 2.5). For $s \in S^+$, we refer to $s = \sum_{i=1}^{l} \alpha_i 1_{A_i}$ with $\alpha_i \in (0, \infty)$, $A_i \in \mathcal{A}$ disjoint, as a representation of s (not unique).

Definition 2.13. For $s \in S^+$ with representation $s = \sum_{i=1}^{l} \alpha_i 1_{A_i}$, the integral of s with respect to μ is

$$\mu(s) \overset{\text{def.}}{=} \int_X s \, \mathrm{d}\mu \overset{\text{def.}}{=} \int s \, \mathrm{d}\mu \overset{\text{def.}}{=} \sum_{i=1}^{l} \alpha_i \mu(A_i) \quad (\in [0, \infty]). \tag{2.22}$$

Remark 2.14. (2.22) is well-defined, i.e. (2.22) does not depend on the choice of representation for s. Indeed if $s = \sum_{i=1}^{l} \alpha_i 1_{A_i} = \sum_{i=1}^{l'} \beta_i 1_{B_i}$ with disjoint sets $A_1, \dots, A_l$ and $B_1, \dots, B_{l'}$, we may assume that $\bigcup_{i=1}^{l} A_i = \bigcup_{i=1}^{l'} B_i = X$, else we add a term $\alpha_{l+1} 1_{A_{l+1}}$ with $\alpha_{l+1} = 0$, $A_{l+1} = X \backslash (\bigcup_{i=1}^{l} A_i)$ to the first representation (and similarly for the second one), which does not contribute in (2.22). Now, by additivity,

$$\sum_{i=1}^{l} \alpha_i \mu(A_i) \overset{(1.10)}{=} \sum_{i=1}^{l} \sum_{j=1}^{l'} \alpha_i \mu(A_i \cap B_j) \underset{\substack{\uparrow \\ s(x) = \alpha_i = \beta_j, \ x \in A_i \cap B_j}}{=} \sum_{i=1}^{l} \sum_{j=1}^{l'} \beta_i \mu(A_i \cap B_j) \overset{(1.10)}{=} \sum_{j=1}^{l'} \beta_i \mu(B_j).$$

Lemma 2.15 $(f, g \in S^+)$**.**

$$\int (\alpha f + \beta g) \, \mathrm{d}\mu = \alpha \int f \, \mathrm{d}\mu + \beta \int g \, \mathrm{d}\mu, \quad \text{for } \alpha, \beta \in [0, \infty). \tag{2.23}$$

$$\int f \, \mathrm{d}\mu \le \int g \, \mathrm{d}\mu, \quad \text{if } f \le g \text{ (i.e. } f(x) \le g(x) \text{ for all } x \in X). \tag{2.24}$$

(Note that $(\alpha f + \beta g) \in S^+$.)

Proof. For (2.23), it is enough to show separately $\int(\alpha f)\,\mathrm{d}\mu = \alpha\int f\,\mathrm{d}\mu$, which is immediate from (2.22), and $\int(f+g)\,\mathrm{d}\mu = \int f\,\mathrm{d}\mu + \int g\,\mathrm{d}\mu$. For the latter, assume $f = \sum_{i=1}^{l}\alpha_i 1_{A_i}$, $g = \sum_{i=1}^{l'}\beta_i 1_{B_i}$ with $\bigcup_{i=1}^{l} A_i = \bigcup_{i=1}^{l'} B_i = X$. Then

$$f + g = \sum_{i=1}^{l}\sum_{j=1}^{l'}(\alpha_i + \beta_j)1_{A_i\cap B_j}, \tag{2.25}$$

hence

$$\begin{aligned}\int (f+g)\,\mathrm{d}\mu &\overset{(2.22)}{\underset{(2.25)}{=}} \sum_{i=1}^{l}\sum_{j=1}^{l'}(\alpha_i+\beta_j)\mu(A_i\cap B_j)\\ &= \sum_{i=1}^{l}\alpha_i\sum_{j=1}^{l'}\mu(A_i\cap B_j) + \sum_{i=1}^{l}\sum_{j=1}^{l'}\beta_i\mu(A_i\cap B_j)\\ &\overset{(1.10)}{=} \sum_{i=1}^{l}\alpha_i\mu(A_i) + \sum_{j=1}^{l'}\beta_i\mu(B_i) \overset{(2.22)}{=} \int f\,\mathrm{d}\mu + \int g\,\mathrm{d}\mu.\end{aligned}$$

The monotonicity property (2.24) is left as an exercise. □

Step 2: integral of a non-negative measurable function.

Definition 2.16. For $f : X \to [0,\infty]$ measurable, the integral of f with respect to μ is defined as

$$\int f\,\mathrm{d}\mu = \sup\left\{\int g\,\mathrm{d}\mu : g\in S^+,\ g\le f\right\}. \tag{2.26}$$

Note: (2.26) is consistent with (2.22) due to (2.24), i.e. if $f\in S^+$, then

$$\sup\left\{\int g\,\mathrm{d}\mu : g\in S^+,\ g\le f\right\} = \int f\,\mathrm{d}\mu.$$

Theorem 2.17.

(i) For $f, g : X\to[0,\infty]$ measurable, (2.23) *and* (2.24) *still hold.*

*(ii) (**Monotone convergence**) If $f_n \nearrow f$ (i.e. $f_n(x)\le f_{n+1}(x)$, for all $x\in X$ and $n\ge 1$ and $\lim_{n\to\infty} f_n(x) = f(x)$ for all $x\in X$) where $f_n : X\to[0,\infty]$ measurable, then $\lim_{n\to\infty}\int f_n\,\mathrm{d}\mu = \int f\,\mathrm{d}\mu$.*

PROOF. (2.24) is immediate by (2.26) and the fact that (2.24) holds for $g \in S^+$ by Lemma 2.15. Suppose we knew (ii). To deduce (2.23), let $(f_n), (g_n)$ be sequences in S^+ such that $f_n \nearrow f$, $g_n \nearrow g$ (using Theorem 2.6). Then $\alpha f_n + \beta g_n \in S^+$ and $\alpha f_n + \beta g_n \nearrow \alpha f + \beta g$, hence

$$\begin{aligned}\int (\alpha f + \beta g)\,\mathrm{d}\mu &\overset{\text{(ii)}}{=} \lim_{n\to\infty} \int (\alpha f_n + \beta g_n)\,\mathrm{d}\mu \\ &\overset{(2.24)}{=} \lim_{n\to\infty} \left(\alpha \int f_n\,\mathrm{d}\mu + \beta \int g_n\,\mathrm{d}\mu\right) \overset{\text{(ii)}\times 2}{=} \alpha \int f\,\mathrm{d}\mu + \beta \int g\,\mathrm{d}\mu.\end{aligned}$$

It remains to show (ii). First note that $\lim_{n\to\infty} \int f_n\,\mathrm{d}\mu$ exists (on $[0,\infty]$) since the limit is monotone by (2.24). Still by (2.24),

$$\lim_{n\to\infty} \int f_n\,\mathrm{d}\mu \le \int f\,\mathrm{d}\mu.$$

We now show the reverse inequality. By (2.26), it's enough to argue that

$$\lim_{n\to\infty} \int f_n\,\mathrm{d}\mu \ge \int g\,\mathrm{d}\mu, \quad \text{for all } g \in S^+ \text{ with } g \le f. \tag{2.27}$$

Let $g = \sum_{i=1}^l \alpha_i 1_{A_i}$, $A_i \in \mathcal{A}$, $1 \le i \le l$ disjoint. For $\varepsilon > 0$, define the set

$$G_n^\varepsilon = \{x \in X : f_n(x) \ge (1-\varepsilon)g(x)\} \overset{\text{abbrev.}}{=} \{f_n \ge (1-\varepsilon)g\} \quad (\in \mathcal{A}).$$

Since $f_n \nearrow f$, $G_n^\varepsilon \subset G_{n+1}^\varepsilon$ and $\bigcup_{n=0}^\infty G_n^\varepsilon = X$, thus

$$\begin{aligned}\int f_n\,\mathrm{d}\mu &\overset{\text{(i)}}{\ge} \int (f_n 1_{G_n^\varepsilon})\,\mathrm{d}\mu \overset{\text{(i)}}{\ge} \int (1-\varepsilon) g\, 1_{G_n^\varepsilon}\,\mathrm{d}\mu \underset{1_A\cdot 1_B = 1_{A\cap B}}{\overset{(2.23)}{=}} (1-\varepsilon)\int \left(\sum_{i=1}^l \alpha_i 1_{A_i\cap G_n^\varepsilon}\right)\mathrm{d}\mu \\ &\overset{(2.22)}{=} (1-\varepsilon)\sum_{i=1}^l \alpha_i \mu(A_i \cap G_n^\varepsilon) \underset{n\to\infty}{\overset{(1.11)}{\longrightarrow}} (1-\varepsilon)\sum_{i=1}^l \alpha_i\mu(A_i) \overset{(2.22)}{=} (1-\varepsilon)\int g\,\mathrm{d}\mu.\end{aligned}$$

Letting $\varepsilon \downarrow 0$ gives (2.27). $\square$

As a consequence of Definition 2.16, one can further show (see exercises):
for $f : X \to [0,\infty]$ measurable,

(iii) $f = 0$ μ-a.e. $\iff \int f\,\mathrm{d}\mu = 0$.

(iv) if $\int f\,\mathrm{d}\mu < \infty$, then $f < \infty$ μ-a.e.

For instance, (iii) can be used to strengthen Theorem 2.17, (i), in that (2.24) holds already if $f \le g$ μ-a.e.. Indeed, let $\widetilde{f} = f \cdot 1_{\{f\le g\}}$. Then $\widetilde{f} = f$ μ-a.e. and $\widetilde{f} \le g$. Hence $\int f\,\mathrm{d}\mu \overset{\text{(iii)}}{=} \int \widetilde{f}\,\mathrm{d}\mu \overset{(2.24)}{\le} \int g\,\mathrm{d}\mu$.

Step 3: real-valued functions.

Definition 2.18. A measurable function $f : X \to [-\infty, \infty]$ is called **μ-integrable** if $\int |f| \, \mathrm{d}\mu < \infty$. We write

(2.28)
$$\mathcal{L}^1(\mu) \stackrel{\text{def.}}{=} \mathcal{L}^1(X, \mathcal{A}, \mu) \stackrel{\text{def.}}{=} \left\{ f : X \to [-\infty, \infty] : f \textit{ measurable and } \int |f| \, \mathrm{d}\mu < \infty \right\}.$$

For $f \in \mathcal{L}^1(\mu)$, the integral of f with respect to μ is defined as

$$\int f \, \mathrm{d}\mu \stackrel{\text{def.}}{=} \int f^+ \, \mathrm{d}\mu - \int f^- \, \mathrm{d}\mu. \qquad \text{(cf. (2.7) for notation)} \tag{2.29}$$

For $A \in \mathcal{A}$, we set

$$\int_A f \, \mathrm{d}\mu \stackrel{\text{def.}}{=} \int (f \, 1_A) \, \mathrm{d}\mu$$

(whence $\int f \, \mathrm{d}\mu = \int_X f \, \mathrm{d}\mu$).

Remark 2.19. (2.29) is well-defined, both $\int f^\pm \, \mathrm{d}\mu$ are integrals of non-negative functions (given by Definition 2.16) and $\int f^\pm \, \mathrm{d}\mu \stackrel{(2.8)}{\leq} \int |f| \, \mathrm{d}\mu < \infty$ if $f \in \mathcal{L}^1(\mu)$ (so one doesn't have "$\infty - \infty$"in (2.29)). Slightly more generally, one can also define $\int f \, \mathrm{d}\mu$ as in (2.29), possibly having value $\pm\infty$, as long as *at least one of* $\int f^\pm \, \mathrm{d}\mu < \infty$.

Theorem 2.20 (Properties of the Integral)**.** *Let $f, g \in \mathcal{L}^1(\mu)$.*

$$f \leq g \ \mu\text{-a.e.} \ \Rightarrow \int f \, \mathrm{d}\mu \leq \int g \, \mathrm{d}\mu. \qquad \text{(Monotonicity)} \tag{2.30}$$

In particular, if $f = g$ μ-a.e. then $\int f \, \mathrm{d}\mu = \int g \, \mathrm{d}\mu$.

$$\left| \int f \, \mathrm{d}\mu \right| \leq \int |f| \, \mathrm{d}\mu. \qquad \text{(Triangle inequality)} \tag{2.31}$$

If $\alpha, \beta \in \mathbb{R}$, then $\alpha f + \beta g \in \mathcal{L}^1(\mu)$ and

$$\int (\alpha f + \beta g) \, \mathrm{d}\mu = \alpha \int f \, \mathrm{d}\mu + \beta \int g \, \mathrm{d}\mu. \qquad \text{(Linearity)} \tag{2.32}$$

Proof. (2.30): $f^+ \leq g^+$ μ-a.e. and $f^- \geq g^-$ μ-a.e.. Hence by Theorem 2.17, (i) and the remark below (iv), $\int f^+ \, \mathrm{d}\mu \leq \int g^+ \, \mathrm{d}\mu$ and $\int f^- \, \mathrm{d}\mu \geq \int g^- \, \mathrm{d}\mu$. It follows that

$$\int f \, \mathrm{d}\mu \stackrel{(2.29)}{=} \int f^+ \, \mathrm{d}\mu - \int f^- \, \mathrm{d}\mu \leq \int g^+ \, \mathrm{d}\mu - \int g^- \, \mathrm{d}\mu \stackrel{(2.29)}{=} \int g \, \mathrm{d}\mu.$$

(2.31):

$$\left|\int f \,\mathrm{d}\mu\right| \overset{(2.29)}{=} \left|\int f^+ \,\mathrm{d}\mu - \int f^- \,\mathrm{d}\mu\right| \overset{\substack{\text{tr.inequ.}\\ \text{on } \mathbb{R}}}{\leq} \int f^+ \,\mathrm{d}\mu + \int f^- \,\mathrm{d}\mu$$
$$\overset{\text{Thm. 2.17,(i)}}{=} \int (f^+ + f^-) \,\mathrm{d}\mu \overset{(2.28)}{=} \int f \,\mathrm{d}\mu.$$

(2.32): One has $|\alpha f + \beta g| \leq |\alpha||f| + |\beta||g|$ and the function $|\alpha||f| + |\beta||g| \in \mathcal{L}^1(\mu)$ since f and g are and by Theorem 2.17 (i) (i.e. (2.23) for non-negative functions). It then follows that $\int |\alpha f + \beta g| \,\mathrm{d}\mu < \infty$ using (2.24) for non-negative functions. To check linearity it suffices to verify separately (a) $\int (f+g) \,\mathrm{d}\mu = \int f \,\mathrm{d}\mu + \int g \,\mathrm{d}\mu$; (b) $\int \alpha f \,\mathrm{d}\mu = \alpha \int f \,\mathrm{d}\mu$. $\alpha \geq 0$; (c) $\int -f \,\mathrm{d}\mu = -\int f \,\mathrm{d}\mu$. Part (b) and (c) follow readily from (2.29) and Theorem 2.17 (i). Part (a) is obtained similarly upon observing that

$$(f+g)^+ - (f+g)^- = f + g = f^+ - f^- + g^+ - g^-.$$

□

Remark 2.21. 1) If $(X, \mathcal{A}, \mu) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \lambda)$, with λ denoting the Lebesgue measure, the corresponding integral $\int f \,\mathrm{d}\lambda$ for $f \in \mathcal{L}^1(\lambda)$ is called the **Lebesgue integral**.
In fact, one commonly "completes the space" by adding all subsets of λ-null sets, i.e. one considers λ on the "complete" σ-algebra

$$\mathcal{B}^*(\mathbb{R}^n) \overset{\text{def.}}{=} \sigma(\mathcal{B}(\mathbb{R}^n) \cup \mathcal{N}) \tag{2.33}$$

where $\mathcal{N} = \{A \subset \mathbb{R}^n : A \subset N \textit{ for some } N \in \mathcal{B}(\mathbb{R}^n) \textit{ with } \lambda(N) = 0\}$.

2) One can show that the Lebesgue integral generalises the Riemann integral in the following sense:

Proposition (Riemann vs. Lebesgue Integral).
Let $f : I \to \mathbb{R}$ be Riemann-integrable (cf. (0.3)*) on $I = [a,b]$ with $|\int_a^b f(x) \,\mathrm{d}x| < \infty$. Then $f\, 1_I$ is Lebesgue-integrable and*

$$\int_I f \,\mathrm{d}\lambda = \int_a^b f(x) \,\mathrm{d}x. \tag{2.34}$$

3) The converse is not true. Consider $f = 1_{\mathbb{Q}\cap[a,b]}$. Then $\int f \,\mathrm{d}\lambda$ is defined since $f \geq 0$, and f is Lebesgue-integrable, but not Riemann-integrable (cf. below (0.4)). Moreover, $\int_{[a,b]} f \,\mathrm{d}\lambda = 0$. Indeed, the functions $f_n : \mathbb{R} \to \mathbb{R}$ defined in (0.5) are simple, $f_n = \sum_{k=0}^n q_k 1_{\{q_k\}}(x)$, $f_n \geq 0$ and $f_n \nearrow f\, 1_{[a,b]}$, hence

$$\int_{[a,b]} f \,\mathrm{d}\lambda \overset{\text{Thm. 2.17(ii)}}{=} \lim_{n\to\infty} \int f_n \,\mathrm{d}\lambda \overset{(2.22)}{=} \lim_{n\to\infty} \sum_{k=0}^n q_k \lambda(\{q_k\}) = 0, \tag{2.35}$$

as

$$\lambda(\{q_k\}) \overset{(1.12)}{=} \lim_{n\to\infty} \lambda([q_k, q_k + 1/n]) \overset{(1.34)}{=} \lim_{n\to\infty} \frac{1}{n} = 0.$$

Note that (2.35) also shows the ease with which problems like (0.6) can be avoided using the Lebeague integral (namely: using the powerful monotone convergence, Theorem 2.17 (ii); see below for further convergence theorems).

4) If $\mu = \mathbb{P}$ is a probability measure and $f \in \mathcal{L}^1(\mathbb{P})$, one usually writes

$$\mathbb{E}[f] \overset{\text{def.}}{=} \int f \,\mathrm{d}\mathbb{P}, \tag{2.36}$$

the **expectation** (or integral) of f w.r.t. $\mathbb{P}$. For instance, let $X = \{1, 2, 3, \ldots\}$, $\mathcal{A} = 2^X$, $\mathbb{P} : X \to [0, 1]$ be defined by $\mathbb{P}(\{k\}) = (1 - p)^{k-1}p$, for $p \in (0, 1)$. ($\mathbb{P}$ is called geometric distribution). Let $f : X \to \mathbb{R}$, $f(x) = x$. Then $f \geq 0$, hence $\mathbb{E}[f]$ is defined. To compute $\mathbb{E}[f]$, one proceeds e.g. as follows: for $n \geq 1$, $x \in X$, let

$$f_n(x) = (f \cdot 1_{\{f\leq n\}})(x) = \sum_{k=1}^{n} f(k) \cdot 1_{\{k\}}(x) = \sum_{k=1}^{n} k \cdot 1_{\{k\}}(x) \quad \Big(\overset{(2.21)}{\in} S^+\Big).$$

Clearly $0 \leq f_n \nearrow f$, hence

$$\begin{aligned}\mathbb{E}[f] &\overset{\text{Thm. 2.17(ii)}}{=} \lim_{n\to\infty} \mathbb{E}[f_n] \overset{(2.22)}{=} \lim_{n\to\infty} \sum_{k=1}^{n} k\mathbb{P}(\{k\}) \\ &= \sum_{k=1}^{\infty} k \cdot (1 - p)^{k-1}p = -p\frac{\mathrm{d}}{\mathrm{d}p}\sum_{k=0}^{\infty}(1 - p)^k \overset{\substack{\text{geometric}\\ \text{series}}}{=} -p\frac{\mathrm{d}}{\mathrm{d}p}\frac{1}{1 - (1 - p)} = \frac{1}{p},\end{aligned}$$

hence $f \in \mathcal{L}^1(\mathbb{P})$.

5) If $X = \mathbb{N}$, $\mathcal{A} = 2^X$, μ = counting measure (see Ex. 1.4,1)), then

$$\ell^1 \overset{\text{def.}}{=} \mathcal{L}^1(\mu) = \Big\{a : \mathbb{N} \to \mathbb{R} : \sum_{n=1}^{\infty} |a_n| < \infty\Big\} \quad \text{and} \quad \int a \,\mathrm{d}\mu = \sum_{n=1}^{\infty} a_n. \tag{2.37}$$

(Note: $a_n = a(n)$.)

2.4 Convergence Theorems

As before, let $(X, \mathcal{A}, \mu)$ be a measure space (e.g. $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$). Convergence theorems regard the interchange of limits and integrals. The first such result is Theorem 2.17 (ii) (monotone convergence). As an application, we obtain:

Theorem 2.22 (Fatou's Lemma)**.**

Let $f_n : X \to [0,\infty]$, $n \in \mathbb{N}$, be a sequence of measurable functions. Then

$$\int (\liminf_{n\to\infty} f_n)\,\mathrm{d}\mu \le \liminf_{n\to\infty} \int f_n\,\mathrm{d}\mu. \tag{2.38}$$

PROOF. For $n \in \mathbb{N}$, define $g_n : X \to [0,\infty]$ by

$$g_n(x) = \inf_{k\ge n} f_k(x), \quad \text{for } x \in X.$$

The function g_n is measurable by (2.5), and plainly $g_n(x) \le f_k(x)$ for all $x \in X$ and $k \ge n$. Hence by monotonicity (Theorem 2.17 (i)),

$$\int g_n\,\mathrm{d}\mu \le \inf_{k\ge n} \int f_k\,\mathrm{d}\mu. \tag{2.39}$$

Now, $g_n(x) \le g_{n+1}(x)$ for all $x \in X$ and $n \in \mathbb{N}$, hence applying Theorem 2.17 (ii), we find that

$$\int (\liminf_{n\to\infty} f_n)\,\mathrm{d}\mu = \int (\lim_{n\to\infty} g_n)\,\mathrm{d}\mu \overset{\text{Thm. 2.17(ii)}}{=} \lim_{n\to\infty} \int g_n\,\mathrm{d}\mu \overset{(2.39)}{\le} \lim_{n\to\infty} \inf_{k\ge n} \int f_k\,\mathrm{d}\mu.$$

$\square$

Example 2.23. Let $E \in \mathcal{A}$ be a measurable set such that $0 < \mu(E) < \mu(X)$. Define $f_n = 1_E$ when n is even and $f_n = 1 - 1_E = 1_{E^c}$ when n is odd. Then $(f_n(x))_n$ is an alternating sequence of 1's and 0's for all $x \in X$ and therefore $\liminf_{n\to\infty} f_n(x) = 0$ for all $x \in X$. Thus

$$\int_X \liminf_{n\to\infty} f_n\,\mathrm{d}\mu = 0 < \min\{\mu(E), \mu(X\backslash E)\} = \liminf_{n\to\infty} \int f_n\,\mathrm{d}\mu,$$

where we used that

$$\int f_n\,\mathrm{d}\mu \overset{(2.22)}{=} \begin{cases} \mu(E), & n \text{ even} \\ \mu(X\backslash E), & n \text{ odd} \end{cases}$$

i.e. the inequality in (2.38) can be strict.

Theorem 2.24 (Lebesgue Dominated Convergence Theorem)**.**

Let $g : X \to [0,\infty]$ be integrable, i.e. $g \in \mathcal{L}^1(\mu)$, and $f, f_n : X \to [-\infty,\infty]$, $n \ge 1$ be measurable functions such that

$$|f_n(x)| \le g(x) \text{ for all } x \in X \quad \text{and} \quad f_n \overset{n\to\infty}{\longrightarrow} f \quad \mu\text{-a.e.} \tag{2.40}$$

Then

$$\left|\int f_n\,\mathrm{d}\mu - \int f\,\mathrm{d}\mu\right| \le \int |f_n - f|\,\mathrm{d}\mu \overset{n\to\infty}{\longrightarrow} 0. \tag{2.41}$$

Proof. The assumptions in (2.40) imply that $|f_n| \leq g$ and $|f| \leq g$ μ-a.e., hence $f_n, f \in \mathcal{L}^1(\mu)$. Moreover $|f_n - f| \leq |f_n| + |f| \leq 2g$, thus $f_n - f \in \mathcal{L}^1(\mu)$. It follows that all integrals appearing in (2.41) are well-defined. The inequality in (2.41) then follows immediately from (2.31) and (2.32).

Applying Fatou's Lemma to the sequence of functions $2g - |f_n - f|$ (≥ 0), we find that

$$\begin{aligned}\int 2g \,\mathrm{d}\mu &\overset{(2.40)}{=} \int \liminf_{n\to\infty} (2g - |f_n - f|) \,\mathrm{d}\mu \overset{(2.38)}{\leq} \liminf_{n\to\infty} \int (2g - |f_n - f|) \,\mathrm{d}\mu \\ &\overset{(2.32)}{=} \liminf_{n\to\infty} \left(\int 2g \,\mathrm{d}\mu - \int |f_n - f| \,\mathrm{d}\mu \right)\end{aligned}$$

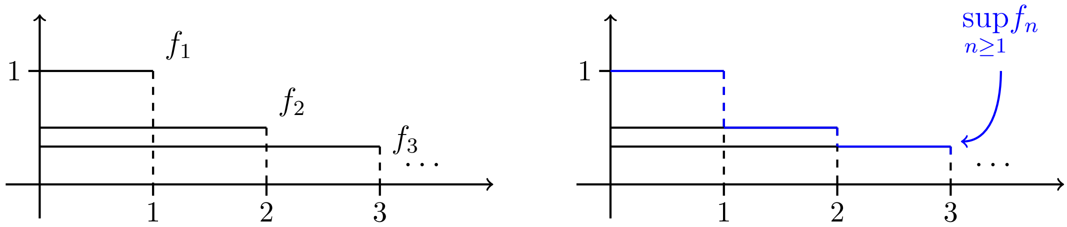
which implies that $0 \leq \limsup_{n\to\infty} \int |f_n - f| \,\mathrm{d}\mu \leq 0$. Hence $\lim_{n\to\infty} \int |f_n - f| \,\mathrm{d}\mu = 0$. $\square$

Example 2.25. Without (2.40), the conclusions (2.41) may fail. Consider the sequence $f_n = \frac{1}{n} 1_{[0,n]}$ (on $X = \mathbb{R}$). Then clearly (2.41) does not hold since $\int f_n \,\mathrm{d}\lambda = \frac{1}{n} \int_0^n 1 \,\mathrm{d}x = 1$ for all n, but $\lim_{n\to\infty} f_n(x) = 0$ for all $x \in \mathbb{R}$, and therefore

$$0 = \int \lim_{n\to\infty} f_n \,\mathrm{d}\lambda \neq \lim_{n\to\infty} \int f_n \,\mathrm{d}\lambda = 1.$$

On the other hand there is no dominating function $g \in \mathcal{L}^1(\lambda)$ such that (2.40) holds. Indeed, such a function necessarily satisfies $g \geq \sup_{n\geq 1} f_n$, whence for all $N \in \mathbb{N}$

$$\int g \,\mathrm{d}\lambda \geq \int \sup_{n\geq 1} f_n \,\mathrm{d}\lambda \underset{\substack{\uparrow \\ \sup_{n\geq 1} f_n = f_k \text{ on } (k-1,k]}}{\geq} \int \left(\sum_{k=1}^N \frac{1}{k} 1_{(k-1,k]} \right) \mathrm{d}\lambda = \sum_{k=1}^N \frac{1}{k} \int 1_{(k-1,k]} \,\mathrm{d}\lambda = \sum_{k=1}^N \frac{1}{k} \overset{n\to\infty}{\longrightarrow} \infty.$$



2.5 Vitali's Theorem

We now discuss an improvement of Theorem 2.24, which is optimal. I.e., the assumptions (weaker than (2.40)) will turn out to be equivalent to (2.41).

We now work towards stating these optimal conditions. Firstly, convergence μ-a.e. (cf. the second part of (2.40)) is unnecessarily strong.

Definition 2.26. Let $f : X \to \mathbb{R}$, $f_n : X \to [-\infty, \infty]$, $n \in \mathbb{N}$, be measurable. The sequence $(f_n)_{n\in\mathbb{N}}$ converges to f **in measure**, denoted $f_n \xrightarrow{\mu} f$, if for all $\varepsilon > 0$:

$$\mu(\{x \in X : |f_n(x) - f(x)| > \varepsilon\}) \to 0, \quad \text{as } n \to \infty. \tag{2.42}$$

Remark 2.27. If $\mu(X) = 1$, (2.42) is also called **convergence in probability**.

Theorem 2.28. *$f, f_n : X \to \mathbb{R}$, $n \in \mathbb{N}$ measurable, $\mu(X) < \infty$.*

(2.43) If $\displaystyle\int |f_n - f| \, \mathrm{d}\mu \overset{n\to\infty}{\longrightarrow} 0$, then $f_n \xrightarrow{\mu} f$.

(2.44) If $f_n \overset{n\to\infty}{\longrightarrow} f$ *μ-a.e., then* $f_n \xrightarrow{\mu} f$.

(2.45) If $f_n \xrightarrow{\mu} f$, there exists a subsequence $\Lambda \subset \mathbb{N}$ s.t. $f_n \to f$ *μ-a.e. as* $n \in \Lambda \to \infty$.

Proof. (2.43) and (2.44) were shown in Exercise 6 a) and b) on Sheet 4, respectively. (2.45) is Exercise 1 on Sheet 6. □

Remark 2.29. The converse of (2.44) is wrong, see solutions to Ex.6 b) on Sheet 4 for a counterexample.

To improve on the first condition in (2.40) (existence of a majorizing function $g \in \mathcal{L}^1(\mu)$), we consider the following.

Let $f : X \to [0, \infty)$ measurable and μ be a measure on $(X, \mathcal{A})$. Then

$$\nu(A) \overset{\text{def.}}{=} \int_A f \, \mathrm{d}\mu \left(= \int f \, 1_A \, \mathrm{d}\mu \right), \text{ for all } A \in \mathcal{A} \tag{2.46}$$

defines a measure on $(X, \mathcal{A})$. If ν has the form (2.46), then ν is said to have a **density with respect to** μ and f is called a **density**. In this case are often writes $\nu = f\mu$ and $f = \frac{\mathrm{d}\mu}{\mathrm{d}\nu}$.

Example 2.30. Let λ denote the Lebesgue measure on $\mathbb{R}$ and $f(x) = (2\pi)^{-\frac{1}{2}} e^{-\frac{x^2}{2}}$. Then $\mu \overset{\text{def.}}{=} f\lambda$ is called the **standard Gaussian distribution** (note: ν is a probability measure on $\mathbb{R}$ ($\nu(\mathbb{R}) = 1$)).

Definition 2.31. Let μ, ν be two measures on $(X, \mathcal{A})$. Then ν is called **absolutely continuous with respect to** μ (denoted by $\nu \ll \mu$) if

$$\nu(A) = 0 \text{ for all } A \in \mathcal{A} \text{ with } \mu(A) = 0. \tag{2.47}$$

The reason for this terminology is given by the following.

Proposition 2.32. *Let $f \in \mathcal{L}^1(\mu)$. Then:*

$$\forall \varepsilon > 0, \ \exists \delta > 0 : \forall A \in \mathcal{A} \ \ \mu(A) < \delta \ \Rightarrow \ \int_A |f| \, \mathrm{d}\mu < \varepsilon. \tag{2.48}$$

Proof. see Exercise 4 b) on Sheet 5. □

Remark 2.33. 1) Clearly $\nu = f\mu$ implies $\nu \ll \mu$. Indeed if $\mu(A) = 0$ then $f\,1_A = 0$ μ-a.e., hence $\int f\,1_A\,\mathrm{d}\mu = 0$ by (iii) above Definition 2.18. In view of (2.46) this means $\nu(A) = 0$, i.e. (2.47) holds.
2) We will later show that any absolutely continuous measure $\nu \ll \mu$ is of the form (2.46) i.e. $\nu = f\mu$ for some density f (Radon-Nikodym theorem). In light of this, (2.48) then implies that if $\nu \ll \mu$ and $\nu(X) < \infty$, one has

$$\forall \varepsilon > 0,\ \exists \delta > 0,\ \forall A \in \mathcal{A}: \quad \mu(A) < \delta \ \Rightarrow\ \nu(A) < \varepsilon. \tag{2.49}$$

We can now formulate an optimal (necessary & sufficient) convergence criterion.

Definition 2.34. Let $\mathcal{F} \subset \mathcal{L}^1(\mu)$. The family $\mathcal{F}$ is said to have **uniformly absolutely continuous integrals** if

$$\forall \varepsilon > 0,\ \exists \delta > 0,\ \forall f \in \mathcal{F},\ \forall A \in \mathcal{A} : \mu(A) < \delta \ \Rightarrow \int_A |f|\,\mathrm{d}\mu < \varepsilon. \tag{2.50}$$

Remark 2.35. By Proposition 2.32, any $\mathcal{F} = \{f\}$ has uniformly abs. cont. integrals. More generally, any finite family of functions $\mathcal{F} = \{f_1, \dots, f_N\}$, $N \in \mathbb{N}$, does.

Theorem 2.36 (Vitali)**.**
Let $\mu(X) < \infty$, $f, f_n : X \to \mathbb{R}$, $n \in \mathbb{N}$, all in $\mathcal{L}^1(\mu)$. The following are equivalent:

$$f_n \xrightarrow{\mu} f \text{ and } \mathcal{F} = \{f_n : n \in \mathbb{N}\} \text{ has uniformly absolutely continuous integrals.} \tag{2.51}$$

$$\int |f_n - f|\,\mathrm{d}\mu \to 0, \quad \text{as } n \to \infty. \tag{2.52}$$

Proof. (2.52) $\Rightarrow$ (2.51). The fact that $f_n \xrightarrow{\mu} f$ follows from (2.43). We now argue that (2.50) holds for $\mathcal{F} = \{f_n : n \in \mathbb{N}\}$. Let $\varepsilon > 0$. First by (2.52) we find $n_0 = n_0(\varepsilon)$ such that

$$\int |f_n - f|\,\mathrm{d}\mu < \frac{\varepsilon}{2} \quad \text{for all } n \geq n_0. \tag{2.53}$$

Then, applying Remark 2.35 to the finite family $\mathcal{F}' = \{f, f_1, \dots, f_{n_0}\}$, we find $\delta > 0$ such that, for all $A \in \mathcal{A}$ with $\mu(A) < \delta$,

$$\int_A |f|\,\mathrm{d}\mu < \frac{\varepsilon}{2} \quad \text{and} \quad \max_{n \leq n_0} \int_A |f_n|\,\mathrm{d}\mu < \frac{\varepsilon}{2}. \tag{2.54}$$

It thus follows for all $n \geq n_0$, $A \in \mathcal{A}$ with $\mu(A) < \delta$ that

$$\int_A |f_n|\,\mathrm{d}\mu \leq \int_A |f|\,\mathrm{d}\mu + \int_A |f_n - f|\,\mathrm{d}\mu \overset{(2.53)(2.54)}{<} \varepsilon. \tag{2.55}$$

Together, (2.55) and the second part of (2.54) yield (2.50) for $\mathcal{F}$.

(2.51) $\Rightarrow$ (2.52). By contradiction, we assume that $\limsup_{n\to\infty} \int |f_n - f| \,\mathrm{d}\mu > 0$. By passing to a subsequence Λ, we may assume that

$$\lim_{\substack{n\to\infty \\ n\in\Lambda}} \int |f_n - f| \,\mathrm{d}\mu > 0. \tag{2.56}$$

and by (2.45) that $f_n \to f$ μ-a.e. as $n \in \Lambda \to \infty$.
Now since $\mathcal{F} \cup \{f\}$ has uniformly absolutely continuous integrals by (2.51) and (2.48), we can find for given $\varepsilon > 0$ a $\delta > 0$ such that, if $A \in \mathcal{A}$ and $\mu(A) < \delta$,

$$\int_A |f| \,\mathrm{d}\mu < \frac{\varepsilon}{3} \quad \text{and} \quad \int_A |f_n| \,\mathrm{d}\mu < \frac{\varepsilon}{3} \quad \text{for all } n \in \mathbb{N}. \tag{2.57}$$

Applying Egorov's Theorem* to $(f_n)_{n\in\Lambda}$, one then finds to this δ a measurable set F such that

$$\mu(X\backslash F) < \delta \quad \text{and} \quad \sup_{x\in F} |f_n(x) - f(x)| \to 0 \;\text{ as } n \in \Lambda \to \infty. \tag{2.58}$$

Thus, choosing $n_0 = n_0(\varepsilon)$ such that

$$\sup_{x\in F} |f_n(x) - f(x)| < \frac{\varepsilon}{3\mu(X)}, \quad \text{for all } n \in \Lambda \text{ with } n \geq n_0, \tag{2.59}$$

it follows that for all $n \geq n_0$ with $n \in \Lambda$,

$$\begin{aligned}
\int |f_n - f| \,\mathrm{d}\mu &\overset{(2.32)}{=} \int |f_n - f| \, 1_F \,\mathrm{d}\mu + \int |f_n - f| \, 1_{X\backslash F} \,\mathrm{d}\mu \\
&\overset{(2.31)}{\leq} \int_F \sup_{x\in F} |f_n(x) - f(x)| \,\mathrm{d}\mu + \int_{X\backslash F} |f_n| \,\mathrm{d}\mu + \int_{X\backslash F} |f| \,\mathrm{d}\mu < \varepsilon,
\end{aligned}$$

as

$$\begin{gathered}
\int_F \sup_{x\in F} |f_n(x) - f(x)| \,\mathrm{d}\mu \overset{(2.59)}{<} \frac{\varepsilon}{3\mu(X)} \mu(F) \leq \frac{\varepsilon}{3}; \\
\int_{X\backslash F} |f_n| \,\mathrm{d}\mu \overset{(2.57)(2.58)}{<} \frac{\varepsilon}{3}; \quad \int_{X\backslash F} |f| \,\mathrm{d}\mu \overset{(2.57)(2.58)}{<} \frac{\varepsilon}{3}.
\end{gathered}$$

Choosing $\varepsilon > 0$ small enough, this contradicts (2.56). □

* By inspection of the proof of Theorem 2.9, one sees that its conclusions still hold for $(X, \mathcal{A}, \lambda)$ any measure space and $\Omega \subset \mathcal{A}$ with $\lambda(\Omega) < \infty$ (cf. (2.12)) if one only requires $F \in \mathcal{A}$ to be measurable (instead of compact) above (2.13).

2.6 L^p-spaces

L^p-spaces are amongst the most important (vector) spaces of functions in analysis. Let $(X, \mathcal{A}, \mu)$ be a measure space and $1 \leq p \leq \infty$.

Definition 2.37. For $f : X \to [-\infty, \infty]$ measurable, let

$$\|f\|_{L^p(\mu)} \stackrel{\text{def.}}{=} \left(\int |f|^p \, \mathrm{d}\mu \right)^{\frac{1}{p}} \quad (\leq \infty) \qquad \text{for } 1 \leq p < \infty \tag{2.60}$$

$$\|f\|_{L^\infty(\mu)} \stackrel{\text{def.}}{=} \operatorname*{esssup}_{x \in X} |f(x)| \stackrel{\text{def.}}{=} \inf\{C \in [0, \infty] : |f| \leq C \ \mu\text{-a.e.}\} \ (\leq \infty), \tag{2.61}$$

and set

$$\mathcal{L}^p = \{f : X \to [-\infty, \infty] : f \ \textit{measurable}, \ \|f\|_{L^p(\mu)} < \infty\} \text{ (consistent with (2.28)).}$$

(Note: $|f| \leq C$ μ-a.e. means $\mu(\{x \in X : |f(x)| > C\}) = 0$ and $\inf \emptyset = \infty$.)

Remark 2.38. 1) For $f \in \mathcal{L}^\infty(\mu)$ we have

$$|f(x)| \leq \|f\|_{L^\infty(\mu)} \quad \text{for } \mu\text{-a.e. } x. \tag{2.62}$$

Indeed consider a sequence $(C_k)_k$ with $C_k \downarrow \|f\|_{L^\infty(\mu)}$ as $k \to \infty$ and set $A_k = \{x : |f(x)| > C_k\}$. Then $\mu(A_k) = 0$ for all $k \in \mathbb{N}$ by (2.61), hence $\mu(A) = 0$ with $A = \bigcup_k A_k$ and $|f(x)| \leq \|f\|_{L^\infty(\mu)}$ holds for all $x \in X \backslash A$.

2) If $X \in \mathcal{B}(\mathbb{R}^n)$, $\mu = \lambda|_X$ = Lebesgue measure on X, one often writes $\mathcal{L}^p(X)$, e.g. $\mathcal{L}^p(\mathbb{R}^n)$. If $(X, \mathcal{A}, \mu)$ is clear from the context, one frequently writes $\|\cdot\|_p = \|\cdot\|_{L^p(\mu)}$.

3) Consider $f(x) = |x|^{-\alpha}$, $\alpha > 0$, $x \in X = [-1, 1]$ endowed with the Lebesgue measure (restricted to X). Then f is measurable and

$$\begin{aligned}
\|f\|^p_{L^p(X)} &= \int_X |f(x)|^p \, \mathrm{d}\lambda \stackrel{\substack{\text{Thm. 2.17} \\ \text{(ii)}}}{=} \lim_{n \to \infty} \int_X |x|^{-\alpha p} \, 1_{\{|x| > \frac{1}{n}\}} \, \mathrm{d}\lambda \\
&= 2 \lim_{n \to \infty} \begin{cases} \left[\frac{x^{-\alpha p + 1}}{-\alpha p + 1} \right]_{x = \frac{1}{n}}^{x=1}, & \alpha p \neq 1 \\ [\log x]_{x=\frac{1}{n}}^{x=1}, & \alpha p = 1 \end{cases} = 2 \lim_{n \to \infty} \begin{cases} c(n^{\alpha p - 1} - 1), & \alpha p \neq 1 \\ \log n, & \alpha p = 1 \end{cases}
\end{aligned}$$

so $f \in \mathcal{L}^p(X)$ if and only if $\alpha p < 1$.

We want to turn $\mathcal{L}^p(\mu)$ into a **normed vector space**. The issue is that $\|\cdot\|_{L^p(\mu)}$ is not definite, i.e. if $f = g$ μ-a.e. then $\|f - g\|_{L^p(\mu)} = 0$ by (iii) above Definition 2.18.

To cure this, we identify functions which are equal μ-a.e. Formally, we introduce the equivalence relation, for $f, g \in \mathcal{L}^p(\mu)$

$$f \sim g \Leftrightarrow f = g \ \mu\text{-a.e.} \quad (\Leftrightarrow \|f - g\|_{L^p(\mu)} = 0 \ ). \tag{2.63}$$

and define the equivalence classes $[f] = \{g \in \mathcal{L}^P(\mu) : g \sim f\}$. We then set

$$L^p(\mu) \overset{\text{def.}}{=} \{[f] : f \in \mathcal{L}^P(\mu)\} \tag{2.64}$$

endowed with $\|[f]\|_{L^p(\mu)} = \inf_{g\sim f} \|g\|_{L^p(\mu)} \overset{(2.63)}{=} \|f\|_{L^p(\mu)}$.

Convention: We often tacitly identify $[f]$ with any of its representatives $f \in [f]$ and write for instance $\|f\|_{L^p(\mu)}$ rather than $\|[f]\|_{L^p(\mu)}$, or $f \in L^p(\mu)$ instead of $[f] \in L^p(\mu)$. In doing so, we understand that f is only defined up to sets of μ-measure 0. To do calculations, one typically works with one specific function f in the equivalence class. For instance, this implies in the example from Remark 2.38,3) above that $f \in L^p(X)$ if and only if $\alpha p < 1$.

The next theorem shows that $L^p(\mu)$ has "good" properties. In particular, i) $\|\cdot\|_{L^p(\mu)}$ defines a *norm* on this space, with respect to which ii) $L^p(\mu)$ is *complete*, in the following sense.

Let $(Y, \|\cdot\|_Y)$ be a normed vector space (soon $Y = L^p(\mu)$). Recall that (f_n) is called a *Cauchy sequence in* Y if $f_n \in Y$ for all $n \in \mathbb{N}$ and for all $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that $\|f_n - f_m\|_Y < \varepsilon$ for all $n, m \geq N$. The space Y is called *complete* (with respect to the norm $\|\cdot\|_Y$) if, for any Cauchy sequence (f_n) in Y, there exists $f \in Y$ such that $\|f_n - f\|_Y \to 0$ as $n \to \infty$. In plain words, $(Y, \|\cdot\|_Y)$ is complete if any Cauchy sequence has a limit. This property is very desirable in practice (for instance, it holds for $Y = \mathbb{R}$ with $\|\cdot\|_Y = |\cdot|$ the usual Euclidean norm, as seen in your introductory analysis course).

Theorem 2.39. *The space $L^p(\mu)$ in* (2.64) *is a complete normed vector space for all $1 \leq p \leq \infty$ (with norm $\|\cdot\|_{L^p(\mu)}$).*

Proof. The definiteness of the norm $\|\cdot\|_p = \|\cdot\|_{L^p(\mu)}$ is clear (i.e. $\|f\|_p = 0 \Leftrightarrow f = 0$, by which are really mean $f \in [0]$ or equivalently $f = 0$ μ-a.e.). Positive homogeneity, i.e. $\|\lambda f\|_p = |\lambda| \, \|f\|_p$ for $\lambda \in \mathbb{R}$ follows from (2.60) and (2.32) if $p < \infty$, and directly from (2.61) if $p = \infty$. The triangle inequality is shown in (2.66) below and the completeness in Lemma 2.43. □

Remark 2.40. The definitions (2.63)–(2.64) remove the issue mentioned below (0.7). Indeed the functions f_n defined in (0.5) satisfy $f \sim 0$, where 0 is the function which is identically 0. In particular, $0 = \|f_n - 0\|_{L^1([0,1])}$ and $f_n \to 1_{\mathbb{Q}\cap[0,1]}$ pointwise but $1_{\mathbb{Q}\cap[0,1]} \sim 0$ since $\lambda(\mathbb{Q} \cap [0,1]) = 0$.

Towards deducing the triangle inequality for $\|\cdot\|_p$ we have:

Proposition 2.41 (Young's Inequality)**.**

Let $1 < p, q < \infty$ *be conjugate, i.e.* $\frac{1}{p} + \frac{1}{q} = 1$*. Then:*

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q} \quad \text{for all } a, b \geq 0.$$

Proof. see Exercises. $\square$

Corollary 2.42.

(i) (Hölder's inequality) Let $1 \leq p, q \leq \infty$ *be conjugate* $f \in L^p(\mu)$*,* $g \in L^q(\mu)$*. Then:*

$$f \cdot g \in L^1(\mu) \quad \text{and} \quad \|fg\|_{L^1(\mu)} \leq \|f\|_{L^p(\mu)} \|g\|_{L^q(\mu)}. \tag{2.65}$$

(ii) (Minkowski's inequality) Let $1 \leq p \leq \infty$ *and* $f, g \in L^p(\mu)$*. Then :*

$$f + g \in L^p(\mu) \quad \text{and} \quad \|f + g\|_{L^p(\mu)} \leq \|f\|_{L^p(\mu)} + \|g\|_{L^p(\mu)}. \tag{2.66}$$

Proof. (i) WLOG, let $p \leq q$. If $p = 1, q = \infty$, (2.65) follows from the fact that $|fg| \leq |f| \cdot \|g\|_\infty$ μ-a.e., which is implied by (2.62).
For $1 < p, q < \infty$, suppose first that $\|f\|_p = \|g\|_q = 1$. Then by Proposition 2.41,

$$\int |fg| \, \mathrm{d}\mu \overset{\substack{\text{Thm. 2.17} \\ \text{(i)}}}{\leq} \int \left(\frac{|f|^p}{p} + \frac{|g|^q}{q} \right) \mathrm{d}\mu = \frac{1}{p} \|f\|_p^p + \frac{1}{q} \|g\|_q^q = \frac{1}{p} + \frac{1}{q} = 1.$$

For general f, g apply the previous reasoning to $\widetilde{f} = \frac{f}{\|f\|_p}$, $\widetilde{g} = \frac{g}{\|g\|_p}$.
(ii) is an application of (i), see Exercises. $\square$

Lemma 2.43 (Completeness of $L^p(\mu)$)**.**

Let (f_n) *be a Cauchy-sequence in* $L^p(\mu)$*, i.e.* $\forall \varepsilon > 0$*,* $\exists N \in \mathbb{N}$*,* $\forall n, m \geq N : \|f_n - f_m\|_{L^p(\mu)} < \varepsilon$*. Then there exists* $f \in L^p(\mu)$ *such that* (f_n) *converges to* f *in* L^p*, i.e.* $\|f_n - f\|_{L^p(\mu)} \to 0$ *as* $n \to \infty$ *(*$f_n \xrightarrow{L^p} f$*).*

Proof. Let $1 \leq p < \infty$. Choose a subsequence $(f_{n_k})_k$, $n_1 < n_2 < \cdots$ such that

$$\|f_n - f_m\|_p \leq 2^{-k} \quad \text{for all } k \geq 1, \; m, n \geq n_k. \tag{2.67}$$

Set $g_l = \sum_{k=1}^{l} |f_{n_{k+1}} - f_{n_k}|$, $g = g_\infty$. By Theorem 2.17(ii), since $0 \leq g_l \nearrow g$ as $l \to \infty$,

$$\|g\|_p = \left(\int |g|^p \, \mathrm{d}\mu \right)^{\frac{1}{p}} = \lim_{l \to \infty} \|g_l\|_p \overset{(2.66)}{\leq} \lim_{l \to \infty} \sum_{k=1}^{l} \|f_{n_{k+1}} - f_{n_k}\|_p \overset{(2.67)}{<} \infty.$$

Hence $g \in L^p(\mu)$ and thus $g < \infty$ μ-a.e. It follows that

$$f(x) = f_{n_j}(x) + \sum_{k=j}^{\infty}(f_{n_{k+1}}(x) - f_{n_k}(x)) \text{ for } j \in \mathbb{N},\ j \geq 1, \quad \text{when } g(x) < \infty \tag{2.68}$$

and $f(x) = 0$ otherwise is well-defined (the series in (2.68) is absolutely convergent and f is given by (2.68) μ-a.e.). We now argue that

$$f \in L^p(\mu) \quad \text{and} \quad \|f_{n_k} - f\|_p \to 0 \text{ as } k \to \infty. \tag{2.69}$$

To see this, note that μ-a.e.

$$|f - f_{n_k}| \overset{(2.68)}{=} \left|\sum_{l=k}^{\infty} f_{n_{l+1}} - f_{n_l}\right| \leq g.$$

This implies $f \in L^p(\mu)$ (use $|f|^p \leq 2^p(|f - f_{n_k}|^p + |f_{n_k}|^p)$) and $\int |f - f_{n_k}|^p \,\mathrm{d}\mu \to 0$ by Theorem 2.24 and (2.69) follows. It is now easy to see that $f_n \overset{L^p}{\to} f$, as for all $k \in \mathbb{N}$ and $n \geq n_k$, one has

$$\|f - f_n\|_p \overset{(2.66)}{\leq} \|f_n - f_{n_k}\|_p + \|f_{n_k} - f\|_p \overset{(2.67)}{\leq} 2^{-k} + \|f_{n_k} - f\|_p$$

whence $\limsup_{n\to\infty} \|f - f_n\|_p = 0$ using (2.69).
For $p = \infty$, $(f_k(x))_k$ is a Cauchy sequence (in $\mathbb{R}$) for μ-a.e. $x \in X$. Hence $\lim_{k\to\infty} f_k(x) \overset{\text{def.}}{=} f(x)$ exists μ-a.e. and

$$|f_k(x) - f(x)| = \lim_{n\to\infty} |f_k(x) - f_n(x)| \leq \limsup_{n\to\infty} \|f_k - f_n\|_\infty \overset{k\to\infty}{\longrightarrow} 0$$

since $(f_k)_k$ is Cauchy in L^∞. □

Remark 2.44. The properties listed in Theorem 2.39 define a **Banach space** (hence Theorem 2.39 can be reformulated as: $L^p(\mu)$ is a Banach space $\forall 1 \leq p \leq \infty$). Such spaces form a central object of study in functional analysis. The space L^p may appear vast but for instance, one has that (no proof)

$$C_c^0(\mathbb{R}^n) \text{ is dense in } L^p(\mathbb{R}^n) \quad \forall 1 \leq p < \infty. \tag{2.70}$$

where

$$C_c^0 = \{f : \mathbb{R}^n \to \mathbb{R} : f \text{ is continuous, } \operatorname{supp}(f) \overset{\text{def.}}{=} \overline{\{x : f(x) \neq 0\}} \text{ is compact}\}$$

and the meaning of (2.70) is that for any $f \in L^p$, there exists a sequence $f_k \in C_c^0$, $k \in \mathbb{N}$ such that $f_k \overset{L^p}{\to} f$, i.e. $\|f_k - f\|_{L^p(\mathbb{R}^n)} \overset{k\to\infty}{\longrightarrow} 0$.

Chapter 3

Product measures and multiple integrals

3.1 Product Measures

Let $(X_i, \mathcal{A}_i, \mu_i)$, $i = 1, 2$ be two measure spaces. We aim to define a measure on their **product space**

$$X = X_1 \times X_2 \quad (= \{(x_1, x_2) : x_i \in X_i,\ i = 1, 2\}). \tag{3.1}$$

We introduce the **coordinate maps**

$$Z_i : X \to X_i, \quad x = (x_1, x_2) \mapsto x_i \quad i = 1, 2 \tag{3.2}$$

and endow X with the **product σ-algebra**

$$\mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2 \overset{\text{def.}}{=} \sigma(Z_1, Z_2) \quad (= \sigma(\{Z_i^{-1}(A_i) : A_i \in \mathcal{A}_i,\ i = 1, 2\})). \tag{3.3}$$

Lemma 3.1.

$$\mathcal{A} = \sigma(\{A_1 \times A_2 : A_i \in \mathcal{A}_i,\ i = 1, 2\}).$$

Proof. This follows immediately from the fact that $Z_1^{-1}(A_1) = A_1 \times X_2$ and $Z_2^{-1}(A_2) = X_1 \times A_2$ on the one hand, and $A_1 \times A_2 = Z_1^{-1}(A_1) \cap Z_2^{-1}(A_2)$ on the other. □

We call a **rectangle** any set of the form $R = A_1 \times A_2$, $A_i \in \mathcal{A}_i$, $i = 1, 2$, and introduce the algebra of elementary figures (cf. (1.33))

$$\mathcal{R} = \left\{ \bigcup_{i=1}^{m} R_i : R_i,\ 1 \le i \le m, \text{ disjoint rectangles} \right\}. \tag{3.4}$$

We then define (cf. (1.34))

$$\rho(R) \overset{\text{def.}}{=} \sum_{i=1}^{m} \mu_1(A_i)\mu_2(B_i) \quad \text{if } R = \bigcup_{i=1}^{m} (A_i \times B_i), \tag{3.5}$$

(where $(A_i \times B_i)$ disjoint).

Lemma 3.2. *ρ defines a pre-measure on $\mathcal{R}$.*

Proof. Evidently $\rho(\emptyset) = 0$ by (3.5) ($\emptyset = \emptyset \times \emptyset$). We need to show (cf. Def. 1.7,i)) that for all disjoint rectangles R_k, $k \in \mathbb{N}$, such that $R = \bigcup_{k=1}^{\infty} R_k \in \mathcal{R}$,

$$\rho(R) = \sum_{k=1}^{\infty} \rho(R_k). \tag{3.6}$$

In view of (3.5), we may assume that R is a rectangle. The assertion (3.6) will follow from a (double) application of the monotone convergence theorem. The key is to note that, if R is a rectangle,

$$\rho(R) = \int \left(\int 1_R(x_1, x_2) \, \mathrm{d}\mu_1(x_1) \right) \mathrm{d}\mu_2(x_2). \tag{3.7}$$

Indeed, if $R = A_1 \times A_2$, then for every $x_2 \in X_2$, $x_1 \mapsto 1_R(x_1, x_2) = 1_{A_1}(x_1) 1_{A_2}(x_2)$ is simple, hence $\mathcal{A}_1$-measurable, and

$$\int 1_R(x_1, x_2) \, \mathrm{d}\mu_1(x_1) = 1_{A_2}(x_2) \int 1_{A_1}(x_1) \, \mathrm{d}\mu_1(x_1) = 1_{A_2}(x_2) \mu_1(A_1).$$

Now $x_2 \mapsto \mu_1(A_1) 1_{A_2}(x_2)$ is $\mathcal{A}_2$-measurable and

$$\int \left(\int 1_R(x_1, x_2) \, \mathrm{d}\mu_1(x_1) \right) \mathrm{d}\mu_2(x_2) = \int \mu_1(A_1) 1_{A_2}(x_2) \, \mathrm{d}\mu_2(x_2) = \mu_1(A_1) \mu_2(A_2)$$

and (3.7) follows on account of (3.5).

Now, if $R = \bigcup_{k=1}^{\infty} R_k$, R_k disjoint rectangles then $1_R = \sum_{k=1}^{\infty} 1_{R_k}$ and so $x_1 \mapsto 1_R(x_1, x_2) = \lim_n \sum_{k=1}^{n} 1_{R_k}(x_1, x_2)$ is $\mathcal{A}_1$-measurable by (2.5), and similarly

$$x_2 \mapsto \sum_{k=1}^{\infty} \int 1_{R_k}(x_1, x_2) \, \mathrm{d}\mu_1(x_1)$$

is $\mathcal{A}_2$-measurable, hence

$$\begin{aligned}
\rho(R) &\overset{(3.7)}{=} \int \left(\int 1_R(x_1, x_2) \, \mathrm{d}\mu_1(x_1) \right) \mathrm{d}\mu_2(x_2) \\
&= \int \left(\int \sum_{k=1}^{\infty} 1_{R_k}(x_1, x_2) \, \mathrm{d}\mu_1(x_1) \right) \mathrm{d}\mu_2(x_2) \\
&\overset{\text{Thm. 2.17(ii)}}{=} \int \left(\sum_{k=1}^{\infty} \int 1_{R_k}(x_1, x_2) \, \mathrm{d}\mu_1(x_1) \right) \mathrm{d}\mu_2(x_2) \\
&\overset{\text{Thm. 2.17(ii)}}{=} \sum_{k=1}^{\infty} \int \left(\int 1_{R_k}(x_1, x_2) \, \mathrm{d}\mu_1(x_1) \right) \mathrm{d}\mu_2(x_2) \overset{(3.7)}{=} \sum_{k=1}^{\infty} \rho(R_k).
\end{aligned}$$

□

With Lemma 3.2 at our disposal, we can make the following

Definition 3.3. Let $(X_i, \mathcal{A}_i, \mu_i)$, $i = 1, 2$, be σ-finite. Then the **product measure** $\mu_1 \otimes \mu_2$ is the (unique) measure on $(X, \mathcal{A})$ (see (3.1),(3.3)) obtained as the Hahn-Carathéodory extension of ρ defined in (3.5). In particular $(\mu_1 \otimes \mu_2)(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$ for all $A_1 \in \mathcal{A}_i$, $i = 1, 2$.

Remark 3.4. 1) The measure $\mu_1 \otimes \mu_2$ is well-defined. Indeed, Theorem 1.13 applies (with $\widetilde{\mu} = \rho)^1$ on account of Lemma 3.2 and $\mu_1 \otimes \mu_2$ is unique by Theorem 1.14, which is in force. To see that ρ is σ-finite (see (1.29)), as required for Theorem 1.14 to apply, one uses that μ_i, $i = 1, 2$ are and considers, if $A_{i,k} \in \mathcal{A}_i$, $k \in \mathbb{N}$ are disjoint sets such that $\mu_i(A_{i,k}) < \infty$ for all k and $X_i = \bigcup_k A_{i,k}$ for $i = 1, 2$, the sets $A_{1,k} \times A_{2,l}$, $k, l \in \mathbb{N}$, which have similar properties relative to ρ. ([1] note also that $\sigma(\mathcal{R}) = \mathcal{A}$).
2) Let λ^n denote the Lebesgue measure on $\mathbb{R}^n$. If $\mu_1 = \mu_2 = \lambda^1$, then $\mu_1 \otimes \mu_2 = \lambda^2$ (on $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$). Indeed, one has

$$\rho([a_1, b_2] \times [a_2, b_2]) \overset{(3.5)}{=} (b_1 - a_1)(b_2 - a_2) \overset{(1.34)}{=} \widetilde{\lambda}^2([a_1, b_1] \times [a_2, b_2])$$

and thus $\rho = \widetilde{\lambda}^2$, hence the claim follows by uniqueness. More generally

$$\lambda^n \otimes \lambda^m = \lambda^{n+m} \quad \text{for all } n, m \geq 1. \tag{3.8}$$

3) With little more effort, for an arbitrary index set I (here $I = \{1, 2\}$) given σ-finite measure spaces $(X_i, \mathcal{A}_i, \mu_i)$, $i \in I$, are defines a measure $\bigotimes_{i \in I} \mu_i$ on $X = \{(x_1, x_2, \dots) : x_i \in X_i, \ \forall i \in I\}$ enclosed with the σ-algebra $\mathcal{A} = \sigma(Z_i, i \in I)$, where $Z_i : X \to X_i$, $(x_1, x_2, \dots) \mapsto x_i$ are the canonical coordinates (cf. (3.2)) such that for all finite $J \subset I$, $A_j \in \mathcal{A}_j$, $j \in J$,

$$\left(\bigotimes_{i \in I} \mu_i \right) (\{Z_j \in A_j, \text{ for all } j \in J\}) = \prod_{j \in J} \mu_j(A_j). \tag{3.9}$$

4)* For instance, let $I = \mathbb{N}$, $X_i = \{-1, 1\}$, $\mathcal{A}_i = 2^{X_i}$, $\mu_i(\{1\}) = \mu_i(\{-1\}) = \frac{1}{2}$ for all $i \in \mathbb{N}$. Define $S = (S_n) : X = \{-1, 1\}^{\mathbb{N}} \to \mathbb{Z}^{\mathbb{N}}$ (enclosed with the product σ-algebra) $S_0 = 0$, $S_n = \sum_{k=1}^n Z_k(x)$ for all $n \in \mathbb{N}$. Then the image measure (see Ex. 5.1) $P = (\bigotimes_{i \in \mathbb{N}} \mu_i) \circ S^{-1}$ is a probability measure, the **canonical law of the simple random walk on $\mathbb{Z}$.**

3.2 Fubini's Theorem

We now discuss Fubini's theorem, which justifies (under suitable assumptions) the exchange of order of integrals.

Theorem 3.5. *Let $(X_i, \mathcal{A}_i, \mu_i)$ be σ-finite measure spaces, $i = 1, 2$. Let $f : X_1 \times X_2 \to [-\infty, \infty]$ be $\mathcal{A}(= \mathcal{A}_1 \otimes \mathcal{A}_2)$-measurable. If $f \geq 0$ or $f \in \mathcal{L}^1(\mu_1 \otimes \mu_2)$, then*

$$\begin{aligned} x_1 &\mapsto \int f(x_1, x_2) \, \mathrm{d}\mu_2(x_2) \quad \text{is } \mathcal{A}_1\text{-measurable} \\ x_2 &\mapsto \int f(x_1, x_2) \, \mathrm{d}\mu_1(x_1) \quad \text{is } \mathcal{A}_2\text{-measurable} \end{aligned} \tag{3.10}$$

and

$$\int f \,\mathrm{d}(\mu_1 \otimes \mu_2) = \int \left(\int f(x_1, x_2) \,\mathrm{d}\mu_2(x_2) \right) \mathrm{d}\mu_1(x_1) \tag{3.11}$$
$$= \int \left(\int f(x_1, x_2) \,\mathrm{d}\mu_1(x_1) \right) \mathrm{d}\mu_2(x_2)$$

Towards proving Theorem 3.5, we first show:

Lemma 3.6. *If $Q \in \mathcal{A}$ and $f : X_1 \times X_2 \to [-\infty, \infty]$ is $\mathcal{A}$-measurable, then $Q_{x_1} = \{x_2 \in X_2 : (x_1, x_2) \in Q\} \in \mathcal{A}_2$ and $f_{x_1} : X_2 \to [-\infty, \infty]$ defined by $f_{x_1}(x_2) = f(x_1, x_2)$ is $\mathcal{A}_2$-measurable for every $x_1 \in X_1$. Similarly, $Q^{x_2} = \{x_1 \in X_1 : (x_1, x_2) \in Q\} \in \mathcal{A}_1$ and $f^{x_2} : X_1 \to [-\infty, \infty]$ defined by $f^{x_2} = f(x_1, x_2)$ is $\mathcal{A}_1$-measurable for every $x_2 \in X_2$.*

Proof. We only show the first case (fixing $x_1 \in X_1$). The other case is analogous. We argue that

$$\mathcal{F} \stackrel{\text{def.}}{=} \{Q \subset X_1 \times X_2 : Q_{x_1} \in \mathcal{A}_2 \text{ for all } x_1 \in X_1\} \tag{3.12}$$

is a σ-algebra. First $X_1 \times X_2 \in \mathcal{F}$ since $(X_1, X_2)_{x_1} = X_2 \in \mathcal{A}_2$. Second if $Q \in \mathcal{F}$ then by (3.12) $Q_{x_1} \in \mathcal{A}_2$ for all $x_1 \in X_1$ hence

$$(Q^c)_{x_1} = \{x_2 \in \mathcal{A} : (x_1, x_2) \notin Q\} = (Q_{x_1})^c \in \mathcal{A}_2.$$

Third if $(Q_n) \subset \mathcal{F}$, $Q = \bigcup_n Q_n$ then $Q_{x_1} = \bigcup_{n=1}^{\infty} (Q_n)_{x_1} \in \mathcal{A}_2$ for all x_1. Thus $\mathcal{F}$ is a σ-algebra and $\mathcal{F} \ni A_1 \times A_2$ for all $A_1 \in \mathcal{A}_1$, $A_2 \in \mathcal{A}_2$, as

$$(A_1 \times A_2)_{x_1} = \begin{cases} A_2, & x_1 \in A_1, \\ \emptyset, & x_1 \notin A_1. \end{cases}$$

Thus $\mathcal{F} \supset \mathcal{A}$ by Lemma 3.1. Now fix $x_1 \in X_1$, $U \subset \mathbb{R}$ open. Then $Q \stackrel{\text{def.}}{=} f^{-1}(U) \in \mathcal{A}$, therefore

$$f_{x_1}^{-1}(U) = \{x_2 \in X_2 : f(x_1, x_2) \in U\} = Q_{x_1} \in \mathcal{A}_2.$$

□

Next we show a special case of Theorem 3.5.

Proposition 3.7. *If $Q \in \mathcal{A}$, then*

$$x_1 \mapsto \mu_2(Q_{x_1}) \text{ is } \mathcal{A}_1\text{-measurable, } x_2 \mapsto \mu_2(Q^{x_2}) \text{ is } \mathcal{A}_2\text{-measurable and} \tag{3.13}$$
$$(\mu_1 \otimes \mu_2)(Q) = \int \mu_2(Q_{x_1}) \,\mathrm{d}\mu_1(x_1) = \int \mu_1(Q^{x_2}) \,\mathrm{d}\mu_2(x_2).$$

We will employ the useful result, see e.g. Durrett, Probability: theory and examples, CUP, Thm. 2.1.2 and App. A.1 for a proof.

Theorem 3.8 (Dynkin's π-λ Theorem)**.**

Let X be a set, $\mathcal{P} \subset 2^X$ be closed under intersections i.e. $A, B \in \mathcal{P} \Rightarrow A \cap B \in \mathcal{P}$ (a π-system) and $\mathcal{L} \subset 2^X$ satisfy (1.1), (1.2) *and* (1.3') *(each with $\mathcal{A} = \mathcal{L}$) for disjoint $A_1, A_2, \ldots$ (a λ-system). Then*

$$\mathcal{P} \subset \mathcal{L} \Longrightarrow \sigma(\mathcal{P}) \subset \mathcal{L}. \tag{3.14}$$

Proof of Proposition 3.7. First assume μ_1, μ_2 are finite, i.e. $\mu_i(X_i) < \infty$, $i = 1, 2$. Let

$$\mathcal{L} = \{Q \in \mathcal{A} : (3.13) \textit{ holds}\} \tag{3.15}$$

and $\mathcal{P} = \{A_1 \times A_2 : A_i \in \mathcal{A}_i,\ i = 1, 2\}$. Clearly $\mathcal{P}$ is a π-system and $\sigma(\mathcal{P}) = \mathcal{A}$ by Lemma 3.1. Thus it suffices to argue that i) $P \subset \mathcal{L}$ and ii) $\mathcal{L}$ is a λ-system, and Proposition 3.7 follows from (3.14).
To see that i) holds, recall that $(A_1 \times A_2)_{x_1} = \begin{cases} A_2, & \text{if } x_1 \in A_1 \\ \emptyset, & \text{else} \end{cases}$, so

$$x_1 \mapsto \mu_2((A_1 \times A_2)_{x_1}) = \mu_2(A_2) 1_{A_1}(x_1) \tag{3.16}$$

is $\mathcal{A}_1$-measurable and

$$\int \mu_2(Q_{x_1})\, \mathrm{d}\mu_1(x_1) \overset{(3.16)}{=} \mu_2(A_2)\mu_1(A_1) \overset{\text{Def.3.3}}{=} (\mu_1 \otimes \mu_2)(A_1 \times A_2).$$

Similar conclusions hold for $\mu_1(Q^{2x})$ and i) follows.
To obtain ii), note that $X = X_1 \times X_2 \in \mathcal{L}$ by i). If $Q \in \mathcal{L}$ then since $(Q^c)_{x_1} = (Q_{x_1})^c$, we get that $x_1 \mapsto \mu_2((Q^c)_{x_1}) = \mu_2(X_2) - \mu_2(Q_{x_1})$ is measurable and that

$$\begin{aligned} \int \mu_2((Q^c)_{x_1})\, \mathrm{d}\mu_1(x_1) &= \mu_1(X_1)\mu_2(X_2) - \int \mu_2(Q_{x_1})\, \mathrm{d}\mu_1(x_1) \\ &\overset{Q \in \mathcal{L}}{=} (\mu_1 \otimes \mu_2)(X_1 \times X_2) - (\mu_1 \otimes \mu_2)(Q) = (\mu_1 \otimes \mu_2)(Q^c). \end{aligned}$$

Lastly if $Q_1, Q_2, \ldots \in \mathcal{L}$ are disjoint then $\bigcup_n Q_n \in \mathcal{A}$ follows using that $(\bigcup_n Q_n)_{x_1} = \bigcup_n (Q_n)_{x_1}$ and σ-additivity of μ_1, μ_2 and $\mu_1 \otimes \mu_2$.
If μ_1, μ_2 are σ-finite one considers an increasing sequence $A_n^1 \times A_n^2 \in \mathcal{A}$ with $\mu_i(A_n^i) < \infty$, $i = 1, 2$. The conclusions of Proposition 3.7 hold for $Q \cap (A_n^1 \times A_n^2)$ by what we already showed, and the claim follows by letting $n \to \infty$ and using monotone convergence. □

We now come to the

Proof of Theorem 3.5. Let $f = 1_Q$ for some $Q \in \mathcal{A}$. Then $1_Q(x_1, x_2) = 1_{Q_{x_1}}(x_2)$ for all x_1, x_2 hence

$$x_1 \mapsto \int f(x_1, x_2)\, \mathrm{d}\mu_2(x_2) = \mu_2(Q_{x_1})$$

is $\mathcal{A}_1$-measurable by (3.13), and (3.10) follows; (3.11) for such f is exactly the second line of (3.13). Using (2.4) and (2.23), (3.10) and (3.11) then follow for arbitrary simple functions $s \in S^+$. For $f \geq 0$, consider $s_n \in S^+$ with $s_n \nearrow f$ as in (2.10). Then by monotone convergence (rel. to μ_2)

$$\int s_n(x_1, x_2)\,\mathrm{d}\mu_2(x_2) \nearrow \int f(x_1, x_2)\,\mathrm{d}\mu_2(x_2) \tag{3.17}$$

for all $x_1 \in X_1$, and measurability of $x_1 \mapsto \int f(x_1, x_2)\,\mathrm{d}\mu_2(x_2)$ follows by (2.5). The first equality in (3.11) now follows by monotone convergence relative to $\mu_1 \otimes \mu_2$, as

$$\begin{aligned}\int f\,\mathrm{d}(\mu_1 \otimes \mu_2) &\overset{\substack{\text{monot.}\\ \text{conv.}}}{=} \lim_{n\to\infty} \int s_n\,\mathrm{d}(\mu_1 \otimes \mu_2) \overset{\substack{\text{(3.11) for}\\ s_n}}{=} \lim_{n\to\infty} \int \left(\int s_n(x_1, x_2)\,\mathrm{d}\mu_2(x_2) \right) \mathrm{d}\mu_1(x_1) \\ &\overset{\substack{\text{monot.}\\ \text{conv.}\\ \text{rel. } \mu_1 + (3.17)}}{=} \int \left(\int f(x_1, x_2)\,\mathrm{d}\mu_2(x_2) \right) \mathrm{d}\mu_1(x_1).\end{aligned}$$

Similar conclusions give the other halves of (3.10) and (3.11) when $f \geq 0$. For $f \in \mathcal{L}^1$ one simply writes $f = f^+ - f^-$ and applies the above to $f^\pm$ separately to conclude. $\square$

Chapter 4

Differentiation of measures

4.1 Differentiability of the Lebesgue Integral

Throughout this section let λ denote the Lebesgue measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, for $n \geq 1$, and write $L^p = L^p(\mathbb{R}^n)$, $p \geq 1$, for the corresponding L^p-spaces. Following standard convention, we will also frequently write $\mathrm{d}x, \mathrm{d}y, \ldots$ instead of $\mathrm{d}\lambda$.

As a motivation, consider the following:

Example 4.1. 1) Let $f : \mathbb{R} \to \mathbb{R}$ be continuous, $x_0 \in \mathbb{R}$,

$$F(x) \overset{\text{def.}}{=} \int_{[x_0,x]} f(y)\,\mathrm{d}y.$$

Then F is continuously differentiable and

$$F'(x) = \lim_{\varepsilon \downarrow 0} \varepsilon^{-1} \int_{[x,x+\varepsilon]} f(y)\,\mathrm{d}y$$

for all $x \in \mathbb{R}$ (cf. Sheet 5, Ex.3 a)).

2) Somewhat more generally, if $f \in C^0(\mathbb{R}^n) = \{g : \mathbb{R}^n \to \mathbb{R} : g \textit{ continuous}\}$ then

$$f(x) = \lim_{\varepsilon \downarrow 0} \frac{1}{\lambda(B(x,\varepsilon))} \int_{B(x,\varepsilon)} f(y)\,\mathrm{d}y, \quad \text{for all } x \in \mathbb{R}^n, \tag{4.1}$$

where $B(x,\varepsilon) = \{y \in \mathbb{R}^n : |y - x| < \varepsilon\}$ is the open ball of radius ε around x.

We ask whether formulae such as (4.1) remain true (λ-a.e.) if we only assume that $f \in L^1_{\text{loc}}(\mathbb{R}^n)$ (i.e. $f \in L^1(\Omega)$ for every bounded open set $\Omega \subset \mathbb{R}^n$). The main result of this section is

Theorem 4.2 (Lebesgue's Differentiation Theorem)**.** *Let $f \in L^1_{\text{loc}}(\mathbb{R}^n)$. Then*

$$f(x) = \lim_{r \downarrow 0} \frac{1}{\lambda(B(x,r))} \int_{B(x,r)} f(y)\,\mathrm{d}y \tag{4.2}$$

holds for λ-a.e. $x \in \mathbb{R}^n$.

Remark 4.3. One can replace balls by similar objects, e.g. cubes $(x-r, x+r) = \{y : x_i - r < y_i < x_i + r,\ 1 \le i \le n\}$.

For the proof of Theorem 4.2 we introduce the function

$$f^*(x) \overset{\text{def.}}{=} \sup_{r>0} \frac{1}{\lambda(B(x,r))} \int_{B(x,r)} |f(y)|\,\mathrm{d}y. \tag{4.3}$$

The function f^* is called **Hardy-Littlewood maximal function** (associated to $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$). By Proposition 2.32, for each $r > 0$ the function

$$x \mapsto \frac{1}{\lambda B(x,r)} \int_{B(x,r)} |f(y)|\,\mathrm{d}y$$

is continuous.[1] Hence f^* is measurable. Note that $f^* \notin L^1(\mathbb{R}^n)$ (except if $\|f\|_{L^1} = 0$). Indeed, choosing $r_0 > 0$ such that $\int_{B(0,r_0)} |f(y)|\,\mathrm{d}y = c_1 > 0$, one obtains for $|x| > r_0$

$$f^*(x) \ge \frac{1}{\lambda(B(x,|x|+r_0))} \int_{B(x,|x|+r_0)} |f(y)|\,\mathrm{d}y \ge^1 \frac{1}{c(|x|+r_0)^n} c_1 \ge c'|x|^{-n}$$

but $g(x) = |x|^{-n} \notin L^1(\mathbb{R}^n)$. (^{1} note $B(0,r_0) \subset B(x,|x|+r_0)$.)

[1]Let $\varepsilon > 0$, $x \in \mathbb{R}^n$ and call $\bar{f}(x)$ this function. Let δ be as supplied by Proposition 2.32 for μ be the Lebesgue measure restricted to $B(x, r+2)$. Then for any point y close enough to x, we have $\lambda(B(x,r)\triangle B(y,r)) < \delta$ and (2.48) applied to $A = B(x,r)\triangle B(y,r)$ readily yields $|\bar{f}(y) - \bar{f}(x)| < \varepsilon$ for such y. Here $A \triangle B = (A\setminus B)\cup(B\setminus A)$ denotes the symmetric difference of two sets.

Our main tool in proving Theorem 4.2 will be

Proposition 4.4 (Maximal Inequality). *Let $f \in L^1(\mathbb{R}^n)$. For all $a > 0$,*

$$\lambda(\{x \in \mathbb{R}^n : f^*(x) > a\}) \le \frac{5^n}{a} \|f\|_{L^1}. \tag{4.4}$$

We will need one more ingredient.

Lemma 4.5. *Let*

$$C^0_c(\mathbb{R}^n) = \{f : \mathbb{R}^n \to \mathbb{R},\ f \text{ is continuous and } \overline{\{x : f(x) \ne 0\}} \text{ is compact}\}.$$

Then $C^0_c(\mathbb{R}^n)$ is dense in $L^1(\mathbb{R}^n)$, i.e. for every $f \in L^1(\mathbb{R}^n)$, there exists a sequence $(f_n) \subset C^0_c(\mathbb{R}^n)$ such that $f_n \overset{L^1}{\to} f$ (i.e. $\|f_n - f\|_{L^1} \to 0$ as $n \to \infty$).

We will first prove Theorem 4.2 assuming Proposition 4.4 and Lemma 4.5 to hold.

Proof of Theorem 4.2: We assume without loss of generality that $f \in L^1(\mathbb{R}^n)$ (else consider $f\,1_{B(x,1)}$ instead of f in the sequel). By Lemma 4.5, there exists $(f_n) \subset C_c^0(\mathbb{R}^n)$ such that

$$\|f_n - f\|_{L^1} \to 0 \quad \text{as } n \to \infty. \tag{4.5}$$

We introduce the shorthand

$$⨍_{B(x,r)} = \frac{1}{\lambda(B(x,r))} \int_{B(x,r)}$$

and bound:

$$\begin{aligned}
\limsup_{r\to 0} \left| ⨍_{B(x,r)} f(y)\,\mathrm{d}y - f(x) \right| &\overset{(2.31)}{\leq} \limsup_{r\to 0} ⨍_{B(x,r)} |f(y) - f(x)|\,\mathrm{d}y \\
&\leq \limsup_{r\to 0} ⨍_{B(x,r)} |f(y) - f_n(y)|\,\mathrm{d}y + \limsup_{r\to 0} ⨍_{B(x,r)} |f_n(y) - f_n(x)|\,\mathrm{d}y + |f_n(x) - f(x)| \\
&\overset{(4.3)}{\leq} (f - f_n)^*(x) + |f_n(x) - f(x)|, \quad \text{for all } n \in \mathbb{N} \text{ and } x \in \mathbb{R}^n,
\end{aligned} \tag{4.6}$$

using that the second term in the penultimate line vanishes by (4.1) since f_n is continuous. Thus for $\varepsilon > 0$ we obtain that

$$\begin{aligned}
A_\varepsilon &\overset{\text{def.}}{=} \left\{ x \in \mathbb{R}^n : \limsup_{r\to 0} ⨍_{B(x,r)} |f(y) - f(x)|\,\mathrm{d}y > 2\varepsilon \right\} \\
&\overset{(4.6)}{\subset} \{x \in \mathbb{R}^n : (f - f_n)^*(x) > \varepsilon\} \cup \{x \in \mathbb{R}^n : |f_n(x) - f(x)| > \varepsilon\}, \quad \text{for all } n \in \mathbb{N}.
\end{aligned}$$

Hence applying subadditivity, (4.4) and Markov's inequality (Sheet 4, Ex.4)) it follows that for all $n \in \mathbb{N}$,

$$\lambda(A_\varepsilon) \leq \frac{c}{\varepsilon} \|f - f_n\|_{L^1} + \frac{1}{\varepsilon} \|f - f_n\|_{L^1} \xrightarrow{(4.5)} 0 \quad \text{as } n \to \infty.$$

Thus $\lambda(A_\varepsilon) = 0$ and the same is true of

$$A \overset{\text{def.}}{=} \left\{ x \in \mathbb{R}^n : \limsup_{r\to 0} ⨍_{B(x,r)} |f(y) - f(x)|\,\mathrm{d}y > 0 \right\} = \bigcup_{k=1}^{\infty} A_{\frac{1}{k}}$$

i.e. $\lambda(A) = 0$. Thus

$$\left| f(x) - \lim_{r\to 0} ⨍_{B(x,r)} f(y)\,\mathrm{d}y \right| \leq \lim_{r\to 0} ⨍_{B(x,r)} |f(y) - f(x)|\,\mathrm{d}y = 0$$

holds for λ-a.e. $x \in \mathbb{R}^n$. □

Remark 4.6. We have actually shown the following stronger statement than Theorem 4.2. If $f \in L^1_{\text{loc}}(\mathbb{R}^n)$, then for λ-a.e. $x \in \mathbb{R}^n$,

$$\lim_{r\to 0} ⨍_{B(x,r)} |f(y) - f(x)|\,\mathrm{d}y = 0. \tag{4.7}$$

A point $x \in \mathbb{R}^n$ satisfying (4.7) is called a **Lebesgue-point of** f.

Towards proving Proposition 4.4, we collect the following useful result.
For a ball $B = B(x,r)$, we write $\operatorname{diam} B = 2r$ for its diameter and $\widehat{B} = B(x,5r)$ for the concentric five times larger ball.

Lemma 4.7 (Vitali's Covering Lemma)**.**
Let $\mathcal{F}$ be a family of (Euclidean) balls in $\mathbb{R}^n$ satisfying $d_0 \overset{\text{def.}}{=} \sup_{B\in\mathcal{F}} \operatorname{diam} B < \infty$. Then there exists a countable family $\mathcal{G} \subset \mathcal{F}$ of pairwise disjoint balls such that

$$\bigcup_{B\in\mathcal{F}} B \subset \bigcup_{B\in\mathcal{G}} \widehat{B}. \tag{4.8}$$

Proof. We define inductively maximal disjoint families $\mathcal{G}_i \subset \mathcal{F}$ as follows.
For $i = 1$ choose

$$\mathcal{G}_1 \subset \left\{ B \in \mathcal{F} : (d_0 \geq) \operatorname{diam} B > \frac{d_0}{2} \overset{\text{def.}}{=} d_1 \right\}$$

of maximal size so that all balls in $\mathcal{G}_1$ are disjoint (note that $\mathcal{G}_1 \neq \emptyset$). For $i = 1, 2, \dots$ then define

$$\mathcal{G}_{i+1} \subset \left\{ B \in \mathcal{F} : d_i \geq \operatorname{diam} B > \frac{d_i}{2} \overset{\text{def.}}{=} d_{i+1},\ B \cap B' = \emptyset \text{ for all } B' \in \mathcal{G}_1 \cup \cdots \cup \mathcal{G}_i \right\}$$

maximal s.t. all balls in $\mathcal{G}_{i+1}$ are disjoint.
Let $\mathcal{G} = \bigcup_{i\geq 1} \mathcal{G}_i$. By construction, $\mathcal{G}$ consists of countably many pairwise disjoint balls in $\mathcal{F}$. It remains to argue that (4.8) holds. Let $B_0 \in \mathcal{F}$ and choose $i \in \mathbb{N}$ so that $d_i < \operatorname{diam} B_0 \leq d_{i-1}$. By maximality of $\mathcal{G}_i$, there exists $B \in \mathcal{G}_1 \cup \cdots \cup \mathcal{G}_i$ with $B \cap B_0 \neq \emptyset$. In particular, $\operatorname{diam} B > d_i$ and since $\operatorname{diam} B_0 \leq d_{i-1} = 2d_i$, it follows that $\widehat{B} \supset B_0$. □

Remark 4.8. In fact, we constructed $\mathcal{G}$ so that for every $B_0 \in \mathcal{F}$, there exists $B \in \mathcal{G}$ such that $B_0 \subset \widehat{B}$, which implies (4.8).

With Lemma 4.7 at hand, we can proceed to the proof of Proposition 4.4.

Proof of Proposition 4.4: Let $A = \{x \in \mathbb{R}^n : f^*(x) > a\}$ for a given $a > 0$. By definition of f^*, see (4.3), for every $x \in A$, there exists $r(x) > 0$ such that

$$⨍_{B(x,r(x))} |f(y)| \,\mathrm{d}y > a.$$

In particular, letting $\omega_n = \lambda(B(0,1))$ denote the (Lebesgue) volume of the unit ball in $\mathbb{R}^n$, we obtain for all $x \in A$,

$$a\omega_n r(x)^n = a\lambda(B(x,r(x))) \overset{(*)}{<} \int_{B(x,r(x))} |f(y)| \,\mathrm{d}y \leq \|f\|_{L^1(\mathbb{R}^n)}, \tag{4.9}$$

which implies a uniform upper bound on $\{r(x) : x \in A\}$:

$$\sup_{x\in A} r(x) \overset{(4.9)}{\leq} \left(a^{-1}\omega_n^{-1}\|f\|_{L^1(\mathbb{R}^n)}\right)^{\frac{1}{n}} \overset{\text{def.}}{=} \frac{d_0}{2}. \tag{4.10}$$

Consider the family $\mathcal{F} = \{B(x, r(x)) : x \in A\}$. Clearly $A \subset \bigcup_{B\in\mathcal{F}} B$, and Lemma 4.7 applies to $\mathcal{F}$ due to (4.10), yielding the existence of a countable family of pairwise disjoint balls $\mathcal{G} \subset \mathcal{F}$ such that

$$(A \subset) \quad \bigcup_{B\in\mathcal{F}} B \subset \bigcup_{B\in\mathcal{G}} \widehat{B}. \tag{4.11}$$

This yields

$$\lambda(A) \overset{(4.11)}{\leq} \lambda\left(\bigcup_{B\in\mathcal{G}} \widehat{B}\right) \overset{\mathcal{G}\text{ countable}}{\leq} \sum_{B\in\mathcal{G}} \lambda(\widehat{B}) = 5^n \sum_{B\in\mathcal{G}} \lambda(B) \underset{\substack{B\in\mathcal{F} \\ B\text{ disjoint}}}{\overset{(4.9)(*)}{\leq}} \frac{5^n\|f\|_{L^1}}{a},$$

as desired. □

Remark 4.9. The statements of Proposition 4.4 (and of Theorem 4.2) can be generalised to measure spaces $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mu)$ such that $\mu(K) < \infty$ for all compact sets $K \subset \mathbb{R}^n$ (such measures are called **Radon measures** on $\mathbb{R}^n$) satisfying a **volume-doubling property**: there exists $c \in (0, \infty)$ such that:

$$\mu(B(x, 2r)) \leq c\mu(B(x, r)) \quad \text{for all } x \in \mathbb{R}^n,\ r > 0.$$

We now give an elementary proof of Lemma 4.5.

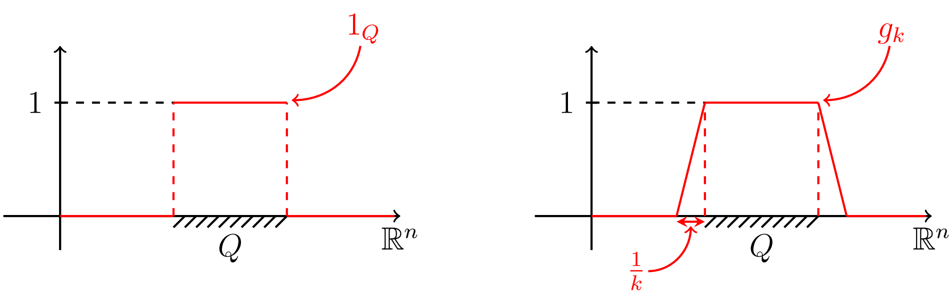
Proof of Lemma 4.5: Consider the (countable) set of cubes $\mathcal{C}$, which consists of all dyadic cubes $Q = x + [0, 2^{-l})$, for some $x \in 2^{-l}\mathbb{Z}^n (\subset \mathbb{R}^n)$ and $l \in \{0, 1, 2, \ldots\}$. Define the set of functions

$$E = \left\{ f = \sum_{k=1}^{N} a_k 1_{Q_k} : N \in \mathbb{N},\ a_k \in \mathbb{Q},\ Q_k \in \mathcal{C} \text{ for all } k = 1, \ldots, N \right\} \ \big(\subset L^1(\mathbb{R}^n)\big). \tag{4.12}$$

We will show that E is dense in $L^1(\mathbb{R}^n)$, i.e.

$$\forall f \in L^1(\mathbb{R}^n),\ \exists (f_k) \subset E \text{ s.t. } \|f - f_k\|_{L^1(\mathbb{R}^n)} \xrightarrow{k\to\infty} 0. \tag{4.13}$$

Since E is a countable set, (4.13) means that $L^1(\mathbb{R}^n)$ is **separable**. Once (4.13) is shown, the claim follows by noticing that for any cube $Q \in \mathcal{C}$ there exist functions $g_k \in C_c^0(\mathbb{R}^n)$ such that $g_k \overset{L^1}{\to} 1_Q$:



We now show (4.13). We may assume that $f \geq 0$. Indeed, if (4.13) holds for such f, then for general f, writing $f = f^+ - f^-$ and denoting by $(f_k^\pm) \subset E$ the sequences such that $f_k^\pm \xrightarrow{L^1} f^\pm$, one has that $f_k \overset{\text{def.}}{=} f_k^+ - f_k^- \in E$ and $\|f - f_k\|_{L_1} \leq \|f^+ - f_k^+\|_1 + \|f^- - f_k^-\|_1 \to 0$ as $k \to \infty$ (Note: $\|\cdot\|_1$ abbrv. $= \|\cdot\|_{L^1(\mathbb{R}^n)}$). If $f \geq 0$, combining Theorem 2.6 and Theorem 2.17(ii), we know that there exists a sequence (f_k) of simple functions such that $f_k \xrightarrow{L^1} f$. Hence, it is enough to argue that

$$\forall \varepsilon > 0,\ \forall A \in \mathcal{B}(\mathbb{R}^n) : \lambda(A) < \infty\ \exists f \in E : \|f - 1_A\|_{L^1(\mathbb{R}^n)} < \varepsilon. \tag{4.14}$$

Let $\varepsilon > 0$ and A be as in (4.14). Applying Proposition 1.21, we find an open set $G \supset A$ such that $\lambda(G \backslash A) < \frac{\varepsilon}{2}$, whence

$$\|1_G - 1_A\|_{L^1(\mathbb{R}^n)} = \int |1_G - 1_A| \,\mathrm{d}\lambda = \lambda(G \backslash A) = \frac{\varepsilon}{2}. \tag{4.15}$$

By the construction below (1.37), there exists a sequence $(Q_k) \subset \mathcal{C}$ of disjoint cubes such that $G = \bigcup_{k=1}^\infty Q_k$. In particular, $f_k \overset{\text{def.}}{=} \sum_{\ell=1}^k 1_{Q_\ell} \in E$ and

$$\|1_G - f_k\|_{L^1(\mathbb{R}^n)} \to 0 \quad \text{as } k \to \infty. \tag{4.16}$$

Choosing k_0 in (4.16) such that $\|1_G - f_{k_0}\|_1 < \frac{\varepsilon}{2}$ and combining with (4.15), (4.14) follows with $f = f_{k_0} \in E$. □

4.2 Lebesgue Decomposition and Radon-Nikodym Theorem

Let $(X, \mathcal{A}, \mu)$ be a σ-finite measure space, $f : X \to [0, \infty]$ measurable. Recall from (2.46) that $\nu = f\mu$ (i.e. $\nu(A) = \int_A f \,\mathrm{d}\mu$) defines a measure on $(X, \mathcal{A})$; indeed $\nu(\emptyset) = 0$ and σ-additivity follows from the monotone convergence theorem. Observe that $\nu \ll \mu$ (ν is absolutely continuous w.r.t. μ) i.e. $\mu(A) \to \nu(A) = 0$, $A \in \mathcal{A}$. In this section we investigate to which extent any absolutely continuous measure $\nu \ll \mu$ is of the form $\nu = f\mu$ for a "density" f.

We begin with a topic that may at first seem unrelated.

Definition 4.10. A map $\alpha : \mathcal{A} \to (-\infty, \infty]$ is called a **signed measure** if $\alpha(\emptyset) = 0$ and

$$\text{for all } (A_n) \subset \mathcal{A},\ A_n \cap A_m = \emptyset \text{ for } n \neq m : \alpha\left(\bigcup_{n=1}^\infty A_n\right) = \sum_{n=1}^\infty \alpha(A_n) \tag{4.17}$$

in the sense that if $\alpha(\bigcup_{n=1}^\infty A_n) < \infty$, the series converges absolutely, and if $\alpha(\bigcup_{n=1}^\infty A_n) = \infty$ then $\sum_{n=1}^\infty(-\alpha(A_n) \vee 0) < \infty$ and $\sum_{n=1}^\infty(\alpha(A_n) \vee 0) = \infty$.

Example 4.11. 1) Let μ be a measure and f be a function with $\int f^- \,\mathrm{d}\mu < \infty$. Let $\alpha(A) = \int_A f \,\mathrm{d}\mu \stackrel{\text{def.}}{=} \int_A f^+ \,\mathrm{d}\mu - \int_A f^- \,\mathrm{d}\mu$. Then α is a signed measure.

2) Let μ_1, μ_2 be measures on $(X, \mathcal{A})$ with $\mu_2(X) < \infty$. Define

$$\mu(A) = \mu_1(A) - \mu_2(A), \ A \in \mathcal{A}. \tag{4.18}$$

Then μ is a signed measure. We will soon see that every signed measure is of the form (4.18).

Definition 4.12. Let α be a signed measure on $(X, \mathcal{A})$. A set $A \in \mathcal{A}$ is called **positive** if

$$\forall B \subset A, \ B \in \mathcal{A} : \alpha(B) \geq 0. \tag{4.19}$$

A set $A \in \mathcal{A}$ is **negative** if

$$\forall B \subset A, \ B \in \mathcal{A} : \alpha(B) \leq 0. \tag{4.20}$$

Example 4.11: 1) continued. In this example,

$$A \in \mathcal{A} \text{ is positive} \quad \Longleftrightarrow \quad \mu(A \cap \{f < 0\}) = 0 \ (= \mu(A \cap \{f^- > 0\})). \tag{4.21}$$

"$\Leftarrow$" For $B \subset A$, one has $\alpha(B) = \int_B f^+ \,\mathrm{d}\mu - \int_B f^- \,\mathrm{d}\mu = \int_B f^+ \,\mathrm{d}\mu \geq 0$.
"$\Rightarrow$" If A is positive then $B \stackrel{\text{def.}}{=} A \cap \{f < 0\} \subset A$ is measurable and $0 \leq \alpha(B) = -\int_B f^- \,\mathrm{d}\mu$, i.e. $\int_B f^- \,\mathrm{d}\mu = 0$. This implies $f = f^+$ μ-a.e. or $\mu(B) = 0$.

Our main result about signed measures is

Theorem 4.13 (Hahn Decomposition)**.** *Let α be a signed measure on $(X, \mathcal{A})$. There exist a positive set A and a negative set B such that $X = A \cup B$ and $A \cap B = \emptyset$.*

Remark 4.14. 1) A Hahn decomposition is not unique. For instance in Example 4.11,1), using (4.21), A can be any set satisfying $\{f > 0\} \subset A \subset \{f \geq 0\}$ μ-a.e., where $B \subset C$ μ-a.e. means $\mu(B \cap C^c) = 0$.

2) If (A_1, B_1) and (A_2, B_2) are two Hahn decompositions for α, then

$$A_1 \triangle A_2 = (A_1 \cap A_2^c) \cup (A_2 \cap A_1^c) = (B_1^c \cap B_2) \cup (B_2^c \cap B_1) = B_1 \triangle B_2,$$

so $\alpha(A_1 \triangle A_2) \geq 0$ as $A_1 \triangle A_2 \subset (A_1 \cup A_2)$, which is positive by (4.23) below, and similarly $\alpha(A_1 \triangle A_2) \leq 0$ as $A_1 \triangle A_2 \subset (B_1 \cup B_2)$. Overall: $\alpha(A_1 \triangle A_2) = \alpha(B_1 \triangle B_2) = 0$.

The proof of the Theorem 4.13 builds on two lemmas.

Lemma 4.15 (Basic properties of positive sets; α a signed measure on $(X, \mathcal{A})$)**.**

(4.22) If A *is positive and* $B \in \mathcal{A},\ B \subset A$ then B *is positive.*

(4.23) If $A_1, A_2, \dots$ *are positive then* $A = \bigcup_{n=1}^{\infty} A_n$ *is positive.*

The conclusions (4.22), (4.23) *remain valid upon replacing positive by negative everywhere.*

Proof. (4.22) follows immediately from (4.19). To show (4.23), given $A_1, A_2, \dots$ positive sets consider $B_1, B_2, \dots$ defined as

$$B_1 = A_1,\ B_n = A_n \backslash \left(\bigcup_{k=1}^{n-1} A_k \right),\ n \geq 2.$$

Since $B_n \subset A_n$ for all n, the sets B_n, $n \in \mathbb{N}$, are positive by (4.22). Moreover they are disjoint by construction and $\bigcup_{n=1}^{\infty} B_n = A$. Hence, for all $B \subset A$,

$$\alpha(B) = \alpha(B \cap A) = \alpha\left(\bigcup_{n=1}^{\infty} B_n \cap B \right) \overset{(4.17)}{=} \sum_{n=1}^{\infty} \underbrace{\alpha(B_n \cap B)}_{\geq 0} \geq 0.$$

The adaptation to negative sets is straightforward. □

Lemma 4.16 (α a signed measure on $(X, \mathcal{A})$)**.** *Let* $B \in \mathcal{A}$ *satisfy* $\alpha(B) < 0$*. Then there exists a negative set* $B' \subset B$*,* $B' \in \mathcal{A}$ *with* $\alpha(B') < 0$*.*

Proof. We remove "chunks" (i.e. sets) spoiling negativity as follows. Let

$$\varepsilon_1 = \sup\{\alpha(A) : A \in \mathcal{A},\ A \subset B\} \in [0, \infty]. \tag{4.24}$$

By (4.24) we can find $A_1 \in \mathcal{A}$, $A_1 \subset B$ with $\alpha(A_1) \geq \frac{\varepsilon_1}{2} \wedge 1$. Proceeding inductively, assuming $\varepsilon_1, A_1, \dots, \varepsilon_{n-1}, A_{n-1}$ have been defined for some $n \geq 2$, let

$$\varepsilon_n \overset{\text{def.}}{=} \sup\left\{ \alpha(A) : A \in \mathcal{A},\ A \subset B \backslash \Big(\bigcup_{i=1}^{n-1} A_i \Big) \right\} \tag{4.25}$$

and choose $A_n \subset B \backslash \left(\bigcup_{i=1}^{n-1} A_i \right)$ such that $\alpha(A_n) \geq \frac{\varepsilon_n}{2} \wedge 1$. Now define

$$B' = B \backslash \Big(\bigcup_{n=1}^{\infty} A_n \Big). \tag{4.26}$$

Observe that the A_n's are disjoint, $\alpha(A_n) \geq 0$, and $\bigcup_{n=1}^{\infty} A_n \subset B$. Hence

$$0 > \alpha(B) = \alpha(B') + \sum_{n=1}^{\infty} \alpha(A_n),$$

which implies that $\alpha(B') < 0$ (since $a(A_n) \geq 0$) as desired. Moreover $\sum_{n=1}^{\infty} \alpha(A_n) < \infty$ since $\alpha(B') > -\infty$. In particular this implies that

$$\varepsilon_n \leq 2\alpha(A_n) \to 0 \text{ as } n \to \infty.$$

Now let $A \subset B'$. then for every n, by (4.26), $A \subset B \backslash (\bigcup_{k=1}^{n-1} A_k)$, thus by (4.25),

$$\alpha(A) \leq \varepsilon_n, \text{ for all } n. \tag{4.27}$$

By (4.27) and since $\varepsilon_n \to 0$, $\alpha(A) \leq 0$, thus B' is negative. □

Proof of Theorem 4.13: Let

$$b = \inf\{\alpha(B) : B \in \mathcal{A},\ B \ \textit{ is a negative set}\} \quad (\in [-\infty, 0] \text{ as } \emptyset \text{ is negative.}) \tag{4.28}$$

By definition, there exist B_n negative sets, $n \geq 1$ such that $\alpha(B_n) \downarrow b$. By (4.23), $B \overset{\text{def.}}{=} \bigcup_{n=1}^{\infty} B_n$ is negative. In particular, due to (4.28), $b \leq \alpha(B)$. Moreover for every n,

$$\alpha(B) \overset{B_n \subseteq B}{=} \alpha(B_n) + \alpha(B \backslash B_n) \overset{B \text{ neg.}}{\leq} \alpha(B_n),$$

and letting $n \to \infty$ and since $\alpha(B_n) \to b$, we deduce that $\alpha(B) \leq b$. Thus, overall,

$$b = \alpha(B). \tag{4.29}$$

In particular, $b > -\infty$ by (4.29) and Def. 4.10. Now let $A \overset{\text{def.}}{=} B^c$. If $A' \subset A$, $A' \in \mathcal{A}$ and $\alpha(A') < 0$, then by Lemma 4.16, A' has a subset A'' which is negative. But then $B \cup A''$ is negative by (4.22) and since A'' and B are disjoint,

$$\alpha(A'' \cup B) = \alpha(A'') + \alpha(B) \overset{\alpha(A'')<0}{<} \alpha(B) = b,$$

which contradicts the minimality of b in (4.28). Hence there exists no such A', i.e. A is positive. □

Definition 4.17. Two measures μ,ν on $(X, \mathcal{A})$ are **mutually singular** if there exists $A \in \mathcal{A}$ such that $\mu(A) = 0$ and $\nu(A^c) = 0$. Often, one also says in this case that μ is singular w.r.t. ν (and vice versa) and writes $\mu \perp \nu$ ($\perp$ is symmetric).

Example 4.18. The measure δ_x, $x \in \mathbb{R}$, and λ, the Lebesgue measure on $\mathbb{R}$, are mutually singular. Indeed, choosing $A = \{x\}$ one has $\delta_x(\mathbb{R} \backslash A) = 0$ whereas $\lambda(A) = 0$.

The following result shows that all signed measures are of the form (4.18).

Theorem 4.19 (Jordan Decomposition)**.** *Let α be a signed measure on $(X, \mathcal{A})$. There exists a unique pair of (α_+, α_-) of mutually singular (positive) measures on $(X, \mathcal{A})$ such that*

$$\alpha = \alpha_+ - \alpha_-.$$

Proof. Let (A, B) be a Hahn decomposition for α and define

$$\alpha_+(E) = \alpha(E \cap A),\ \alpha_-(E) = -\alpha(E \cap B),\ E \in \mathcal{A}. \tag{4.30}$$

Since $A \cup B = X$ and A, B disjoint, it follows that $\alpha_+ - \alpha_- = \alpha$. Moreover since A and B are positive, resp. negative, (4.30) implies that $\alpha_\pm$ are (positive) measures, and $\alpha_+(A^c) = 0$ whereas $\alpha_-(A) = \alpha_-(B^c) = 0$, so the measures are singular.

It remains to show uniqueness. Suppose $\alpha = \beta_+ - \beta_-$ with $\beta_\pm$ mutually singular. Let D be such that $\beta_+(D) = 0 = \beta_-(D^c)$. Due to this and since $\beta_\pm$ are measures, it follows that D is negative and $C \overset{\text{def.}}{=} D^c$ is positive, i.e. (C, D) is a Hahn decomposition for α. By Remark 4.14,2), $\alpha(A \triangle C) = 0$ hence for all $E \in \mathcal{A}$,

$$\alpha_+(E) = \alpha(E \cap A) = \alpha(E \cap (A \cup C)) = \alpha(E \cap C) \overset{\beta_-(C)=0}{=} \beta_+(E)$$

and similarly $\alpha_-(E) = \beta_-(E)$, yielding the asserted uniqueness. □

We now proceed to the main decomposition result of this section.

Theorem 4.20 (Lebesgue Decomposition). *Let μ, ν be two (positive) σ-finite measures on $(X, \mathcal{A})$. Then there exists a unique pair of measures (ν_s, ν_{ac}), the Lebesgue decomposition of ν w.r.t. μ, with the following properties:*

$$\nu = \nu_s + \nu_{ac} \quad \text{and} \tag{4.31}$$

$$\nu_s \perp \mu,\ \nu_{ac}(A) = \int_A f \, \mathrm{d}\mu,\ \textit{for some } f : X \to [0, \infty] \text{ measurable.} \tag{4.32}$$

Remark 4.21. ν_s is often called the singular part of the decomposition, ν_{ac} the absolutely continuous one. Indeed in view of (2.47) and (4.32), one has that $\nu_{ac} \ll \mu$.

The proof of Theorem 4.20 will rely on the following lemma.

Lemma 4.22. *Let μ, ν be two positive finite measures on $(X, \mathcal{A})$. Then either $\mu \perp \nu$ or there exist $\varepsilon > 0$ and $E \in \mathcal{A}$ such that $\nu(E) > 0$ and E is positive for $\mu - \varepsilon\nu$.*

Proof. For $n \in \mathbb{N}$, let (A_n, B_n) be a Hahn decomposition for $\mu - \frac{1}{n}\nu$. Set $A = \bigcup_{n=1}^\infty A_n$, $B = \bigcap_{n=1}^\infty B_n$. Since $B \subset B_n$ for each n, B is a negative set for $\mu - \frac{1}{n}\nu$, i.e.

$$\mu(B) \leq \frac{1}{n}\nu(B), \text{ for all } n \in \mathbb{N}. \tag{4.33}$$

Since ν is finite, letting $n \to \infty$ in (4.33), it follows that $\mu(B) = 0$. If $\mu \not\perp \nu$ then $\nu(B^c) = \nu(A) > 0$, whence $\nu(A_{n_0}) > 0$ for some $n_0 \in \mathbb{N}$. Choosing $\varepsilon = \frac{1}{n_0}$, $E = A_{n_0}$, it follows that E is positive for $\mu - \varepsilon\nu$. □

Proof of Theorem 4.20: We first show existence. We may assume that μ and ν are finite. Indeed in the general case, writing $X = \bigcup_{i=1}^\infty S_i$ for disjoint sets $S_i \in \mathcal{A}$

with $\mu(S_i), \nu(S_i) < \infty$ (the sets S_i exist by σ-finiteness), letting $\mu_i(\cdot) = \mu(\,\cdot \cap S_i)$, $\nu_i(\cdot) = \nu(\,\cdot \cap S_i)$, which are finite measures, and decomposing $\nu_i = \nu_{i,s} + \nu_{i,ac}$, with $\nu_{i,ac}(A) = \int_A f_i \, \mathrm{d}\mu_i$, one obtains that

$$\nu_s \stackrel{\text{def.}}{=} \sum_{i=1}^{\infty} \nu_{i,s},\ \nu_{ac} = \sum_{i=1}^{\infty} \nu_{i,ac} \text{ and } f = \sum_{i=1}^{\infty} f_i$$

satisfy (4.31) and (4.32).

With μ and ν finite, we introduce

$$\mathcal{G} = \left\{ g : X \to [0,\infty] \textit{ meas.} : \int_A g \, \mathrm{d}\mu \le \nu(A) \text{ for all } A \in \mathcal{A} \right\}. \tag{4.34}$$

We will choose f in (4.32) as the "largest" element in $\mathcal{G}$.

We first make one observation about $\mathcal{G}$:

$$g, h \in \mathcal{G} \Longrightarrow g \vee h \in \mathcal{G}. \tag{4.35}$$

Indeed if $g, h \in \mathcal{G}$ then for all $A \in \mathcal{A}$,

$$\begin{aligned} \int_A (g \vee h) \, \mathrm{d}\mu &= \int_{A \cap \{g > h\}} g \, \mathrm{d}\mu + \int_{A \cap \{g \le h\}} h \, \mathrm{d}\mu \\ &\stackrel{(4.34)}{\le} \nu(A \cap \{g > h\}) + \nu(A \cap \{g \le h\}) = \nu(A). \end{aligned}$$

Now using that $0 \in \mathcal{G}$ and that ν is finite, it follows that

$$M \stackrel{\text{def.}}{=} \sup \left\{ \int g \, \mathrm{d}\mu : g \in \mathcal{G} \right\} \in [0, \infty). \tag{4.36}$$

(Indeed for every $g \in \mathcal{G}$, one has $\int g \, \mathrm{d}\mu \le \nu(X) < \infty$ by (4.34).) By definition of M in (4.36), there exists $(g_n) \subset \mathcal{G}$ such that

$$\lim_{n \to \infty} \int g_n \, \mathrm{d}\mu = M. \tag{4.37}$$

Set $f_n = g_1 \vee \cdots \vee g_n$ for all $n \in \mathbb{N}$. Then $f_n \in \mathcal{G}$ by (4.35) and (f_n) is monotone increasing, hence $f \stackrel{\text{def.}}{=} \lim_{n \to \infty} f_n$ exists, and by monotone convergence,

$$\int_A f \, \mathrm{d}\mu = \lim_{n \to \infty} \int_A f_n \, \mathrm{d}\mu \stackrel{\substack{f_n \in \mathcal{G} \\ g_n \in \mathcal{G}}}{\le} \nu(A), \text{ for all } A \in \mathcal{A}, \tag{4.38}$$

i.e.

$$f \in \mathcal{G}. \tag{4.39}$$

By (4.38),

$$\nu_s \stackrel{\text{def.}}{=} \nu - f\,\mathrm{d}\mu$$

is a (positive) measure, and it remains to argue that $\nu_s \perp \mu$. If not, then by Lemma 4.22 there exists $E \in \mathcal{A}$ with $\mu(E) > 0$ such that $\nu - f\,\mathrm{d}\mu - \varepsilon 1_E\,\mathrm{d}\mu \geq 0$ on E, i.e. for all $A \in \mathcal{A}$

$$\int_{A\cap E} (f+\varepsilon)\,\mathrm{d}\mu \leq \nu(A\cap E). \tag{4.40}$$

But this implies in turn that for all $A \in \mathcal{A}$,

$$\int_A (f+\varepsilon 1_E)\,\mathrm{d}\mu = \int_{A\cap E} (f+\varepsilon)\,\mathrm{d}\mu + \int_{A\cap E^c} f\,\mathrm{d}\mu \stackrel{(4.39)\,(4.40)}{\leq} \nu(A)$$

whence $f + \varepsilon 1_E \in \mathcal{G}$. But by (4.37) and (4.39) $\int f\,\mathrm{d}\mu =^1 M$, and

$$\int (f+\varepsilon 1_E)\,\mathrm{d}\mu = M + \varepsilon\mu(E) > M,$$

which by (4.36) contradicts the fact that $f + \varepsilon 1_E \in \mathcal{G}$. Hence $\nu_s \perp \mu$.

We now show uniqueness. Assume $\nu = \nu_s' + \nu_{ac}'$ with ν_s', ν_{ac}' having the properties (4.31), (4.32). Then $\nu_s - \nu_s' = \nu_{ac}' - \nu_{ac}$ holds as an equality between signed measures. Since $\nu_s \perp \mu$ and $\nu_s' \perp \mu$, there exists[2] $B \in \mathcal{A}$ such that $\nu_s(B^c) = \nu_s'(B^c) = 0$ and $\mu(B) = 0$, whence

$$(\nu_s - \nu_s')(A) = (\nu_s - \nu_s')(A\cap B) = (\nu_{ac}' - \nu_{ac})(A \cap B) = 0,\ A \in \mathcal{A},$$

where the last equality follows because $\mu(A\cap B) = 0$ and ν_{ac}', ν_{ac} are both absolutely continuous w.r.t. μ.
Note: [1]because

$$\int g_n\,\mathrm{d}\mu \leq \int f_n\,\mathrm{d}\mu \stackrel{\forall n}{\leq} \int f\,\mathrm{d}\mu \leq M \quad \text{and} \quad \int g_n\,\mathrm{d}\mu \to M.$$

[2]There exists $B_1, B_2 \in \mathcal{A}$ such that $\nu_s(B_1^c) = 0$ and $\nu_s'(B_2^c) = 0$ with $\mu(B_1) = \mu(B_2) = 0$. Then take $B = B_1 \cup B_2$. □

As a final consequence of the Lebesgue decomposition we obtain the desired characterisation of absolute continuity (cf. Def. 2.31).

Corollary 4.23 (Radon-Nikodym Theorem)**.**

Let μ, ν be two σ-finite measures on $(X, \mathcal{A})$. The following are equivalent:

(i) $\nu \ll \mu$;

(ii) $\nu = f\mu$ *for some measurable function* $f : X \to [0, \infty]$.

Moreover if $g \geq 0$ is another function such that $\nu = g\mu$ then $f = g$ μ-a.e. One writes $f = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}$, which is called the **Radon-Nikodym derivative** *of ν with respect to μ.*

Proof. (ii) $\Rightarrow$ (i): see Remark 2.33,1).
Assume (i) holds. Let $\nu = \nu_s + \nu_{ac}$ be the Lebesgue decomposition of ν w.r.t. μ supplied by Theorem 4.20. Since $\nu_s \perp \mu$, there exists $A \in \mathcal{A}$ such that $\nu_s(A^c) = 0$ and $\mu(A) = 0$. Thus

$$0 = \mu(A) \overset{\nu \ll \mu}{=} \nu(A) \overset{(4.31)}{\geq} \nu_s(A),$$

whence $\nu_s(X) = 0$, i.e. $\nu_s \equiv 0$. Thus, $\nu = \nu_{ac}$ and (ii) follows from (4.32). The fact that $f = g$ μ-a.e. if $\nu = g\,\mathrm{d}\mu$ follows from (iii) on p.31. $\square$

That's it.

Happy Holidays!