Numerical Analysis of Stochastic Ordinary Differential Equations

Andrea Barth, Arnulf Jentzen,
Annika Lang, and Christoph Schwab

December 15, 2018

Preface to the first version

These lecture notes originate from the master course *Numerical Analysis of Stochastic Ordinary Differential Equations* — also called *Computational Methods in Quantitative Finance I* — which was taught at ETH Zürich from 2008 to 2011. It includes a review on probability theory, basics on the generation of random numbers, and an introduction to Monte Carlo methods in the first part. Then stochastic processes and stochastic differential equations are introduced and solutions are approximated with strong and weak approximation schemes.

Zürich, December 2011

Andrea Barth, Annika Lang, Christoph Schwab

Preface to the second version

These lectures notes are a rewritten version of the lecture notes written by Andrea Barth, Annika Lang and Christoph Schwab. They have been written for the course "401-4657-00L Numerical Analysis of Stochastic Ordinary Differential Equations" in the Autumn Semester 2012. These lectures notes would be in a much worser shape without the help of a number of people. I am particulary indebted to Sonja Cox and Raphael Kruse for various valuable suggestions, for proofreading the lecture notes and for their very helpful advice. Moreover, I am very grateful to Martin Hutzenthaler for his permission to use parts of the material in [Hutzenthaler and Jentzen(2012)] for these lecture notes. Finally, special thanks are due to the students of the course "401-4657-00L Numerical Analysis of Stochastic Ordinary Differential Equations" for pointing out a number of misprints and for various useful remarks and questions that helped to improve and correct these lecture notes.

Zürich, December 2012

Arnulf Jentzen

Preface to the third version

Special thanks are due to Sonja Cox and Raphael Kruse for placing a number of excercises and their permission to use these excercises as a part of these lecture notes. Stefan Geiss, Johannes Muhle-Karbe and Josef Teichmann are also gratefully acknowledged for a number of insightful comments concerning mathematical finance. The students of the course "401-4657-00L Numerical Analysis of Stochastic Ordinary Differential Equations" are also gratefully acknowledged for pointing out a number of misprints and for asking a number of questions that helped to improve these lecture notes. Finally, I am particulary indebted to Florian Müller-Reiter from the swissQuant Group AG for a number of quite instructive demonstrations from the financial pratice.

These lecture notes are still under construction. In particular, these lecture notes do not yet contain an appropriate classification and an appropriate comparison of the presented material with the relevant material from the literature. This will be the subject of a later version of these lecture notes.

Zürich, December 2013

Arnulf Jentzen

Preface to the fourth version

These lecture notes are still under construction. In particular, these lecture notes do not yet contain an appropriate classification and an appropriate comparison of the presented material with the relevant material from the literature. This will be the subject of a later version of these lecture notes. Special thanks are due to Lukas Herrmann and Ryan Kurniawan for their substantial help with the solutions of the exercises.

Zürich, September 2014

Arnulf Jentzen

Preface to the fifth version

These lecture notes are still under construction. In particular, these lecture notes do not yet contain an appropriate classification and an appropriate comparison of the presented material with the relevant material from the literature. This will be the subject of a later version of these lecture notes.

Zürich, September 2015

Arnulf Jentzen

Preface to the sixth version

These lecture notes are still under construction. In particular, these lecture notes do not yet contain an appropriate classification and an appropriate comparison of the presented material with the relevant material from the literature. This will be the subject of a later version of these lecture notes.

Zürich, September 2015

Arnulf Jentzen

Exercises

Exercises series

Solutions to the exercises can be handed in before the start of the lecture or in the designated mailbox in front of room HG G 53.2. Please also submit your Matlab code and your figures at `https://sam-up.math.ethz.ch`.

Series	Exercises	Deadline	Solutions
1	Exercises 0.2.8, 0.2.43, 0.4.7, & 0.4.14	03.10.2018, 13:15 PM	Chapter 8
2	Exercises 1.2.16, 1.2.24, 1.2.25, & 1.2.26	10.10.2018, 13:15 PM	Chapter 8
3	Exercises 1.2.27, 1.2.30, 1.2.38, & 1.2.39	17.10.2018, 13:15 PM	Chapter 8
4*	Exercises 1.3.7, 1.3.9, 1.3.11, & 2.1.12	24.10.2018, 13:15 PM	Chapter 8
5*	Exercises 2.1.13, 2.2.8, 2.2.10, & 2.2.13	31.10.2018, 13:15 PM	Chapter 8
6*	Exercises 2.3.9, 2.3.10, 2.4.10, & 3.1.9	07.11.2018, 13:15 PM	Chapter 8
7*	Exercises 3.1.10, 3.2.23, 3.3.9, & 3.3.10	14.11.2018, 13:15 PM	Chapter 8
8	Exercises 3.3.11, 3.3.15, 3.4.20, & 3.4.22	21.11.2018, 13:15 PM	Chapter 8
9*	Exercises 5.2.4, 5.2.7, 5.3.3, & 5.3.4	05.12.2018, 13:15 PM	Chapter 8
10*	Exercises 5.5.6, 5.5.9, 5.6.7, & 5.6.8	12.12.2018, 13:15 PM	Chapter 8

(..)* The PhD students of the Department of Mathematics of ETH Zurich must solve this exercise series successfully to get the credit points of the course.

Rules for the Matlab exercises

For each Matlab exercise we expect that you hand in:

(i) A printout of the .m-file containing the Matlab commands.

(ii) A printout of the exact output of your .m-file (where all outputs are properly labelled). This printout should also contain all graphics produced by your .m-file. Your .m-file should only output those values for which you are asked for in the Matlab exercise. You can use ";" at the end of your Matlab code lines to suppress the output of intermediate results.

In addition, we also expect you to submit your .m-file at `https://sam-up.math.ethz.ch`.

Contents

0 Preliminaries from measure and probability theory

These lecture notes use a number of concepts and results from measure and probability theory. Some of these notions and results are briefly reviewed in this preliminary chapter (see Sections 0.2, 0.3, and 0.4 below). Further concepts and results from measure and probability theory can, for example, be found in [Klenke(2008)] and [Bauer(1991)]. In addition, several important probability distributions from the literature are considered in Section 0.4 of this preliminary chapter.

0.1 Basic notation

Definition 0.1.1 (Set of numbers)**.** *We denote by* $\mathbb{R}$ *the set of real numbers, we denote by* $\mathbb{C}$ *the set of complex numbers, and we denote by* $\mathbb{N}$*,* $\mathbb{N}_0$*, and* $\bar{\mathbb{R}}$ *the sets given by*

$$\begin{aligned} \mathbb{N} &= \{1, 2, \dots\}, \\ \mathbb{N}_0 &= \mathbb{N} \cup \{0\} = \{0, 1, 2, \dots\}, \\ \bar{\mathbb{R}} &= \mathbb{R} \cup \{-\infty\} \cup \{\infty\}. \end{aligned} \tag{0.1}$$

Definition 0.1.2 (Intervals of extended real numbers)**.** *Let* $a, b \in \bar{\mathbb{R}}$*. Then we denote by* $[a, b]$*,* $(a, b]$*,* $[a, b)$*, and* (a, b) *the sets given by*

$$\begin{aligned} [a, b] &= \{x \in \bar{\mathbb{R}} \colon a \leq x \leq b\}, \\ (a, b] &= \{x \in \bar{\mathbb{R}} \colon a < x \leq b\}, \\ [a, b) &= \{x \in \bar{\mathbb{R}} \colon a \leq x < b\}, \\ (a, b) &= \{x \in \bar{\mathbb{R}} \colon a < x < b\}. \end{aligned} \tag{0.2}$$

Definition 0.1.3 (Absolute value)**.** *We denote by* $|\cdot|_{\mathbb{R}} : \mathbb{R} \to [0, \infty)$ *the function which satisfies for all* $a \in [0, \infty)$ *that*

$$|a|_{\mathbb{R}} = |-a|_{\mathbb{R}} = a, \tag{0.3}$$

we denote by $|\cdot|_{\mathbb{C}} : \mathbb{C} \to [0, \infty)$ *the function which satisfies for all* $a, b \in \mathbb{R}$ *that*

$$|a + \mathbf{i}b|_{\mathbb{C}} = \sqrt{a^2 + b^2}, \tag{0.4}$$

and we denote by $|\cdot| : \mathbb{C} \to [0, \infty)$ *the function which satisfies for all* $a \in \mathbb{C}$ *that* $|a| = |a|_{\mathbb{C}}$*.*

Definition 0.1.4 (Euclidean norm and Euclidean scalar product)**.** *Let $n \in \mathbb{N}$. Then we denote by $\|\cdot\|_{\mathbb{R}^n} : \mathbb{R}^n \to [0, \infty)$ and $\langle \cdot, \cdot \rangle_{\mathbb{R}^n} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ the functions which satisfy for all $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$, $w = (w_1, \ldots, w_n) \in \mathbb{R}^n$ that*

$$\|v\|_{\mathbb{R}^n} = \left[\sum_{i=1}^{n} |v_i|^2 \right]^{1/2} \qquad \textit{and} \qquad \langle v, w \rangle_{\mathbb{R}^n} = \sum_{i=1}^{n} v_i w_i. \tag{0.5}$$

Note that for every $n \in \mathbb{N}$ it holds that $(\mathbb{R}^n, \|\cdot\|_{\mathbb{R}^n}, \langle \cdot, \cdot \rangle_{\mathbb{R}^n})$ is an inner product space.

Definition 0.1.5 (Identity function)**.** *Let A be a set. Then we denote by $\mathrm{id}_A \colon A \to A$ the function which satisfies for all $a \in A$ that*

$$\mathrm{id}_A(a) = a \tag{0.6}$$

and we call $\mathrm{id}_A \colon A \to A$ the identity function on A.

Definition 0.1.6 (Identity matrix)**.** *Let $d \in \mathbb{N}$. Then we denote by $I_{\mathbb{R}^d} \in \mathbb{R}^{d \times d}$ the $d \times d$-matrix given by*

$$I_{\mathbb{R}^d} = \begin{pmatrix} 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & 1 & 0 \\ 0 & 0 & \ldots & 0 & 1 \end{pmatrix}. \tag{0.7}$$

Definition 0.1.7 (Transpose)**.** *Let $n, m \in \mathbb{N}$ be natural numbers and let $A \in \mathbb{R}^{n \times m}$ be an $n \times m$-matrix. Then we denote by $A^\top \in \mathbb{R}^{m \times n}$ the transpose of A.*

Definition 0.1.8 (Power set)**.** *Let Ω be a set. Then we denote by $\mathcal{P}(\Omega)$ the power set of Ω (the set of all subsets of Ω).*

Definition 0.1.9 (Set of functions)**.** *Let A and B be sets. Then we denote by $\mathbb{M}(A, B)$ the set of all functions from A to B.*

Definition 0.1.10 (Extended composition of functions)**.** *Let A, $\tilde{A}$, B, and $\tilde{B}$ be sets and let $f \colon A \to \tilde{A}$ and $g \colon B \to \tilde{B}$ be functions which satisfy that $g(B) \subseteq A$. Then we denote by $f \circ g \colon B \to \tilde{A}$ the function which satisfies for all $b \in B$ that*

$$(f \circ g)(b) = f(g(b)) \tag{0.8}$$

and we call $f \circ g$ the composition of f and g.

0.2 Measure theory

0.2.1 Measurable spaces

Definition 0.2.1 (Sigma-algebra)**.** *We say that $\mathcal{A}$ is a sigma-algebra on Ω if and only if it holds that*

(i) $\emptyset \in \mathcal{A} \subseteq \mathcal{P}(\Omega)$,

(ii) $\forall\, A \in \mathcal{A}\colon A^c = \Omega \backslash A \in \mathcal{A}$ *($\mathcal{A}$ is closed under complementation), and*

(iii) $\forall\, A_1, A_2, \ldots \in \mathcal{A}\colon \cup_{n\in\mathbb{N}} A_n \in \mathcal{A}$ *($\mathcal{A}$ is closed under countable unions).*

Item (i) in Definition 0.2.1 is equivalent to the assumption that $\mathcal{A}$ is a non-empty subset of $\mathcal{P}(\Omega)$. More precisely, if $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ is a non-empty set which is closed under complementation in the sense of Item (ii) in Definition 0.2.1 and which is closed under countable unions in the sense of Item (iii) in Definition 0.2.1, then there exists an $A \in \mathcal{A}$ and Item (ii) and Item (iii) ensure that $\mathcal{A} \ni (A \cup A^c)^c = \Omega^c = \emptyset$.

For every set Ω it holds that the power set $\mathcal{P}(\Omega)$ is the largest sigma-algebra on Ω. Moreover, for every set Ω it holds that the set $\{\emptyset, \Omega\}$ is the smallest sigma-algebra on Ω.

Definition 0.2.2 (Sigma-algebra)**.** *We say that $\mathcal{A}$ is a sigma-algebra if and only if there exists a set Ω such that $\mathcal{A}$ is a sigma-algebra on Ω.*

Definition 0.2.3 (Measurable space)**.** *We say that $\mathbf{\Omega}$ is a measurable space if and only if there exist sets Ω and $\mathcal{A}$ such that it holds*

(i) that $\mathbf{\Omega} = (\Omega, \mathcal{A})$ and

(ii) that $\mathcal{A}$ is a sigma-algebra on Ω.

Definition 0.2.4 (Measurable set)**.** *We say that A is measurable with respect to $\mathbf{\Omega}$ (we say that A is measurable) if and only if there exist sets Ω and $\mathcal{A}$ such that it holds*

(i) that $\mathcal{A}$ is a sigma-algebra on Ω,

(ii) that $\mathbf{\Omega} = (\Omega, \mathcal{A})$, and

(iii) that $A \in \mathcal{A}$.

Let $(\Omega, \mathcal{A})$ be a measurable space and let $A \subseteq \Omega$ be an arbitrary subset of Ω which is not necessarily measurable with respect to $(\Omega, \mathcal{A})$. In some situations one is interested to equip the set A with a suitable sigma-algebra. For this the following two concepts, Definition 0.2.5 and Definition 0.2.9, are useful.

Definition 0.2.5 (Trace set). *Let A and $\mathcal{A}$ be sets. Then we denote by $A \Cap \mathcal{A}$ the set given by*

$$A \Cap \mathcal{A} = \{A \cap B \in \mathcal{P}(A) \colon B \in \mathcal{A}\} = \{C \in \mathcal{P}(A) \colon (\exists\, B \in \mathcal{A} \colon A \cap B = C)\} \tag{0.9}$$

and we call $A \Cap \mathcal{A}$ the trace set of A in $\mathcal{A}$ (we call $A \Cap \mathcal{A}$ the trace set).

Class exercise 0.2.6. *What is $\{1, 2\} \Cap \mathcal{P}(\mathbb{N})$?*

Class exercise 0.2.7. *Prove or disprove the following statement: For all sets A and $\mathcal{A}$ it holds that*

$$(A \Cap \mathcal{A}) \neq \emptyset \tag{0.10}$$

if and only if $\mathcal{A} \neq \emptyset$.

Exercise 0.2.8. *Let $(\Omega, \mathcal{A})$ be a measurable space and let $A \subseteq \Omega$ be a subset of Ω. Prove that $(A, A \Cap \mathcal{A})$ is a measurable space.*

Definition 0.2.5 and Exercise 0.2.8 suggest the following notion.

Definition 0.2.9 (Trace sigma-algebra). *Let $(\Omega, \mathcal{A})$ be a measurable space and let $A \subseteq \Omega$ be a subset of Ω. Then we call $A \Cap \mathcal{A}$ the trace sigma-algebra of A in $\mathcal{A}$ (the trace sigma-algebra).*

Observe that for every measurable space $(\Omega, \mathcal{A})$ and every $A \in \mathcal{A}$ it holds that

$$A \Cap \mathcal{A} = \{B \in \mathcal{A} \colon B \subseteq A\} = \mathcal{P}(A) \cap \mathcal{A}. \tag{0.11}$$

Definition 0.2.10 (Generation of a sigma-algebra). *Let Ω be a set and let $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ be a subset of the power set of Ω. Then we denote by $\sigma_\Omega(\mathcal{A})$ the set given by*

$$\sigma_\Omega(\mathcal{A}) = \bigcap_{\substack{\mathcal{B} \text{ is a sigma-algebra} \\ \text{on } \Omega \text{ with } \mathcal{B} \supseteq \mathcal{A}}} \mathcal{B} \tag{0.12}$$

and we call $\sigma_\Omega(\mathcal{A})$ the sigma-algebra on Ω generated by $\mathcal{A}$ (we call $\sigma_\Omega(\mathcal{A})$ the sigma-algebra generated by $\mathcal{A}$).

Note that for every set Ω and every subset $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ of the power set of Ω it holds that $\sigma_\Omega(\mathcal{A})$ is the smallest sigma-algebra on Ω that contains $\mathcal{A}$.

Lemma 0.2.11 (Generator of a trace sigma-algebra). *Let Ω be a set, let $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ be a subset of the power set of Ω, and let $A \subseteq \Omega$ be a subset of Ω. Then*

$$A \Cap \sigma_\Omega(\mathcal{A}) = \sigma_A(A \Cap \mathcal{A}). \tag{0.13}$$

Lemma 0.2.11 is, e.g., proved as Corollary 1.83 in [Klenke(2008)].

0.2.2 Topological spaces and their sigma-algebras

We first recall the notion of a topology.

Definition 0.2.12 (Topology)**.** *We say that $\mathcal{E}$ is a topology on Ω if and only if*

(i) it holds that $\{\emptyset, \Omega\} \subseteq \mathcal{E} \subseteq \mathcal{P}(\Omega)$,

(ii) it holds for all $A, B \in \mathcal{E}$ that $A \cap B \in \mathcal{E}$ ($\mathcal{E}$ is closed under finite intersections), and

(iii) it holds for all sets I and all families $(A_i)_{i \in I} \subseteq \mathcal{E}$ that $\cup_{i \in I} A_i \in \mathcal{E}$ ($\mathcal{E}$ is closed under arbitrary unions).

Definition 0.2.13. *We say that $\mathcal{E}$ is a topology if and only if there exists a set Ω such that $\mathcal{E}$ is a topology on Ω.*

Definition 0.2.14 (Topological space)**.** *We say that $\mathbf{\Omega}$ is a topological space if and only if there exist sets Ω and $\mathcal{E}$ such that it holds*

(i) that $\mathbf{\Omega} = (\Omega, \mathcal{E})$ and

(ii) that $\mathcal{E}$ is a topology on Ω.

Definition 0.2.15 (Open set)**.** *We say that E is open in $\mathbf{\Omega}$ (we say that E is open) if and only if there exist sets Ω and $\mathcal{E}$ such that it holds*

(i) that $\mathcal{E}$ is a topology on Ω,

(ii) that $\mathbf{\Omega} = (\Omega, \mathcal{E})$, and

(iii) that $E \in \mathcal{E}$.

In the next step we briefly recall the notion of a metric.

Definition 0.2.16 (Metric)**.** *We say that d is a metric on E if and only if*

(i) it holds that $d \in \mathbb{M}(E \times E, [0, \infty))$ is a function from $E \times E$ to $[0, \infty)$,

*(ii) it holds for all $x, y \in E$ that $\big(d(x, y) = 0$ if and only if $x = y\big)$ (*positive definiteness*),*

*(iii) it holds for all $x, y \in E$ that $d(x, y) = d(y, x)$ (*symmetry*), and*

*(iv) it holds for all $x, y, z \in E$ that $d(x, z) \leq d(x, y) + d(y, z)$ (*triangle inequality*).*

Definition 0.2.17. *We say that d is a metric if and only if there exists a set E such that d is a metric on E.*

Definition 0.2.18 (Metric space)**.** *We say that $\mathbf{E}$ is a metric space if and only if there exist a set E and a metric d on E such that $\mathbf{E} = (E, d)$.*

Observe that for every normed $\mathbb{R}$-vector space $(V, \|\cdot\|_V)$ it holds that the function

$$V \times V \ni (v, w) \mapsto \|v - w\|_V \in [0, \infty) \tag{0.14}$$

is a metric on V. In particular, note that for every $d \in \mathbb{N}$ it holds that the function

$$\mathbb{R}^d \times \mathbb{R}^d \ni (x, y) \mapsto \|x - y\|_{\mathbb{R}^d} \in [0, \infty) \tag{0.15}$$

is a metric on $\mathbb{R}^d$. In the next well-known lemma we recall that a metric space (E, d) (see Definition 0.2.18) induces a topology on E (see Definition 0.2.12).

Lemma 0.2.19 (The topology induced by a metric). *Let (E, d) be a metric space. Then it holds that*

$$\big\{A \in \mathcal{P}(E) \colon \big(\forall\, a \in A \colon \exists\, \varepsilon \in (0, \infty) \colon \{b \in E \colon d(a, b) < \varepsilon\} \subseteq A\big)\big\} \tag{0.16}$$

is a topology on E.

Definition 0.2.20 (The topology induced by a norm). *Let $(V, \|\cdot\|_V)$ be a normed $\mathbb{R}$-vector space. Then we denote by $\mathcal{E}(V, \|\cdot\|_V)$ the set given by*

$$\mathcal{E}(V, \|\cdot\|_V) = \{U \subseteq V \colon (\forall\, u \in U \colon \exists\, \varepsilon \in (0, \infty) \colon \{y \in V \colon \|u - y\|_V < \varepsilon\} \subseteq U)\}. \tag{0.17}$$

Observe that for every normed $\mathbb{R}$-vector space $(V, \|\cdot\|_V)$ it holds that the pair $\big(V, \mathcal{E}(V, \|\cdot\|_V)\big)$ is a topological space.

Definition 0.2.21 (Borel sigma-algebra). *Let $(E, \mathcal{E})$ be a topological space. Then we denote by $\mathcal{B}(E)$ the set given by $\mathcal{B}(E) = \sigma_E(\mathcal{E})$ and we call $\mathcal{B}(E)$ the* Borel sigma-algebra of *$(E, \mathcal{E})$.*

Definition 0.2.22 (Borel set). *We say that A is a Borel set with respect to $\mathbf{E}$ if and only if there exist a set E and a topology $\mathcal{E}$ on E such that*

$$\mathbf{E} = (E, \mathcal{E}) \qquad \text{and} \qquad A \in \mathcal{B}(E) = \sigma_E(\mathcal{E}). \tag{0.18}$$

Observe that for all $d \in \mathbb{N}$ it holds that

$$\mathcal{E}(\mathbb{R}^d, \|\cdot\|_{\mathbb{R}^d}) \subseteq \mathcal{B}(\mathbb{R}^d) \subseteq \mathcal{P}(\mathbb{R}^d). \tag{0.19}$$

0.2.3 Measure spaces

Definition 0.2.23 (Measure)**.** *We say that μ is a measure on $\mathbf{\Omega}$ if and only if there exist sets Ω and $\mathcal{A}$ such that*

(i) it holds that $\mathcal{A}$ is a sigma-algebra on Ω,

(ii) it holds that $\mathbf{\Omega} = (\Omega, \mathcal{A})$,

(iii) it holds that $\mu \in \mathbb{M}(\mathcal{A}, [0, \infty])$ is a function from $\mathcal{A}$ to $[0, \infty)$,

(iv) it holds that $\mu(\emptyset) = 0$, and

(v) it holds for every function $A\colon \mathbb{N} \to \mathcal{A}$ with $\forall\, n \in \mathbb{N}, m \in \mathbb{N}\backslash\{n\}\colon A_n \cap A_m = \emptyset$ that

$$\mu\Bigg(\bigcup_{n\in\mathbb{N}} A_n\Bigg) = \sum_{n=1}^{\infty} \mu(A_n). \qquad \text{(sigma additivity)}$$

Definition 0.2.24. *We say that μ is a measure if and only if there exists a measurable space $\mathbf{\Omega}$ such that μ is a measure on $\mathbf{\Omega}$.*

Definition 0.2.25. *We say that $\mathbf{\Omega}$ is a measure space if and only if there exist a measurable space $(\Omega, \mathcal{A})$ and a measure μ on $(\Omega, \mathcal{A})$ such that $\mathbf{\Omega} = (\Omega, \mathcal{A}, \mu)$.*

Definition 0.2.26 (On sets of measure zero (null sets) and negligible sets)**.** *We say that A is a μ-negligible set if and only if there exist Ω, $\mathcal{A}$, and B such that it holds*

(i) that μ is a measure on $(\Omega, \mathcal{A})$,

(ii) that $A \subseteq B \in \mathcal{A}$, and

(iii) that $\mu(B) = 0$.

Definition 0.2.27 (Completeness of a measure)**.** *Let μ be a measure. Then we say that μ is complete if and only if there exist a set Ω and a sigma-algebra $\mathcal{A}$ on Ω such that μ is a measure on $(\Omega, \mathcal{A})$ and such that for every μ-negligible set A it holds that $A \in \mathcal{A}$.*

Definition 0.2.28 (Completeness of a measure space)**.** *Let $\mathbf{\Omega} = (\Omega, \mathcal{A}, \mu)$ be a measure space. Then we say that $\mathbf{\Omega}$ is complete if and only if μ is a complete measure.*

Any arbitrary (not necessarily complete) measure can be extended to a complete measure. This is the subject of the next concept.

Definition 0.2.29 (Completion of a measure)**.** *Let* $(\Omega, \mathcal{A}, \mu)$ *be a measure space. Then we denote by* $\mathcal{A}^*$ *the set given by*

$$\mathcal{A}^* = \{A \in \mathcal{P}(\Omega)\colon\ (\exists\, B, C \in \mathcal{A}\colon B \subseteq A \subseteq C \text{ and } \mu(C \backslash B) = 0)\}, \tag{0.20}$$

we denote by $\mu^*\colon \mathcal{A}^* \to [0, \infty]$ *the function which satisfies for all* $A \in \mathcal{P}(\Omega)$*,* $B, C \in \mathcal{A}$ *with* $B \subseteq A \subseteq C$ *and* $\mu(C \backslash B) = 0$ *that*

$$\mu^*(A) = \mu(B), \tag{0.21}$$

we call μ^* *the completion of* μ*, and we call the triple* $(\Omega, \mathcal{A}^*, \mu^*)$ *the completion of* $(\Omega, \mathcal{A}, \mu)$*.*

Observe that for every measure space $(\Omega, \mathcal{A}, \mu)$ it holds that the triple $(\Omega, \mathcal{A}^*, \mu^*)$ is a complete measure space.

Definition 0.2.30 (Lebesgue-Borel measure)**.** *Let* $d \in \mathbb{N}$ *and let* $A \in \mathcal{B}(\mathbb{R}^d)$*. Then we denote by* $\mathrm{Borel}_{\mathbb{R}^d}\colon \mathcal{B}(\mathbb{R}^d) \to [0, \infty]$ *the measure which satisfies for all* $a_1, \ldots, a_d, b_1, \ldots, b_d \in \mathbb{R}$ *with* $a_1 \leq b_1, \ \ldots, \ a_d \leq b_d$ *that*

$$\mathrm{Borel}_{\mathbb{R}^d}\big((a_1, b_1] \times \cdots \times (a_d, b_d]\big) = \prod_{i=1}^{d} (b_i - a_i), \tag{0.22}$$

we denote by $B_A\colon \mathcal{B}(A) \to [0, \infty]$ *the measure given by* $B_A = \mathrm{Borel}_{\mathbb{R}^d}|_{\mathcal{B}(A)}$*, and we call* B_A *the Lebesgue-Borel measure on* A*.*

Definition 0.2.31 (Lebesgue measure)**.** *Let* $d \in \mathbb{N}$*. Then we denote by* $\lambda_{\mathbb{R}^d}\colon \mathcal{B}(\mathbb{R}^d)^* \to [0, \infty]$ *the measure given by* $\lambda_{\mathbb{R}^d} = B^*_{\mathbb{R}^d}$ *and we call* $\lambda_{\mathbb{R}^d}$ *the Lebesgue measure on* $\mathbb{R}^d$*.*

Definition 0.2.32. *We say that* P *is a probability measure on* $\mathbf{\Omega}$ *if and only if there exists a measurable space* $(\Omega, \mathcal{A})$ *such that it holds*

(i) *that* P *is a measure on* $(\Omega, \mathcal{A})$*,*

(ii) *that* $P(\Omega) = 1$*, and*

(iii) *that* $\mathbf{\Omega} = (\Omega, \mathcal{A})$*.*

Definition 0.2.33 (Probability measure)**.** *We say that* P *is a probability measure if and only if there exists a measurable space* $\mathbf{\Omega}$ *such that* P *is a probability measure on* $\mathbf{\Omega}$*.*

Definition 0.2.34. *We say that* $\mathbf{\Omega}$ *is a probability space if and only if there exist* Ω*,* $\mathcal{A}$*, and* P *such that* $\mathbf{\Omega} = (\Omega, \mathcal{A}, P)$ *and such that* P *is a probability measure on* $(\Omega, \mathcal{A})$*.*

Definition 0.2.35 (Indicator function)**.** *Let Ω be a set and let $A \in \mathcal{P}(\Omega)$ be a subset of Ω. Then we denote by $\mathbb{1}_A^\Omega\colon \Omega \to \{0,1\}$ the function which satisfies for all $x \in \Omega$ that*

$$\mathbb{1}_A^\Omega(x) = \begin{cases} 1 & : x \in A \\ 0 & : x \notin A \end{cases} \tag{0.23}$$

and we call $\mathbb{1}_A^\Omega\colon \Omega \to \{0,1\}$ the indicator function of A in Ω.

Definition 0.2.36. *Let x and Ω be sets. Then we denote by $\mathbb{1}_\Omega(x) \in \mathbb{R}$ the real number given by*

$$\mathbb{1}_\Omega(x) = \begin{cases} 1 & : x \in \Omega \\ 0 & : x \notin \Omega \end{cases}. \tag{0.24}$$

Definition 0.2.37 (Dirac measure)**.** *Let Ω be a set and let $x \in \Omega$. Then we denote by $\delta_x^\Omega\colon \mathcal{P}(\Omega) \to [0,\infty]$ the function which satisfies for all $A \in \mathcal{P}(\Omega)$ that*

$$\delta_x^\Omega(A) = \mathbb{1}_A^\Omega(x) \tag{0.25}$$

and we call δ_x^Ω the Dirac measure associated to x in Ω.

Definition 0.2.38 (Counting measure)**.** *Let Ω be a set. Then we denote by $\#^\Omega\colon \mathcal{P}(\Omega) \to [0,\infty]$ the measure given by*

$$\#^\Omega = \sum_{\omega \in \Omega} \delta_\omega^\Omega \tag{0.26}$$

and we call $\#^\Omega$ the counting measure on Ω.

Definition 0.2.39 (Number of elements of a set)**.** *Let Ω be a set. Then we denote by $\#_\Omega \in \mathbb{N}_0 \cup \{\infty\}$ the extended real number given by*

$$\#_\Omega = \#^\Omega(\Omega). \tag{0.27}$$

Definition 0.2.40 (Support of a measure)**.** *Let $(E,\mathcal{E})$ be a topological space and let $\mu\colon \mathcal{B}(E) \to [0,\infty]$ be a measure on $(E,\mathcal{B}(E))$. Then we denote by $\operatorname{supp}(\mu)$ the set given by*

$$\operatorname{supp}(\mu) = \{x \in E\colon (\forall U \in \mathcal{E}\colon x \in U \Rightarrow \mu(U) > 0)\} \tag{0.28}$$

and we call $\operatorname{supp}(\mu)$ the support of μ.

Class exercise 0.2.41. *Let $x \in \mathbb{R}$. What is $\operatorname{supp}\big(\delta_x^\mathbb{R}|_{\mathcal{B}(\mathbb{R})}\big)$?*

Class exercise 0.2.42. *Let $d \in \mathbb{N}$. What is $\operatorname{supp}(B_{\mathbb{R}^d})$?*

Exercise 0.2.43. *Let $(E,\mathcal{E})$ be a topological space and let $\mu\colon \mathcal{B}(E) \to [0,\infty]$ be a measure on $(E,\mathcal{B}(E))$. Prove that $\operatorname{supp}(\mu)$ is a closed set in $(E,\mathcal{E})$, i.e., prove that $E \backslash \operatorname{supp}(\mu) \in \mathcal{E}$.*

0.2.4 Measurable functions

Definition 0.2.44 (Measurable function)**.** *We say that X is an $\mathcal{A}/\tilde{\mathcal{A}}$-measurable function (we say that X is a measurable function) if and only if it holds*

(i) that X is a function,

(ii) that $\mathcal{A}$ is a sigma-algebra on domain(X),

(iii) that $\tilde{\mathcal{A}}$ is a sigma-algebra on codomain(X), *and*

(iv) that $\forall\, A \in \tilde{\mathcal{A}} \colon X^{-1}(A) = \{x \in \text{domain}(X) \colon X(x) \in A\} \in \mathcal{A}$.

Lemma 0.2.45 (Measurability of functions)**.** *Let A and B be sets, let $\mathcal{B} \subseteq \mathcal{P}(B)$ be a subset of the power set of B, and let $X \colon A \to B$ be a function. Then*

$$\big\{X^{-1}(S) \colon S \in \sigma_B(\mathcal{B})\big\} = \sigma_A\big(\big\{X^{-1}(S) \colon S \in \mathcal{B}\big\}\big)\,. \tag{0.29}$$

Proof of Lemma 0.2.45. Throughout this proof let $\mathcal{A}$, $\tilde{\mathcal{A}}$, and $\tilde{\mathcal{B}}$ be the sets given by

$$\begin{aligned} \mathcal{A} &= \big\{X^{-1}(S) \colon S \in \sigma_B(\mathcal{B})\big\}\,, \\ \tilde{\mathcal{A}} &= \sigma_A\big(\big\{X^{-1}(S) \colon S \in \mathcal{B}\big\}\big)\,, \\ \tilde{\mathcal{B}} &= \big\{S \in \mathcal{P}(B) \colon X^{-1}(S) \in \tilde{\mathcal{A}}\big\}. \end{aligned} \tag{0.30}$$

Next observe that the fact that $\sigma_B(\mathcal{B})$ is a sigma-algebra on B, the fact that

$$\forall\, S_1, S_2, \ldots \in \mathcal{P}(B) \colon X^{-1}(\cup_{n\in\mathbb{N}} S_n) = \cup_{n\in\mathbb{N}} X^{-1}(S_n), \tag{0.31}$$

and the fact that

$$\forall\, S \in \mathcal{P}(B) \colon X^{-1}(B \backslash S) = X^{-1}(B) \backslash X^{-1}(S) = A \backslash X^{-1}(S) \tag{0.32}$$

ensure that $\mathcal{A}$ is a sigma-algebra on A. This and the fact that

$$\mathcal{A} \supseteq \{X^{-1}(S) \colon S \in \mathcal{B}\} \tag{0.33}$$

imply that $\mathcal{A} \supseteq \tilde{\mathcal{A}}$. It thus remains to prove that $\mathcal{A} \subseteq \tilde{\mathcal{A}}$. For this note that (0.31), (0.32), and the fact that $\tilde{\mathcal{A}}$ is a sigma-algebra on A establish that $\tilde{\mathcal{B}}$ is a sigma-algebra on B. Next note that the fact that

$$\tilde{\mathcal{A}} = \sigma_A\big(\big\{X^{-1}(S) \colon S \in \mathcal{B}\big\}\big) \supseteq \big\{X^{-1}(S) \colon S \in \mathcal{B}\big\} \tag{0.34}$$

shows that for all $S \in \mathcal{B} \subseteq \mathcal{P}(B)$ it holds that

$$X^{-1}(S) \in \tilde{\mathcal{A}}. \tag{0.35}$$

Hence, we obtain that

$$\mathcal{B} \subseteq \tilde{\mathcal{B}}. \tag{0.36}$$

This and the fact that $\tilde{\mathcal{B}}$ is a sigma-algebra on B prove that

$$\sigma_B(\mathcal{B}) \subseteq \tilde{\mathcal{B}}. \tag{0.37}$$

Therefore, we get that for all $S \in \sigma_B(\mathcal{B})$ it holds that $S \in \tilde{\mathcal{B}}$. Hence, we obtain that for all $S \in \sigma_B(\mathcal{B})$ it holds that

$$X^{-1}(S) \in \tilde{\mathcal{A}}. \tag{0.38}$$

This implies that $\mathcal{A} \subseteq \tilde{\mathcal{A}}$. The proof of Lemma 0.2.45 is thus completed. □

Lemma 0.2.45 above is, for instance, also proved as Theorem 1.81 in [Klenke(2008)]. The next result, Corollary 0.2.46, is an immediate consequence of Lemma 0.2.45 above.

Corollary 0.2.46 (Measurability of functions on generators)**.** *Let Ω_1 and Ω_2 be sets, let $\mathcal{A}_1 \subseteq \mathcal{P}(\Omega_1)$ be a subset of the power set of Ω_1, let $\mathcal{A}_2 \subseteq \mathcal{P}(\Omega_2)$ be a subset of the power set of Ω_2, and let $X\colon \Omega_1 \to \Omega_2$ be a function. Then it holds that X is $\sigma_{\Omega_1}(\mathcal{A}_1)/\sigma_{\Omega_2}(\mathcal{A}_2)$-measurable if and only if it holds for all $A \in \mathcal{A}_2$ that*

$$X^{-1}(A) \in \sigma_{\Omega_1}(\mathcal{A}_1). \tag{0.39}$$

Definition 0.2.47 (Sigma-algebra generated by a function)**.** *Let Ω be a set, let $(\tilde{\Omega}, \tilde{\mathcal{A}})$ be a measurable space, and let $X\colon \Omega \to \tilde{\Omega}$ be a function. Then we denote by $\sigma_\Omega(X)$ the set given by*

$$\sigma_\Omega(X) = \left\{X^{-1}(A) \in \mathcal{P}(\Omega)\colon A \in \tilde{\mathcal{A}}\right\} \tag{0.40}$$

and we call $\sigma_\Omega(X)$ the sigma-algebra generated by X.

Note that for every set Ω, every measurable space $(\tilde{\Omega}, \tilde{\mathcal{A}})$, and every function $X\colon \Omega \to \tilde{\Omega}$ it holds that $\sigma_\Omega(X)$ is the smallest sigma-algebra $\mathcal{A}$ on Ω with respect to which X is $\mathcal{A}/\tilde{\mathcal{A}}$-measurable. A measurable function on a measure space naturally induces a measure on its range. This is the subject of the next definition.

Definition 0.2.48 (Image measure/Push forward measure)**.** *Let $(A, \mathcal{A}, \mu)$ be a measure space, let $(B, \mathcal{B})$ be a measurable space, and let $X\colon A \to B$ be an $\mathcal{A}/\mathcal{B}$-measurable function. Then we denote by $X(\mu)_{\mathcal{B}}\colon \mathcal{B} \to [0, \infty]$ the function which satisfies for all $S \in \mathcal{B}$ that*

$$X(\mu)_{\mathcal{B}}(S) = \mu(X \in S) = \mu\big(X^{-1}(S)\big) \tag{0.41}$$

and we call $X(\mu)_{\mathcal{B}}$ the image measure associated to X and $\mathcal{B}$ (we call $X(\mu)_{\mathcal{B}}$ the push forward measure associated to X and $\mathcal{B}$).

Definition 0.2.49 (Random variable)**.** *We say that X is a random variable on $\mathbf{A}$ with respect to $\mathbf{B}$ (we say that X is a random variable on $\mathbf{A}$, we say that X is a random variable) if and only if there exist $A, \mathcal{A}, P$, B, and $\mathcal{B}$ such that it holds*

(i) that $\mathbf{A} = (A, \mathcal{A}, P)$ and $\mathbf{B} = (B, \mathcal{B})$,

(ii) that $\mathbf{A}$ is a probability space,

(iii) that $\mathbf{B}$ is a measurable space, and

(iv) that X is an $\mathcal{A}/\mathcal{B}$-measurable function.

Definition 0.2.50 (Probability distribution of a random variable). *We say that μ is the probability distribution of X under P with respect to $\mathcal{B}$ (we say that μ is the distribution of X under P with respect to $\mathcal{B}$, we say that μ is the probability distribution of X under P, we say that μ is the distribution of X under P, we say that μ is the probability distribution of X, we say that μ is the distribution of X) if and only if there exist A, $\mathcal{A}$, and B such that it holds*

(i) that P is a probability measure on $(A, \mathcal{A})$,

(ii) that $(B, \mathcal{B})$ is a measurable space,

(iii) that X is an $\mathcal{A}/\mathcal{B}$-measurable function, and

(iv) that $\mu = X(P)_{\mathcal{B}}$.

0.2.5 Products of measurable spaces and measure spaces

Definition 0.2.51 (Product sigma-algebra). *Let I be a non-empty set and let $(\Omega_i, \mathcal{F}_i)$, $i \in I$, be measurable spaces. Then*

(i) we denote by $\pi^{(i)}_{\Omega_j, j\in I,}\colon [\times_{j\in I}\Omega_j] \to \Omega_i$, $i \in I$, the functions which satisfy for all $i \in I$, $(\omega_j)_{j\in I} \in [\times_{j\in I}\Omega_j]$ that

$$\pi^{(i)}_{\Omega_j, j\in I,}\big((\omega_j)_{j\in I}\big) = \omega_i, \tag{0.42}$$

(ii) for every $i \in I$ we call $\pi^{(i)}_{\Omega_j, j\in I,}\colon [\times_{j\in I}\Omega_j] \to \Omega_i$ the i-th projection function for Ω_j, $j \in I$,

(iii) we denote by $\otimes_{i\in I}\mathcal{F}_i$ the sigma-algebra given by

$$\otimes_{i\in I}\, \mathcal{F}_i = \sigma_{\times_{j\in I}\Omega_j}\Big(\cup_{i\in I}\, \sigma_{\times_{j\in I}\Omega_j}\big(\pi^{(i)}_{\Omega_j, j\in I,}\big)\Big), \tag{0.43}$$

and

(iv) we call $\otimes_{i\in I}\mathcal{F}_i$ the product sigma-algebra of $\mathcal{F}_i$, $i \in I$.

Note that the product sigma-algebra is the smallest sigma-algebra so that every projection function is a measurable function. Moreover, observe that for all measurable spaces $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ it holds that

$$\otimes_{i\in\{1,2\}}\, \mathcal{F}_i = \mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma_{\Omega_1\times\Omega_2}(\{A_1 \times A_2 \subseteq \Omega_1 \times \Omega_2 \colon A_1 \in \mathcal{F}_1 \text{ and } A_2 \in \mathcal{F}_2\}). \tag{0.44}$$

We also briefly recall the notion of the product measure (see, e.g., Section 38 in [Halmos(1950)]). To do so, we first introduce the following concept.

Definition 0.2.52 (Sigma-finiteness). *Let $\mathbf{\Omega}$ be a measure space. Then we say that $\mathbf{\Omega}$ is sigma-finite if and only if there exist sets Ω, $\mathcal{F}$, and μ and a function $\varpi\colon \mathbb{N} \to \mathcal{F}$ such that it holds*

(i) that $\mathbf{\Omega} = (\Omega, \mathcal{F}, \mu)$,

(ii) that $\forall\, n \in \mathbb{N}\colon \mu(\varpi_n) < \infty$, and

(iii) that $\cup_{n\in\mathbb{N}} \varpi_n = \Omega$.

Definition 0.2.53 (Products of measures). *Let I be a non-empty finite set and let $(\Omega_i, \mathcal{F}_i, \mu_i)$, $i \in I$, be sigma-finite measure spaces. Then we denote by*

$$\otimes_{i\in I}\, \mu_i\colon\ \otimes_{i\in I}\, \mathcal{F}_i \to [0, \infty] \tag{0.45}$$

the measure which satisfies for all $n \in \mathbb{N}$, $i_1, \ldots, i_n \in I$, $A_{i_1} \in \mathcal{F}_{i_1}$, $\ldots$, $A_{i_n} \in \mathcal{F}_{i_n}$ with $\#_{\{i_1,\ldots,i_n\}} = n$ that

$$(\otimes_{i\in I}\mu_i)\Big(\Big\{\pi^{(i_1)}_{\Omega_j, j\in I,} \in A_{i_1}\Big\} \cap \ldots \cap \Big\{\pi^{(i_n)}_{\Omega_j, j\in I,} \in A_{i_n}\Big\}\Big) = \prod_{k=1}^{n} \mu_{i_k}(A_{i_k}) \tag{0.46}$$

and we call $\otimes_{i\in I}\mu_i$ the product measure of μ_i, $i \in I$.

Observe that (0.46) ensures that for all probability spaces $(\Omega_1, \mathcal{F}_1, P_1)$ and $(\Omega_2, \mathcal{F}_2, P_2)$ and all $A_1 \in \mathcal{F}_1$, $A_2 \in \mathcal{F}_2$ it holds that

$$(P_1 \otimes P_2)(A_1 \times A_2) = P_1(A_1) \cdot P_2(A_2). \tag{0.47}$$

Definition 0.2.54 (Powers of a sigma-algebra). *Let $(\Omega, \mathcal{F})$ be a measurable space and let $n \in \mathbb{N}$. Then we denote by $\mathcal{F}^{\otimes n} \subseteq \mathcal{P}(\Omega^n)$ the sigma-algebra given by*

$$\mathcal{F}^{\otimes n} = \otimes_{k\in\{1,2,\ldots,n\}}\, \mathcal{F} = \underbrace{\mathcal{F} \otimes \ldots \otimes \mathcal{F}}_{n\text{-times}}. \tag{0.48}$$

Definition 0.2.55 (Powers of a measure). *Let $(\Omega, \mathcal{F}, \mu)$ be a sigma-finite measure space and let $n \in \mathbb{N}$. Then we denote by $\mu^{\otimes n}\colon \mathcal{F}^{\otimes n} \to [0, \infty]$ the measure given by*

$$\mu^{\otimes n} = \otimes_{k\in\{1,2,\ldots,n\}}\, \mu = \underbrace{\mu \otimes \ldots \otimes \mu}_{n\text{-times}}. \tag{0.49}$$

0.2.6 Integration of measurable functions

Definition 0.2.56 (Lebesgue integral for nonnegative functions)**.** *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space and let $X\colon \Omega \to \mathbb{R}$ be an $\mathcal{A}/\mathcal{B}(\mathbb{R})$-measurable function with $X(\Omega) \subseteq [0,\infty)$. Then we denote by $\int_\Omega X\,d\mu \in [0,\infty]$ the extended real number given by*

$$\int_\Omega X\,d\mu = \sup\left\{ \sum_{y \in Y(\Omega)} y \cdot \mu(Y^{-1}(\{y\})) \in [0,\infty] \colon \begin{array}{c} Y\colon \Omega \to [0,\infty) \text{ is an} \\ \mathcal{A}/\mathcal{B}([0,\infty))\text{-measurable function} \\ \text{with } Y \le X \text{ and } \#_{Y(\Omega)} < \infty \end{array} \right\}. \tag{0.50}$$

Definition 0.2.57 (Lebesgue integral for real valued functions)**.** *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space and let $X\colon \Omega \to \mathbb{R}$ be an $\mathcal{A}/\mathcal{B}(\mathbb{R})$-measurable function with $\min\{ \int_\Omega \max\{X,0\}\,d\mu, \int_\Omega \max\{-X,0\}\,d\mu\} < \infty$. Then we denote by $\int_\Omega X\,d\mu \in [-\infty,\infty]$ the extended real number given by*

$$\int_\Omega X\,d\mu = \int_\Omega \max\{X,0\}\,d\mu - \int_\Omega \max\{-X,0\}\,d\mu. \tag{0.51}$$

Definition 0.2.58 (Lebesgue integral for vector/matrix valued functions)**.** *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, let $n, m \in \mathbb{N}$, and let $X = (X_{i,j})_{(i,j)\in\{1,\dots,n\}\times\{1,\dots,m\}} \colon \Omega \to \mathbb{R}^{n\times m}$ be an $\mathcal{A}/\mathcal{B}(\mathbb{R}^{n\times m})$-measurable function with $\int_\Omega \|X\|_{\mathbb{R}^{n\times m}}\,d\mu < \infty$. Then we denote by $\int_\Omega X\,d\mu \in \mathbb{R}^{n\times m}$ the $n\times m$-matrix given by*

$$\int_\Omega X\,d\mu = \left(\int_\Omega X_{i,j}\,d\mu \right)_{(i,j)\in\{1,\dots,n\}\times\{1,\dots,m\}}. \tag{0.52}$$

The next result is known as *change of variables formula* in the literature.

Theorem 0.2.59 (Change of variables formula)**.** Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, let $(\tilde{\Omega}, \tilde{\mathcal{A}})$ be a measurable space, and let $X\colon \Omega \to \tilde{\Omega}$ be an $\mathcal{A}/\tilde{\mathcal{A}}$-measurable function. Then

(i) it holds for all $\tilde{\mathcal{A}}/\mathcal{B}(\mathbb{R})$-measurable functions $f\colon \tilde{\Omega} \to \mathbb{R}$ that

$$\int_\Omega |f(X)|\,d\mu = \int_{\tilde{\Omega}} |f|\,dX(\mu)_{\tilde{\mathcal{A}}} \tag{0.53}$$

and

(ii) it holds for all $\tilde{\mathcal{A}}/\mathcal{B}(\mathbb{R})$-measurable functions $f\colon \tilde{\Omega} \to \mathbb{R}$ with $\int_\Omega |f(X)|\,d\mu < \infty$ that

$$\int_\Omega f(X)\,d\mu = \int_{\tilde{\Omega}} f\,dX(\mu)_{\tilde{\mathcal{A}}}. \tag{0.54}$$

0.2.7 Absolute continuity of measures

Definition 0.2.60 (Measures with densities)**.** *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space and let $f\colon \Omega \to [0, \infty)$ be an $\mathcal{A}/\mathcal{B}([0, \infty))$-measurable function. Then we denote by $f \odot \mu\colon \mathcal{A} \to [0, \infty]$ the function which satisfies for all $A \in \mathcal{A}$ that*

$$(f \odot \mu)(A) = \int_{\Omega} f(\omega) \cdot \mathbb{1}_A^{\Omega}(\omega)\, \mu(d\omega) = \int_{\Omega} f \cdot \mathbb{1}_A^{\Omega}\, d\mu. \tag{0.55}$$

Lemma 0.2.61 (Measures with densities)**.** *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space and let $f\colon \Omega \to [0, \infty)$ be an $\mathcal{A}/\mathcal{B}([0, \infty))$-measurable function. Then $(\Omega, \mathcal{A}, f \odot \mu)$ is a measure space.*

Definition 0.2.62 (Densities and absolute continuity of measures)**.** *We say that ν is absolutely continuous with respect to μ with density f if and only if there exists a measurable space $(\Omega, \mathcal{A})$ such that it holds*

(i) that μ is a measure on $(\Omega, \mathcal{A})$,

(ii) that f is an $\mathcal{A}/\mathcal{B}([0, \infty))$-measurable function, and

(iii) that $\nu = f \odot \mu$.

Definition 0.2.63. *We say that ν is absolutely continuous with respect to μ if and only if there exists a f such that ν is absolutely continuous with respect to μ with density f.*

Definition 0.2.64. *We say that f is a density of ν with respect to μ if and only if ν is absolutely continuous with respect to μ with density f.*

Definition 0.2.65 (Absolute continuity with respect to the Lebesgue-Borel measure)**.** *We say that ν is absolutely continuous with density f if and only if there exists a natural number $d \in \mathbb{N}$ such that ν is absolutely continuous with respect to $B_{\mathbb{R}^d}$ with density f.*

Definition 0.2.66 (Absolute continuity with respect to the Lebesgue-Borel measure)**.** *We say that ν is absolutely continuous if and only if there exists a natural number $d \in \mathbb{N}$ such that ν is absolutely continuous with respect to $B_{\mathbb{R}^d}$.*

Definition 0.2.67 (Density)**.** *We say that f is a density of ν if and only if ν is absolutely continuous with density f.*

Several important probability measures from the literature are absolutely continuous. Probability measures that are discrete in a certain sense are not absolutely continuous. Discrete measures are the subject of the next definition.

Definition 0.2.68 (Discrete measure)**.** *Let μ be a measure. Then we say that μ is discrete if and only if there exist sets Ω and $\mathcal{A}$ and an at most countable set $A \in \mathcal{A}$ such that μ is a measure on $(\Omega, \mathcal{A})$ which satisfies*

$$\mu(\Omega \backslash A) = 0. \tag{0.56}$$

Observe that for every measure space $(\Omega, \mathcal{A}, \mu)$ it holds that μ is discrete if and only if there exist an at most countable set I, a family $x_i \in \Omega$, $i \in I$, of elements in Ω, and a family $p_i \in [0, \infty]$, $i \in I$, of real numbers such that

$$\mu = \sum_{i \in I} p_i \, \delta^{\Omega}_{x_i}, \tag{0.57}$$

i.e., such that for all $A \in \mathcal{A}$ it holds that

$$\mu(A) = \sum_{i \in I} p_i \cdot \delta^{\Omega}_{x_i}(A). \tag{0.58}$$

0.3 Random variables

0.3.1 Expectation and covariance

Definition 0.3.1. *Let $(\Omega, \mathcal{A}, P)$ be a probability space and let $X \colon \Omega \to \mathbb{R}$ be an $\mathcal{A}/\mathcal{B}(\mathbb{R})$-measurable function with $\min\big\{ \int_\Omega \max\{X, 0\} \, dP, \int_\Omega \max\{-X, 0\} \, dP \big\} < \infty$. Then we denote by $\mathbb{E}_P[X] \in [-\infty, \infty]$ the extended real number given by*

$$\mathbb{E}_P[X] = \int_\Omega X \, dP \tag{0.59}$$

and we call $\mathbb{E}_P[X]$ the P-expectation of X (the expectation of X).

The expectations of vector valued and matrix valued random variables are defined analogously; cf. Section 0.2.6.

Definition 0.3.2. *Let $(\Omega, \mathcal{A}, P)$ be a probability space, let $n, m \in \mathbb{N}$, and let $X \colon \Omega \to \mathbb{R}^{n \times m}$ be an $\mathcal{A}/\mathcal{B}(\mathbb{R}^{n \times m})$-measurable function with $\mathbb{E}_P\big[\|X\|_{\mathbb{R}^{n \times m}}\big] < \infty$. Then we denote by $\mathbb{E}_P[X] \in \mathbb{R}^{n \times m}$ the $n \times m$-matrix given by*

$$\mathbb{E}_P[X] = \int_\Omega X \, dP \tag{0.60}$$

and we call $\mathbb{E}_P[X]$ the expectation of X (the P-expectation of X).

Definition 0.3.3. *Let $(\Omega, \mathcal{A}, P)$ be a probability space and let $X \colon \Omega \to \mathbb{R}$ be an $\mathcal{A}/\mathcal{B}(\mathbb{R})$-measurable function with $\mathbb{E}_P\big[|X|_{\mathbb{R}}\big] < \infty$. Then we denote by $\mathsf{Var}_P(X) \in [0, \infty]$ the extended real number given by*

$$\mathsf{Var}_P(X) = \mathbb{E}_P\big[(X - \mathbb{E}_P[X])^2\big] \tag{0.61}$$

and we call $\mathsf{Var}_P(X)$ the P-variance of X (the variance of X).

Definition 0.3.4. *Let $(\Omega, \mathcal{A}, P)$ be a probability space, let $n \in \mathbb{N}$, and let $X = (X_1, \dots, X_n)\colon \Omega \to \mathbb{R}^n$ and $Y = (Y_1, \dots, Y_n)\colon \Omega \to \mathbb{R}^n$ be $\mathcal{A}/\mathcal{B}(\mathbb{R}^n)$-measurable functions with $\mathbb{E}_P\big[\|X\|^2_{\mathbb{R}^n} + \|Y\|^2_{\mathbb{R}^n}\big] < \infty$. Then we denote by $\mathsf{Cov}_P(X, Y) \in \mathbb{R}^{n \times n}$ the $n \times n$-matrix given by*

$$\begin{aligned} \mathsf{Cov}_P(X,Y) &= \mathbb{E}_P\big[(X - \mathbb{E}_P[X])(Y - \mathbb{E}_P[Y])^\top\big] \\ &= \Big(\mathbb{E}_P\big[(X_i - \mathbb{E}_P[X_i])(Y_j - \mathbb{E}_P[Y_j])\big]\Big)_{(i,j)\in\{1,\dots,n\}\times\{1,\dots,n\}} \end{aligned} \tag{0.62}$$

and we call $\mathsf{Cov}_P(X, Y)$ the P-covariance of X and Y (the covariance of X and Y).

Definition 0.3.5. *Let $(\Omega, \mathcal{A}, P)$ be a probability space, let $n, m \in \mathbb{N}$, and let $X = (X_1, \dots, X_n)\colon \Omega \to \mathbb{R}^n$ be an $\mathcal{A}/\mathcal{B}(\mathbb{R}^n)$-measurable function with $\mathbb{E}_P\big[\|X\|^2_{\mathbb{R}^n}\big] < \infty$. Then we denote by $\mathsf{Cov}_P(X) \in \mathbb{R}^{n\times n}$ the $n \times n$-matrix given by*

$$\mathsf{Cov}_P(X) = \mathsf{Cov}_P(X, X) \tag{0.63}$$

and we call $\mathsf{Cov}_P(X)$ the P-covariance of X (the covariance of X).

Definition 0.3.6 (Uncorrelated)**.** *We say that X and Y are P-uncorrelated (we say that X and Y are uncorrelated) if and only if there exists a probability space $(\Omega, \mathcal{A}, P)$ such that it holds*

(i) that X is a random variable on $(\Omega, \mathcal{A}, P)$ with respect to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$,

(ii) that Y is a random variable on $(\Omega, \mathcal{A}, P)$ with respect to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$,

(iii) that $\mathbb{E}\big[X^2 + Y^2\big] < \infty$, and

(iv) that $\mathsf{Cov}_P(X, Y) = 0$.

In the next step we record a useful identity for the covariance matrix of a random variable. More precisely, observe that for every probability space $(\Omega, \mathcal{A}, P)$, every $d \in \mathbb{N}$, and every $\mathcal{A}/\mathcal{B}(\mathbb{R}^d)$-measurable function $X\colon \Omega \to \mathbb{R}^d$ with $\mathbb{E}_P\big[\|X\|^2_{\mathbb{R}^d}\big] < \infty$ it holds that

$$\begin{aligned} \mathsf{Cov}_P(X) &= \mathbb{E}_P\Big[(X - \mathbb{E}_P[X])\,(X - \mathbb{E}_P[X])^\top\Big] \\ &= \mathbb{E}_P\big[XX^\top\big] - \mathbb{E}_P[X]\,\mathbb{E}_P\big[X^\top\big] - \mathbb{E}_P[X]\,\mathbb{E}_P\big[X^\top\big] + \mathbb{E}_P[X]\,\mathbb{E}_P\big[X^\top\big] \\ &= \mathbb{E}_P\big[XX^\top\big] - \mathbb{E}_P\big[X\big]\,\mathbb{E}_P\big[X^\top\big]. \end{aligned} \tag{0.64}$$

0.3.2 Distribution functions

An important instrument to describe probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ are distribution functions. They are the subject of the next definition.

Definition 0.3.7 (Distribution function)**.** *We say that F is a distribution function if and only if it holds*

(i) that $F \in \mathbb{M}(\mathbb{R}, [0, 1])$ is a function from $\mathbb{R}$ to $[0, 1]$,

(ii) that F is non-decreasing (it holds for all $x_1, x_2 \in \mathbb{R}$ with $x_1 \leq x_2$ that $F(x_1) \leq F(x_2)$),

(iii) that $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$, and

(iv) that F is a càdlàg (continue à droite, limite à gauche) function (it holds for all $x \in \mathbb{R}$ that $\lim_{y \nearrow x} F(y)$ and $\lim_{y \searrow x} F(y)$ exist and that $\lim_{y \searrow x} F(y) = F(x)$).

As announced above, a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ induces a distribution function. This is the subject of the next lemma.

Lemma 0.3.8. *Let $(\Omega, \mathcal{A}, P)$ be a probability space with $\Omega \Cap \mathcal{B}(\mathbb{R}) \subseteq \mathcal{A}$ and $P(\Omega \cap \mathbb{R}) = 1$. Then it holds that the function*

$$\mathbb{R} \ni x \mapsto P\big((-\infty, x] \cap \Omega\big) \in [0, 1] \tag{0.65}$$

is a distribution function.

The proof of Lemma 0.3.8 is elementary and left to the reader. Lemma 0.3.8 motivates the next definition.

Definition 0.3.9 (Distribution function of a probability measure)**.** *We say that F is the distribution function of P if and only if there exists a measurable space $(\Omega, \mathcal{A})$ such that it holds*

(i) that P is a probability measure on $(\Omega, \mathcal{A})$,

(ii) that $\Omega \Cap \mathcal{B}(\mathbb{R}) = \mathcal{A}$, and

(iii) that $F = \big(\mathbb{R} \ni x \mapsto P((-\infty, x] \cap \Omega) \in [0, 1]\big)$.

We also present the definition of a distribution function of a random variable.

Definition 0.3.10 (Distribution function of a random variable)**.** *We say that F is the distribution function of X if and only if there exists a probability space $(\Omega, \mathcal{A}, P)$ and a set $B \in \mathcal{P}(\mathbb{R})$ such that it holds*

(i) that X is an $\mathcal{A}/(B \Cap \mathcal{B}(\mathbb{R}))$-measurable function and

(ii) that $F = \big(\mathbb{R} \ni x \mapsto P(X \leq x) = P(\{\omega \in \Omega \colon X(\omega) \leq x\}) \in [0, 1]\big)$.

Definition 0.3.11 (Equality in distribution)**.** *Let $(\Omega, \mathcal{A}, P)$ and $(\hat{\Omega}, \hat{\mathcal{A}}, \hat{P})$ be probability spaces, let $(S, \mathcal{S})$ be a measurable space, let $X \colon \Omega \to S$ be an $\mathcal{A}/\mathcal{S}$-measurable function, and let $\hat{X} \colon \hat{\Omega} \to S$ be an $\hat{\mathcal{A}}/\mathcal{S}$-measurable function. Then we write $X = \hat{X}$ in distribution on $\mathcal{S}$ (we write $X = \hat{X}$ in distribution) if and only if*

$$X(P)_{\mathcal{S}} = \hat{X}(\hat{P})_{\mathcal{S}}. \tag{0.66}$$

Definition 0.3.12 (Distributed according to a given probability measure)**.** *We say that X is $\hat{P}$-distributed on $\mathbf{\Omega}$ (we say that X is $\hat{P}$-distributed) if and only if there exist a probability space $(\Omega, \mathcal{F}, P)$ and a measurable space $(\hat{\Omega}, \hat{\mathcal{F}})$ such that it holds*

(i) that $\mathbf{\Omega} = (\Omega, \mathcal{F}, P)$,

(ii) that X is an $\mathcal{F}/\hat{\mathcal{F}}$-measurable function, and

(iii) that $X(P)_{\hat{\mathcal{F}}} = \hat{P}$.

0.4 Examples of probability distributions

This section briefly reviews a few important probability distributions from the literature.

0.4.1 Discrete probability distributions

0.4.1.1 Discrete uniform distribution

Definition 0.4.1 (Discrete uniform distribution)**.** *Let Ω be a non-empty finite set. Then we denote by $\mathit{Unif}_\Omega \colon \mathcal{P}(\Omega) \to [0,\infty]$ the probability measure given by*

$$\mathit{Unif}_\Omega = \frac{1}{\#_\Omega} \left(\sum_{\omega \in \Omega} \delta_\omega^\Omega \right) \tag{0.67}$$

and we call Unif_Ω the discrete uniform distribution on Ω.

0.4.1.2 Bernoulli distribution

Definition 0.4.2. *Let $p \in [0,1]$. Then we denote by $\mathit{Ber}_p \colon \mathcal{B}(\mathbb{R}) \to [0,\infty]$ the probability measure given by*

$$\mathit{Ber}_p = (1-p) \cdot \delta_0^{\mathbb{R}}|_{\mathcal{B}(\mathbb{R})} + p \cdot \delta_1^{\mathbb{R}}|_{\mathcal{B}(\mathbb{R})} \tag{0.68}$$

and we call Ber_p the Bernoulli distribution with parameter p.

Observe that for every $p \in [0,1]$ and every $x \in \mathbb{R}$ it holds that

$$\mathrm{Ber}_p((-\infty, x]) = \begin{cases} 0 & : x < 0, \\ 1-p & : 0 \leq x < 1, \\ 1 & : x \geq 1. \end{cases} \tag{0.69}$$

Moreover, note that for every $p \in [0,1]$, every probability space $(\Omega, \mathcal{F}, P)$, and every Ber_p-distributed random variable $X \colon \Omega \to \mathbb{R}$ it holds that

$$\mathbb{E}_P[X] = p \qquad \text{and} \qquad \mathrm{Var}_P(X) = p\,(1-p)\,. \tag{0.70}$$

0.4.1.3 Binomial distribution

Definition 0.4.3 (Binomial distribution). *Let $p \in [0,1]$, $n \in \mathbb{N}$. Then we denote by $b_{n,p} \colon \mathcal{B}(\mathbb{R}) \to [0,\infty]$ the probability measure given by*

$$b_{n,p} = \sum_{k=0}^{n} \binom{n}{k} p^k \, (1-p)^{(n-k)} \, \delta_k^{\mathbb{R}}|_{\mathcal{B}(\mathbb{R})} \tag{0.71}$$

and we call $b_{n,p}$ the binomial distribution with parameters n and p.

Observe that for all $n \in \mathbb{N}$, $p \in [0,1]$, all probability spaces $(\Omega, \mathcal{F}, P)$, and all P-independent Ber_p-distributed random variables $X_1, \dots, X_n \colon \Omega \to \mathbb{R}$ on $(\Omega, \mathcal{F}, P)$ it holds that the function

$$\Omega \ni \omega \mapsto \sum_{k=1}^{n} X_k(\omega) \in \mathbb{R} \tag{0.72}$$

(number of successes in n independent Bernoulli experiments with parameter p) is $b_{n,p}$-distributed (binomially distributed with parameters n and p).

0.4.1.4 Geometric distribution

Definition 0.4.4 (Geometric distribution). *Let $p \in (0,1]$. Then we denote by $geom_p \colon \mathcal{B}(\mathbb{R}) \to [0,\infty]$ the probability measure given by*

$$geom_p = \sum_{n=0}^{\infty} p \, (1-p)^n \, \delta_n^{\mathbb{R}}|_{\mathcal{B}(\mathbb{R})} \tag{0.73}$$

and we call $geom_p$ the geometric distribution with parameter p.

Note that for all $p \in (0,1]$, all probability spaces $(\Omega, \mathcal{F}, P)$, and all P-independent Ber_p-distributed random variables $X_1, X_2, \ldots \colon \Omega \to \mathbb{R}$ on $(\Omega, \mathcal{F}, P)$ it holds that the function

$$\Omega \ni \omega \mapsto \mathbb{1}^{\Omega}_{\cup_{k \in \mathbb{N}} \{X_k = 1\}}(\omega) \cdot \min(\{k \in \mathbb{N}_0 \colon X_{k+1}(\omega) = 1\} \cup \{\infty\}) \in \mathbb{R} \tag{0.74}$$

(waiting time for the first success minus 1/number of failures before the first sucess) is geom_p-distributed (geometrically distributed with parameter p).

0.4.1.5 Shifted geometric distribution

Definition 0.4.5 (Shifted geometric distribution). *Let $p \in (0,1]$. Then we denote by $sgeom_p \colon \mathcal{B}(\mathbb{R}) \to [0,\infty]$ the probability measure given by*

$$sgeom_p = \sum_{n=1}^{\infty} p \, (1-p)^{(n-1)} \, \delta_n^{\mathbb{R}}|_{\mathcal{B}(\mathbb{R})} \tag{0.75}$$

and we call $sgeom_p$ the shifted geometric distribution with parameter p.

Note that for all $p \in (0,1]$, all probability spaces $(\Omega, \mathcal{F}, P)$, and all P-independent Ber_p-distributed random variables $X_1, X_2, \ldots \colon \Omega \to \mathbb{R}$ on $(\Omega, \mathcal{F}, P)$ it holds that the function

$$\Omega \ni \omega \mapsto \mathbb{1}^{\Omega}_{\cup_{k \in \mathbb{N}} \{X_k = 1\}}(\omega) \cdot \min(\{k \in \mathbb{N} \colon X_k(\omega) = 1\} \cup \{\infty\}) \in \mathbb{R} \tag{0.76}$$

(waiting time for the first success) is sgeom_p-distributed (shifted geometrically distributed with parameter p).

0.4.1.6 Poisson distribution

Definition 0.4.6 (Poisson distribution)**.** *Let $\lambda \in (0, \infty)$. Then we denote by $\mathrm{Poi}_\lambda \colon \mathcal{B}(\mathbb{R}) \to [0, \infty]$ the probability measure given by*

$$\mathrm{Poi}_\lambda = \sum_{n=0}^{\infty} \frac{e^{-\lambda} \lambda^n}{n!} \delta_n^{\mathbb{R}}|_{\mathcal{B}(\mathbb{R})} \tag{0.77}$$

and we call Poi_λ the Poisson distribution with parameter λ.

The Poisson distribution with parameter $\lambda \in (0, \infty)$ appears, for example, as an approximation of the binomial distribution with parameters $n \in \mathbb{N}$ and $p \in (0,1)$ with n large, p small, and $np \approx \lambda$ in a suitable sense. Details can be found in Theorem 0.4.8 below, which is sometimes also referred to as *law of rare events* in the literature. The proof of Theorem 0.4.8 uses the following exercise on approximations of the exponential function.

Exercise 0.4.7 (Approximations of the exponential function)**.** *Let $a_l \in \mathbb{R}$, $l \in \mathbb{N}$, be a convergent sequence and let $n_l \in \mathbb{N}$, $l \in \mathbb{N}$, satisfy $\liminf_{l \to \infty} n_l = \infty$. Prove that*

$$\lim_{l \to \infty} \left[\left[1 + \tfrac{a_l}{n_l} \right]^{n_l} \right] = \exp\!\left(\lim_{l \to \infty} a_l \right). \tag{0.78}$$

We now present the promised law of rare events in the following result, Theorem 0.4.8. The proof of Theorem 0.4.8 uses Exercise 0.4.7.

Theorem 0.4.8 (Poisson approximation: Law of rare events)**.** *Let $\lambda \in (0, \infty)$, $k \in \mathbb{N}$ and let $n_l \in \mathbb{N}$, $l \in \mathbb{N}$, and $p_l \in [0,1]$, $l \in \mathbb{N}$, satisfy that $\liminf_{l \to \infty} n_l = \infty$ and $\limsup_{l \to \infty} |n_l p_l - \lambda| = 0$. Then*

$$\limsup_{l \to \infty} |p_l| = 0 \qquad \textit{and} \qquad \limsup_{l \to \infty} |b_{n_l, p_l}(\{k\}) - \mathrm{Poi}_\lambda(\{k\})| = 0. \tag{0.79}$$

Proof of Theorem 0.4.8. First, we note that the assumptions $\liminf_{l \to \infty} n_l = \infty$ and

$\limsup_{l\to\infty} |n_l p_l - \lambda| = 0$ ensure that $\limsup_{l\to\infty} |p_l| = 0$. Next observe that

$$\begin{aligned}
&\lim_{l\to\infty} b_{n_l,p_l}(\{k\}) = \lim_{l\to\infty} \left[\binom{n_l}{k} |p_l|^k (1-p_l)^{(n_l-k)} \right] \\
&= \lim_{l\to\infty} \left[\frac{n_l (n_l-1) \cdot \ldots \cdot (n_l-k+1)}{(n_l)^k} \cdot \frac{1}{k!} \cdot (n_l p_l)^k \cdot (1-p_l)^{(n_l-k)} \right] \\
&= \frac{1}{k!} \cdot \lim_{l\to\infty} \left[\frac{n_l (n_l-1) \cdot \ldots \cdot (n_l-k+1)}{(n_l)^k} \right] \cdot \lim_{l\to\infty} \left[(n_l p_l)^k \right] \cdot \lim_{l\to\infty} \left[(1-p_l)^{(n_l-k)} \right] \\
&= \frac{\lambda^k}{k!} \cdot \lim_{l\to\infty} \left[(1-p_l)^{n_l} \right] \cdot \lim_{l\to\infty} \left[(1-p_l)^{-k} \right] = \frac{\lambda^k}{k!} \cdot \lim_{l\to\infty} \left[\left[1 - \frac{n_l p_l}{n_l} \right]^{n_l} \right].
\end{aligned} \tag{0.80}$$

Exercise 0.4.7 hence proves that

$$\lim_{l\to\infty} b_{n_l,p_l}(\{k\}) = \frac{\lambda^k}{k!} \cdot \exp\!\Big(- \lim_{l\to\infty} n_l p_l \Big) = \frac{e^{-\lambda} \lambda^k}{k!}. \tag{0.81}$$

The proof of Theorem 0.4.8 is thus completed. □

0.4.2 Absolutely continuous probability distributions

0.4.2.1 Continuous uniform distribution

Definition 0.4.9 (Continuous uniform distribution). *Let $d \in \mathbb{N}$ and let $A \in \mathcal{B}(\mathbb{R}^d)$ be a set with $0 < \lambda_{\mathbb{R}^d}(A) < \infty$. Then we denote by $\mathcal{U}_A \colon \mathcal{B}(\mathbb{R}^d) \to [0,\infty]$ the function which satisfies for all $B \in \mathcal{B}(\mathbb{R}^d)$ that*

$$\mathcal{U}_A(B) = \frac{\lambda_{\mathbb{R}^d}(B \cap A)}{\lambda_{\mathbb{R}^d}(A)} \tag{0.82}$$

and we call $\mathcal{U}_A$ the uniform distribution on A.

0.4.2.2 Exponential distribution

Definition 0.4.10 (Exponential distribution). *Let $\lambda \in (0,\infty)$. Then we denote by $exp_\lambda \colon \mathcal{B}(\mathbb{R}) \to [0,\infty]$ the function which satisfies for all $B \in \mathcal{B}(\mathbb{R})$ that*

$$exp_\lambda(B) = \int_B \lambda\, e^{-\lambda x}\, \mathbb{1}^{\mathbb{R}}_{(0,\infty)}(x)\, dx \tag{0.83}$$

and we call exp_λ the exponential distribution with parameter λ.

0.4.2.3 Cauchy distribution

Definition 0.4.11 (Cauchy distribution). *Let $\mu \in \mathbb{R}$ and $\lambda \in (0, \infty)$. Then we denote by $Cau_{\mu,\lambda} \colon \mathcal{B}(\mathbb{R}) \to [0, \infty]$ the function which satisfies for all $B \in \mathcal{B}(\mathbb{R})$ that*

$$Cau_{\mu,\lambda}(B) = \int_B \frac{1}{\pi\lambda\left(1 + \frac{(x-\mu)^2}{\lambda^2}\right)}\,dx \tag{0.84}$$

and we call $Cau_{\mu,\lambda}$ the Cauchy distribution with parameters μ and λ.

0.4.2.4 Laplace distribution

Definition 0.4.12 (Laplace distribution). *Let $\lambda \in (0, \infty)$. Then we denote by $Laplace_\lambda \colon \mathcal{B}(\mathbb{R}) \to [0, \infty]$ the function which satisfies for all $B \in \mathcal{B}(\mathbb{R})$ that*

$$Laplace_\lambda(B) = \frac{\lambda}{2}\int_B e^{-\lambda|x|}\,dx$$

and we call $Laplace_\lambda$ the Laplace distribution with parameter λ (we call $Laplace_\lambda$ the double exponential distribution with parameter λ).

0.4.3 Normal distribution

In the next definition we introduce the normal distribution. We achieve this by using the explicit density in the case of the standard normal distribution and by means of an affine transformation in the case of the general normal distribution.

Definition 0.4.13 (Normal distribution). *Let $d \in \mathbb{N}$, $v \in \mathbb{R}^d$ and let $Q \in \mathbb{R}^{d\times d}$ be a nonnegative symmetric $d \times d$-matrix. Then we denote by $\mathcal{N}_{0,I_{\mathbb{R}^d}} \colon \mathcal{B}(\mathbb{R}^d) \to [0, \infty]$ the function which satisfies for all $B \in \mathcal{B}(\mathbb{R}^d)$ that*

$$\mathcal{N}_{0,I_{\mathbb{R}^d}}(B) = \frac{1}{(2\pi)^{d/2}}\int_B e^{\frac{-1}{2}\|x\|^2_{\mathbb{R}^d}}\,dx, \tag{0.85}$$

we call $\mathcal{N}_{0,I_{\mathbb{R}^d}}$ the d-dimensional standard normal distribution, we denote by $\mathcal{N}_{v,Q} \colon \mathcal{B}(\mathbb{R}^d) \to [0, \infty]$ the function which satisfies for all $B \in \mathcal{B}(\mathbb{R}^d)$ that

$$\mathcal{N}_{v,Q}(B) = \mathcal{N}_{0,I_{\mathbb{R}^d}}\big(\sqrt{Q}\,\mathrm{id}_{\mathbb{R}^d} + v \in B\big) = \mathcal{N}_{0,I_{\mathbb{R}^d}}\left(\left\{x \in \mathbb{R}^d \colon \sqrt{Q}x + v \in B\right\}\right), \tag{0.86}$$

and we call $\mathcal{N}_{v,Q}$ the normal distribution with mean v and covariance Q.

We note that for all $d \in \mathbb{N}$, $v \in \mathbb{R}^d$ and all symmetric nonnegative $d \times d$-matrices $Q \in \mathbb{R}^{d\times d}$ it holds that $\mathcal{N}_{v,Q}(\mathbb{R}^d) = 1$. It thus holds for every $d \in \mathbb{N}$, $v \in \mathbb{R}^d$ and every nonnegative $d \times d$-matrices $Q \in \mathbb{R}^{d\times d}$ that $\mathcal{N}_{v,Q}$ is indeed a probability measure.

Furthermore, we observe that for all $d \in \mathbb{N}$, $v \in \mathbb{R}^d$ and all symmetric nonnegative $d \times d$-matrices $Q \in \mathbb{R}^{d\times d}$ it holds that the probability measure $\mathcal{N}_{v,Q}$ is absolutely continuous if and only if Q is invertible. Moreover, we note that for all $d \in \mathbb{N}$, $v \in \mathbb{R}^d$ and all symmetric nonnegative $d \times d$-matrices $Q \in \mathbb{R}^{d\times d}$ it holds that the probability measure $\mathcal{N}_{v,Q}$ is discrete if and only if $Q = 0 \in \mathbb{R}^{d\times d}$. In the next exercise, Exercise 0.4.14, a simple property of the probability measure $\mathcal{N}_{0,I_{\mathbb{R}^d}}$ is formulated.

Exercise 0.4.14. *Prove that for all $d \in \mathbb{N}$, $i, j \in \{1, \dots, d\}$ it holds that*

$$\int_{\mathbb{R}^d} x_i \, \mathcal{N}_{0,I_{\mathbb{R}^d}}(dx_1, \dots, dx_d) = 0, \quad \int_{\mathbb{R}^d} x_i \cdot x_j \, \mathcal{N}_{0,I_{\mathbb{R}^d}}(dx_1, \dots, dx_d) = \begin{cases} 1 & : i = j \\ 0 & : else \end{cases}.$$

The normal distributed is preserved under affine linear transformations. This is the subject of the next proposition.

Proposition 0.4.15 (Affine linear transformations of the normal distribution)**.** Let $(\Omega, \mathcal{A}, P)$ be a probability space, let $d, m \in \mathbb{N}$, $v \in \mathbb{R}^m$, $b \in \mathbb{R}^d$, $A \in \mathbb{R}^{d\times m}$, let $Q \in \mathbb{R}^{m\times m}$ be a nonnegative symmetric $m \times m$-matrix, and let

$$X : \Omega \to \mathbb{R}^m \tag{0.87}$$

be an $\mathcal{N}_{v,Q}$-distributed random variable. Then the random variable

$$\Omega \ni \omega \mapsto AX(\omega) + b \in \mathbb{R}^d \tag{0.88}$$

is $\mathcal{N}_{Av+b,AQA^\top}$-distributed.

Proof of Proposition 0.4.15. We prove Proposition 0.4.15 by using *characteristic functions*. More formally, note that for all $x \in \mathbb{R}^m$ it holds that

$$\mathbb{E}\Big[e^{i\langle X,x\rangle_{\mathbb{R}^m}}\Big] = e^{i\langle v,x\rangle_{\mathbb{R}^m} - \frac{1}{2}\langle x,Qx\rangle_{\mathbb{R}^m}} \tag{0.89}$$

(see, e.g., Remark 15.54 in Klenke [Klenke(2008)]). This implies that for all $y \in \mathbb{R}^d$ it holds that

$$\begin{aligned} \mathbb{E}\Big[e^{i\langle AX+b,y\rangle_{\mathbb{R}^d}}\Big] &= e^{i\langle b,y\rangle_{\mathbb{R}^d}} \cdot \mathbb{E}\Big[e^{i\langle AX,y\rangle_{\mathbb{R}^d}}\Big] = e^{i\langle b,y\rangle_{\mathbb{R}^d}} \cdot \mathbb{E}\Big[e^{i\left\langle X,A^\top y\right\rangle_{\mathbb{R}^m}}\Big] \\ &= e^{i\langle b,y\rangle_{\mathbb{R}^d}} \cdot e^{i\left\langle v,A^\top y\right\rangle_{\mathbb{R}^m} - \frac{1}{2}\left\langle A^\top y,QA^\top y\right\rangle_{\mathbb{R}^m}} \\ &= e^{i\langle Av+b,y\rangle_{\mathbb{R}^d} - \frac{1}{2}\left\langle y,AQA^\top y\right\rangle_{\mathbb{R}^d}}. \end{aligned} \tag{0.90}$$

This shows that

$$\Omega \ni \omega \mapsto AX(\omega) + b \in \mathbb{R}^d \tag{0.91}$$

is $\mathcal{N}_{Av+b,AQA^\top}$-distributed. The proof of Proposition 0.4.15 is thus completed. □

The next result, Corollary 0.4.16, is an immediate consequence from Proposition 0.4.15 and Exercise 0.4.14.

Corollary 0.4.16. *Let $(\Omega, \mathcal{A}, P)$ be a probability space, let $d \in \mathbb{N}$, $v \in \mathbb{R}^d$, let $Q \in \mathbb{R}^{d\times d}$ be a nonnegative symmetric $d \times d$-matrix, and let $X\colon \Omega \to \mathbb{R}^d$ be an $\mathcal{N}_{v,Q}$-distributed random variable. Then*

$$\mathbb{E}_P[X] = v \qquad \text{and} \qquad \mathsf{Cov}_P(X) = Q. \tag{0.92}$$

Corollary 0.4.16 motivates the following notion.

Definition 0.4.17 (Normally distributed random variable)**.** *We say that X is normally distributed on $\mathbf{\Omega}$ (we say that X is normally distributed, we say that X is jointly normally distributed, we say that X is Gaussian distributed, we say that X is jointly Gaussian distributed) if and only if there exist a natural number $n \in \mathbb{N}$ and a probability space $(\Omega, \mathcal{F}, P)$ such that it holds*

(i) that $\mathbf{\Omega} = (\Omega, \mathcal{F}, P)$,

(ii) that X is an $\mathcal{F}/\mathcal{B}(\mathbb{R}^n)$-measurable function,

(iii) that $\mathbb{E}_P\big[\|X\|^2_{\mathbb{R}^n}\big] < \infty$, and

(iv) that $X(P)_{\mathcal{B}(\mathbb{R}^n)} = \mathcal{N}_{\mathbb{E}_P[X],\mathsf{Cov}_P(X)}$.

Definition 0.4.18 (Standard normal random variable)**.** *We say that X is a standard normal random variable on $\mathbf{\Omega}$ (we say that X is a standard normal random variable) if and only if there exists a probability space $(\Omega, \mathcal{F}, P)$ such that it holds*

(i) that $\mathbf{\Omega} = (\Omega, \mathcal{F}, P)$,

(ii) that X is an $\mathcal{F}/\mathcal{B}(\mathbb{R})$-measurable function, and

(iii) that $X(P)_{\mathcal{B}(\mathbb{R})} = \mathcal{N}_{0,I_{\mathbb{R}}}$.

1 Generation of random numbers

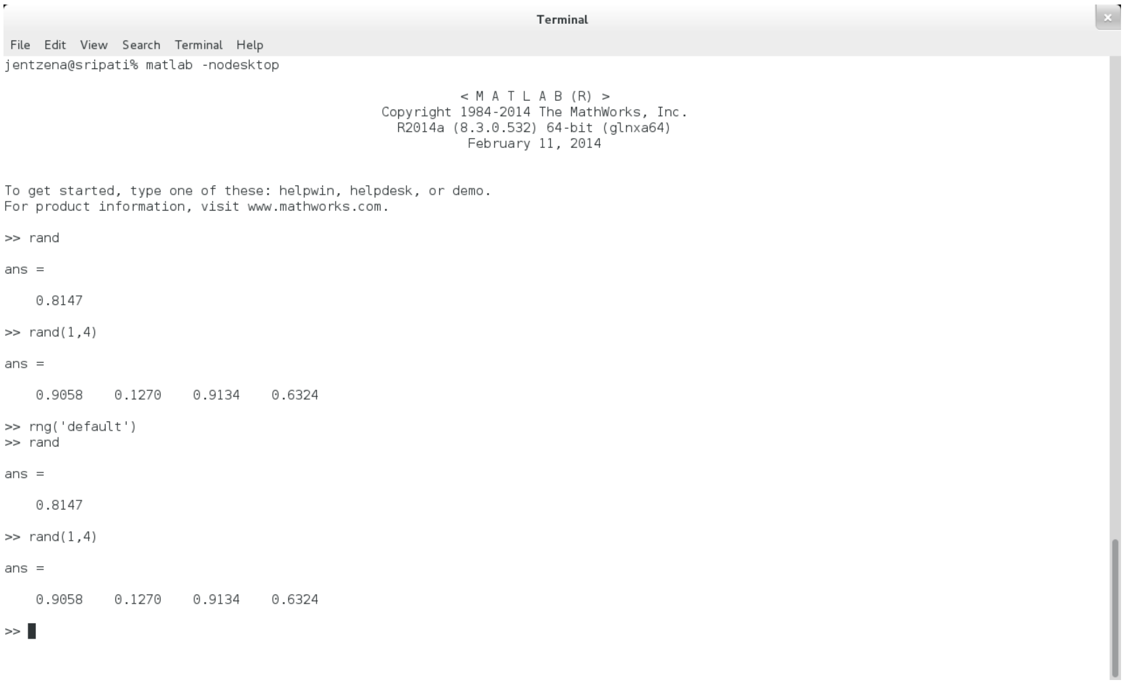
1.1 Pseudo random number generators

If one talks about random numbers, first of all the question arises what randomness means. Does randomness exist? What do we mean by *random numbers*? Discussing these questions, one very quickly ends up in a philosophical discussion. One possible *real* random number generator is a USB stick which returns zeros and ones according to physical phenomena like. Here we do not focus on this discussion but we will briefly sketch the concept and the generation of $\mathcal{U}_{(0,1)}$*-pseudo random numbers*.

Definition 1.1.1. *Let* $(\Omega, \mathcal{A}, P)$ *be a probability space and let* $U_n\colon \Omega \to \mathbb{R}$*,* $n \in \mathbb{N}$*, be a sequence of* P*-independent* $\mathcal{U}_{(0,1)}$*-distributed random variables.* $\mathcal{U}_{(0,1)}$-pseudo random numbers *are sequences of real numbers that are calculated by a deterministic algorithm and that have – in an appropriate sense – similiar statistical properties as* $(U_n)_{n\in\mathbb{N}}$*.*

Clearly, this is a very vague definition and the reader is referred to the literature (e.g., [Knuth(1998), Kloeden and Platen(1992), Higham(2004), Glasserman(2004)] and the references mentioned therein) for a more elaborate treatment of $\mathcal{U}_{(0,1)}$-pseudo random numbers. An advantage of $\mathcal{U}_{(0,1)}$-pseudo random numbers compared to "real" random numbers is the fact that they can be reproduced and one can repeat experiments or variants of experiments with the same pseudo random input (cf. the command "*rng('default')*" in Matlab). The algorithms that produce $\mathcal{U}_{(0,1)}$-pseudo random numbers are called $\mathcal{U}_{(0,1)}$-pseudo random number generators ($\mathcal{U}_{(0,1)}$-PRNGs for short). Examples of classical $\mathcal{U}_{(0,1)}$-PRNGs are Knuth's *Algorithm K*, Lehmer's linear congruent pseudo random number generator, which imitates Roulette, and IBM's *RANDU* (see, e.g., [Knuth(1998)]). These generators should not be used nowadays. When many random numbers are needed, $\mathcal{U}_{(0,1)}$-PRNGs with large periods have to be used. Today examples of in this sense "good" $\mathcal{U}_{(0,1)}$-PRNGs are *Marsaglia's Mother*, *Mersenne Twister*, *Kiss*, ... The function "rand" in MATLAB allows to use different $\mathcal{U}_{(0,1)}$-PRNGs, where Mersenne Twister is used in the R2011a release by default.

In the following we assume that we are given a method that generates independent $\mathcal{U}_{(0,1)}$-distributed random numbers (compare with the command "*rand*" in Matlab). We then present different methods how to transform these $\mathcal{U}_{(0,1)}$-distributed random numbers into random numbers with other distributions. First, in Section 1.2, we present methods that work for a general class of distributions. Later, in Section 1.3, specific methods for the normal distribution are presented.



The content of this chapter is well known in the literature on the generation of random numbers. It can in a similar form be found in several books and lectures notes containing a section on the generation of random numbers; cf., e.g., in [Kloeden and Platen(1992)], [Wichura(2001)], [Glasserman(2004)], [Ross(2006)], [Asmussen and Glynn(2007)] and [Müller-Gronbach et al.(2012)Müller-Gronbach, Novak, and Ritter].

1.2 Methods for general distributions

1.2.1 Inversion method

This subsection presents the *inversion method* (also known as *inverse transformation method*) which transforms uniformly distributed random numbers to those of an arbitrary distribution by using a suitable generalized inverse of the distribution function. This generalized inverse of the distribution function is the subject of the next definition.

Definition 1.2.1 (Generalized inverse distribution function associated to a distribution function)**.** *Let* $F\colon \mathbb{R} \to [0,1]$ *be a distribution function. Then we denote by* $I_F\colon (0,1) \to \mathbb{R}$ *the function which satisfies for all* $y \in (0,1)$ *that*

$$I_F(y) = \inf\{x \in \mathbb{R}\colon F(x) \geq y\} = \inf\bigl(F^{-1}([y,1])\bigr) \tag{1.1}$$

and we call I_F *the generalized inverse distribution function associated to* F*.*

Definition 1.2.1 plays a key role in the inversion method which we present in Theorem 1.2.7 below. In the following we present a few comments regarding Definition 1.2.1.

(i) Let $F\colon \mathbb{R} \to [0,1]$ be a distribution function. Then note that for every $y \in (0,1)$ it holds that the set

$$\{x \in \mathbb{R}\colon F(x) \geq y\} \tag{1.2}$$

(see Definition 1.2.1) is not empty and bounded from below. Indeed, the fact that F satisfies

$$\lim_{x\to\infty} F(x) = 1 \tag{1.3}$$

ensures that for every $y \in (0,1)$ it holds that $\{x \in \mathbb{R}\colon F(x) \geq y\}$ is not empty. Moreover, the fact that F satisfies

$$\lim_{x\to-\infty} F(x) = 0 \tag{1.4}$$

implies that for every $y \in (0,1)$ it holds that $\{x \in \mathbb{R}\colon F(x) \geq y\}$ is bounded from below. Therefore, the function $I_F\colon (0,1) \to \mathbb{R}$ in Definition 1.2.1 is well defined.

(ii) Let $F\colon \mathbb{R} \to [0,1]$ be a distribution function. The *generalized inverse distribution function associated to* F is sometimes also referred to as *quantile function associated to* F (cf. Item (iii) below).

(iii) Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $X\colon \Omega \to \mathbb{R}$ be an $\mathcal{F}/\mathcal{B}(\mathbb{R})$-measurable function (we think of X as a model for the change of the value of a given portfolio of financial assets within a given time period), let $F\colon \mathbb{R} \to [0,1]$ be the distribution function of $-X$ and let $\alpha \in (0,1)$ (we think of α as a confidence level for the statement below, which is typically a number close to 1 such as 99% or 95%, and we think of $1-\alpha$ as a small number such as 1% or 5%). Then the real number

$$\begin{aligned} I_F(\alpha) &= \min\{x \in \mathbb{R}\colon F(x) \geq \alpha\} = \min\{x \in \mathbb{R}\colon P(-X \leq x) \geq \alpha\} \\ &= \min\{x \in \mathbb{R}\colon P(\neg\,[-X > x]) \geq \alpha\} \\ &= \min\{x \in \mathbb{R}\colon P(-X > x) \leq 1-\alpha\} \end{aligned} \tag{1.5}$$

is sometimes referred to as the *value at risk with confidence level* α *associated to* $-X$ (*VaR with confidence level* α *associated to* $-X$) in the financial risk management literature. Observe that (1.5) ensures that

$$P\big(-X \leq I_F(\alpha)\big) = P\big(\neg\,[-X > I_F(\alpha)]\,\big) \geq \alpha. \tag{1.6}$$

It thus holds with a probability of at least α that the loss $-X$ will not exceed the number $I_F(\alpha)$ (*the value at risk with confidence level* α).

In the next step we illustrate Definition 1.2.1 by means of a simple example.

Example 1.2.2 (Generalized inverse distribution function for the Bernoulli distribution)**.** *Let* $p \in [0,1]$ *be a real number and let* $F\colon \mathbb{R} \to [0,1]$ *be the distribution function* *of the Bernoulli distribution with parameter $p \in [0,1]$, i.e., assume that for all $x \in \mathbb{R}$ it holds that*

$$F(x) = \mathrm{Ber}_p((-\infty,x]) = (1-p)\cdot\delta_0^{\mathbb{R}}((-\infty,x]) + p\cdot\delta_1^{\mathbb{R}}((-\infty,x]) \\ = \begin{cases} 0 & : x < 0, \\ 1-p & : 0 \le x < 1, \\ 1 & : x \ge 1 \end{cases} \tag{1.7}$$

(see Subsection 0.4.1.2). Then the generalized inverse distribution function $I_F\colon (0,1) \to \mathbb{R}$ associated to F satisfies that for all $y \in (0,1)$ it holds that

$$I_F(y) = \inf\{x \in [0,\infty)\colon F(x) \ge y\} = \begin{cases} 0 & : 0 < y \le 1-p \\ 1 & : 1-p < y < 1 \end{cases}. \tag{1.8}$$

Class exercise 1.2.3. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $c \in \mathbb{R}$, let $X\colon \Omega \to \mathbb{R}$ be an $\mathcal{F}/\mathcal{B}(\mathbb{R})$-measurable function which satisfies for all $\omega \in \Omega$ that*

$$X(\omega) = c, \tag{1.9}$$

and let $F\colon \mathbb{R} \to [0,1]$ be the distribution function of X. What is $I_F(y)$, $y \in (0,1)$?

A few essentially well-known properties of the inverse distribution function are collected in the following two lemmas, Lemma 1.2.4 and Lemma 1.2.5. Most of the statements and most of the proofs of Lemma 1.2.4 and Lemma 1.2.5 can, for example, be found in Theorems 2 and 3 in the first section in [Wichura(2001)]. In particular, we also follow [Wichura(2001)] by referring to the properties in Item (iv) and Item (v) in Lemma 1.2.4 as "switching formulas".

Lemma 1.2.4 (Properties of generalized inverse distribution functions)**.** *Let $F\colon \mathbb{R} \to [0,1]$ be a distribution function and let $I_F\colon (0,1) \to \mathbb{R}$ be the generalized inverse distribution function associated to F. Then I_F fulfills the following properties:*

(i) I_F is non-decreasing, i.e., for all $y_1, y_2 \in (0,1)$ with $y_1 \le y_2$: $I_F(y_1) \le I_F(y_2)$,

(ii) for all $y \in (0,1)$: $F(I_F(y)) \ge y$, $F^{-1}([y,1]) = [I_F(y), \infty)$, and $I_F(y) = \min\big(F^{-1}([y,1])\big)$

(iii) for all $x \in F^{-1}((0,1)) = \{z \in \mathbb{R}\colon F(z) \in (0,1)\}$: $I_F(F(x)) \le x$,

(iv) for all $x \in \mathbb{R}$, $y \in (0,1)$: $I_F(y) \le x$ if and only if $y \le F(x)$ (switching formula),

(v) for all $x \in \mathbb{R}$, $y \in (0,1)$: $I_F(y) > x$ if and only if $y > F(x)$ (switching formula), and

(vi) for all open sets $D \subseteq \mathbb{R}$ with the property that $F|_D\colon D \to [0,1]$ is injective and all $y \in F(D)$: $F|_D^{-1}(y) = I_F(y)$.

Proof of Lemma 1.2.4. First of all, observe that for all $y_1, y_2 \in \mathbb{R}$ with $y_1 \leq y_2$ it holds that

$$\{x \in \mathbb{R} \colon F(x) \geq y_1\} \supseteq \{x \in \mathbb{R} \colon F(x) \geq y_2\}. \tag{1.10}$$

This proves the monotonicity of I_F as asserted in Item (i). Next let $y \in (0,1)$ be arbitrary and let $(x_n)_{n\in\mathbb{N}} \subseteq F^{-1}([y,1])$ be a non-increasing sequence of real numbers which satisfies

$$\lim_{n\to\infty} x_n = \inf\{x \in \mathbb{R} \colon F(x) \geq y\} = I_F(y). \tag{1.11}$$

The definition of the infimum ensures that such a sequence does indeed exist. The right continuity of F assures that

$$y \leq \lim_{n\to\infty} F(x_n) = F\Big(\lim_{n\to\infty} x_n\Big) = F(I_F(y)). \tag{1.12}$$

This proves Item (ii). In the next step we observe that for all $x \in F^{-1}((0,1)) \subseteq \mathbb{R}$ it holds that

$$I_F\big(F(x)\big) = \inf\{z \in \mathbb{R} \colon F(z) \geq F(x)\} \leq \inf\{x\} = x. \tag{1.13}$$

This establishes Item (iii). The switching formula in Item (iv) follows immediately from Item (ii), from Item (iii), from the monotonicity of F, and from the monotonicity of I_F (see Item i). Clearly, the switching formula in Item (iv) is equivalent to the switching formula in Item (v). Next let $D \subseteq \mathbb{R}$ be an open set such that $F|_D \colon D \to [0,1]$ is injective (which is equivalent to the assumption that $F|_D \colon D \to [0,1]$ is strictly increasing). The fact that $F \colon \mathbb{R} \to [0,1]$ is non-decreasing and the fact that D is an open set ensure that for all $z \in F(D)$ it holds that

$$F^{-1}(\{z\}) = \{x \in \mathbb{R} \colon F(x) = z\} = \{F|_D^{-1}(z)\}. \tag{1.14}$$

This implies that for all $z \in F(D)$ it holds that

$$\begin{aligned} I_F(z) &= \inf\{x \in \mathbb{R} \colon F(x) \geq z\} = \inf\{x \in \mathbb{R} \colon F(x) = z\} \\ &= \inf\{F|_D^{-1}(z)\} = F|_D^{-1}(z). \end{aligned} \tag{1.15}$$

The proof of Lemma 1.2.4 is thus completed. □

Distribution functions are càdlàg (continue à droite, limitée à gauche) functions and are, in particular, right continuous; recall Definition 0.3.7 in Chapter 0. The generalized inverse distribution function associated to a given distribution function is, in turn, a càglàd (continue à gauche, limitée à droite) function and is, in particular, left continuous. This is the subject of the next lemma.

Lemma 1.2.5 (Continuity properties of generalized inverse distribution functions)**.** *Let $F\colon \mathbb{R} \to [0,1]$ be a distribution function and let $I_F\colon (0,1) \to \mathbb{R}$ be the generalized inverse distribution function associated to F. Then I_F fulfills the following properties:*

(i) I_F is a càglàd (continue à gauche, limitée à droite) function, i.e., for all $y \in (0,1)$: $\lim_{z\nearrow y} I_F(z)$ and $\lim_{z\searrow y} I_F(z)$ exist and $\lim_{z\nearrow y} I_F(z) = I_F(y)$,

(ii) for all $y \in (0,1)$: $\big[I_F(y), \lim_{z\searrow y} I_F(z)\big) \subseteq F^{-1}(\{y\}) \subseteq \big[I_F(y), \lim_{z\searrow y} I_F(z)\big]$,

(iii) for all $y \in (0,1)$: I_F is continuous in $y \in (0,1)$ if and only if $\#_{\mathbb{R}}\big(F^{-1}(\{y\})\big) = \#_{\mathbb{R}}\big(\{x \in \mathbb{R}\colon F(x) = y\}\big) \leq 1$.

Proof of Lemma 1.2.5. First of all, observe that the fact that I_F is non-decreasing (see Item (i) in Lemma 1.2.5) implies that for every $y \in (0,1)$ it holds that the limits

$$\lim_{z\nearrow y} I_F(z) \qquad \text{and} \qquad \lim_{z\searrow y} I_F(z) \tag{1.16}$$

exist. The monotonicity of I_F also proves that for all $y \in (0,1)$ it holds that

$$\lim_{z\nearrow y} I_F(z) \leq I_F(y). \tag{1.17}$$

It thus remains to establish that for all $y \in (0,1)$ it holds that

$$\lim_{z\nearrow y} I_F(z) \geq I_F(y). \tag{1.18}$$

For this observe that Item (ii) in Lemma 1.2.4 and the fact that F and I_F are non-decreasing imply that for all $y \in (0,1)$, $\varepsilon \in (0,y)$ it holds that

$$F\left(\lim_{z\nearrow y} I_F(z)\right) \geq F(I_F(y-\varepsilon)) \geq y - \varepsilon. \tag{1.19}$$

This, in turn, ensures that for all $y \in (0,1)$ it holds that

$$F\left(\lim_{z\nearrow y} I_F(z)\right) \geq y. \tag{1.20}$$

The definition of I_F hence establishes that for all $y \in (0,1)$ it holds that

$$I_F(y) = \inf\{x \in \mathbb{R}\colon F(x) \geq y\} \leq \inf\left\{\lim_{z\nearrow y} I_F(z)\right\} = \lim_{z\nearrow y} I_F(z). \tag{1.21}$$

This proves Item (i). In the next step we note that for all $y \in (0,1)$ and all $x \in F^{-1}(\{y\})$ it holds that

$$F(x) \geq y. \tag{1.22}$$

The fact that

$$\forall\, y \in (0,1)\colon I_F(y) = \min\{z \in \mathbb{R}\colon F(z) \geq y\} \tag{1.23}$$

hence implies that for all $y \in (0,1)$ and all $x \in F^{-1}(\{y\})$ it holds that

$$x \geq I_F(y). \tag{1.24}$$

This ensures that for all $y \in (0,1)$ it holds that

$$F^{-1}(\{y\}) \subseteq [I_F(y), \infty). \tag{1.25}$$

Next we note that for all $x \in \mathbb{R}$, $y \in (0,1)$ with $\lim_{z \searrow y} I_F(z) < x$ it holds that there exists a real number $z \in (y,1)$ such that

$$I_F(z) \leq x. \tag{1.26}$$

The switching formula in Item (iv) in Lemma 1.2.4 therefore implies that for every $x \in \mathbb{R}$, $y \in (0,1)$ with $\lim_{z \searrow y} I_F(z) < x$ it holds that there exists a real number $z \in (y,1)$ such that

$$z \leq F(x). \tag{1.27}$$

This proves that for every $x \in \mathbb{R}$, $y \in (0,1)$ with $\lim_{z \searrow y} I_F(z) < x$ it holds that

$$y < F(x). \tag{1.28}$$

This implies that for all $y \in (0,1)$ it holds that

$$\left(\lim_{z \searrow y} I_F(z), \infty\right) \subseteq \{x \in \mathbb{R} \colon F(x) > y\} = F^{-1}((y,1]) \,. \tag{1.29}$$

This, in turn, ensures that for every $y \in (0,1)$ it holds that

$$\begin{split} F^{-1}(\{y\}) \subseteq F^{-1}([0,y]) = F^{-1}\big([0,1] \backslash (y,1]\big) = \mathbb{R} \backslash F^{-1}((y,1]) \\ \subseteq \mathbb{R} \backslash \left(\lim_{z \searrow y} I_F(z), \infty\right) = \left(-\infty, \lim_{z \searrow y} I_F(z)\right] . \end{split} \tag{1.30}$$

Combining (1.25) and (1.30) proves that for every $y \in (0,1)$ it holds that

$$F^{-1}(\{y\}) \subseteq \Big[I_F(y), \infty\Big) \cap \left(-\infty, \lim_{z \searrow y} I_F(z)\right] = \left[I_F(y), \lim_{z \searrow y} I_F(z)\right] . \tag{1.31}$$

Next we note that for every $x \in \mathbb{R}$, $y \in (0,1)$ with $y < F(x)$ it holds that there exists a real number $z \in (y,1)$ such that

$$z \leq F(x). \tag{1.32}$$

The switching formula in Item (iv) in Lemma 1.2.4 hence implies that for every $x \in \mathbb{R}$, $y \in (0,1)$ with $y < F(x)$ it holds that there exists a real number $z \in (y,1)$ such that

$$I_F(z) \leq x. \tag{1.33}$$

This ensures that for every $x \in \mathbb{R}$, $y \in (0,1)$ with $y < F(x)$ it holds that

$$\lim_{z \searrow y} I_F(z) \leq x. \tag{1.34}$$

This proves that for every $y \in (0,1)$ it holds that

$$F^{-1}((y,1]) = \{x \in \mathbb{R}\colon F(x) > y\} \subseteq \Big[\lim_{z \searrow y} I_F(z), \infty\Big) \tag{1.35}$$

Hence, we obtain that for all $y \in (0,1)$ it holds that

$$\Big(-\infty, \lim_{z \searrow y} I_F(z)\Big) = \mathbb{R}\backslash \Big[\lim_{z \searrow y} I_F(z), \infty\Big) \subseteq \mathbb{R}\backslash F^{-1}((y,1]) = F^{-1}\big([0,y]\big). \tag{1.36}$$

This together with the fact that

$$\forall\, y \in (0,1)\colon [I_F(y), \infty) = F^{-1}([y,1]) \tag{1.37}$$

(see Item (ii) in Lemma 1.2.4) implies that for every $y \in (0,1)$ it holds that

$$\begin{aligned} \Big[I_F(y), \lim_{z \searrow y} I_F(z)\Big) &= \Big(-\infty, \lim_{z \searrow y} I_F(z)\Big) \cap [I_F(y), \infty) \\ &\subseteq F^{-1}\big([0,y]\big) \cap F^{-1}([y,1]) \\ &= F^{-1}([0,y] \cap [y,1]) = F^{-1}(\{y\}). \end{aligned} \tag{1.38}$$

This proves Item (ii). Item (iii) is an immediate consequence of Item (ii). The proof of Lemma 1.2.5 is thus completed. □

Class exercise 1.2.6. *Prove or disprove the following statement: For every distribution function $F\colon \mathbb{R} \to [0,1]$ and every $y \in (0,1)$, $x \in \mathbb{R}$ it holds that $F(x) > y$ if and only if $x > I_F(y)$.*

In the following we assume that there exists a method to generate $\mathcal{U}_{(0,1)}$-distributed random numbers. Then the next proposition results in a method to generate realizations of a real valued random variable with an arbitrary given distribution function. This method is referred to as *inversion method* or *inverse transformation method* in the literature.

Theorem 1.2.7 (Inversion method)**.** *Let $F\colon \mathbb{R} \to [0,1]$ be a distribution function, let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $U\colon \Omega \to \mathbb{R}$ be an $\mathcal{U}_{(0,1)}$-distributed random variable with $U(\Omega) \subseteq (0,1)$. Then F is the distribution function of the $\mathcal{F}/\mathcal{B}(\mathbb{R})$-measurable function $I_F(U) = I_F \circ U\colon \Omega \to \mathbb{R}$, i.e., it holds for all $x \in \mathbb{R}$ that*

$$P\big(I_F(U) \leq x\big) = F(x). \tag{1.39}$$

Proof of Theorem 1.2.7. Observe that the switching formula in Item (iv) of Lemma 1.2.4 implies that for all $x \in \mathbb{R}$ it holds that

$$P\big(I_F(U) \leq x\big) = P\big(U \leq F(x)\big) = F(x). \tag{1.40}$$

This completes the proof of Theorem 1.2.7. □

We add some remarks concerning Theorem 1.2.7. Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $U\colon \Omega \to \mathbb{R}$ be an $\mathcal{U}_{(0,1)}$-distributed random variable with $U(\Omega) \subseteq (0,1)$, let $X\colon \Omega \to \mathbb{R}$ be an arbitrary $\mathcal{F}/\mathcal{B}(\mathbb{R})$-measurable function with distribution function $F\colon \mathbb{R} \to [0,1]$ (i.e., assume for all $x \in \mathbb{R}$ that $F(x) = P(X \leq x)$), and let $\hat{X}\colon \Omega \to \mathbb{R}$ be the function given by $\hat{X} = I_F(U)$. Theorem 1.2.7 proves that $\hat{X}$ and X have the same distribution on $\mathcal{B}(\mathbb{R})$, i.e., Theorem 1.2.7 proves that

$$X(P)_{\mathcal{B}(\mathbb{R})} = \hat{X}(P)_{\mathcal{B}(\mathbb{R})}. \tag{1.41}$$

If we thus want to simulate a realization from X, it is thus sufficient to calculate the generalized inverse distribution function $I_F\colon (0,1) \to \mathbb{R}$ associated to F, to simulate a realization from U, and then to put this realization as an argument of I_F.

Class exercise 1.2.8. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $U\colon \Omega \to \mathbb{R}$ be an $\mathcal{U}_{(0,1)}$-distributed random variable, let $X\colon \Omega \to \mathbb{R}$ be a function which satisfies for all $\omega \in \Omega$ that*

$$X(\omega) = \sin(U(\omega)), \tag{1.42}$$

and let $F\colon \mathbb{R} \to [0,1]$ be the distribution function of X. What is $I_F(y)$, $y \in (0,1)$?

We now calculate the generalized inverse distribution function for a few example probability distributions which are absolutely continuous with respect to the Lebesgue measure.

Example 1.2.9 (Absolutely continuous distributions)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $X\colon \Omega \to \mathbb{R}$ be an $\mathcal{F}/\mathcal{B}(\mathbb{R})$-measurable function with distribution function $F\colon \mathbb{R} \to [0,1]$, and let $U\colon \Omega \to \mathbb{R}$ be an $\mathcal{U}_{(0,1)}$-distributed random variable with $U(\Omega) \subseteq (0,1)$.*

(i) In this item let $n \in \mathbb{N}$ be a natural number and assume that F satisfies for all $x \in (0,1)$ that

$$F(x) = x^n. \tag{1.43}$$

Then $F|_{(0,1)}\colon (0,1) \to [0,1]$ is injective and it holds for all $y \in (0,1)$ that

$$F|_{(0,1)}^{-1}(y) = y^{\frac{1}{n}}. \tag{1.44}$$

Item (vi) *in Lemma 1.2.4 and Theorem 1.2.7 hence prove that*

$$X = U^{\frac{1}{n}} \tag{1.45}$$

in distribution on $\mathcal{B}(\mathbb{R})$.

(ii) In this item let $\lambda \in (0,\infty)$ be a real number and assume that X is $\exp_\lambda$*-distributed (*exponentially distributed with parameter λ*; see Subsection 0.4.2.2). Then it holds for all $x \in \mathbb{R}$ that*

$$F(x) = \exp_\lambda\big((-\infty, x]\big) = \begin{cases} 0 & : x < 0 \\ \int_0^x \lambda\, e^{-\lambda y}\, dy = \big[-e^{-\lambda y}\big]_{y=0}^{y=x} = 1 - e^{-\lambda x} & : x \geq 0 \end{cases}. \tag{1.46}$$

In particular, it holds that $F|_{(0,\infty)} \colon (0,\infty) \to [0,1]$ *is injective and Item* (vi) *in Lemma 1.2.4 therefore shows that for all* $y \in (0,1)$ *it holds that*

$$I_F(y) = \frac{-\ln(1-y)}{\lambda}. \tag{1.47}$$

Theorem 1.2.7 hence proves that

$$X = \frac{-\ln(U)}{\lambda} \tag{1.48}$$

in distribution on $\mathcal{B}(\mathbb{R})$*.*

```
N = 10^5;
lambda = 0.1;
X=-log(rand(1,N))/lambda;
hist(X,10^3);
```

Matlab code 1.1: A Matlab code which plots 10^5 realizations of a pseudo $\exp_{0.1}$-distributed random variable in an histogram with 1000 bins.

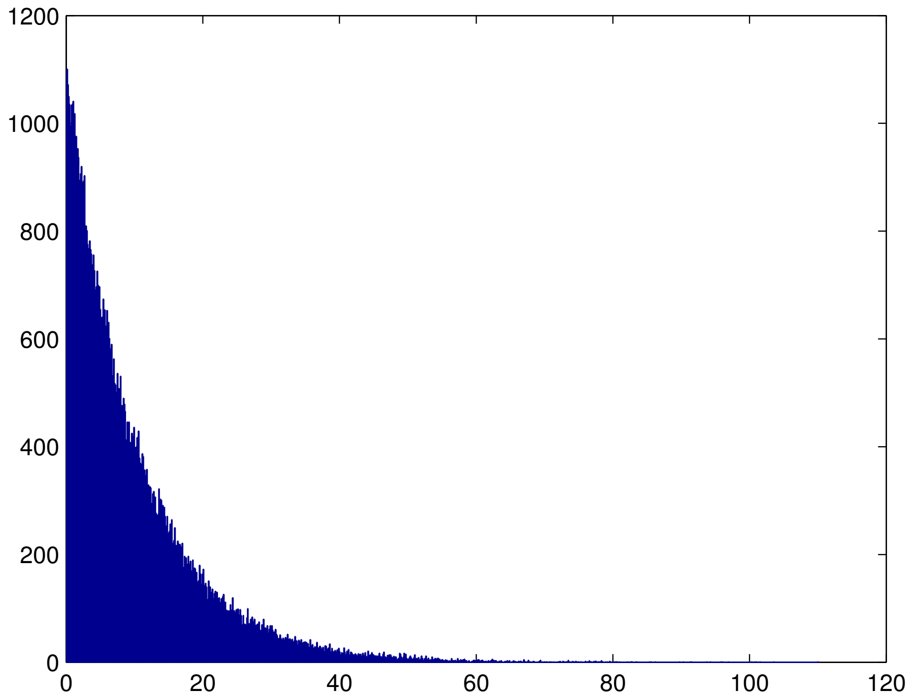


Figure 1.1: Result of a call of the Matlab code 1.1.

(iii) In this item let $\mu \in \mathbb{R}$ and $\lambda \in (0,\infty)$ be real numbers and assume that X is $\mathrm{Cau}_{\mu,\lambda}$*-distributed (*Cauchy distributed with parameters μ and λ*; see Subsection 0.4.2.3). Then it holds for all $x \in \mathbb{R}$ that*

$$\begin{aligned} F(x) = Cau_{\mu,\lambda}\big((-\infty,x]\big) &= \int_{-\infty}^{x} \frac{1}{\pi\lambda\left(1+\frac{(y-\mu)^2}{\lambda^2}\right)}\,dy \\ &= \frac{1}{\pi\lambda}\int_{-\infty}^{x} arctan'\big(\tfrac{y-\mu}{\lambda}\big)\,dy = \frac{1}{\pi}\int_{-\infty}^{\frac{x-\mu}{\lambda}} arctan'(y)\,dy \\ &= \frac{1}{\pi}\left[arctan(\tfrac{x-\mu}{\lambda}) + \frac{\pi}{2}\right] = \frac{arctan(\frac{x-\mu}{\lambda})}{\pi} + \frac{1}{2}. \end{aligned} \tag{1.49}$$

This shows that $F\colon \mathbb{R} \to [0,1]$ is injective and that for all $y \in (0,1)$ it holds that

$$F^{-1}(y) = \lambda\tan\!\big(\pi(y-\tfrac{1}{2})\big) + \mu. \tag{1.50}$$

Item (vi) *in Lemma 1.2.4 and Theorem 1.2.7 hence show that*

$$X = \lambda\tan\!\big(\pi(U-\tfrac{1}{2})\big) + \mu \tag{1.51}$$

in distribution on $\mathcal{B}(\mathbb{R})$.

After having presented the inverse transformation method for a few distributions that are absolutely continuous with respect to the Lebesgue-Borel measure, we now intend to illustrate the inverse transformation method in Theorem 1.2.7 in the case of a few discrete distributions. For this the following two notions are used.

Definition 1.2.10 (Round down to the grid)**.** *We denote by $\lfloor\cdot\rfloor_h\colon \mathbb{R}\to\mathbb{R}$, $h \in (0,\infty)$, the functions which satisfy for all $h \in (0,\infty)$, $x \in \mathbb{R}$ that*

$$\lfloor x\rfloor_h = \max((-\infty,x] \cap \{0,h,-h,2h,-2h,\dots\}). \tag{1.52}$$

Definition 1.2.11 (Round up to the grid)**.** *We denote by $\lceil\cdot\rceil_h\colon \mathbb{R}\to\mathbb{R}$, $h \in (0,\infty)$, the functions which satisfy for all $h \in (0,\infty)$, $x \in \mathbb{R}$ that*

$$\lceil x\rceil_h = \min([x,\infty) \cap \{0,h,-h,2h,-2h,\dots\}). \tag{1.53}$$

Class exercise 1.2.12. *Let $x \in \mathbb{R}$, $h \in (0,\infty)$. What is $\max\{x,\lfloor x\rfloor_h,\lceil x\rceil_h\}$ and $\min\{x,\lfloor x\rfloor_h,\lceil x\rceil_h\}$?*

Class exercise 1.2.13. *What is $\lfloor 1/8\rfloor_{1/2}$, $\lceil -2\rceil_{0.3}$, and $\lfloor -2\rfloor_{0.3}$?*

Example 1.2.14 (Discrete distributions)**.** *Let $(p_n)_{n\in\mathbb{N}_0} \subseteq [0,1]$ be a family of real numbers with*

$$\sum_{n=0}^{\infty} p_n = 1, \tag{1.54}$$

let $(\Omega, \mathcal{F}, P)$ *be a probability space, let* $U\colon \Omega \to \mathbb{R}$ *be an* $\mathcal{U}_{(0,1)}$*-distributed random variable with* $U(\Omega) \subseteq (0,1)$*, let* $X\colon \Omega \to \mathbb{R}$ *be an* $\mathcal{F}/\mathcal{B}(\mathbb{R})$*-measurable function with*

$$X(P)_{\mathcal{B}(\mathbb{R})} = \sum_{n=0}^{\infty} p_n \delta_n^{\mathbb{R}}|_{\mathcal{B}(\mathbb{R})}, \tag{1.55}$$

and let $F\colon \mathbb{R} \to [0,1]$ *be the distribution function of* X*. Note that for all* $x \in \mathbb{R}$ *it holds that*

$$F(x) = \sum_{n=0}^{\infty} \mathbb{1}_{[n,\infty)}(x) \cdot p_n = \sum_{n=0}^{\infty} \mathbb{1}_{[n,n+1)}(x) \left(\sum_{k=0}^{n} p_k \right) = \sum_{n=0}^{\lfloor x \rfloor_1} p_n. \tag{1.56}$$

Hence, we get for all $y \in (0,1)$ *that*

$$I_F(y) = \min\{n \in \mathbb{N}_0 \colon F(n) \geq y\} = \min\left\{ n \in \mathbb{N}_0 \colon \sum_{k=0}^{n} p_k \geq y \right\}. \tag{1.57}$$

We illustrate (1.57) *through a few more specific examples.*

(i) In this item let $p \in (0,1)$ *be a real number and assume that* X *is* Ber_p*-distributed (*Bernoulli distributed with parameter p*; see Subsection 0.4.1.2). Then it holds for all* $n \in \mathbb{N}$ *that*

$$p_0 = 1 - p, \qquad p_1 = p, \qquad \text{and} \qquad p_{1+n} = 0. \tag{1.58}$$

The distribution function $F\colon \mathbb{R} \to [0,1]$ *of* Ber_p *and the generalized inverse distribution function* $I_F\colon \mathbb{R} \to (0,1)$ *associated to* F *are presented in Example 1.2.2 above. According to Theorem 1.2.7, the following algorithm returns as an output a realization of a random variable which is Bernoulli distributed with parameter* p*.*

```
Output: Realization x of X ~ Ber_p
Generate realization u of U ~ U_(0,1)
if u ≤ 1 − p then
  x = 0
else
  x = 1
end if
```

Using the observations that

$$U = 1 - U \tag{1.59}$$

in distribution on $\mathcal{B}(\mathbb{R})$ *and that for all* $u \in (0,1)$ *it holds that*

$$u \leq 1 - p \qquad \text{if and only if} \qquad p \leq 1 - u \tag{1.60}$$

results in the following alternative algorithm.

Output*: Realization x of $X \sim \mathrm{Ber}_p$*
Generate realization u of $U \sim \mathcal{U}_{(0,1)}$
if $u < p$ ***then***
  $x = 1$
else
  $x = 0$
end if

In Matlab *the above algorithm can be implemented through the command "rand¡p" (see Figure 1.2 below).*

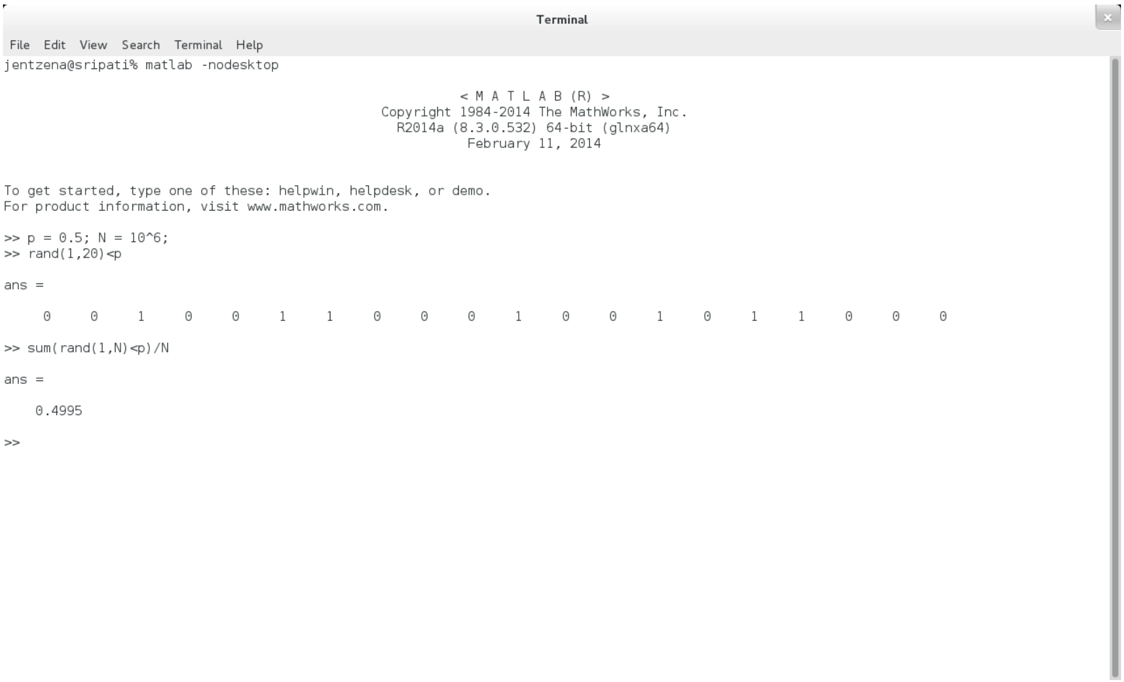


Figure 1.2: Simulating realizations of a pseudo $\mathrm{Ber}_{0.5}$-distributed random variable.

*(ii) Let $\lambda \in (0, \infty)$ be a real number and assume that X is Poi_λ-distributed (*Poisson distributed with parameter λ*; see Subsection 0.4.1.6). Then it holds for all $n \in \mathbb{N}_0$ that $p_n = \frac{\lambda^n}{n!\, e^\lambda}$. Hence, we obtain for all $n \in \mathbb{N}_0$ that*

$$p_{n+1} = \frac{\lambda^{(n+1)}}{(n+1)!\, e^\lambda} = \frac{\lambda}{(n+1)}\, p_n\,. \tag{1.61}$$

Equation (1.57)*, equation* (1.61)*, and Theorem 1.2.7 result in the following algorithm for generating realizations of* X*.*

Output*: Realization* x *of* $X \sim Poi_\lambda$
Generate realization u *of* $U \sim \mathcal{U}_{(0,1)}$
$n = 0$
$p = e^{-\lambda}$
$F = p$
while $u > F$ ***do***
  $p = p \cdot \lambda/(n+1)$
  $F = F + p$
  $n = n + 1$
end while
$x = n$

If the parameter λ *is large, then much more efficient algorithms can, e.g., be found in Section 5 in [Ahrens and Dieter(1974)].*

(iii) In this item let $n \in \mathbb{N}$ *and* $p \in (0,1)$ *be real numbers and assume that* X *is* $b_{n,p}$*-distributed (* binomial distributed with parameters n and p*; see Subsection 0.4.1.3). Then it holds for all* $k \in \{0, 1, \dots, n\}$ *that*

$$p_k = \binom{n}{k} p^k (1-p)^{(n-k)} \tag{1.62}$$

and it holds for all $k \in \{n+1, n+2, \dots\}$ *that* $p_k = 0$*. Note that the coefficients* $(p_k)_{k \in \{0,1,\dots,n\}}$ *satisfy the recursion that for all* $k \in \{0, 1, \dots, n-1\}$ *it holds that*

$$p_0 = (1-p)^n \tag{1.63}$$

and

$$\begin{aligned} p_{k+1} &= \binom{n}{k+1} p^{(k+1)} (1-p)^{(n-(k+1))} \\ &= \frac{p}{(1-p)} \frac{n!}{(n-k-1)!\,(k+1)!} p^k (1-p)^{(n-k)} \\ &= \frac{p}{(1-p)} \frac{(n-k)}{(k+1)} p_k. \end{aligned} \tag{1.64}$$

Exploiting (1.57)*,* (1.64)*, and Theorem 1.2.7 results in the following algorithm for generating binomial distributed random numbers.*

Output*: Realization* x *of* $X \sim b_{n,p}$
Generate realization u *of* $U \sim \mathcal{U}_{(0,1)}$
$k = 0$
$r = p/(1-p)$
$q = (1-p)^n$
$F = q$
while $u > F$ ***do***

$q = r \cdot q \cdot (n-k)/(k+1)$
$F = F + q$
$k = k + 1$
end while
$x = k$

(iv) In this item let $p \in (0,1)$ be a real number and assume that X is geom_p*-distributed (*geometrically distributed with parameter *p; see Subsection 0.4.1.4). Then it holds for all $n \in \mathbb{N}_0$ that*

$$p_n = p\,(1-p)^n. \tag{1.65}$$

This implies for all $n \in \{-1, 0, 1, 2, \dots\}$ that

$$F(n) = \sum_{k=0}^{n} p_k = p\left(\sum_{k=0}^{n} (1-p)^k\right) = \frac{p\left(1-(1-p)^{(n+1)}\right)}{(1-(1-p))} = 1-(1-p)^{(n+1)}.$$

Therefore, we obtain that for every $u \in (0,1)$ and every $n \in \mathbb{N}_0$ it holds that

$$\begin{aligned} & F(n-1) = 1-(1-p)^n < u \leq F(n) = 1-(1-p)^{(n+1)} \\ \Leftrightarrow\ & (1-p)^n > 1-u \geq (1-p)^{(n+1)} \\ \Leftrightarrow\ & n \cdot \ln(1-p) > \ln(1-u) \geq (n+1)\cdot\ln(1-p) \\ \Leftrightarrow\ & n < \frac{\ln(1-u)}{\ln(1-p)} \leq n+1 \\ \Leftrightarrow\ & \left\lceil \frac{\ln(1-u)}{\ln(1-p)} \right\rceil_1 = n+1. \end{aligned} \tag{1.66}$$

This shows for all $u \in (0,1)$ that

$$I_F(u) = \left\lceil \frac{\ln(1-u)}{\ln(1-p)} \right\rceil_1 - 1. \tag{1.67}$$

Hence, we get

$$X = \left\lfloor \frac{\ln(U)}{\ln(1-p)} \right\rfloor_1 \tag{1.68}$$

in distribution on $\mathcal{B}(\mathbb{R})$. Please compare (1.68) *for the geometric distribution with* (1.48) *for the exponential distribution.*

Above we have used the inversion method (see Theorem 1.2.7 above) for the simulation of real valued random variables. The inversion method also has an interesting purely analytical consequence. More precisely, Lemma 0.3.8 in Subsection 0.3.2 illustrates that every probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ induces a distribution function. Corollary 1.2.18 below, in turn, shows that every distribution function also induces a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Corollary 1.2.18 is a consequence of Theorem 1.2.7. In our proof of Corollary 1.2.18 we also employ the unique theorem for probability measures, see

Theorem 1.2.15 below, as well as Lemma 1.2.17 below. Theorem 1.2.15 is, e.g., proved as a special case of Lemma 1.42 in [Klenke(2008)].

Theorem 1.2.15 (Uniqueness theorem for probability measures)**.** *Let Ω be a set, let $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ be a set which satisfies for all $A, B \in \mathcal{A}$ that*

$$A \cap B \in \mathcal{A}, \tag{1.69}$$

and let $\mu_k \colon \sigma_\Omega(\mathcal{A}) \to [0, \infty]$, $k \in \{1, 2\}$, be probability measures which satisfy that

$$\mu_1|_\mathcal{A} = \mu_2|_\mathcal{A}. \tag{1.70}$$

Then

$$\mu_1 = \mu_2. \tag{1.71}$$

Exercise 1.2.16 (An $\cap$-stable generating system for the Borel sigma-algebra)**.** *Let $d \in \mathbb{N}$ and let $\mathcal{A} \subseteq \mathcal{P}(\mathbb{R}^d)$ be the set given by*

$$\mathcal{A} = \cup_{x_1,\dots,x_d \in \mathbb{R}} \{(-\infty, x_1) \times \cdots \times (-\infty, x_d)\}. \tag{1.72}$$

(i) Prove for all $A, B \in \mathcal{A}$ that

$$A \cap B \in \mathcal{A}. \tag{1.73}$$

(ii) Prove that

$$\mathcal{B}(\mathbb{R}^d) = \sigma_{\mathbb{R}^d}(\mathcal{A}). \tag{1.74}$$

The following lemma can, e.g., be proved analogously as Exercise 1.2.16.

Lemma 1.2.17 (Another $\cap$-stable generating system for the Borel sigma-algebra)**.** *Let $d \in \mathbb{N}$ and let $\mathcal{A} \subseteq \mathcal{P}(\mathbb{R}^d)$ be the set given by*

$$\mathcal{A} = \cup_{x_1,\dots,x_d \in \mathbb{R}} \{(-\infty, x_1] \times \cdots \times (-\infty, x_d]\}. \tag{1.75}$$

Then

(i) it holds for all $A, B \in \mathcal{A}$ that

$$A \cap B \in \mathcal{A} \tag{1.76}$$

and

(ii) it holds that

$$\mathcal{B}(\mathbb{R}^d) = \sigma_{\mathbb{R}^d}(\mathcal{A}). \tag{1.77}$$

We now present the promised corollary of Theorem 1.2.22, Corollary 1.2.18 below. Observe that Lemma 0.3.8 in Subsection 0.3.2 above ensures that Φ in (1.78)–(1.79) in Corollary 1.2.18 below does indeed exist.

Corollary 1.2.18 (Bijection between probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and distribution functions)**.** *Let*

$$\Phi\colon \left\{ \begin{array}{c} P\colon \mathcal{B}(\mathbb{R}) \to [0,\infty]\colon \\ P \textit{ is a probability measure} \end{array} \right\} \to \left\{ \begin{array}{c} F\colon \mathbb{R} \to [0,1]\colon \\ F \textit{ is a distribution function} \end{array} \right\} \tag{1.78}$$

be the function which satisfies for all probability measures $P\colon \mathcal{B}(\mathbb{R}) \to [0,\infty]$ and all $x \in \mathbb{R}$ that

$$\big(\Phi(P)\big)(x) = P\big((-\infty, x]\big). \tag{1.79}$$

Then Φ is bijective.

Proof of Corollary 1.2.18. Theorem 1.2.15 and Lemma 1.2.17 ensure that Φ is injective. It thus remains to prove that Φ is surjective. For this let $F\colon \mathbb{R} \to [0,1]$ be an arbitrary distribution function, let $\Omega = (0,1)$, let $\mathcal{A} = \mathcal{B}((0,1))$, let $P\colon \mathcal{A} \to [0,\infty]$ be the probability measure given by

$$P = \mathcal{U}_{(0,1)}|_{\mathcal{A}}, \tag{1.80}$$

and let $U\colon \Omega \to \mathbb{R}$ be the function which satisfies for all $x \in \Omega$ that

$$U(x) = x. \tag{1.81}$$

Then it holds that the triple $(\Omega, \mathcal{A}, P)$ is a probability space and that $U\colon \Omega \to \mathbb{R}$ is an $\mathcal{U}_{(0,1)}$-distributed random variable with $U(\Omega) = (0,1)$. We can thus apply Theorem 1.2.7 to obtain that $F\colon \mathbb{R} \to [0,1]$ is the distribution function of the random variable $I_F \circ U = I_F(U)\colon \Omega \to \mathbb{R}$. We hence obtain that

$$\Phi\Big(\big(I_F \circ U\big)(P)_{\mathcal{B}(\mathbb{R})}\Big) = F. \tag{1.82}$$

This completes the proof of Corollary 1.2.18. □

In Theorem 1.2.22 below we intend to develop a deeper understanding of the relation between a probability measure, a distribution function, and the generalized inverse distribution associated to it. The proof of Theorem 1.2.22 uses the following lemma.

Lemma 1.2.19 (Co-domain of the generalized inversion function)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $U\colon \Omega \to \mathbb{R}$ be an $\mathcal{U}_{(0,1)}$-distributed random variable with $U(\Omega) \subseteq (0,1)$, and let $G\colon (0,1) \to \mathbb{R}$ be a non-decreasing and left continuous function. Then it holds that*

$$G = I_{\mathbb{R} \ni x \mapsto P(G(U) \leq x) \in [0,1]}, \tag{1.83}$$

i.e., it holds for all $y \in (0,1)$ that

$$G(y) = \inf\big\{x \in \mathbb{R}\colon P\big(G(U) \leq x\big) \geq y\big\} = I_{\mathbb{R} \ni x \mapsto P(G(U) \leq x) \in [0,1]}(y). \tag{1.84}$$

Proof of Lemma 1.2.19. Throughout this proof let $F\colon \mathbb{R} \to [0,1]$ be the distribution function of the measure $(G \circ U)(P)_{\mathcal{B}(\mathbb{R})}$, i.e., assume that for all $x \in \mathbb{R}$ it holds that

$$F(x) = \big((G \circ U)(P)_{\mathcal{B}(\mathbb{R})}\big)\big((-\infty, x]\big) = P\big(G(U) \le x\big). \tag{1.85}$$

Observe that the fact that G is non-decreasing ensures that for all $x \in \mathbb{R}$ it holds that

$$\begin{aligned} F(x) &= P(\{\omega \in \Omega\colon G(U(\omega)) \le x\}) = P\big(\big\{\omega \in \Omega\colon U(\omega) \in G^{-1}((-\infty, x])\big\}\big) \\ &= \lambda_{\mathbb{R}}\big(G^{-1}((-\infty, x])\big) = \sup\big(G^{-1}((-\infty, x])\big)\,. \end{aligned} \tag{1.86}$$

The left continuity of G hence proves that for all $x \in F^{-1}((0,1))$ it holds that

$$G(F(x)) = G\big(\sup\big(G^{-1}((-\infty, x])\big)\big) = \sup\big(G\big(G^{-1}((-\infty, x])\big)\big) \le x. \tag{1.87}$$

This and the fact that G is non-decreasing imply that for all $y \in (0,1)$, $x \in \mathbb{R}$ with $y \le F(x) < 1$ it holds that

$$G(y) \le G(F(x)) \le x. \tag{1.88}$$

Moreover, we note that (1.86) proves that for all $y \in (0,1)$, $x \in \mathbb{R}$ with $F(x) = 1$ it holds that

$$G(y) \le x. \tag{1.89}$$

This and (1.88) show that for all $y \in (0,1)$, $x \in \mathbb{R}$ with $y \le F(x)$ it holds that

$$G(y) \le x. \tag{1.90}$$

Hence, we obtain that for all $y \in (0,1)$ it holds that

$$G(y) \le \min\{x \in \mathbb{R}\colon F(x) \ge y\} = I_F(y). \tag{1.91}$$

It thus remains to prove that for all $y \in (0,1)$ it holds that

$$G(y) \ge I_F(y). \tag{1.92}$$

For this note that (1.86) implies that for all $y \in (0,1)$ it holds that

$$\begin{aligned} F(G(y)) &= \sup\Big(G^{-1}\big(\,(-\infty, G(y)]\,\big)\Big) = \sup\{z \in (0,1)\colon G(z) \le G(y)\} \\ &\ge \sup\{y\} = y. \end{aligned} \tag{1.93}$$

The switching formula in Item (iv) in Lemma 1.2.4 hence implies that for all $y \in (0,1)$ it holds that

$$G(y) \ge I_F(y). \tag{1.94}$$

The proof of Lemma 1.2.19 is thus completed. □

Combining Item (i) in Lemma 1.2.4 and Item (i) in Lemma 1.2.5 with Lemma 1.2.19 motivates the following definition.

Definition 1.2.20 (Generalized inverse distribution function). *We say that G is a generalized inverse distribution function if and only if it holds*

(i) that $G \in \mathbb{M}((0,1),\mathbb{R})$ is a function from $(0,1)$ to $\mathbb{R}$,

(ii) that G is non-decreasing (it holds for all $y_1, y_2 \in (0,1)$ with $y_1 \leq y_2$ that $G(y_1) \leq G(y_2)$), and

(iii) that G is left-continuous (it holds for all $y \in (0,1)$ that $\lim_{z \nearrow y} G(z) = G(y)$).

Lemma 1.2.21. *Let $G\colon (0,1) \to \mathbb{R}$ be a function. Then G is a generalized inverse distribution function if and only if there exists a distribution function $F\colon \mathbb{R} \to [0,1]$ such that*

$$I_F = G. \tag{1.95}$$

Proof. Item (i) in Lemma 1.2.4 and Item (i) in Lemma 1.2.5 prove the "⇐" statement in Lemma 1.2.21. Lemma 1.2.19 proves the "⇒" statement in Lemma 1.2.21. The proof of Lemma 1.2.21 is thus completed. □

Theorem 1.2.22. *Let $(\Omega, \mathcal{A}, Q)$ be a probability space, let $U\colon \Omega \to \mathbb{R}$ be an $\mathcal{U}_{(0,1)}$-distributed random variable with $U(\Omega) \subseteq (0,1)$, let*

$$\Phi\colon \left\{ \begin{array}{c} P \in \mathbb{M}(\mathcal{B}(\mathbb{R}), [0,\infty])\colon \\ P \textit{ is a probability measure} \end{array} \right\} \to \left\{ \begin{array}{c} F \in \mathbb{M}(\mathbb{R}, [0,1])\colon \\ F \textit{ is a distribution function} \end{array} \right\} \tag{1.96}$$

be the function which satisfies for all probability measures $P\colon \mathcal{B}(\mathbb{R}) \to [0,\infty]$ and all $x \in \mathbb{R}$ that

$$\big(\Phi(P)\big)(x) = P\big((-\infty, x]\big), \tag{1.97}$$

let

$$\mathcal{I}\colon \left\{ \begin{array}{c} F \in \mathbb{M}(\mathbb{R}, [0,1])\colon \\ F \textit{ is a distribution function} \end{array} \right\} \to \left\{ \begin{array}{c} G \in \mathbb{M}((0,1), \mathbb{R})\colon \\ G \textit{ is a generalized inverse} \\ \textit{distribution function} \end{array} \right\} \tag{1.98}$$

be the function which satisfies for all distribution functions $F\colon \mathbb{R} \to [0,1]$ that

$$\mathcal{I}(F) = I_F, \tag{1.99}$$

and let

$$\Psi\colon \left\{ \begin{array}{c} G \in \mathbb{M}((0,1), \mathbb{R})\colon \\ G \textit{ is a generalized inverse} \\ \textit{distribution function} \end{array} \right\} \to \left\{ \begin{array}{c} P \in \mathbb{M}(\mathcal{B}(\mathbb{R}), [0,\infty])\colon \\ P \textit{ is a probability measure} \end{array} \right\} \tag{1.100}$$

be the function which satisfies for all generalized inverse distribution functions $G\colon (0,1) \to \mathbb{R}$ that

$$\Psi(G) = (G \circ U)(Q)_{\mathcal{B}(\mathbb{R})}. \tag{1.101}$$

Then

(i) it holds that Φ, $\mathcal{I}$, and Ψ are bijective and

(ii) it holds that

$$\Phi \circ \Psi \circ \mathcal{I} = \mathrm{Id}_{\mathrm{dom}(\mathcal{I})}, \quad \mathcal{I} \circ \Phi \circ \Psi = \mathrm{Id}_{\mathrm{dom}(\Psi)}, \quad \textit{and} \quad \Psi \circ \mathcal{I} \circ \Phi = \mathrm{Id}_{\mathrm{dom}(\Phi)}. \tag{1.102}$$

Proof of Theorem 1.2.22. First, recall that Corollary 1.2.18 ensures that Φ is bijective. Moreover, note that Theorem 1.2.7 proves that

$$\Phi \circ \Psi \circ \mathcal{I} = \mathrm{Id}_{\mathrm{dom}(\mathcal{I})}. \tag{1.103}$$

This implies that

$$\Psi \circ \mathcal{I} = \Phi^{-1}. \tag{1.104}$$

This and the fact that Φ^{-1} is bijective prove that

$$\Psi \circ \mathcal{I} \tag{1.105}$$

is bijective. This implies that $\mathcal{I}$ is injective. Next note that Lemma 1.2.19 ensures that $\mathcal{I}$ is surjective. This together with the fact that $\mathcal{I}$ is injective ensures that $\mathcal{I}$ is bijective. Furthermore, observe that (1.104) proves that

$$\Psi = \Phi^{-1} \circ \mathcal{I}^{-1}. \tag{1.106}$$

This together with the fact that Φ and $\mathcal{I}$ are bijective ensures that Ψ is bijective too. The second and the third identity in (1.102) follows from the first identity in (1.102). The proof of Theorem 1.2.22 is thus completed. □

Class exercise 1.2.23. *Let $F\colon \mathbb{R} \to [0,1]$ be a distribution function, let $(\Omega, \mathcal{A}, Q)$ be a probability space, and let $U\colon \Omega \to \mathbb{R}$ be an $\mathcal{U}_{(0,1)}$-distributed random variable with $U(\Omega) \subseteq (0,1)$. What is*

$$\Phi\big((I_F \circ U)(Q)_{\mathcal{B}(\mathbb{R})}\big)? \tag{1.107}$$

Exercise 1.2.24. *Let $a, b \in \mathbb{R}$ be real numbers with $a < b$ and let $F\colon \mathbb{R} \to [0,1]$ be a distribution function which satisfies for all $y \in (0,1)$ that*

$$I_F(y) = yb + (1-y)a. \tag{1.108}$$

Specify $F(x)$, $x \in \mathbb{R}$, explicitly and prove that your result is correct.

Exercise 1.2.25. *In this exercise we do not distinguish between pseudo random numbers and actual random numbers. Write a* Matlab *function* `Cauchy(N,μ,λ)` *with input $N \in \mathbb{N}$, $\mu \in \mathbb{R}$, $\lambda \in (0,\infty)$ and output a realization of an $(\mathrm{Cau}_{\mu,\lambda})^{\otimes N}$-distributed random variable generated with the inversion method. The* Matlab *function* `Cauchy(N,μ,λ)` *may use at most N realizations of an $\mathcal{U}_{(0,1)}$-distributed random variable. Type* `Cauchy(10,0,1)` *to test your implementation.*

Exercise 1.2.26. *Let $\lambda \in (0,\infty)$ and let $F\colon \mathbb{R} \to [0,1]$ be the distribution function of the Laplace distribution with parameter λ.*

(i) Specify $F(x)$, $x \in \mathbb{R}$, explicitly.

(ii) Specify $I_F(y)$, $y \in (0,1)$, explicitly.

Exercise 1.2.27. *In this exercise we do not distinguish between pseudo random numbers and actual random numbers.*

(i) Write a Matlab *function* `Laplace(N,λ)` *with input $N \in \mathbb{N}$, $\lambda \in (0,\infty)$ and output a realization of an $(\mathrm{Laplace}_\lambda)^{\otimes N}$-distributed random variable generated with the inversion method. The* Matlab *function* `Laplace(N,λ)` *may use at most N realizations of an $\mathcal{U}_{(0,1)}$-distributed random variable. Type* `Laplace(10,0.5)` *to test your implementation.*

(ii) Write a Matlab *function* `LaplacePlot()` *which plots 10^5 realizations of an $\mathrm{Laplace}_{0.1}$-distributed random variable generated with your* Matlab *function* `Laplace(N,λ)` *in a histogram with 1000 bins.*

1.2.2 Acceptance-rejection method

This subsection presents the *acceptance-rejection method* (also known as *rejection sampling*). It is a method to simulate from a complicated distribution by using simulations from a simpler distribution from which one assumes to be able to simulate from. To get an idea of the acceptance-rejection method, we first consider the special situation of the uniform distribution.

Lemma 1.2.28 (Acceptance-rejection method in the case of the continuous uniform distribution)**.** *Let $(\Omega, \mathcal{A}, P)$ be a probability space, let $d \in \mathbb{N}$, let $A, B \in \mathcal{B}(\mathbb{R}^d)$ with $A \subseteq B$ and $0 < \lambda_{\mathbb{R}^d}(A) \leq \lambda_{\mathbb{R}^d}(B) < \infty$, let $Y_n \colon \Omega \to \mathbb{R}^d$, $n \in \mathbb{N}_0$, be P-independent $\mathcal{U}_B$-distributed random variables, and let $L \colon \Omega \to \mathbb{N}_0$ be the function which satisfies for all $\omega \in \Omega$ that*

$$L(\omega) = \begin{cases} \min(\{n \in \mathbb{N}_0 \colon Y_n(\omega) \in A\}) & \colon \omega \in (\cup_{n \in \mathbb{N}_0} \{Y_n \in A\}) \\ 0 & \colon \omega \in \Omega \backslash (\cup_{n \in \mathbb{N}_0} \{Y_n \in A\}) \end{cases}. \tag{1.109}$$

Then it holds that $Y_L \colon \Omega \to \mathbb{R}^d$ is $\mathcal{U}_A$-distributed.

Proof of Lemma 1.2.28. Note that for all $C \in \mathcal{B}(\mathbb{R}^d)$ it holds that

$$\begin{aligned} P(Y_L \in C) &= \sum_{n=0}^{\infty} P(\{Y_L \in C\} \cap \{L = n\}) \\ &= \sum_{n=0}^{\infty} P(\{Y_n \in C \cap A\} \cap \{Y_0, Y_1, \ldots, Y_{n-1} \in B \backslash A\}) \\ &= \sum_{n=0}^{\infty} P(Y_n \in C \cap A) \cdot P(Y_0, Y_1, \ldots, Y_{n-1} \in B \backslash A) \\ &= \frac{P(Y_1 \in C \cap A)}{P(Y_1 \in A)} \sum_{n=0}^{\infty} P(Y_n \in A) \cdot P(Y_0, Y_1, \ldots, Y_{n-1} \in B \backslash A) \\ &= \frac{\mathcal{U}_B(C \cap A)}{\mathcal{U}_B(A)} \left[\sum_{n=0}^{\infty} P(\{Y_n \in A\} \cap \{Y_0, Y_1, \ldots, Y_{n-1} \in B \backslash A\}) \right] \\ &= \frac{\lambda_{\mathbb{R}^d}(C \cap A)}{\lambda_{\mathbb{R}^d}(A)} \left[\sum_{n=0}^{\infty} P(L = n) \right] = \frac{\lambda_{\mathbb{R}^d}(C \cap A)}{\lambda_{\mathbb{R}^d}(A)} = \mathcal{U}_A(C) \end{aligned} \tag{1.110}$$

This completes the proof of Lemma 1.2.28. □

Lemma 1.2.28 motivates the following algorithm. Let $(\Omega, \mathcal{A}, P)$ be a probability space, let $d \in \mathbb{N}$, let $A, B \in \mathcal{B}(\mathbb{R}^d)$ with $A \subseteq B$ and $0 < \lambda_{\mathbb{R}^d}(A) \leq \lambda_{\mathbb{R}^d}(B) < \infty$, let $Y \colon \Omega \to \mathbb{R}^d$ be an $\mathcal{U}_B$-distributed random variable, and let $X \colon \Omega \to \mathbb{R}^d$ be an $\mathcal{U}_A$-distributed random variable. Then the following algorithm generates a realization from X.

Output: Realization x of $X \sim \mathcal{U}_A$

Generate realization y of $Y \sim \mathcal{U}_B$
if $y \in A$ **then**
  $x = y$ (ACCEPT)
else
  Restart the algorithm (REJECT)
end if

Remark 1.2.29. *Let us point out that the above presented algorithm is not completely correct since it may happen that the algorithm never terminates and thus does not return a realization of an $\mathcal{U}_A$-distributed random variable. In a* Matlab *implementation this is, however, often not a problem since the $\mathcal{U}_{(0,1)}$-pseudo random number generator in* Matlab *will in the case of many choices for A and B always produce at some point a realization which is in A. In the following acceptance-rejection algorithms this point is neglected.*

Exercise 1.2.30. *In this exercise we do not distinguish between pseudo random numbers and actual random numbers. Let $A \subseteq \mathbb{R}^2$ be the set given by*

$$A = \left\{(x, y) \in \mathbb{R}^2 \colon \frac{x^2}{8} + y^2 \leq 2\right\}. \tag{1.111}$$

(i) Write a Matlab *function* `AcceptanceRejection(N)` *with input $N \in \mathbb{N}$ and output a realization of an $(\mathcal{U}_A)^{\otimes N}$-distributed random variable generated with the acceptance-rejection method. Type* `AcceptanceRejection(6)` *to test your implementation.*

(ii) Write a Matlab *function* `AcceptanceRejectionPlot()` *which uses your* Matlab *function* `AcceptanceRejection(N)` *from Item* (i) *and the built-in* Matlab *function* `plot(...)` *to plot 10^5 realizations of an $\mathcal{U}_A$-distributed random variable in a coordinate plane.*

In the next step we extend the idea in Lemma 1.2.28 to more complicated distributions. For this we need the notion of the *subgraph of a nonnegative function*. This is the subject of the next definition.

Definition 1.2.31 (Subgraph of a real valued nonnegative function)**.** *Let $d \in \mathbb{N}$ and let $f \colon \mathbb{R}^d \to [0, \infty)$ be a function. Then we denote by* $\operatorname{subgraph}(f)$ *the set given by*

$$\operatorname{subgraph}(f) = \{(x, y) \in \mathbb{R}^d \times \mathbb{R} = \mathbb{R}^{d+1} \colon 0 \leq y \leq f(x)\} \tag{1.112}$$

and we call $\operatorname{subgraph}(f)$ *the* subgraph of f.

For every $d \in \mathbb{N}$ and every $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}([0,\infty))$-measurable function $f \colon \mathbb{R}^d \to [0, \infty)$ it holds that the set

$$\operatorname{subgraph}(f) \subseteq \mathbb{R}^{d+1} \tag{1.113}$$

is Borel measurable too. This is the subject of the next lemma.

Lemma 1.2.32. *Let $d \in \mathbb{N}$ and let $f \colon \mathbb{R}^d \to [0, \infty)$ be an $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}([0,\infty))$-measurable function. Then*

$$\operatorname{subgraph}(f) \in \mathcal{B}\big(\mathbb{R}^{d+1}\big). \tag{1.114}$$

Proof of Lemma 1.2.32. Throughout this proof let $\hat{f}\colon \mathbb{R}^{d+1} \to \mathbb{R}$ be the function which satisfies for all $x \in \mathbb{R}^d$, $y \in \mathbb{R}$ that

$$\hat{f}(x, y) = y - f(x). \tag{1.115}$$

Observe that the assumption that f is $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}([0,\infty))$-measurable ensures that $\hat{f}$ is $\mathcal{B}(\mathbb{R}^{d+1})/\mathcal{B}(\mathbb{R})$-measurable. This implies that

$$\operatorname{subgraph}(f) = \underbrace{\hat{f}^{-1}\big((-\infty, 0]\big)}_{\in \mathcal{B}(\mathbb{R}^{d+1})} \cap \underbrace{\big(\mathbb{R}^d \times [0,\infty)\big)}_{\in \mathcal{B}(\mathbb{R}^{d+1})} \in \mathcal{B}(\mathbb{R}^{d+1}). \tag{1.116}$$

The proof of Proposition 1.2.35 is thus completed. □

Observe that for all $d \in \mathbb{N}$ and all $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}([0,\infty))$-measurable functions $f\colon \mathbb{R}^d \to [0,\infty)$ it holds that

$$\lambda_{\mathbb{R}^{d+1}}\big(\operatorname{subgraph}(f)\big) = \int_{\mathbb{R}^d} \int_0^{f(x)} dy \, dx = \int_{\mathbb{R}^d} f(x) \, dx. \tag{1.117}$$

In the next step the acceptance-rejection method is presented and analyzed for more general distributions. For this the following notion is used.

Definition 1.2.33 (Unnormalized density functions with respect to the Lebesgue-Borel measure)**.** *We say that f is an unnormalized density of μ if and only if there exists a natural number $d \in \mathbb{N}$ such that it holds*

(i) that μ is a measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$,

(ii) that f is an $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}([0,\infty))$-measurable function,

(iii) that $\int_{\mathbb{R}^d} f(x) \, \lambda_{\mathbb{R}^d}(dx) \in (0,\infty)$, and

(iv) that for all $A \in \mathcal{B}(\mathbb{R}^d)$ it holds that

$$\mu(A) = \frac{\int_A f(x) \, \lambda_{\mathbb{R}^d}(dx)}{\int_{\mathbb{R}^d} f(x) \, \lambda_{\mathbb{R}^d}(dx)}. \tag{1.118}$$

Lemma 1.2.34 (Properties of $\mathcal{U}_{\operatorname{subgraph}(f)}$-distributed random variables)**.** *Let $d \in \mathbb{N}$, let $f\colon \mathbb{R}^d \to [0,\infty)$ be an $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}([0,\infty))$-measurable function with $0 < \int_{\mathbb{R}^d} f(x) \, dx < \infty$, let $(\Omega, \mathcal{A}, P)$ be a probability space, and let $X = (X_1, \ldots, X_d, X_{d+1})\colon \Omega \to \mathbb{R}^{d+1}$ be an $\mathcal{U}_{subgraph(f)}$-distributed random variable. Then it holds that $f\colon \mathbb{R}^d \to [0,\infty)$ is an unnormalized density of $(X_1, \ldots, X_d)(P)_{\mathcal{B}(\mathbb{R}^d)}$.*

Proof of Lemma 1.2.34. Note that for all $A \in \mathcal{B}(\mathbb{R}^d)$ it holds that

$$\begin{aligned} P((X_1, \ldots, X_d) \in A) &= P((X_1, \ldots, X_d, X_{d+1}) \in A \times \mathbb{R}) \\ &= \frac{\lambda_{\mathbb{R}^{d+1}}\big((A \times \mathbb{R}) \cap \operatorname{subgraph}(f)\big)}{\lambda_{\mathbb{R}^{d+1}}\big(\operatorname{subgraph}(f)\big)} = \frac{\int_A f(x) \, dx}{\int_{\mathbb{R}^d} f(x) \, dx}. \end{aligned} \tag{1.119}$$

This completes the proof of Lemma 1.2.34. □

The following proposition results in a method to generate $\mathcal{U}_{\mathrm{subgraph}(f)}$-distributed random variables.

Proposition 1.2.35 (Generation of realizations of an $\mathcal{U}_{\mathrm{subgraph}(f)}$-distributed random variable)**.** *Let $d \in \mathbb{N}$, let $(\Omega, \mathcal{A}, P)$ be a probability space, let $X : \Omega \to \mathbb{R}^d$ be an $\mathcal{A}/\mathcal{B}(\mathbb{R}^d)$-measurable function, let $U : \Omega \to \mathbb{R}$ be an $\mathcal{U}_{(0,1)}$-distributed random variable, let $f : \mathbb{R}^d \to [0, \infty)$ be an unnormalized density of $X(P)_{\mathcal{B}(\mathbb{R}^d)}$, and assume that X and U are P-independent. Then the function $Y : \Omega \to \mathbb{R}^{d+1}$ given by*

$$Y = (X, f(X) \cdot U) \tag{1.120}$$

is $\mathcal{U}_{subgraph(f)}$-distributed.

Proof of Proposition 1.2.35. Let $A \in \mathcal{B}(\mathbb{R}^{d+1})$ and let $A_x \in \mathcal{B}(\mathbb{R})$, $x \in \mathbb{R}^d$, be the sets which satisfy for all $x \in \mathbb{R}^d$ that

$$A_x = \{y \in \mathbb{R} \colon (x, y) \in A\}. \tag{1.121}$$

Next observe that

$$\begin{aligned}
P\big(Y \in A\big) &= P\big((X, f(X) \cdot U) \in A\big) \\
&= \int_{\mathbb{R}^d \times (0,1)} \mathbb{1}_A(x, f(x) \cdot u) \, \big((X, U)(P)_{\mathcal{B}(\mathbb{R}^{d+1})}\big)(dx, du) \\
&= \int_{\mathbb{R}^d} \int_0^1 \mathbb{1}_A(x, f(x) \cdot u) \, du \, \big(X(P)_{\mathcal{B}(\mathbb{R}^d)}\big)(dx) \\
&= \int_{\mathbb{R}^d} \int_0^1 \mathbb{1}_{A_x}(f(x) \cdot u) \, du \, \big(X(P)_{\mathcal{B}(\mathbb{R}^d)}\big)(dx) \\
&= \frac{1}{\int_{\mathbb{R}^d} f(x) \, dx} \int_{\mathbb{R}^d} \int_0^1 \mathbb{1}_{A_x}(f(x) \cdot u) \, du \, f(x) \, dx.
\end{aligned} \tag{1.122}$$

This shows that

$$\begin{aligned}
&P\big(Y \in A\big) \\
&= \frac{1}{\lambda_{\mathbb{R}^{(d+1)}}\big(\mathrm{subgraph}(f)\big)} \int_{\{y \in \mathbb{R}^d \colon f(y) \neq 0\}} \int_0^1 \mathbb{1}_{A_x}(f(x) \cdot u) \, du \, f(x) \, dx \\
&= \frac{1}{\lambda_{\mathbb{R}^{(d+1)}}\big(\mathrm{subgraph}(f)\big)} \int_{\{y \in \mathbb{R}^d \colon f(y) \neq 0\}} \int_0^{f(x)} \mathbb{1}_{A_x}(u) \, du \, dx \\
&= \frac{1}{\lambda_{\mathbb{R}^{(d+1)}}\big(\mathrm{subgraph}(f)\big)} \int_{\mathbb{R}^d} \int_0^{f(x)} \mathbb{1}_{A_x}(u) \, du \, dx.
\end{aligned} \tag{1.123}$$

This implies that

$$\begin{aligned} P\big(Y \in A\big) &= \frac{1}{\lambda_{\mathbb{R}^{(d+1)}}\big(\operatorname{subgraph}(f)\big)} \int_{\mathbb{R}^d} \int_0^{f(x)} \mathbb{1}_A(x,u)\,du\,dx \\ &= \frac{1}{\lambda_{\mathbb{R}^{(d+1)}}\big(\operatorname{subgraph}(f)\big)} \int_{\mathbb{R}^d} \int_{\mathbb{R}} \mathbb{1}_A(x,u) \cdot \mathbb{1}_{[0,f(x)]}(u)\,du\,dx \\ &= \frac{1}{\lambda_{\mathbb{R}^{(d+1)}}\big(\operatorname{subgraph}(f)\big)} \int_{\mathbb{R}^{(d+1)}} \mathbb{1}_A(x,u) \cdot \mathbb{1}_{\operatorname{subgraph}(f)}(x,u)\,\lambda_{\mathbb{R}^{(d+1)}}(du,dx) \\ &= \mathcal{U}_{\operatorname{subgraph}(f)}(A). \end{aligned} \tag{1.124}$$

The proof of Proposition 1.2.35 is thus completed. □

We now formulate the acceptance-rejection method. Let $d \in \mathbb{N}$, let $f, g \colon \mathbb{R}^d \to [0,\infty)$ be $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}([0,\infty))$-measurable functions which satisfy for all $y \in \mathbb{R}^d$ that $f(y) \le g(y)$ and

$$0 < \int_{\mathbb{R}^d} f(x)\,dx \le \int_{\mathbb{R}^d} g(x)\,dx < \infty, \tag{1.125}$$

let $(\Omega, \mathcal{A}, P)$ be a probability space, let $X \colon \Omega \to \mathbb{R}^d$ be an $\mathcal{A}/\mathcal{B}(\mathbb{R}^d)$-measurable function which satisfies for all $A \in \mathcal{B}(\mathbb{R}^d)$ that

$$X(P)_{\mathcal{B}(\mathbb{R}^d)}(A) = \frac{\int_A f(x)\,dx}{\int_{\mathbb{R}^d} f(x)\,dx}, \tag{1.126}$$

and let $U \colon \Omega \to \mathbb{R}$ and $Y \colon \Omega \to \mathbb{R}^d$ be P-independent random variables which satisfy for all $A \in \mathcal{B}(\mathbb{R}^d)$ that U is $\mathcal{U}_{(0,1)}$-distributed, that $U(\Omega) \subseteq (0,1)$, and that

$$Y(P)_{\mathcal{B}(\mathbb{R}^d)}(A) = \frac{\int_A g(x)\,dx}{\int_{\mathbb{R}^d} g(x)\,dx}. \tag{1.127}$$

The goal of the acceptance-rejection method is to generate realizations from X where it is assumed that one can generate realizations from (Y, U). Thus one can also generate realizations from the $\mathcal{U}_{\operatorname{subgraph}(g)}$-distributed random variable $(Y, g(Y) \cdot U)$ (see Proposition 1.2.35). The condition $f \le g$ ensures that $\operatorname{subgraph}(f) \subseteq \operatorname{subgraph}(g)$. Using Lemma 1.2.28 one can then obtain realizations from an $\mathcal{U}_{\operatorname{subgraph}(f)}$-distributed random variable. For this observe that for every $u \in (0,1)$ and every $y \in \mathbb{R}^d$ it holds that

$$(y, g(y) \cdot u) \in \operatorname{subgraph}(f) \qquad \text{if and only if} \qquad g(y) \cdot u \le f(y). \tag{1.128}$$

According to Lemma 1.2.34, the first d-components of the $\mathcal{U}_{\operatorname{subgraph}(f)}$-distributed random variable are then realizations from an $X(P)_{\mathcal{B}(\mathbb{R}^d)}$-distributed random variable. The algorithm thus reads as follows.

Acceptance-rejection method

Output: Realization x of $X \sim X(P)_{\mathcal{B}(\mathbb{R}^d)}$ (with unnormalized density f)
Generate realization y of $Y \sim Y(P)_{\mathcal{B}(\mathbb{R}^d)}$ (with unnormalized density g)

Generate realization u of $U \sim \mathcal{U}_{(0,1)}$
if $g(y) \cdot u \leq f(y)$ **then**
  $x = y$ (ACCEPT)
else
  Restart the algorithm (REJECT)
end if

Remark 1.2.36. *Note that in the acceptance-rejection algorithm it is not assumed that f is a probability density function, i.e., that $\int_{\mathbb{R}^d} f(x)\,dx = 1$ and it is also not assumed that g is a probability density function, i.e., that $\int_{\mathbb{R}^d} g(x)\,dx = 1$.*

Lemma 1.2.37 (Number of rejections before acceptance)**.** *Let $d \in \mathbb{N}$, let $f, g \colon \mathbb{R}^d \to [0,\infty)$ be $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}([0,\infty))$-measurable functions which satisfy for all $y \in \mathbb{R}^d$ that $f(y) \leq g(y)$ and*

$$0 < \int_{\mathbb{R}^d} f(x)\,dx \leq \int_{\mathbb{R}^d} g(x)\,dx < \infty, \tag{1.129}$$

let $(\Omega, \mathcal{A}, P)$ be a probability space, let $U_n \colon \Omega \to \mathbb{R}$, $n \in \mathbb{N}_0$, and $Y_n \colon \Omega \to \mathbb{R}^d$, $n \in \mathbb{N}_0$, be P-independent random variables which satisfy for all $n \in \mathbb{N}_0$, $A \in \mathcal{B}(\mathbb{R}^d)$ that U_n is $\mathcal{U}_{(0,1)}$-distributed, that $U_n(\Omega) \subseteq (0,1)$, and that

$$\big(Y_n(P)_{\mathcal{B}(\mathbb{R}^d)}\big)(A) = \frac{\int_A g(x)\,dx}{\int_{\mathbb{R}^d} g(x)\,dx}, \tag{1.130}$$

and let $L \colon \Omega \to \mathbb{N}_0$ be the function which satisfies for all $\omega \in \Omega$ that

$$L(\omega) = \tag{1.131}$$
$$\begin{cases} \min\big(\big\{n \in \mathbb{N}_0 \colon g(Y_n(\omega))\,U_n(\omega) \leq f(Y_n(\omega))\big\}\big) & \colon \omega \in (\cup_{n\in\mathbb{N}_0}\{g(Y_n)\,U_n \leq f(Y_n)\}) \\ 0 & \colon \omega \in \Omega \backslash (\cup_{n\in\mathbb{N}_0}\{g(Y_n)\,U_n \leq f(Y_n)\}) \end{cases}.$$

Then it holds that the function $L \colon \Omega \to \mathbb{N}_0$ is $\operatorname{geom}_{\int_{\mathbb{R}^d} f(x)\,dx / \int_{\mathbb{R}^d} g(x)\,dx}$-distributed (geometrically distributed with parameter $p = \frac{\int_{\mathbb{R}^d} f(x)\,dx}{\int_{\mathbb{R}^d} g(x)\,dx}$).

Proof of Lemma 1.2.37. By definition it is clear that there exists a real number $p \in (0,1]$ such that L is geom_p-distributed (geometrically distributed with parameter $p \in (0,1]$). Moreover, observe that Proposition 1.2.35 ensures that

$$\begin{aligned} p = P(L = 0) = P\Big(g(Y_0)U_0 \leq f(Y_0)\Big) = P\Big(\big(Y_0, g(Y_0)U_0\big) \in \operatorname{subgraph}(f)\Big) \\ = \mathcal{U}_{\operatorname{subgraph}(g)}\big(\operatorname{subgraph}(f)\big) = \frac{\lambda_{\mathbb{R}^{d+1}}\big(\operatorname{subgraph}(f)\big)}{\lambda_{\mathbb{R}^{d+1}}\big(\operatorname{subgraph}(g)\big)} = \frac{\int_{\mathbb{R}^d} f(x)\,dx}{\int_{\mathbb{R}^d} g(x)\,dx}. \end{aligned} \tag{1.132}$$

This completes the proof of Lemma 1.2.37. □

Exercise 1.2.38. *In this exercise we do not distinguish between pseudo random numbers and actual random numbers. Let* $f, \tilde{f} \colon \mathbb{R} \to [0, \infty)$ *be the functions which satisfy for all* $x \in \mathbb{R}$ *that*

$$f(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2} \qquad \textit{and} \qquad \tilde{f}(x) = \frac{1}{\pi\,(1 + x^2)} \tag{1.133}$$

and let $\tilde{A} \subseteq \mathbb{R}$ *be the set given by*

$$\tilde{A} = \Big\{ C \in \mathbb{R} \colon \big(\forall\, x \in \mathbb{R} \colon f(x) \leq C \tilde{f}(x)\big) \Big\}. \tag{1.134}$$

Note that f *is the density of* $\mathcal{N}_{0, I_{\mathbb{R}}}$ *(1-dimensional standard normal distribution) and that* $\tilde{f}$ *is the density of* $\operatorname{Cau}_{0,1}$ *(Cauchy distribution with parameters* 0 *and* 1*).*

(i) Prove that for all $C \in \mathbb{R}$ *it holds that* $C \in \tilde{A}$ *if and only if for all* $y \in [0, \infty)$ *it holds that*

$$1 + 2y \leq \frac{\sqrt{2}\, C\, e^{y}}{\sqrt{\pi}}. \tag{1.135}$$

(ii) Specify $\tilde{A}$ *explicitly and prove that your result is correct.*

(iii) Specify

$$\frac{1}{\int_{\mathbb{R}} f(x)\, dx} \int_{\mathbb{R}} \left[\sup_{y \in \mathbb{R}} \frac{f(y)}{\tilde{f}(y)} \right] \tilde{f}(x)\, dx \tag{1.136}$$

explicitly and prove that your result is correct.

(iv) Write a M*atlab function* ***AcceptanceRejectionGaussianCauchy(N)*** *with input* $N \in \mathbb{N}$ *and output a realization of an* $(\mathcal{N}_{0, I_{\mathbb{R}}})^{\otimes N}$*-distributed random variable generated with the acceptance-rejection method with* f *as the density of the target distribution and*

$$\mathbb{R} \ni x \mapsto \left[\sup_{y \in \mathbb{R}} \tfrac{f(y)}{\tilde{f}(y)} \right] \tilde{f}(x) \in [0, \infty) \tag{1.137}$$

as the unnormalized density of the proposal distribution $\operatorname{Cau}_{0,1}$*. Your* M*atlab function* ***AcceptanceRejectionGaussianCauchy(N)*** *should use the inversion method for the generation of realizations of an* $\operatorname{Cau}_{0,1}$*-distributed random variable. Type* ***AcceptanceRejectionGaussianCauchy(6)*** *to test your implementation.*

(v) Write a M*atlab function* ***AcceptanceRejectionGaussianCauchyPlot()*** *which plots* 10^5 *realizations of an* $\mathcal{N}_{0, I_{\mathbb{R}}}$*-distributed random variable generated with your* M*atlab function* ***AcceptanceRejectionGaussianCauchy(N)*** *in a histogram with* 1000 *bins.*

Exercise 1.2.39. *In this exercise we do not distinguish between pseudo random numbers and actual random numbers. Let* $f, \hat{f} \colon \mathbb{R} \to [0, \infty)$ *be the functions which satisfy for all* $x \in \mathbb{R}$ *that*

$$f(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2} \qquad \textit{and} \qquad \hat{f}(x) = \frac{e^{-|x|}}{2} \tag{1.138}$$

and let $\hat{A} \subseteq \mathbb{R}$ be the set given by

$$\hat{A} = \left\{ C \in \mathbb{R} \colon \big(\forall\, x \in \mathbb{R} \colon f(x) \leq C \hat{f}(x) \big) \right\}. \tag{1.139}$$

Note that f is the density of $\mathcal{N}_{0,I_{\mathbb{R}}}$ (1-dimensional standard normal distribution) and that $\hat{f}$ is the density of Laplace$_1$ *(Laplace distribution with parameter 1).*

(i) Specify $\hat{A}$ explicitly and prove that your result is correct.

(ii) Specify

$$\frac{1}{\int_{\mathbb{R}} f(x)\, dx} \int_{\mathbb{R}} \left[\sup_{y \in \mathbb{R}} \frac{f(y)}{\hat{f}(y)} \right] \hat{f}(x)\, dx \tag{1.140}$$

explicitly and prove that your result is correct.

(iii) Write a Matlab *function* `AcceptanceRejectionGaussianLaplace(N)` *with input $N \in \mathbb{N}$ and output a realization of an $(\mathcal{N}_{0,I_{\mathbb{R}}})^{\otimes N}$-distributed random variable generated with the acceptance-rejection method with f as the density of the target distribution and*

$$\mathbb{R} \ni x \mapsto \left[\sup_{y \in \mathbb{R}} \tfrac{f(y)}{\hat{f}(y)} \right] \hat{f}(x) \in [0, \infty) \tag{1.141}$$

as the unnormalized density of the proposal distribution Laplace$_1$. *Your* Matlab *function* `AcceptanceRejectionGaussianLaplace(N)` *should use the inversion method for the generation of realizations of an* Laplace$_1$*-distributed random variable. Type* `AcceptanceRejectionGaussianLaplace(6)` *to test your implementation.*

(iv) Write a Matlab *function* `AcceptanceRejectionGaussianLaplacePlot()` *which plots 10^5 realizations of an $\mathcal{N}_{0,I_{\mathbb{R}}}$-distributed random variable generated with your* Matlab *function* `AcceptanceRejectionGaussianLaplace(N)` *in a histogram with 1000 bins.*

1.3 Methods for the normal distribution

This section considers several methods for the generation of realizations of independent normally distributed random variables. We first consider methods for the generation of (approximative) realizations of independent standard normal random variables; see Subsection 1.3.1–Subsection 1.3.3. Then in Subsection 1.3.4 we consider methods for the generation of realizations of independent normally distributed random variables with mean $v \in \mathbb{R}^d$ and covariance matrix $Q \in \mathbb{R}^{d \times d}$ where $v \in \mathbb{R}^d$ is a vector, where $Q \in \mathbb{R}^{d \times d}$ is a nonnegative symmetric matrix, and where $d \in \mathbb{N}$ is a natural number.

One possibility to generate realizations of standard normal random variables is to use the inverse transform method presented in Subsection 1.2.1. For this the generalized inverse distribution function associated to the one-dimensional normal distribution has to be calculated (cf. the function "*erfinv*" in Matlab). This is typically computationally very expensive. Further methods for the generation of realizations of independent standard normal random variables are the *Box-Muller method* which will be considered in Subsection 1.3.2 and the *Marsaglia polar method* which is the subject of Subsection 1.3.3. Another method which will not be considered here is the *Ziggurat method*. It is nowadays used in the Matlab function "*randn*" and it is based on the acceptance-rejection method presented in Subsection 1.2.2.

1.3.1 Central limit theorem

Before we present the Box-Muller method and the Marsaglia polar method for the generation of realizations of standard normal random variables, we briefly consider a method for the generation of realizations of random variables that are approximatively normally distributed in a suitable sense. For this we recall the central limit theorem.

Theorem 1.3.1 (Central limit theorem – scalar case)**.** *Let $(\Omega, \mathcal{A}, P)$ be a probability space, let $Y_n\colon \Omega \to \mathbb{R}$, $n \in \mathbb{N}$, be P-independent and identically distributed random variables (i.i.d. random variables) which satisfy $\mathbb{E}_P[|Y_1|^2] < \infty$ and $\mathsf{Var}_P(Y_1) > 0$. Then it holds that*

$$\frac{Y_1 + \ldots + Y_n - n \cdot \mathbb{E}_P[Y_1]}{\sqrt{n \, \mathsf{Var}_P(Y_1)}}, \quad n \in \mathbb{N}, \tag{1.142}$$

converges in distribution to $\mathcal{N}_{0,I_\mathbb{R}}$, i.e., it holds for all $x \in \mathbb{R}$ that

$$\lim_{n\to\infty} P\left(\frac{Y_1 + \ldots + Y_n - n \cdot \mathbb{E}_P[Y_1]}{\sqrt{n \, \mathsf{Var}_P(Y_1)}} \leq x\right) = \mathcal{N}_{0,I_\mathbb{R}}\big((-\infty, x]\big). \tag{1.143}$$

The proof can, e.g., be found in Theorem 15.37 in [Klenke(2008)]. Now let $(\Omega, \mathcal{A}, P)$ be a probability space and let $U_n\colon \Omega \to \mathbb{R}$, $n \in \mathbb{N}$, be P-independent $\mathcal{U}_{(0,1)}$-distributed random variables. Note that for all $n \in \mathbb{N}$ it holds that

$$\mathbb{E}\big[U_n\big] = \frac{1}{2} \qquad \text{and} \qquad \mathsf{Var}_P(U_n) = \mathbb{E}\Big[\big(U_n - \tfrac{1}{2}\big)^2\Big] = \frac{1}{12}. \tag{1.144}$$

Next let $S_n\colon \Omega \to \mathbb{R}$, $n \in \mathbb{N}$, be the functions which satisfy for all $n \in \mathbb{N}$ that

$$S_n = \frac{U_1 + \ldots + U_n - \mathbb{E}\big[U_1 + \ldots + U_n\big]}{\sqrt{\mathsf{Var}_P(U_1 + \ldots + U_n)}} = \frac{U_1 + \ldots + U_n - \frac{n}{2}}{\sqrt{\frac{n}{12}}}. \tag{1.145}$$

The central limit theorem then proves that S_n, $n \in \mathbb{N}$, converges in distribution to $\mathcal{N}_{0,I_\mathbb{R}}$. Thus, if $n \in \mathbb{N}$ is large, then $S_n(P)_{\mathcal{B}(\mathbb{R})}$ is a good approximation of $\mathcal{N}_{0,I_\mathbb{R}}$ in the sense of Theorem 1.3.1. In computer programs sometimes realizations of S_n for large $n \in \mathbb{N}$ are used as approximative realizations of a standard normal random variable. For example, in the case $n = 12$ we obtain $S_{12} = U_1 + \ldots + U_{12} - 6 \in (-6, 6)$ and in that case the algorithm reads as follows.

Output: Realization x of $X \sim S_{12}(P)_{\mathcal{B}(\mathbb{R})} \approx \mathcal{N}_{0,I_\mathbb{R}}$

$s = 0$
for $n = 1 \to 12$ **do**
  Generate realization u of $U_n \sim \mathcal{U}_{(0,1)}$
  $s = s + u$
end for
$x = s - 6$

In Matlab the command "`sum(rand(1,12))-6`" generates a realization of a pseudo $S_{12}(P)_{\mathcal{B}(\mathbb{R})}$-distributed random variable; cf. Figures 1.3 and 1.4 below.

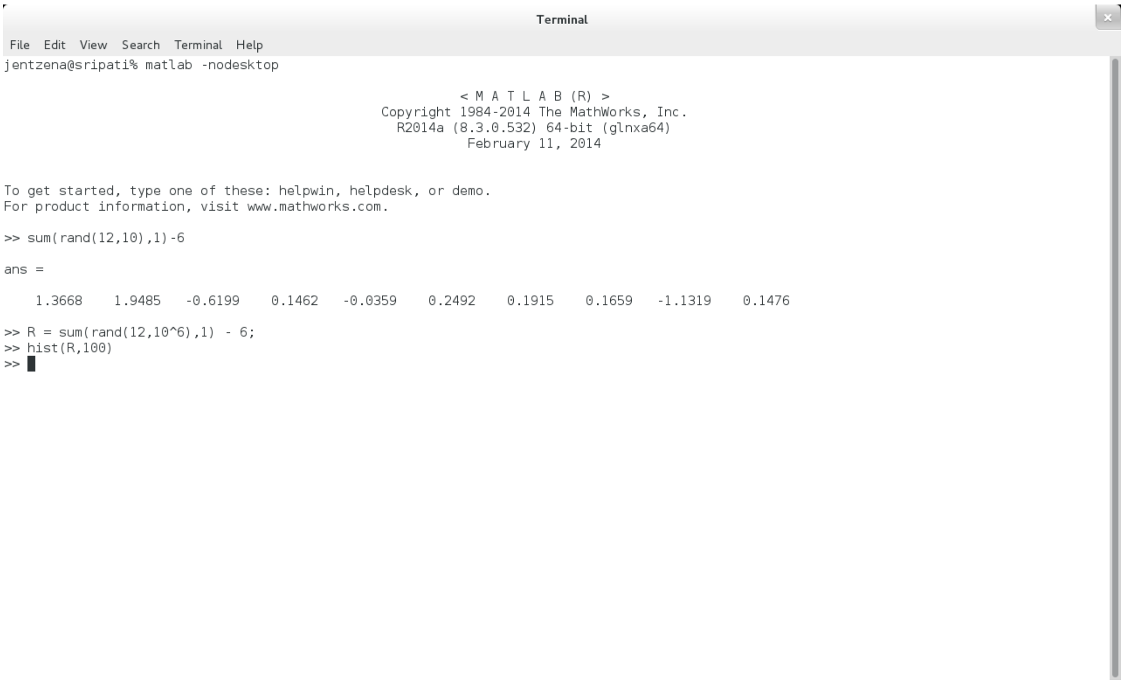


Figure 1.3: Matlab commands for generating realizations of an approximatively pseudo $\mathcal{N}_{0,I_{\mathbb{R}}}$-distributed random variable.

1.3.2 Box-Muller method

In this subsection the Box-Muller method is presented and analyzed and in the next subsection the Marsaglia polar method is investigated. Both methods are based on a polar representation result for the 2-dimensional standard normal distribution. To present this result, we use the following definition.

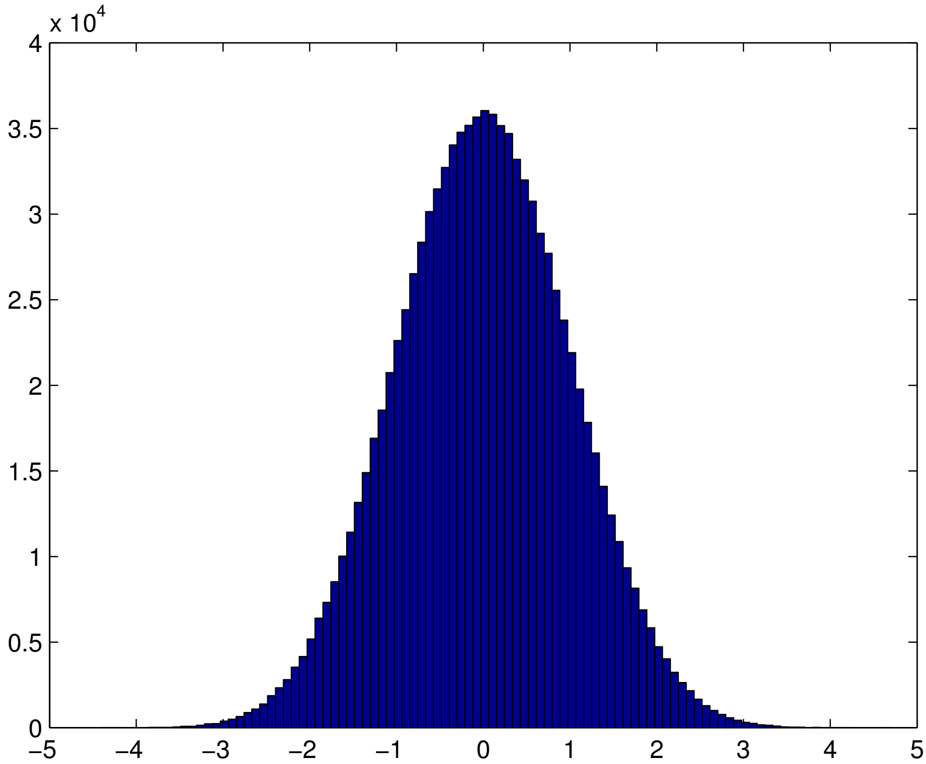


Figure 1.4: Output of the *hist* Matlab command in Figure 1.3.

Definition 1.3.2 (arg-function). *We denote by* $\arg\colon \{(x,y)\in\mathbb{R}^2\colon x^2+y^2=1\}\to[0,2\pi)$ *the function which satisfies for all* $\alpha\in[0,2\pi)$, $(x,y)\in\mathbb{R}^2$ *with* $x^2+y^2=1$ *that*

$$\arg\big(\cos(\alpha),\sin(\alpha)\big)=\alpha \tag{1.146}$$

and

$$\big(\cos\big(\arg(x,y)\big),\sin\big(\arg(x,y)\big)\big)=(x,y). \tag{1.147}$$

The function arg introduced in Definition 1.3.2 is thus the inverse of the continuous and bijective function

$$[0,2\pi)\ni\alpha\mapsto(\cos(\alpha),\sin(\alpha))\in\{(x,y)\in\mathbb{R}^2\colon x^2+y^2=1\}. \tag{1.148}$$

Observe that arg is an $\mathcal{B}(\{(x,y)\in\mathbb{R}^2\colon x^2+y^2=1\})/\mathcal{B}([0,2\pi))$-measurable function. We are now ready to present the polar representation result.

Proposition 1.3.3 (Polar representation for the two dimensional standard normal distribution). *Let* $(\Omega,\mathcal{A},P)$ *be a probability space, let* $R\colon\Omega\to[0,\infty)$ *be an* $\mathcal{A}/\mathcal{B}([0,\infty))$-*measurable function, and let* $S=(S_1,S_2)\colon\Omega\to\{(x,y)\in\mathbb{R}^2\colon x^2+y^2=1\}$ *be an* $\mathcal{A}/\mathcal{B}(\{(x,y)\in\mathbb{R}^2\colon x^2+y^2=1\})$-*measurable function. Then it holds that*

$$\Omega\ni\omega\mapsto R(\omega)\cdot S(\omega)=\big(R(\omega)\cdot S_1(\omega),R(\omega)\cdot S_2(\omega)\big)\in\mathbb{R}^2 \tag{1.149}$$

is $\mathcal{N}_{0,I_{\mathbb{R}^2}}$-*distributed if and only if*

$$\Omega\ni\omega\mapsto\big(|R(\omega)|^2,\arg\big(S(\omega)\big)\big)\in\mathbb{R}^2 \tag{1.150}$$

is $(\exp_{1/2}\otimes\,\mathcal{U}_{(0,2\pi)})$-*distributed.*

Proof of Proposition 1.3.3. Let $(\bar\Omega,\bar{\mathcal{A}},\bar P)$ be a probability space, let $\bar R\colon\bar\Omega\to[0,\infty)$ be an $\bar{\mathcal{A}}/\mathcal{B}([0,\infty))$-measurable function, let $\bar S=(\bar S_1,\bar S_2)\colon\bar\Omega\to\{(x,y)\in\mathbb{R}^2\colon x^2+y^2=1\}$ be an $\bar{\mathcal{A}}/\mathcal{B}(\{(x,y)\in\mathbb{R}^2\colon x^2+y^2=1\})$-measurable function, assume that

$$\bar\Omega\ni\omega\mapsto|\bar R(\omega)|^2\in\mathbb{R} \tag{1.151}$$

is $\exp_{1/2}$-distributed, assume that

$$\bar\Omega\ni\omega\mapsto\arg(\bar S(\omega))\in\mathbb{R} \tag{1.152}$$

is $\mathcal{U}_{(0,2\pi)}$-distributed, and assume that $\bar R$ and $\bar S$ are $\bar P$-independent. (Observe that such a probability space does indeed exist.) Next let $\bar X\colon\bar\Omega\to\mathbb{R}^2$ be the function which satisfies for all $\omega\in\Omega$ that

$$\begin{aligned}&\bar X(\omega)\\&=\begin{cases}\bar R(\omega)\cdot\bar S(\omega)=\Big(\bar R(\omega)\cdot\cos\big(\arg(\bar S(\omega))\big),\bar R(\omega)\cdot\sin\big(\arg(\bar S(\omega))\big)\Big)&\colon\bar R(\omega)>0\\\bar S(\omega)&\colon\bar R(\omega)=0\end{cases}.\end{aligned} \tag{1.153}$$

The $\bar{P}$-independency of $\bar{R}^2$ and $\arg(\bar{S})$, the integral transformation theorem, and (1.153) then prove that for all bounded $\mathcal{B}(\mathbb{R}^2)/\mathcal{B}(\mathbb{R})$-measurable functions $g\colon \mathbb{R}^2 \to \mathbb{R}$ it holds that

$$\begin{aligned}
&\int_{\bar{\Omega}} g(\bar{X})\, d\bar{P} = \int_{\bar{\Omega}} g(\bar{R}\cdot\bar{S})\, d\bar{P} \\
&= \int_{\bar{\Omega}} g(|\bar{R}|\cdot\bar{S})\, d\bar{P} = \int_{\bar{\Omega}} g\Big(\sqrt{\bar{R}^2}\cdot\bar{S}\Big)\, d\bar{P} \\
&= \int_{\bar{\Omega}} g\Big(\sqrt{\bar{R}^2}\cdot\cos(\arg(\bar{S})), \sqrt{\bar{R}^2}\cdot\sin(\arg(\bar{S}))\Big)\, d\bar{P} \\
&= \int_{(0,\infty)\times[0,2\pi)} g(\sqrt{r}\cdot\cos(\alpha), \sqrt{r}\cdot\sin(\alpha))\,\big(\bar{R}^2, \arg(\bar{S})\big)(\bar{P})_{\mathcal{B}([0,\infty)\times[0,2\pi))}(dr, d\alpha) \qquad (1.154)\\
&= \frac{1}{2\pi}\int_{[0,2\pi)}\int_{(0,\infty)} g(\sqrt{r}\cdot\cos(\alpha), \sqrt{r}\cdot\sin(\alpha))\,\big((\bar{R}^2)(\bar{P})\big)_{\mathcal{B}([0,\infty))}(dr)\, d\alpha \\
&= \frac{1}{4\pi}\int_{[0,2\pi)}\int_{(0,\infty)} g(\sqrt{r}\cdot\cos(\alpha), \sqrt{r}\cdot\sin(\alpha))\, e^{\frac{-r}{2}}\, dr\, d\alpha \\
&= \frac{1}{2\pi}\int_0^{2\pi}\int_0^{\infty} g(r\cdot\cos(\alpha), r\cdot\sin(\alpha))\, e^{\frac{-r^2}{2}}\, r\, dr\, d\alpha.
\end{aligned}$$

A polar coordinate transform hence gives that for all bounded $\mathcal{B}(\mathbb{R}^2)/\mathcal{B}(\mathbb{R})$-measurable functions $g\colon \mathbb{R}^2 \to \mathbb{R}$ it holds that

$$\int_{\bar{\Omega}} g(\bar{X})\, d\bar{P} = \frac{1}{2\pi}\int_{\mathbb{R}^2} g(x)\, e^{\frac{-1}{2}\|x\|^2_{\mathbb{R}^2}}\, dx = \int_{\mathbb{R}^2} g(x)\, \mathcal{N}_{0,I_{\mathbb{R}^2}}(dx)\,. \qquad (1.155)$$

This, in particular, implies that for all $A \in \mathcal{B}(\mathbb{R}^2)$ it holds that

$$\big(\bar{X}(\bar{P})_{\mathcal{B}(\mathbb{R}^2)}\big)(A) = \mathcal{N}_{0,I_{\mathbb{R}^2}}(A). \qquad (1.156)$$

This proves the "$\Leftarrow$" statement in Proposition 1.3.3. Next assume that

$$\Omega \ni \omega \mapsto R(\omega)\cdot S(\omega) = \big(R(\omega)\cdot S_1(\omega), R(\omega)\cdot S_2(\omega)\big) \in \mathbb{R}^2 \qquad (1.157)$$

is $\mathcal{N}_{0,I_{\mathbb{R}^2}}$-distributed and let $X\colon \Omega \to \mathbb{R}^2$ be the function with the property that for all $\omega \in \Omega$ it holds that

$$X(\omega) = \begin{cases} R(\omega)\cdot S(\omega) & : R(\omega) \neq 0 \\ S(\omega) & : R(\omega) = 0 \end{cases}. \qquad (1.158)$$

Then we obtain that

$$X(P)_{\mathcal{B}(\mathbb{R}^2)} = \mathcal{N}_{0,I_{\mathbb{R}^2}} = \bar{X}(\bar{P})_{\mathcal{B}(\mathbb{R}^2)}. \qquad (1.159)$$

Hence, we get that

$$\begin{aligned}
&\big(\Omega \ni \omega \mapsto \big(|R(\omega)|^2, \arg(S(\omega))\big) \in \mathbb{R}^2\big)(P)_{\mathcal{B}(\mathbb{R}^2)} \\
&= \Big(\Omega \ni \omega \mapsto \Big(\|X(\omega)\|^2_{\mathbb{R}^2}, \arg\Big(\tfrac{X(\omega)}{\|X(\omega)\|_{\mathbb{R}^2}}\Big)\Big) \in \mathbb{R}^2\Big)(P)_{\mathcal{B}(\mathbb{R}^2)} \\
&= \Big(\bar{\Omega} \ni \omega \mapsto \Big(\|\bar{X}(\omega)\|^2_{\mathbb{R}^2}, \arg\Big(\tfrac{\bar{X}(\omega)}{\|\bar{X}(\omega)\|_{\mathbb{R}^2}}\Big)\Big) \in \mathbb{R}^2\Big)(\bar{P})_{\mathcal{B}(\mathbb{R}^2)} \\
&= \Big(\bar{\Omega} \ni \omega \mapsto \Big(\big|\bar{R}(\omega)\big|^2, \arg\big(\bar{S}(\omega)\big)\Big) \in \mathbb{R}^2\Big)(\bar{P})_{\mathcal{B}(\mathbb{R}^2)} = \exp_{1/2}\otimes\,\mathcal{U}_{(0,2\pi)}.
\end{aligned} \qquad (1.160)$$

This proves the "⇒" statement in Proposition 1.3.3. The proof of Proposition 1.3.3 is thus completed. □

The following result, Corollary 1.3.4, presents a property for the 2-dimensional normal distribution and is a slightly different perspective on Proposition 1.3.3. Corollary 1.3.4 follows immediately from Proposition 1.3.3.

Corollary 1.3.4 (A property for the 2-dimensional normal distribution)**.** *Let $(\Omega, \mathcal{A}, P)$ be a probability space, let $X\colon \Omega \to \mathbb{R}^2$ be an $\mathcal{N}_{0,I_{\mathbb{R}^2}}$-distributed random variable, and let $Q, U\colon \Omega \to \mathbb{R}$ be the functions which satisfy for all $\omega \in \Omega$ that*

$$Q(\omega) = \|X(\omega)\|_{\mathbb{R}^2}^2 \qquad \text{and} \qquad U(\omega) = \begin{cases} \arg\!\Big(\frac{X(\omega)}{\|X(\omega)\|_{\mathbb{R}^2}}\Big) & : X(\omega) \neq 0 \\ 0 & : \text{otherwise} \end{cases}. \tag{1.161}$$

Then it holds that Q and U are P-independent, that Q is $\exp_{1/2}$-distributed, and that U is $\mathcal{U}_{(0,2\pi)}$-distributed.

Proof of Corollary 1.3.4. Throughout this proof let $\bar{X}\colon \Omega \to \mathbb{R}^2$, $R\colon \Omega \to [0,\infty)$, and $S\colon \Omega \to \{(x,y) \in \mathbb{R}^2\colon x^2 + y^2 = 1\}$ be the functions which satisfy for all $\omega \in \Omega$ that

$$\bar{X}(\omega) = \begin{cases} X(\omega) & : X(\omega) \neq 0 \\ (1,0) & : X(\omega) = 0 \end{cases}, \qquad R(\omega) = \big\|\bar{X}(\omega)\big\|_{\mathbb{R}^2}, \qquad S(\omega) = \frac{\bar{X}(\omega)}{R(\omega)}. \tag{1.162}$$

The fact that X is $\mathcal{N}_{0,I_{\mathbb{R}^2}}$-distributed then shows that it holds P-a.s. that

$$X = \bar{X}. \tag{1.163}$$

Hence, we obtain that $\bar{X}$ is $\mathcal{N}_{0,I_{\mathbb{R}^2}}$-distributed too. Combining this and the fact that for all $\omega \in \Omega$ it holds that

$$\bar{X}(\omega) = R(\omega) \cdot S(\omega) \tag{1.164}$$

allows us to apply Proposition 1.3.3 to obtain that the function

$$\Omega \ni \omega \mapsto \big(|R(\omega)|^2, \arg(S(\omega))\big) = \Big(\|\bar{X}(\omega)\|_{\mathbb{R}^2}^2, \arg\!\Big(\tfrac{\bar{X}(\omega)}{\|\bar{X}(\omega)\|_{\mathbb{R}^2}}\Big)\Big) \in \mathbb{R}^2 \tag{1.165}$$

is $(\exp_{1/2} \otimes \mathcal{U}_{(0,2\pi)})$-distributed. This together with (1.163) completes the proof of Corollary 1.3.4. □

Class exercise 1.3.5. *Does there exist a set $A \in \mathcal{B}(\mathbb{R})$ which satisfies that $\mathcal{U}_{(0,1)}(A) \neq \mathcal{U}_{[0,1]}(A)$?*

The next result, Corollary 1.3.6, results in a method for the generation of realizations of two independent standard normal random variables. The method is referred to as the *Box-Muller method* in the literature and has been proposed in [Box and Muller(1958)]. Corollary 1.3.6 follows from Proposition 1.3.3.

Corollary 1.3.6 (Box-Muller method)**.** *Let $(\Omega, \mathcal{A}, P)$ be a probability space and let $U_1, U_2 \colon \Omega \to \mathbb{R}$ be two P-independent $\mathcal{U}_{(0,1)}$-distributed random variables which satisfy $U_1(\Omega) \subseteq (0,1)$ and $U_2(\Omega) \subseteq (0,1)$. Then the function $X = (X_1, X_2) \colon \Omega \to \mathbb{R}^2$ given by*

$$\begin{aligned} X_1 &= \sqrt{-2\ln(U_1)}\cos(2\pi U_2), \\ X_2 &= \sqrt{-2\ln(U_1)}\sin(2\pi U_2) \end{aligned} \tag{1.166}$$

is $\mathcal{N}_{0,I_{\mathbb{R}^2}}$-distributed.

Proof of Corollary 1.3.6. We prove Corollary 1.3.6 through an application of Proposition 1.3.3. For this let $R \colon \Omega \to \mathbb{R}$ and $S \colon \Omega \to \{(x,y) \in \mathbb{R}^2 \colon x^2 + y^2 = 1\}$ be the functions given by

$$R = \sqrt{-2\ln(U_1)} \tag{1.167}$$

and

$$S = (\cos(2\pi U_2), \sin(2\pi U_2)). \tag{1.168}$$

Note that the assumption that U_1 and U_2 are P-independent ensures that R and S are P-independent. Moreover, observe that for all $\omega \in \Omega$ it holds that

$$\arg(S(\omega)) = 2\pi U_2(\omega). \tag{1.169}$$

This and the assumption that U_2 is $\mathcal{U}_{(0,1)}$-distributed imply that

$$\Omega \ni \omega \mapsto \arg(S(\omega)) \in \mathbb{R} \tag{1.170}$$

is $\mathcal{U}_{(0,2\pi)}$-distributed. Next note that for all $\omega \in \Omega$ it holds that

$$[R(\omega)]^2 = -2\ln(U_1(\omega)) = \frac{-\ln(U_1(\omega))}{1/2}. \tag{1.171}$$

This and (1.48) prove that

$$\Omega \ni \omega \mapsto [R(\omega)]^2 \in \mathbb{R} \tag{1.172}$$

is $\exp_{1/2}$-distributed. We can thus apply Proposition 1.3.3 to obtain that $(X_1, X_2) \colon \Omega \to \mathbb{R}^2$ is $\mathcal{N}_{0,I_{\mathbb{R}^2}}$-distributed. The proof of Corollary 1.3.6 is thus completed. ☐

Corollary 1.3.6 results in the following algorithm for the generation of two independent standard normal distributed random variables.

Box-Muller method

Output: Realization (x_1, x_2) of $(X_1, X_2) \sim \mathcal{N}_{0,I_{\mathbb{R}^2}}$
Generate realization (u_1, u_2) of $(U_1, U_2) \sim \mathcal{U}_{(0,1)^2}$
$x_1 = \sqrt{-2\ln(u_1)}\cos(2\pi u_2)$
$x_2 = \sqrt{-2\ln(u_1)}\sin(2\pi u_2)$

Exercise 1.3.7. *In this exercise we do not distinguish between pseudo random numbers and actual random numbers.*

(i) *Write a* Matlab *function* `BoxMuller(N)` *with input* $N \in \mathbb{N}$ *and output a realization of an* $\mathcal{N}_{0,I_{\mathbb{R}^N}}$*-distributed random variable generated with the Box-Muller method. Your* Matlab *function* `BoxMuller(N)` *may use at most* $N+1$ *realizations of an* $\mathcal{U}_{(0,1)}$*-distributed random variable. Type* `BoxMuller(11)` *to test your implementation.*

(ii) *Write a* Matlab *function* `BoxMullerPlot()` *which plots* 10^5 *realizations of an* $\mathcal{N}_{0,I_{\mathbb{R}}}$*-distributed random variable generated with your* Matlab *function* `BoxMuller(N)` *from* (i) *in a* normalized *histogram with* 1000 *bins and which also plots the density of* $\mathcal{N}_{0,I_{\mathbb{R}}}$ *in this histogram.*

Hint: Use the built-in Matlab *function* `hist(...)` *to obtain the raw data of the histogram, then normalize it, and then create the plot, for example, with the built-in* Matlab *function* `bar(...)`*. For the plot of the density function of* $\mathcal{N}_{0,I_{\mathbb{R}}}$ *use the built-in* Matlab *function* `plot(...)` *and the built-in* Matlab *command* `hold on`*.*

1.3.3 Marsaglia polar method

The Marsaglia polar method is a slight modification of the Box-Muller method. It avoids the computationally expensive evalutions of the sine- and the cosine-function and is therefore typically faster than the Box-Muller method. The theorectical justification of the method is provided through Lemma 1.2.28 and through the next proposition.

Proposition 1.3.8 (Marsaglia polar method)**.** *Let* $(\Omega, \mathcal{A}, P)$ *be a probability space, let* $U\colon \Omega \to \mathbb{R}^2$ *be an* $\mathcal{U}_{\{(x,y)\in\mathbb{R}^2\colon x^2+y^2\in(0,1)\}}$*-distributed random variable with* $U(\Omega) \subseteq \{(x,y)\in\mathbb{R}^2\colon x^2+y^2\in(0,1)\}$*, and let* $X\colon \Omega \to \mathbb{R}^2$ *be the function given by*

$$X = \frac{U\sqrt{-2\ln\big(\|U\|_{\mathbb{R}^2}^2\big)}}{\|U\|_{\mathbb{R}^2}}. \tag{1.173}$$

Then X *is* $\mathcal{N}_{0,I_{\mathbb{R}^2}}$*-distributed.*

Proof of Proposition 1.3.8. We prove Proposition 1.3.8 through an application of Proposition 1.3.3. For this let $R\colon \Omega \to [0,\infty)$ and $S\colon \Omega \to \{(x,y)\in\mathbb{R}^2\colon x^2+y^2=1\}$ be the functions given by

$$R = \sqrt{-2\ln\big(\|U\|_{\mathbb{R}^2}^2\big)} = \sqrt{\frac{-\ln\big(\|U\|_{\mathbb{R}^2}^2\big)}{1/2}} \qquad \text{and} \qquad S = \frac{U}{\|U\|_{\mathbb{R}^2}}. \tag{1.174}$$

Next we observe that for all $x, \alpha \in \mathbb{R}$ it holds that

$$\begin{aligned}
&P\big(\big(\|U\|_{\mathbb{R}^2}^2, \arg(S)\big) \in (-\infty, x) \times (-\infty, \alpha)\big)\\
&= P\Big(\|U\|_{\mathbb{R}^2}^2 < x,\ \arg(S) < \alpha\Big)\\
&= \big(U(P)_{\mathcal{B}(\{u\in\mathbb{R}^2\colon \|u\|_{\mathbb{R}^2}\in(0,1)\})}\big)\Big(\Big\{u \in \mathbb{R}^2 \colon 0 < \|u\|_{\mathbb{R}^2}^2 < x,\ \arg\Big(\tfrac{u}{\|u\|_{\mathbb{R}^2}}\Big) < \alpha\Big\}\Big)\\
&= \frac{\lambda_{\mathbb{R}^2}\Big(\Big\{u \in \mathbb{R}^2 \colon 0 < \|u\|_{\mathbb{R}^2}^2 < \min\{1, x\},\ \arg\Big(\tfrac{u}{\|u\|_{\mathbb{R}^2}}\Big) < \alpha\Big\}\Big)}{\lambda_{\mathbb{R}^2}(\{u \in \mathbb{R}^2 \colon \|u\|_{\mathbb{R}^2} \in (0,1)\})}\\
&= \frac{\lambda_{\mathbb{R}^2}\Big(\Big\{u \in \mathbb{R}^2 \colon 0 < \|u\|_{\mathbb{R}^2} < \sqrt{\min\{1, \max\{0, x\}\}},\ \arg\Big(\tfrac{u}{\|u\|_{\mathbb{R}^2}}\Big) < \alpha\Big\}\Big)}{\lambda_{\mathbb{R}^2}(\{u \in \mathbb{R}^2 \colon \|u\|_{\mathbb{R}^2} \in (0,1)\})}\\
&= \frac{\pi \min\{1, \max\{0, x\}\} \cdot \frac{\min\{2\pi, \max\{0,\alpha\}\}}{2\pi}}{\pi}\\
&= \min\{1, \max\{0, x\}\} \cdot \frac{\min\{2\pi, \max\{0, \alpha\}\}}{2\pi}\\
&= \mathcal{U}_{(0,1)}((-\infty, x)) \cdot \mathcal{U}_{(0,2\pi)}((-\infty, \alpha)) = \big(\mathcal{U}_{(0,1)} \otimes \mathcal{U}_{(0,2\pi)}\big)\big((-\infty, x) \times (-\infty, \alpha)\big).
\end{aligned} \tag{1.175}$$

Combining this with Exercise 1.2.16 and Theorem 1.2.15 implies for all $A \in \mathcal{B}(\mathbb{R}^2)$ that

$$P\big(\big(\|U\|_{\mathbb{R}^2}^2, \arg(S)\big) \in A\big) = \big(\mathcal{U}_{(0,1)} \otimes \mathcal{U}_{(0,2\pi)}\big)(A). \tag{1.176}$$

This proves that

$$\Omega \ni \omega \mapsto \|U(\omega)\|_{\mathbb{R}^2}^2 \in \mathbb{R} \tag{1.177}$$

is $\mathcal{U}_{(0,1)}$-distributed, that

$$\Omega \ni \omega \mapsto \arg(S(\omega)) \in \mathbb{R} \tag{1.178}$$

is $\mathcal{U}_{(0,2\pi)}$-distributed, and that (1.177) and (1.178) are P-independent. This together with (1.48), in turn, implies that

$$\Omega \ni \omega \mapsto [R(\omega)]^2 \in \mathbb{R} \tag{1.179}$$

is $\exp_{1/2}$-distributed and that (1.179) and (1.178) are P-independent. We can thus apply Proposition 1.3.3 to obtain that

$$\Omega \ni \omega \mapsto R(\omega) \cdot S(\omega) = X(\omega) \in \mathbb{R}^2 \tag{1.180}$$

is $\mathcal{N}_{0,I_{\mathbb{R}^2}}$-distributed. The proof of Proposition 1.3.8 is thus completed. □

Combining Lemma 1.2.28 and Proposition 1.3.8 results in the following algorithm in which $V = (V_1, V_2) \colon \Omega \to \mathbb{R}^2$ is an $\mathcal{U}_{(0,1)^2}$-distributed random variable with $V(\Omega) \subseteq (0,1)^2$.

Marsaglia polar method

Output: Realization (x_1, x_2) of $(X_1, X_2) \sim \mathcal{N}_{0,I_{\mathbb{R}^2}}$
repeat
Generate realization (v_1, v_2) of $(V_1, V_2) \sim \mathcal{U}_{(0,1)^2}$
$q = (2v_1 - 1)^2 + (2v_2 - 1)^2$
until $q \in (0, 1)$
$w = \sqrt{-2\ln(q)/q}$
$x_1 = (2v_1 - 1)w$
$x_2 = (2v_2 - 1)w$

Finally, observe that the acceptance probability in the acceptance-rejection algorithm in the Marsaglia polar method is

$$\mathcal{U}_{(-1,1)^2}\big(\{u \in \mathbb{R}^2 \colon \|u\|_{\mathbb{R}^2} \in (0,1)\}\big) = \frac{\lambda_{\mathbb{R}^2}(\{u \in \mathbb{R}^2 \colon \|u\|_{\mathbb{R}^2} \in (0,1)\})}{4} = \frac{\pi}{4} \approx 0.78 \tag{1.181}$$

Thus on average the algorithm in the Marsaglia polar method runs $\frac{4}{\pi} \approx 1.27$-times through the loop.

Exercise 1.3.9. *In this exercise we do not distinguish between pseudo random numbers and actual random numbers.*

(i) Write a Matlab *function* `MarsagliaPolar(N)` *with input* $N \in \mathbb{N}$ *and output a realization of an* $\mathcal{N}_{0,I_{\mathbb{R}^N}}$*-distributed random variable generated with the Marsaglia polar method. Type* `MarsagliaPolar(11)` *to test your implementation.*

(ii) Write a Matlab *function* `MarsagliaPolarPlot()` *which plots* 10^5 *realizations of an* $\mathcal{N}_{0,I_{\mathbb{R}}}$*-distributed random variable generated with the* Matlab *function* `MarsagliaPolar(N)` *in a* normalized *histogram with* 1000 *bins and which also plots the density of* $\mathcal{N}_{0,I_{\mathbb{R}}}$ *in this histogram.*

Hint: Use the built-in Matlab *function* `hist(...)` *to obtain the raw data of the histogram, then normalize it, and then create the plot, for example, with the built-in* Matlab *function* `bar(...)`*. For the plot of the density function of* $\mathcal{N}_{0,I_{\mathbb{R}}}$ *use the built-in* Matlab *function* `plot(...)` *and the built-in* Matlab *command* `hold on`*.*

1.3.4 Normally distributed random variables with general mean and general covariance matrix

In this subsection we illustrate how realizations of a normally distributed random variable with mean $v \in \mathbb{R}^d$ and covariance matrix $Q \in \mathbb{R}^{d\times d}$ can be generated, where $v \in \mathbb{R}^d$ is a vector, where $Q \in \mathbb{R}^{d\times d}$ is a nonnegative symmetric matrix, and where $d \in \mathbb{N}$ is a natural number. For this we use the following special case of Proposition 0.4.15.

Corollary 1.3.10. *Let* $(\Omega, \mathcal{A}, P)$ *be a probability space, let* $d \in \mathbb{N}$*,* $b \in \mathbb{R}^d$*,* $A \in \mathbb{R}^{d\times d}$*, and let* $X \colon \Omega \to \mathbb{R}^d$ *be an* $\mathcal{N}_{0,I_{\mathbb{R}^d}}$*-distributed random variable. Then the function*

$$\Omega \ni \omega \mapsto AX(\omega) + b \in \mathbb{R}^d \tag{1.182}$$

is $\mathcal{N}_{b,AA^\top}$-distributed.

Corollary 1.3.10 illustrates for every $d \in \mathbb{N}$, $v \in \mathbb{R}^d$ and every strictly positive symmetric $d \times d$-matrix $Q \in \mathbb{R}^{d\times d}$ that the Matlab command

$$\texttt{chol(Q)' * randn(d,1) + v} \tag{1.183}$$

generates a realization of a pseudo $\mathcal{N}_{v,Q}$-distributed random variable.

Exercise 1.3.11 (Approximative realizations of a one-dimensional standard Brownian motion)**.** *In this exercise we do not distinguish between pseudo random numbers and actual random numbers. Let A be the set given by*

$$A = \cup_{n=1}^{\infty} \{\mathbf{t} = (t_1, \dots, t_n) \in [0,\infty)^n \colon \#_{\mathbb{R}}(\{t_1, \dots, t_n\}) = n\}, \tag{1.184}$$

let length: $A \to \mathbb{N}$ *be the function which satisfies for all $n \in \mathbb{N}$, $\mathbf{t} = (t_1, \dots, t_n) \in [0,\infty)^n \cap A$ that*

$$\text{length}(\mathbf{t}) = n, \tag{1.185}$$

and let $Q \colon A \to (\cup_{n=1}^{\infty} \mathbb{R}^{n\times n})$ be the function which satisfies for all $n \in \mathbb{N}$, $\mathbf{t} = (t_1, \dots, t_n) \in [0,\infty)^n \cap A$ that

$$Q(\mathbf{t}) = (\min\{t_i, t_j\})_{(i,j)\in\{1,\dots,n\}^2}. \tag{1.186}$$

Write a Matlab *function* ***StandardBrownianMotion(t)*** *with input $\mathbf{t} \in A$ and output a realization of an $\mathcal{N}_{0,Q(\mathbf{t})}$-distributed random variable. The* Matlab *function* ***StandardBrownianMotion(t)*** *may use at most* length($\mathbf{t}$) *realizations of an $\mathcal{N}_{0,I_{\mathbb{R}}}$-distributed random variable. Call the* Matlab *commands*

```
rng('default');
N=10^3;
preimage = (0:1/N:1);
X=StandardBrownianMotion(preimage);
plot(preimage,X);
hold on
X=StandardBrownianMotion(preimage);
plot(preimage,X,'r');
X=StandardBrownianMotion(preimage);
plot(preimage,X,'g');
```

to test your implementation.

2 Monte Carlo integration methods

Let $(A, \mathcal{A}, \mu)$ be a finite measure space with $\mu(A) \neq 0$ and let $f\colon A \to \mathbb{R}$ be an $\mathcal{A}/\mathcal{B}(\mathbb{R})$-measurable function with $\int_A |f(x)|_{\mathbb{R}}\, \mu(dx) < \infty$. This chapter presents numerical methods for the approximative computation of the real number

$$\int_A f(x)\, \mu(dx) \in \mathbb{R}. \tag{2.1}$$

If $d \in \mathbb{N}$, $a, b \in \mathbb{R}$ with $a < b$, if $A = [a, b]^d \subseteq \mathbb{R}^d$, and if $\mu = B_A$ is the Lebesgue-Borel measure on A in (2.1), then (2.1) reduces to

$$\int_{[a,b]^d} f(x)\, dx \in \mathbb{R} \tag{2.2}$$

and in that case, classical deterministic numerical methods can be used for the approximative computation of (2.2) and (2.1) respectively. These deterministic numerical integration methods are briefly reviewed in Section 2.2. Sections 2.3–2.6 present and analyze a random method, the so-called *Monte Carlo method*, for the approximative computation of (2.1).

The convergence speed of numerical integration methods for (2.1) and (2.2) often depends on the regularity of the integrand function $f\colon A \to \mathbb{R}$. To study these regularities, a bit more notation is introduced in the next section, Section 2.1.

The content of this chapter can be found in a similar form in diverse books on numerical integration methods and Monte Carlo methods respectively. We refer, e.g., to [Atkinson(1989)], [Fishman(1996)], [Kloeden and Platen(1992)], [Glasserman(2004)] and [Müller-Gronbach et al.(2012)Müller-Gronbach, Novak, and Ritter].

2.1 Regularity of functions

Definition 2.1.1 (Separability)**.** *Let $\mathbf{E}$ be a topological space. Then we say that $\mathbf{E}$ is separable if and only if there exist E, $\mathcal{E}$, and F such that*

(i) it holds that $\mathbf{E} = (E, \mathcal{E})$,

(ii) it holds that F is at most countable, and

(iii) it holds that $\overline{E \cap F}^{\mathcal{E}} = E$.

Definition 2.1.2. *Let $(A, \mathcal{A})$ and $(B, \mathcal{B})$ be measurable spaces. Then we denote by $\mathcal{M}(\mathcal{A}, \mathcal{B})$ the set of all $\mathcal{A}/\mathcal{B}$-measurable functions.*

Definition 2.1.3 ($\mathcal{L}^p$-spaces for $p \in [0, \infty)$)**.** *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, let $q \in (0, \infty)$, and let $(V, \|\cdot\|_V)$ be a separable normed $\mathbb{R}$-vector space. Then we denote by $\mathcal{L}^0(\mu; \|\cdot\|_V)$ the set given by*

$$\mathcal{L}^0(\mu; \|\cdot\|_V) = \mathcal{M}\big(\mathcal{A}, \mathcal{B}(V)\big), \tag{2.3}$$

we denote by $\|\cdot\|_{\mathcal{L}^q(\mu;\|\cdot\|_V)} : \mathcal{L}^0(\mu; \|\cdot\|_V) \to [0, \infty]$ the function which satisfies for all $f \in \mathcal{L}^0(\mu; \|\cdot\|_V)$ that

$$\|f\|_{\mathcal{L}^q(\mu;\|\cdot\|_V)} = \left[\int_\Omega \|f(\omega)\|_V^q \, \mu(d\omega)\right]^{1/q} \in [0, \infty], \tag{2.4}$$

and we denote by $\mathcal{L}^q(\mu; \|\cdot\|_V)$ the set given by

$$\mathcal{L}^q(\mu; \|\cdot\|_V) = \big\{f \in \mathcal{L}^0(\mu; \|\cdot\|_V) \colon \|f\|_{\mathcal{L}^q(\mu;\|\cdot\|_V)} < \infty\big\}. \tag{2.5}$$

Observe that $\mathcal{L}^0(\mu; \|\cdot\|_V)$ and $\mathcal{L}^p(\mu; \|\cdot\|_V)$ in Definition 2.1.3 are $\mathbb{R}$-vector spaces. However, for every $q \in [1, \infty)$ it is in general not true that in the setting of Definition 2.1.3 it holds that the function $\|\cdot\|_{\mathcal{L}^q(\mu;\|\cdot\|_V)}$ is a norm on $\mathcal{L}^q(\mu; \|\cdot\|_V)$. This lack of being definite brings us to the next definition.

Definition 2.1.4 (Equivalence classes)**.** *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, let $(E, \mathcal{E})$ be a measurable space, let R be a set, and let $f \colon \Omega \to R$ be a function. Then we denote by $[f]_{\mu,\mathcal{E}} \subseteq \mathcal{M}(\mathcal{A}, \mathcal{E})$ the set given by*

$$[f]_{\mu,\mathcal{E}} = \Big\{g \in \mathcal{M}(\mathcal{A}, \mathcal{E}) \colon \big(\exists\, A \in \mathcal{A} \colon \mu(A) = 0 \text{ and } \big\{\omega \in \Omega \colon f(\omega) \neq g(\omega)\big\} \subseteq A\big)\Big\}. \tag{2.6}$$

Note that for every measure space $(\Omega, \mathcal{A}, \mu)$, every separable normed $\mathbb{R}$-vector space $(V, \|\cdot\|_V)$, every $p \in [0, \infty)$, and every $f \in \mathcal{L}^p(\mu; \|\cdot\|_V)$ it holds that

$$[f]_{\mu,\mathcal{B}(V)} \subseteq \mathcal{L}^p(\mu; \|\cdot\|_V). \tag{2.7}$$

Definition 2.1.5 (L^p-spaces for $p \in [0, \infty)$)**.** *Let $p \in [0, \infty)$, $q \in (0, \infty)$, let $(\Omega, \mathcal{A}, \mu)$ be a measure space, and let $(V, \|\cdot\|_V)$ be a separable normed $\mathbb{R}$-vector space. Then we denote by $L^p(\mu; \|\cdot\|_V)$ the set given by*

$$L^p(\mu; \|\cdot\|_V) = \big\{\, [f]_{\mu,\mathcal{B}(V)} \in \mathcal{P}(\mathcal{L}^p(\mu; \|\cdot\|_V)) \colon f \in \mathcal{L}^p(\mu; \|\cdot\|_V)\big\} \tag{2.8}$$

and we denote by

$$\|\cdot\|_{L^q(\mu;\|\cdot\|_V)} : L^0(\mu; \|\cdot\|_V) \to [0, \infty] \tag{2.9}$$

the function which satisfies for all $f \in \mathcal{L}^0(\mu; \|\cdot\|_V)$ that

$$\big\|[f]_{\mu,\mathcal{B}(V)}\big\|_{L^q(\mu;\|\cdot\|_V)} = \|f\|_{\mathcal{L}^q(\mu;\|\cdot\|_V)} \in [0, \infty]. \tag{2.10}$$

Class exercise 2.1.6. *Specify all possible relations ($\subseteq$, $\supseteq$, $\not\subseteq$, $\not\supseteq$, $\subsetneq$, and $\supsetneq$) between the sets* $\mathbb{M}([0,3],\mathbb{R})$, $C([0,3],\mathbb{R})$, $\mathcal{L}^{0.5}(B_{[0,3]};|\cdot|_{\mathbb{R}})$, $\mathcal{L}^{1.5}(B_{[0,3]};|\cdot|_{\mathbb{R}})$, $\mathcal{L}^{3}(B_{[0,3]};|\cdot|_{\mathbb{R}})$, $\mathbb{M}(\mathbb{N},\mathbb{R})$, $\mathcal{L}^{0.5}(\#^{\mathbb{N}};|\cdot|_{\mathbb{R}})$, $\mathcal{L}^{1.5}(\#^{\mathbb{N}};|\cdot|_{\mathbb{R}})$, $\mathcal{L}^{3}(\#^{\mathbb{N}};|\cdot|_{\mathbb{R}})$, $\mathcal{L}^{0.5}(B_{\mathbb{R}};|\cdot|_{\mathbb{R}})$, $\mathcal{L}^{1.5}(B_{\mathbb{R}};|\cdot|_{\mathbb{R}})$, *and* $\mathcal{L}^{3}(B_{\mathbb{R}};|\cdot|_{\mathbb{R}})$.

Let $p \in [0,\infty)$, $q \in [1,\infty)$, let $(\Omega,\mathcal{A},\mu)$ be a measure space, and let $(V,\|\cdot\|_V)$ be a separable normed $\mathbb{R}$-vector space. Then observe that the pair

$$\big(L^q(\mu;\|\cdot\|_V), \|\cdot\|_{L^q(\mu;\|\cdot\|_V)}\big) \tag{2.11}$$

is a normed $\mathbb{R}$-vector space. Moreover, note that if $(V,\|\cdot\|_V)$ is complete, then so is

$$\big(L^q(\mu;\|\cdot\|_V), \|\cdot\|_{L^q(\mu;\|\cdot\|_V)}\big) \tag{2.12}$$

(see, e.g., Theorem 7.3 in [Klenke(2008)]). Moreover, as it is usual in the literature, we do in the following often not distinguish between

$$f \in \mathcal{L}^p(\mu;\|\cdot\|_V) \tag{2.13}$$

and its equivalence class

$$[f]_{\mu,\mathcal{B}(V)} \in L^p(\mu;\|\cdot\|_V) \tag{2.14}$$

in $L^p(\mu;\|\cdot\|_V) \subseteq L^0(\mu;\|\cdot\|_V)$ and, in particular, sometimes we simply write f instead of $[f]_{\mu,\mathcal{B}(V)}$. In the next step we introduce a tool to investigate continuity properties of a function.

Definition 2.1.7 (Modulus of continuity)**.** *Let (E,d_E) and (F,d_F) be metric spaces and let $f\colon E\to F$ be a function. Then we denote by*

$$w_f\colon [0,\infty] \to [0,\infty] \tag{2.15}$$

the function which satisfies for all $h \in [0,\infty]$ that

$$w_f(h) = \sup\Big(\Big\{d_F\big(f(x),f(y)\big) \in [0,\infty)\colon x,y\in E,\, d_E(x,y)\le h\Big\} \cup \{0\}\Big) \tag{2.16}$$

and we call w_f the modulus of continuity of f.

Observe that Definition 2.1.7 ensures that for all metric spaces (E,d_E) and (F,d_F) and all $f \in \mathbb{M}(E,F)$, $x,y\in E$ it holds that

$$d_F(f(x),f(y)) \le w_f(d_E(x,y))\,. \tag{2.17}$$

Moreover, note that for all $d,k\in\mathbb{N}$, $A\in\mathcal{P}(\mathbb{R}^d)$, $f\in\mathbb{M}(A,\mathbb{R}^k)$, $x,y\in A$, $h\in[0,\infty]$ it holds that

$$w_f(h) = \sup_{\substack{a,b\in A,\\ \|a-b\|_{\mathbb{R}^d}\le h}} \|f(a)-f(b)\|_{\mathbb{R}^k} \in [0,\infty] \tag{2.18}$$

and

$$\|f(x) - f(y)\|_{\mathbb{R}^k} \leq w_f(\|x - y\|_{\mathbb{R}^d}). \tag{2.19}$$

We also use spaces of Hölder continuous functions. This is the subject of the next definition.

Definition 2.1.8 (Hölder continuous functions)**.** *Let (E, d_E) and (F, d_F) be metric spaces and let $\alpha \in (0, 1]$. Then we denote by*

$$|\cdot|_{\mathcal{C}^\alpha(E,F)} \colon \mathbb{M}(E, F) \to [0, \infty] \tag{2.20}$$

the function which satisfies for all $f \in \mathbb{M}(E, F)$ that

$$|f|_{\mathcal{C}^\alpha(E,F)} = \sup\left(\left\{ \frac{d_F(f(x), f(y))}{|d_E(x, y)|_{\mathbb{R}}^\alpha} \in [0, \infty) \colon x, y \in E,\, x \neq y \right\} \cup \{0\}\right) \in [0, \infty] \tag{2.21}$$

and we denote by $\mathcal{C}^\alpha(E, F)$ the set given by

$$\mathcal{C}^\alpha(E, F) = \left\{ f \in \mathbb{M}(E, F) \colon\ |f|_{\mathcal{C}^\alpha(E,F)} < \infty \right\}. \tag{2.22}$$

Definition 2.1.9. *We say that f is d/δ-α-Hölder continuous (we say that f is α-Hölder continuous) if and only if there exist D and $\mathcal{D}$ such that*

(i) it holds that d is a metric on D,

(ii) it holds that δ is a metric on $\mathcal{D}$,

(iii) it holds that $\alpha \in (0, 1]$, and

(iv) it holds that $f \in \mathcal{C}^\alpha(D, \mathcal{D})$.

Class exercise 2.1.10. *Let $\alpha \in (0, 1]$ and let $f \colon [0, 2] \to \mathbb{R}$ be an α-Hölder continuous function. Is it true that $\sup_{x \in [0,2]} |f(x)| < \infty$?*

Class exercise 2.1.11. *Let $f \colon [0, 2] \to \mathbb{R}$ be a Lipschitz continuous function. Is it true that f is $1/2$-Hölder continuous?*

Exercise 2.1.12. *Let (E, d_E) and (F, d_F) be metric spaces and let $f \colon E \to F$ be a function. Prove that f is uniformly continuous if and only if*

$$\lim_{h \searrow 0} w_f(h) = w_f(0). \tag{2.23}$$

Exercise 2.1.13. *Let (E, d_E) and (F, d_F) be metric spaces, let $\alpha \in (0, 1]$, and let $f \colon E \to F$ be a function. Prove that*

$$|f|_{\mathcal{C}^\alpha(E,F)} = \sup_{h \in (0,\infty)} \left[\frac{w_f(h)}{h^\alpha} \right]. \tag{2.24}$$

Exercise 2.1.13, in particular, ensures that for all metric spaces (E, d_E) and (F, d_F) and all $\alpha \in (0,1]$, $f \in \mathbb{M}(E,F)$, $h \in (0,\infty)$ it holds that

$$w_f(h) \leq |f|_{\mathcal{C}^\alpha(E,F)}\, h^\alpha. \tag{2.25}$$

Definition 2.1.14 (Hölder continuity of derivatives)**.** *Let $k \in \mathbb{N}_0$, $l, d \in \mathbb{N}$, $\alpha \in (0,1]$, $a, b \in \mathbb{R}$ with $a < b$. Then we denote by $C^{k,\alpha}([a,b]^d, \mathbb{R}^l)$ the set given by*

$$C^{k,\alpha}([a,b]^d, \mathbb{R}^l) = \left\{ f \in C^k([a,b]^d, \mathbb{R}^l) \colon |f^{(k)}|_{\mathcal{C}^\alpha([a,b]^d, L^{(k)}(\mathbb{R}^d,\mathbb{R}^l))} < \infty \right\}. \tag{2.26}$$

There are a number of relations between the above introduced spaces. Some of them are illustrated in the following example.

Example 2.1.15. *Let $m, n \in \{2, 3, \dots\}$, $a, b \in \mathbb{R}$, $\alpha, \beta \in (0,1]$, $p, q \in [0,\infty)$ with $m < n$, $a < b$, $\alpha \leq \beta$ and $p \leq q$. Then*

$$\begin{split} &C^\infty([a,b],\mathbb{R}) \subseteq C^{n,\beta}([a,b],\mathbb{R}) \subseteq C^{n,\alpha}([a,b],\mathbb{R}) \subseteq C^{m,\beta}([a,b],\mathbb{R}) \\ &\subseteq C^{m,\alpha}([a,b],\mathbb{R}) \subseteq C^{1,\beta}([a,b],\mathbb{R}) \subseteq C^{1,\alpha}([a,b],\mathbb{R}) \subseteq C^1([a,b],\mathbb{R}) \\ &\subseteq C^{0,\beta}([a,b],\mathbb{R}) \subseteq C^{0,\alpha}([a,b],\mathbb{R}) \subseteq C([a,b],\mathbb{R}) \subseteq \mathcal{L}^q(B_{[a,b]}; |\cdot|_{\mathbb{R}}) \\ &\subseteq \mathcal{L}^p(B_{[a,b]}; |\cdot|_{\mathbb{R}}) \subseteq \mathcal{L}^0(B_{[a,b]}; |\cdot|_{\mathbb{R}}) = \mathcal{M}(\mathcal{B}([a,b]), \mathcal{B}(\mathbb{R})) \subseteq \mathbb{M}([a,b],\mathbb{R}). \end{split} \tag{2.27}$$

2.2 Deterministic numerical integration methods

In this section some basic deterministic methods for the approximative calculation of integrals of the form (2.2) are considered.

Definition 2.2.1 (Quadrature formula)**.** *We say that Q is a quadrature formula on A with quadrature nodes x and quadrature weights w if and only if there exist a natural number $d \in \mathbb{N}$ and a finite set I such that*

(i) it holds that $A \in \mathcal{B}(\mathbb{R}^d)$,

(ii) it holds that $x \in \mathbb{M}(I, A)$,

(iii) it holds that $w \in \mathbb{M}(I, \mathbb{R})$,

(iv) it holds that $Q \in \mathbb{M}(\mathcal{L}^1(B_A; |\cdot|_{\mathbb{R}}), \mathbb{R})$, and

(v) it holds for all $f \in \mathcal{L}^1(B_A; |\cdot|_{\mathbb{R}})$ that $Q[f] = \sum_{i \in I} w_i\, f(x_i)$.

Definition 2.2.2. *We say that Q is a quadrature formula if and only if there exist A, x, and w such that Q is a quadrature formula on A with quadrature nodes x and quadrature weights w.*

The quadrature nodes and the quadrature weights are typically chosen so that the quadrature formula $Q\colon \mathcal{L}^1(B_A; |\cdot|_{\mathbb{R}}) \to \mathbb{R}$ in Definition 2.2.1 is – in a suitable sense – a good approximation of the function

$$\mathcal{L}^1(B_A; |\cdot|_{\mathbb{R}}) \ni f \mapsto \int_A f(x)\,dx \in \mathbb{R}, \tag{2.28}$$

see Propositions 2.2.4 and 2.2.9 below for more details.

2.2.1 Rectangle method

Definition 2.2.3 (d-dimensional left rectangle method)**.** *Let $d \in \mathbb{N}$, $a, b \in \mathbb{R}$ with $a < b$. Then we denote by*

$$R^n_{[a,b]^d}\colon \mathcal{L}^1(B_{[a,b]^d}; |\cdot|_{\mathbb{R}}) \to \mathbb{R},\ n \in \mathbb{N}, \tag{2.29}$$

the functions which satisfy for all $n \in \mathbb{N}$, $f \in \mathcal{L}^1(B_{[a,b]^d}; |\cdot|_{\mathbb{R}})$ that

$$R^n_{[a,b]^d}[f] = \frac{(b-a)^d}{n^d}\left(\sum_{i_1,\dots,i_d \in \{0,1,\dots,n-1\}} f\big(a + \tfrac{i_1}{n}(b-a), \dots, a + \tfrac{i_d}{n}(b-a)\big)\right) \tag{2.30}$$

and we call the sequence $R^n_{[a,b]^d}$, $n \in \mathbb{N}$, the d-dimensional left rectangle method.

Observe that for every $d, n \in \mathbb{N}$, $a, b \in \mathbb{R}$ with $a < b$ it holds that $R^n_{[a,b]^d}$ is a quadrature formula on $[a,b]^d$ with quadrature nodes

$$\big(a + \tfrac{i_1}{n}(b-a), \dots, a + \tfrac{i_d}{n}(b-a)\big), \qquad (i_1, \dots, i_d) \in \{0, 1, \dots, n-1\}^d, \tag{2.31}$$

and quadrature weights

$$\tfrac{(b-a)^d}{n^d}, \qquad (i_1, \dots, i_d) \in \{0, 1, \dots, n-1\}^d. \tag{2.32}$$

Moreover, note that for all $a \in \mathbb{R}$, $b \in (a, \infty)$, $n \in \mathbb{N}$, $f \in \mathcal{L}^1(B_{[a,b]}; |\cdot|_{\mathbb{R}})$ it holds that

$$R^n_{[a,b]}[f] = \frac{(b-a)}{n}\left(\sum_{i=0}^{n-1} f\big(a + \tfrac{i}{n}(b-a)\big)\right). \tag{2.33}$$

In addition, observe that for all $d \in \mathbb{N}$, $a, b \in \mathbb{R}$ with $a < b$ and all $f\colon [a,b]^d \to \mathbb{R}$ with $\forall\, i = (i_1, \dots, i_d) \in \{0, 1, \dots, n-1\}^d$:

$$\begin{split}&\left[a + \frac{(b-a)i_1}{n}, a + \frac{(b-a)(i_1+1)}{n}\right) \times \dots \times \left[a + \frac{(b-a)i_d}{n}, a + \frac{(b-a)(i_d+1)}{n}\right) \\ &\subseteq f^{-1}\left(\left\{f\left(\begin{pmatrix} a \\ \vdots \\ a \end{pmatrix} + \tfrac{(b-a)i}{n}\right)\right\}\right)\end{split} \tag{2.34}$$

it holds that

$$R^n_{[a,b]^d}[f] = \int_{[a,b]^d} f(x)\,dx. \tag{2.35}$$

It thus holds for all $d \in \mathbb{N}$ that the d-dimensional left rectangle method is *exact* for all functions that are piecewise constant on the corresponding grid. The error of the d-dimensional left rectangle method is analyzed in the next proposition.

Proposition 2.2.4 (Error estimate for the d-dimensional left rectangle method)**.** *Let $d, n \in \mathbb{N}$, $\alpha \in (0,1]$, $a, b \in \mathbb{R}$ with $a < b$ and let $f \in \mathcal{L}^1(B_{[a,b]^d}; |\cdot|_{\mathbb{R}})$. Then*

$$\left| R^n_{[a,b]^d}[f] - \int_{[a,b]^d} f(x)\,dx \right|_{\mathbb{R}} \leq (b-a)^d\, w_f\big(\tfrac{(b-a)\sqrt{d}}{n}\big) \leq \frac{(b-a)^{(d+\alpha)}\, d^{\frac{\alpha}{2}}\, |f|_{\mathcal{C}^\alpha([a,b]^d,\mathbb{R})}}{n^\alpha}. \tag{2.36}$$

Proof of Proposition 2.2.4. Throughout this proof let $\mathbf{1} \in \mathbb{R}^d$ be the vector given by $\mathbf{1} = (1, 1, \ldots, 1)$. Note that

$$\begin{aligned}
&\left| R^n_{[a,b]^d}[f] - \int_{[a,b]^d} f(x)\,dx \right|_{\mathbb{R}} \\
&= \left| \frac{(b-a)^d}{n^d} \left[\sum_{i=(i_1,\ldots,i_d)\in\{0,1,\ldots,n-1\}^d} f\Big(a\mathbf{1} + \tfrac{(b-a)}{n} i\Big) \right] - \int_{[a,b]^d} f(x)\,dx \right|_{\mathbb{R}} \\
&= \Bigg| \sum_{\substack{i=(i_1,\ldots,i_d)\in \\ \{0,1,\ldots,n-1\}^d}} \int_{\left[a+\frac{(b-a)}{n}i_1, a+\frac{(b-a)}{n}(i_1+1)\right]\times\cdots\times\left[a+\frac{(b-a)}{n}i_d, a+\frac{(b-a)}{n}(i_d+1)\right]} f\Big(a\mathbf{1} + \tfrac{(b-a)}{n} i\Big)\,dx \\
&\quad - \int_{[a,b]^d} f(x)\,dx \Bigg|.
\end{aligned} \tag{2.37}$$

The triangle inequality and (2.19) therefore prove that

$$\begin{aligned}
&\left| R^n_{[a,b]^d}[f] - \int_{[a,b]^d} f(x)\,dx \right|_{\mathbb{R}} \\
&\leq \sum_{\substack{i=(i_1,\ldots,i_d)\in \\ \{0,1,\ldots,n-1\}^d}} \int_{\left[a+\frac{(b-a)}{n}i_1, a+\frac{(b-a)}{n}(i_1+1)\right]\times\cdots\times\left[a+\frac{(b-a)}{n}i_d, a+\frac{(b-a)}{n}(i_d+1)\right]} \left| f\Big(a\mathbf{1} + \tfrac{(b-a)}{n} i\Big) - f(x) \right|_{\mathbb{R}} dx \\
&\leq \sum_{\substack{i=(i_1,\ldots,i_d)\in \\ \{0,1,\ldots,n-1\}^d}} \int_{\left[a+\frac{(b-a)}{n}i_1, a+\frac{(b-a)}{n}(i_1+1)\right]\times\cdots\times\left[a+\frac{(b-a)}{n}i_d, a+\frac{(b-a)}{n}(i_d+1)\right]} w_f\Big(\big\|a\mathbf{1} + \tfrac{(b-a)}{n} i - x\big\|_{\mathbb{R}^d}\Big)\,dx.
\end{aligned} \tag{2.38}$$

Inequality (2.25) hence shows that

$$\begin{aligned}
&\left| R^n_{[a,b]^d}[f] - \int_{[a,b]^d} f(x)\,dx \right|_{\mathbb{R}} \\
&\leq \sum_{\substack{i=(i_1,\dots,i_d)\in\\ \{0,1,\dots,n-1\}^d}} \int_{\left[a+\frac{(b-a)}{n}i_1, a+\frac{(b-a)}{n}(i_1+1)\right]\times\dots\times\left[a+\frac{(b-a)}{n}i_d, a+\frac{(b-a)}{n}(i_d+1)\right]} w_f\Big(\big\|\tfrac{(b-a)}{n}\mathbf{1}\big\|_{\mathbb{R}^d}\Big)\,dx \\
&= \sum_{i\in\{0,1,\dots,n-1\}^d} \tfrac{(b-a)^d}{n^d}\cdot w_f\Big(\tfrac{(b-a)}{n}\|\mathbf{1}\|_{\mathbb{R}^d}\Big) = (b-a)^d\, w_f\Big(\tfrac{(b-a)\sqrt{d}}{n}\Big) \\
&\leq (b-a)^d\, |f|_{\mathcal{C}^\alpha([a,b]^d,\mathbb{R})} \left[\tfrac{(b-a)\sqrt{d}}{n}\right]^\alpha = \frac{(b-a)^{(d+\alpha)}\, d^{\frac{\alpha}{2}}\, |f|_{\mathcal{C}^\alpha([a,b]^d,\mathbb{R})}}{n^\alpha}.
\end{aligned} \tag{2.39}$$

This completes the proof of Proposition 2.2.4. □

Proposition 2.2.4 proves that for every $d \in \mathbb{N}$, $a, b \in \mathbb{R}$ with $a < b$, every $\alpha \in (0, 1]$, and every $f \in C^{0,\alpha}([a, b]^d, \mathbb{R})$ it holds that the approximation errors

$$\left| R^n_{[a,b]^d}[f] - \int_{[a,b]^d} f(x)\,dx \right|_{\mathbb{R}} \tag{2.40}$$

for $n \in \mathbb{N}$ of the d-dimensional left rectangle method converge with rate α to zero as n tends to infinity. The convergence rate of the d-dimensional left rectangle method does in general not improve if $f\colon [a, b] \to \mathbb{R}$ in (2.40) enjoys more differentiability regularity. This is illustrated in the next example.

Example 2.2.5. *Let $d \in \mathbb{N}$ and let $f\colon [0,1]^d \to \mathbb{R}$ be the function which satisfies for all $x_1, \ldots, x_d \in [0,1]$ that*

$$f(x_1, \ldots, x_d) = x_1. \tag{2.41}$$

Then we observe that f is infinitely often differentiable and we note that for all $n \in \mathbb{N}$ it holds that

$$\begin{aligned}
&\left| R^n_{[0,1]^d}[f] - \int_{[0,1]^d} f(x)\,dx \right|_{\mathbb{R}} \\
&= \int_{[0,1]^d} f(x)\,dx - \sum_{\substack{i=(i_1,\ldots,i_d)\in \\ \{0,1,\ldots,n-1\}^d}} \int_{\frac{i_1}{n}}^{\frac{i_1+1}{n}} \ldots \int_{\frac{i_d}{n}}^{\frac{i_d+1}{n}} f\left(\tfrac{i}{n}\right) dx_d \ldots dx_1 \\
&= \sum_{\substack{i=(i_1,\ldots,i_d)\in \\ \{0,1,\ldots,n-1\}^d}} \int_{\frac{i_1}{n}}^{\frac{i_1+1}{n}} \ldots \int_{\frac{i_d}{n}}^{\frac{i_d+1}{n}} \left[f(x) - f\left(\tfrac{i}{n}\right) \right] dx_d \ldots dx_1 \\
&= \sum_{\substack{i=(i_1,\ldots,i_d)\in \\ \{0,1,\ldots,n-1\}^d}} \int_{\frac{i_1}{n}}^{\frac{i_1+1}{n}} \ldots \int_{\frac{i_d}{n}}^{\frac{i_d+1}{n}} \left[x_1 - \tfrac{i_1}{n} \right] dx_d \ldots dx_1 \\
&= n^d \left[\frac{1}{n} \right]^{(d-1)} \int_0^{\frac{1}{n}} x\,dx = n \left[\frac{x^2}{2} \right]_{x=0}^{x=\frac{1}{n}} = \frac{n}{2n^2} = \frac{1}{2n}.
\end{aligned} \tag{2.42}$$

The sequence $|R^n_{[0,1]^d}[f] - \int_{[0,1]^d} f(x)\,dx|$, $n \in \mathbb{N}$, thus converges to zero with rate 1 but not with any higher rate.

Class exercise 2.2.6. *Does there exist a function $f \in \mathcal{L}^1(B_{[0,1]}; |\cdot|_{\mathbb{R}})$ and a real number $c \in \mathbb{R}$ such that for all $n \in \mathbb{N}$ it holds that*

$$\left| R^n_{[0,1]}[f] - \int_0^1 f(x)\,dx \right|_{\mathbb{R}} \leq \frac{c}{n^2}? \tag{2.43}$$

2.2.2 Trapezoidal rule

In Example 2.2.5 above it is illustrated that the convergence rate of the d-dimensional left rectangle method does in general not overcome the rate of convergence 1 even if $f\colon [a,b] \to \mathbb{R}$ in (2.40) is infinitely often differentiable. However, the rate of convergence does improve if $f\colon [a,b] \to \mathbb{R}$ in (2.40) is smooth and if another quadrature method is used. This is briefly illustrated in this subsection in the case of the 1-dimensional *trapezoidal method*.

Definition 2.2.7 (Trapezoidal method). *Let $a, b \in \mathbb{R}$ with $a < b$. Then we denote by $T^n_{[a,b]} \colon \mathcal{L}^1(B_{[a,b]}; |\cdot|_{\mathbb{R}}) \to \mathbb{R}$, $n \in \mathbb{N}$, the quadrature formulas which satisfy for all $n \in \mathbb{N}$, $f \in \mathcal{L}^1(B_{[a,b]}; |\cdot|_{\mathbb{R}})$ that*

$$\begin{aligned} T^n_{[a,b]}[f] &= \frac{(b-a)}{n}\left(\frac{f(a)+f(b)}{2} + \sum_{i=1}^{n-1} f\big(a + \tfrac{i}{n}(b-a)\big)\right) \\ &= \frac{(b-a)}{n}\left(\sum_{i=0}^{n-1} \frac{f\big(a+\tfrac{i}{n}(b-a)\big) + f\Big(a + \tfrac{(i+1)}{n}(b-a)\Big)}{2}\right) \end{aligned} \tag{2.44}$$

and we call the sequence $T^n_{[a,b]}$, $n \in \mathbb{N}$, the 1-dimensional trapezoidal method.

Error estimates for the 1-dimensional trapezoidal method are given in the next exercise and in Proposition 2.2.9 below.

Exercise 2.2.8. *Let $\alpha \in (0,1]$, $n \in \mathbb{N}$, $a, b \in \mathbb{R}$ with $a < b$ and let $f \in \mathcal{L}^1(B_{[a,b]}; |\cdot|_{\mathbb{R}})$. Prove that*

$$\left| T^n_{[a,b]}[f] - \int_a^b f(x)\,dx \right|_{\mathbb{R}} \leq (b-a) \cdot w_f\big(\tfrac{(b-a)}{2n}\big) \leq \frac{(b-a)^{(1+\alpha)}\,|f|_{\mathcal{C}^\alpha([a,b],\mathbb{R})}}{(2n)^\alpha}. \tag{2.45}$$

Proposition 2.2.9 (Error estimates for the trapezoidal method). *Let $n \in \mathbb{N}$, $\alpha \in (0,1]$, $a, b \in \mathbb{R}$ with $a < b$. Then*

(i) it holds for all $f \in \mathcal{L}^1(B_{[a,b]}; |\cdot|_{\mathbb{R}})$ that

$$\left| T^n_{[a,b]}[f] - \int_a^b f(x)\,dx \right|_{\mathbb{R}} \leq (b-a) \cdot w_f\big(\tfrac{(b-a)}{2n}\big) \leq \frac{(b-a)^{(1+\alpha)}\,|f|_{\mathcal{C}^\alpha([a,b],\mathbb{R})}}{(2n)^\alpha} \tag{2.46}$$

and

(ii) it holds for all $f \in C^1([a,b],\mathbb{R})$ that

$$\left| T^n_{[a,b]}[f] - \int_a^b f(x)\,dx \right|_{\mathbb{R}} \leq \tfrac{(b-a)^2}{n} \cdot w_{f'}\big(\tfrac{(b-a)}{2n}\big) \leq \frac{(b-a)^{(2+\alpha)}\,|f'|_{\mathcal{C}^\alpha([a,b],\mathbb{R})}}{2^\alpha\, n^{(1+\alpha)}}. \tag{2.47}$$

Proof of Proposition 2.2.9. Estimate (2.46) follows from Exercise 2.2.8. It thus remains to show (2.47) to complete the proof of Proposition 2.2.9. For this observe that for all

$f \in C^1([a,b],\mathbb{R})$ it holds that

$$
\begin{aligned}
&T^n_{[a,b]}[f] - \int_a^b f(x)\,dx \\
&= \sum_{i=0}^{n-1} \int_{a+\frac{i(b-a)}{n}}^{a+\frac{(i+1)(b-a)}{n}} \left[\frac{f\big(a+\frac{i(b-a)}{n}\big) + f\big(a+\frac{(i+1)(b-a)}{n}\big)}{2} - f(x) \right] dx \\
&= \sum_{i=0}^{n-1} \int_{a+\frac{i(b-a)}{n}}^{a+\frac{(i+1)(b-a)}{n}} \left[f\big(a+\tfrac{i(b-a)}{n}\big) + \frac{1}{2}\int_{a+\frac{i(b-a)}{n}}^{a+\frac{(i+1)(b-a)}{n}} f'(y)\,dy \right] dx \\
&- \sum_{i=0}^{n-1} \int_{a+\frac{i(b-a)}{n}}^{a+\frac{(i+1)(b-a)}{n}} \left[f\big(a+\tfrac{i(b-a)}{n}\big) + \int_{a+\frac{i(b-a)}{n}}^{x} f'(y)\,dy \right] dx \\
&= \sum_{i=0}^{n-1} \int_{a+\frac{i(b-a)}{n}}^{a+\frac{(i+1)(b-a)}{n}} \left[\frac{1}{2}\int_{a+\frac{i(b-a)}{n}}^{a+\frac{(i+1)(b-a)}{n}} f'(y) - f'\big(a+\tfrac{(i+1/2)(b-a)}{n}\big)\,dy \right] dx \\
&- \sum_{i=0}^{n-1} \int_{a+\frac{i(b-a)}{n}}^{a+\frac{(i+1)(b-a)}{n}} \left[\int_{a+\frac{i(b-a)}{n}}^{x} f'(y) - f'\big(a+\tfrac{(i+1/2)(b-a)}{n}\big)\,dy \right] dx.
\end{aligned}
\tag{2.48}
$$

Therefore, we obtain that for all $f \in C^1([a,b],\mathbb{R})$ it holds that

$$
\begin{aligned}
&\left| T^n_{[a,b]}[f] - \int_a^b f(x)\,dx \right|_{\mathbb{R}} \\
&\leq \frac{1}{2}\sum_{i=0}^{n-1} \int_{a+\frac{i(b-a)}{n}}^{a+\frac{(i+1)(b-a)}{n}} \int_{a+\frac{i(b-a)}{n}}^{a+\frac{(i+1)(b-a)}{n}} \left| f'(y) - f'\big(a+\tfrac{(i+1/2)(b-a)}{n}\big)\right|_{\mathbb{R}} dy\,dx \\
&+ \sum_{i=0}^{n-1} \int_{a+\frac{i(b-a)}{n}}^{a+\frac{(i+1)(b-a)}{n}} \int_{a+\frac{i(b-a)}{n}}^{x} \left| f'(y) - f'\big(a+\tfrac{(i+1/2)(b-a)}{n}\big)\right|_{\mathbb{R}} dy\,dx \\
&\leq \tfrac{(b-a)^2}{n} \cdot w_{f'}\big(\tfrac{(b-a)}{2n}\big).
\end{aligned}
\tag{2.49}
$$

Combining this with (2.25) proves (2.47). The proof of Proposition 2.2.9 is thus completed. □

Exercise 2.2.10. *Prove or disprove the following statement: For all infinitely often differentiable functions $f\colon [0,1] \to \mathbb{R}$ it holds that*

$$
\inf_{n\in\mathbb{N}} \left(n^2 \left| T^n_{[0,1]}[f] - \textstyle\int_0^1 f(x)\,dx \right| \right) = 0. \tag{2.50}
$$

Class exercise 2.2.11. *Let $a,b \in \mathbb{R}$ with $a < b$. Does there exists a function $f \in \mathcal{L}^2(B_{[a,b]};|\cdot|_{\mathbb{R}})$ such that*

$$
\Big[\inf\nolimits_{n\in\mathbb{N}} \left| R^n_{[a,b]}[f] - \textstyle\int_a^b f(x)\,dx \right| > 0 < \inf\nolimits_{n\in\mathbb{N}} \left| T^n_{[a,b]}[f] - \textstyle\int_a^b f(x)\,dx \right| \Big]? \tag{2.51}
$$

2.2.3 Curse of dimensionality

In the next step we analyze the number of function evaluations needed to compute the multi-dimensional rectangle method. For this let $d \in \mathbb{N}$, $\alpha \in (0,1]$, $a, b \in \mathbb{R}$ with $a < b$ and let $f \in C^{0,\alpha}([a,b]^d, \mathbb{R})$. Then observe that for every $n \in \mathbb{N}$ it holds that the computation of

$$R^n_{[a,b]^d}[f] \tag{2.52}$$

requires $N = n^d \in \mathbb{N}$ evaluations of the function f. Moreover, we get from Proposition 2.2.4 that for all $N \in \{1^d, 2^d, 3^d, \dots\}$ it holds that

$$\left| R^{N^{1/d}}_{[a,b]^d}[f] - \int_{[a,b]^d} f(x)\,dx \right|_{\mathbb{R}} \leq \frac{(b-a)^{(d+\alpha)}\, d^{\frac{\alpha}{2}}\, |f|_{\mathcal{C}^{\alpha}([a,b]^d,\mathbb{R})}}{N^{\frac{\alpha}{d}}}. \tag{2.53}$$

The quantity $R^{N^{1/d}}_{[a,b]^d}[f]$ thus converges to $\int_{[a,b]^d} f(x)\,dx$ with order $\frac{\alpha}{d}$ as the number of function evaluations $N \in \{1^d, 2^d, 3^d, \dots\}$ tends to infinity. Note that if $d \in \mathbb{N}$ increases, then the convergence order $\frac{\alpha}{d} \leq \frac{1}{d}$ decreases. We have thus sketched that the convergence speed of quadrature formulas such as (2.30) and (2.44) may be very poor

(i) if the dimension $d \in \mathbb{N}$ is large (see (2.53)) or

(ii) if the integrand function $f\colon [a,b]^d \to \mathbb{R}$ has low regularity properties (see, e.g, (2.36) and (2.46)).

In both cases, Monte Carlo methods provide a *competitive* alternative. This is illustrated in the next sections.

Remark 2.2.12 (Sparse grids). *A deterministic numerical approximation method which can be used in moderately high dimensional problems can, e.g., be found in [Garcke(2008)] and in the references mentioned therein.*

Exercise 2.2.13.

(i) Write a MATLAB function **`RecRule(a,b,d,n,f)`** *with input $a \in \mathbb{R}$, $b \in (a,\infty)$, $d, n \in \mathbb{N}$, $f\colon [a,b]^d \to \mathbb{R} \in \mathcal{L}^1(B_{[a,b]^d}; |\cdot|_{\mathbb{R}})$ and output $R^n_{[a,b]^d}[f]$.*

(ii) Let I be the set given by

$$I = \big(\{1,2,\dots,8\} \times \{1,2,\dots,10\}\big) \cup \big(\{9\} \times \{1,2,\dots,8\}\big) \cup \big(\{10\} \times \{1,2,\dots,6\}\big). \tag{2.54}$$

Test your Matlab *function* **`RecRule(a,b,d,n,f)`** *in the cases $a = 0$, $b = 2$, $(d,n) \in I$, and $f = [0,2]^d \ni (x_1,\dots,x_d) \mapsto x_1 \in \mathbb{R}$ and measure the runtime of your* Matlab *function* **`RecRule(a,b,d,n,f)`** *in these cases.*

2.3 Monte Carlo methods

Let $(A, \mathcal{A}, \mu)$ be a finite measure space with $\mu(A) \neq 0$ and let $f \in \mathcal{L}^1(\mu; |\cdot|_{\mathbb{R}})$. This section presents the Monte Carlo method for the approximative computation of the real number

$$\int_A f(x)\, \mu(dx) \in \mathbb{R}. \tag{2.55}$$

It is based *on the interpretation of* (2.55) *as an expectation of a random variable*. More precisely, let $(\Omega, \mathcal{F}, P)$ be a probability space and let $Y \colon \Omega \to A$ be an $\mathcal{F}/\mathcal{A}$-measurable function which satisfies for all $B \in \mathcal{A}$ that

$$\big(Y(P)_{\mathcal{A}}\big)(B) = \frac{\mu(B)}{\mu(A)}. \tag{2.56}$$

(Observe that such a probability space and such a random variable do indeed exist. For example, define $(\Omega, \mathcal{F}) := (A, \mathcal{A})$, define $P \colon \mathcal{F} \to [0, \infty]$ through $P(B) := \frac{\mu(B)}{\mu(A)}$ for all $B \in \mathcal{F}$ and define $Y \colon \Omega \to A$ through $Y(\omega) := \omega$ for all $\omega \in \Omega$.) Next let $X \colon \Omega \to \mathbb{R}$ be given by

$$X = \mu(A) \cdot f(Y) \tag{2.57}$$

and observe that (2.1) reduces to

$$\begin{aligned}
&\int_A f(x)\, \mu(dx) = \mu(A) \cdot \int_A \frac{f(x)}{\mu(A)}\, \mu(dx) = \mu(A) \cdot \int_A f(x)\, \big(Y(P)_{\mathcal{A}}\big)(dx) \\
&= \mathbb{E}_P\big[\mu(A) \cdot f(Y)\big] = \mathbb{E}_P\big[X\big] \in \mathbb{R}.
\end{aligned} \tag{2.58}$$

The Monte Carlo method then uses realizations of the random variable X to obtain an approximation of $\mathbb{E}_P\big[X\big]$ and thereby produces an approximation of the quantity

$$\int_A f(x)\, \mu(dx) = \mathbb{E}_P\big[X\big] \tag{2.59}$$

which we actually want to approximate. Let us formulate this more precisely in the next definitions.

Definition 2.3.1 (A Monte Carlo approximation with order n)**.** *Let $c \in \mathbb{R}$. Then we say that X is an n-Monte Carlo approximation of c on $\mathbf{\Omega}$ (we say that X is a Monte Carlo approximation of c on $\mathbf{\Omega}$ with order n, we say that X is an n-Monte Carlo approximation of c, we say that X is a Monte Carlo approximation of c with order n) if and only if there exist Ω, $\mathcal{F}$, P such that it holds*

(i) that $\mathbf{\Omega} = (\Omega, \mathcal{F}, P)$ is a probability space,

(ii) that $X \in \mathcal{L}^1(P; |\cdot|_{\mathbb{R}})$,

(iii) that $n \in \mathbb{N}$, and

(iv) that there exist exist $Z_1, Z_2, \ldots, Z_n \in \mathcal{L}^1(P; |\cdot|_{\mathbb{R}})$ such that Z_1, Z_2, $\ldots$, Z_n are P-independent and such that

$$\mathbb{E}_P[Z_1] = c, \quad Z_1(P)_{\mathcal{B}(\mathbb{R})} = \ldots = Z_n(P)_{\mathcal{B}(\mathbb{R})}, \quad \text{and} \quad X = \frac{Z_1 + \ldots + Z_n}{n}. \tag{2.60}$$

Definition 2.3.2 (A Monte Carlo approximation)**.** *Let $c \in \mathbb{R}$. Then we say that X is a Monte Carlo approximation of c on $\mathbf{\Omega}$ (we say that X is a Monte Carlo approximation of c, we say that X is a Monte Carlo approximation) if and only if there exist $n \in \mathbb{N}$ such that X is an n-Monte Carlo approximation of c on $\mathbf{\Omega}$.*

Definition 2.3.3 (A Monte Carlo approximation sequence)**.** *Let $c \in \mathbb{R}$. Then we say that $\mathbf{X}$ is a Monte Carlo approximation sequence of c on $\mathbf{\Omega}$ (we say that $\mathbf{X}$ is a Monte Carlo approximation sequence of c) if and only if there exist Ω, $\mathcal{F}$, P such that it holds*

(i) that $\mathbf{\Omega} = (\Omega, \mathcal{F}, P)$ is a probability space,

(ii) that $\mathbf{X} \in \mathbb{M}(\mathbb{N}, \mathcal{L}^1(P; |\cdot|_{\mathbb{R}}))$, and

(iii) that there exists a sequence $Z_n \in \mathcal{L}^1(P; |\cdot|_{\mathbb{R}})$, $n \in \mathbb{N}$, such that Z_n, $n \in \mathbb{N}$, are P-independent and

$$\forall\, n \in \mathbb{N}\colon \quad \mathbb{E}_P[Z_1] = c, \quad Z_1(P)_{\mathcal{B}(\mathbb{R})} = Z_n(P)_{\mathcal{B}(\mathbb{R})}, \quad \text{and} \quad \mathbf{X}_n = \tfrac{Z_1 + \ldots + Z_n}{n}. \tag{2.61}$$

Let us illustrate this definition through the following simple example (see, e.g., Section 21.4 in [Higham(2004)] for a similar example).

Example 2.3.4. *Suppose in this example that we want to compute the integral*

$$\int_{-1}^{1} \exp\Big(\sqrt{|x|}\Big)\, dx. \tag{2.62}$$

In view of (2.58)*, we rewrite* (2.62) *by*

$$\int_{-1}^{1} \exp\Big(\sqrt{|x|_{\mathbb{R}}}\Big)\, dx = \frac{1}{2} \int_{-1}^{1} 2\, \exp\Big(\sqrt{|x|}\Big)\, dx = \int_{\mathbb{R}} 2\, \exp\Big(\sqrt{|x|}\Big)\, \mathcal{U}_{(-1,1)}(dx). \tag{2.63}$$

Next let $(\Omega, \mathcal{F}, P)$ be a probabilty space and let $Y_n \colon \Omega \to \mathbb{R}$, $n \in \mathbb{N}$, be a sequence of independent $\mathcal{U}_{(-1,1)}$-distributed random variables. Then it holds that

$$\frac{2}{N}\left(e^{\sqrt{|Y_1|}} + \ldots + e^{\sqrt{|Y_N|}}\right) \tag{2.64}$$

for $N \in \mathbb{N}$ is a Monte Carlo approximation sequence of $\int_{-1}^{1} \exp\Big(\sqrt{|x|}\Big)\, dx$. If $U_n \colon \Omega \to \mathbb{R}$, $n \in \mathbb{N}$, are independent $\mathcal{U}_{(0,1)}$-distributed random variables, then (2.64) suggests in the case $N = 100$ the following algorithm.

Monte Carlo approximations

Output: *Realization x of $X \sim \big(\frac{1}{50}[\exp(\sqrt{|Y_1|}) + \ldots + \exp(\sqrt{|Y_{100}|})]\big)(P)_{\mathcal{B}(\mathbb{R})} \approx \int_{-1}^{1} \exp(\sqrt{|x|})\, dx$*

$s = 0$

for $n = 1 \to 100$ ***do***

  Generate realization u of $U_n \sim \mathcal{U}_{(0,1)}$

  $y = 2u - 1$

  $s = s + \exp(\sqrt{|y|})$

end for

$x = s/50$

In Matlab the above algorithm can be implemented through the command "`sum(exp(sqrt(abs(2*rand(1,100)-1))))/50`".

2.3.1 Bias of an estimator

In this subsection we investigate a certain property of Monte Carlo approximations. For this the following definition is used.

Definition 2.3.5 (Bias)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $c \in \mathbb{R}$ be a real number, and let $X \in \mathcal{L}^1(P; |\cdot|_{\mathbb{R}})$. Then we denote by $\operatorname{Bias}_{P,c}(X)$ the real number given by*

$$\operatorname{Bias}_{P,c}(X) = \mathbb{E}_P\big[X\big] - c \tag{2.65}$$

*and we call $\operatorname{Bias}_{P,c}(X)$ the P-*bias of X with respect to c.

Definition 2.3.6 (Unbiased)**.** *Let $c \in \mathbb{R}$. Then we say that X is P-unbiased with respect to c (we say that X is P-unbiased for c, we say that X is unbiased with respect to c, we say that X is unbiased for c) if and only if it holds*

(i) that P is a probability measure,

(ii) that $X \in \mathcal{L}^1(P; |\cdot|_{\mathbb{R}})$, and

(iii) that $\operatorname{Bias}_{P,c}(X) = 0$.

Definition 2.3.7 (Biased)**.** *Let $c \in \mathbb{R}$. Then we say that X is P-biased with respect to c (we say that X is P-biased for c, we say that X is biased with respect to c, we say that X is biased for c) if and only if it holds*

(i) that P is a probability measure,

(ii) that $X \in \mathcal{L}^1(P; |\cdot|_{\mathbb{R}})$, and

(iii) that $\operatorname{Bias}_{P,c}(X) \neq 0$.

Lemma 2.3.8 (Unbiasedness of Monte Carlo approximations)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $X_n \in \mathcal{L}^1(P; |\cdot|_{\mathbb{R}})$, $n \in \mathbb{N}$, be i.i.d. random variables, and let $N \in \mathbb{N}$. Then it holds that $\frac{1}{N}(X_1 + \ldots + X_N)$ is P-unbiased with respect to $\mathbb{E}_P[X_1]$, i.e., it holds that*

$$\mathbb{E}_P\big[\tfrac{1}{N}(X_1 + \ldots + X_N)\big] = \mathbb{E}_P\big[X_1\big]. \tag{2.66}$$

Lemma 2.3.8 follows immediately from the linearity of the expectation.

Exercise 2.3.9. *Let $A, B \subseteq \mathbb{R}^2$ be the sets given by*

$$A = \left\{(x, y) \in \mathbb{R}^2 \colon\ |x-2|_{\mathbb{R}}^2 + y^2 \leq 4\right\}, \tag{2.67}$$

$$B = \left\{(x, y) \in \mathbb{R}^2 \colon x^2 + |y-2|_{\mathbb{R}}^2 \leq 4\right\}, \tag{2.68}$$

let $f \colon \mathbb{R}^2 \to \mathbb{R}$ be the function with the property that for all $x, y \in \mathbb{R}$ it holds that

$$f(x, y) = \mathbb{1}_{(A \cap B)}(x, y) \cdot |x|_{\mathbb{R}}^{2/3}, \tag{2.69}$$

let $(\Omega, \mathcal{F}, P)$ be a probability space, let $Y_n, Z_n \colon \Omega \to \mathbb{R}$, $n \in \mathbb{N}$, be independent $\mathcal{U}_{(0,1)}$-distributed random variables, and let $I_N \colon \Omega \to \mathbb{R}$, $N \in \mathbb{N}$, be the functions with the property that for all $N \in \mathbb{N}$ it holds that

$$I_N = \frac{4}{N}\left[\sum_{n=1}^{N} f(2Y_n, 2Z_n)\right]. \tag{2.70}$$

The random variables I_N, $N \in \mathbb{N}$, are thus Monte Carlo approximations of $\mathbb{E}_P\big[4f(2Y_1, 2Z_1)\big]$.

(i) Prove or disprove the following statement: I_N, $N \in \mathbb{N}$, are P-unbiased with respect to $\int_0^2 \int_0^2 f(x, y)\, dx\, dy$.

(ii) Write a Matlab *function `MonteCarlo(N)` with input $N \in \mathbb{N}$ and output a realization of I_N.*

(iii) Write a Matlab *function `MonteCarloPlot()` which plots for every $k \in \{2, 3, 4, 5, 6\}$ five realizations of I_{10^k}, each marked by a blue star, in a coordinate plane. Plot $k \in \{2, 3, 4, 5, 6\}$ on the x-axis and realizations of I_{10^k} on the y-axis. N.B.: Your plot should thus contain a total of 25 blue stars.*

Exercise 2.3.10. *In this exercise we do not distinguish between pseudo random numbers and actual random numbers. Let* $d \in \mathbb{N}$, $a \in \mathbb{R}$, $b \in (a, \infty)$, $f \in \mathcal{L}^1(B_{[a,b]^d}; |\cdot|_{\mathbb{R}})$, *let* $(\Omega, \mathcal{F}, P)$ *be a probability space, let* $X_j \colon \Omega \to \mathbb{R}^d$, $j \in \mathbb{N}$, *be independent* $\mathcal{U}_{[a,b]^d}$*-distributed random variables with* $\forall\, j \in \mathbb{N} \colon X_j(\Omega) \subseteq [a,b]^d$, *and let* $I_N \colon \Omega \to \mathbb{R}$, $N \in \mathbb{N}$, *be the functions with the property that for all* $N \in \mathbb{N}$ *it holds that*

$$I_N = \frac{(b-a)^d}{N} \left[\sum_{j=1}^{N} f(X_j) \right]. \tag{2.71}$$

Write a Matlab *function* **`intMC(a,b,d,f,N)`** *with input* $a \in \mathbb{R}$, $b \in (a, \infty)$, $d \in \mathbb{N}$, $f \in \mathcal{L}^1(B_{[a,b]^d}; |\cdot|_{\mathbb{R}})$, $N \in \mathbb{N}$ *and output a realization of* I_N. *Test your* Matlab *function* **`intMC(a,b,d,f,N)`** *in the cases* $a = 0$, $b = 2$, $(d, N) \in \cup_{l \in \{5,10\}} \cup_{k \in \{3,5,7\}} \{(k, l^k)\}$, $f = [a,b]^d \ni x = (x_1, \ldots, x_d) \mapsto x_1 \in \mathbb{R}$.

2.4 Error analysis of the Monte Carlo method

2.4.1 Consistency of the Monte Carlo method

This section presents consistency properties for Monte Carlo approximations.

Definition 2.4.1 (Consistency)**.** *Let* $c \in \mathbb{R}$. *We say that* X *is* P*-consistent for* c *(we say that* X *is consistent for* c*, we say that* X *is consistent) if and only if there exist a measurable space* $(\Omega, \mathcal{F})$ *such that it holds*

(i) that P *is a probability measure on* $(\Omega, \mathcal{F})$,

(ii) that $X \in \mathbb{M}(\mathbb{N}, \mathcal{M}(\mathcal{F}, \mathcal{B}(\mathbb{R})))$, *and*

(iii) that

$$\forall\, \varepsilon \in (0, \infty) \colon \qquad \limsup_{n \to \infty} P(|X_n - c|_{\mathbb{R}} \geq \varepsilon) = 0. \tag{2.72}$$

Definition 2.4.2 (Strong consistency)**.** *Let* $c \in \mathbb{R}$. *We say that* X *is strongly* P*-consistent for* c *(we say that* X *is strongly consistent for* c*, we say that* X *is strongly consistent) if and only if there exist a measurable space* $(\Omega, \mathcal{F})$ *such that it holds*

(i) that P *is a probability measure on* $(\Omega, \mathcal{F})$,

(ii) that $X \in \mathbb{M}(\mathbb{N}, \mathcal{M}(\mathcal{F}, \mathcal{B}(\mathbb{R})))$, *and*

(iii) that

$$\forall\, \varepsilon \in (0, \infty) \colon \qquad P\Big(\limsup_{n \to \infty} |X_n - c|_{\mathbb{R}} \geq \varepsilon \Big) = 0. \tag{2.73}$$

The convergence property in (2.73) is equivalent to almost sure convergence. This is the subject of the next lemma.

Lemma 2.4.3 (A characterization for almost sure convergence)**.** *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space and let* $X_n \colon \Omega \to \mathbb{R}$, $n \in \mathbb{N}_0$, *be* $\mathcal{F}/\mathcal{B}(\mathbb{R})$*-measurable functions. Then*

$$\begin{aligned}
&\sup_{\varepsilon\in(0,\infty)} P\bigg(\limsup_{n\to\infty} |X_n - X_0|_{\mathbb{R}} \geq \varepsilon\bigg) = \sup_{\varepsilon\in(0,\infty)} P\bigg(\limsup_{n\to\infty} |X_n - X_0|_{\mathbb{R}} > \varepsilon\bigg) \\
&= \lim_{\varepsilon\searrow 0} P\bigg(\limsup_{n\to\infty} |X_n - X_0|_{\mathbb{R}} \geq \varepsilon\bigg) = \lim_{\varepsilon\searrow 0} P\bigg(\limsup_{n\to\infty} |X_n - X_0|_{\mathbb{R}} > \varepsilon\bigg) \\
&= P\bigg(\limsup_{n\to\infty} |X_n - X_0|_{\mathbb{R}} > 0\bigg).
\end{aligned} \tag{2.74}$$

Lemma 2.4.3 is an immediate consequence from the fact that probability measures are continuous from below. The next lemma provides a well-known relation between consistency and strong consistency.

Lemma 2.4.4 (Almost sure convergence implies convergence in probability)**.** *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space and let* $X_n \colon \Omega \to \mathbb{R}$, $n \in \mathbb{N}_0$, *be* $\mathcal{F}/\mathcal{B}(\mathbb{R})$*-measurable functions which satisfy that*

$$P\bigg(\limsup_{n\to\infty} |X_n - X_0|_{\mathbb{R}} = 0\bigg) = 1. \tag{2.75}$$

Then it holds for all $\varepsilon \in (0, \infty)$ *that*

$$\limsup_{n\to\infty} P(|X_n - X_0|_{\mathbb{R}} \geq \varepsilon) = \limsup_{n\to\infty} P(|X_n - X_0|_{\mathbb{R}} > \varepsilon) = 0. \tag{2.76}$$

Proof of Lemma 2.4.4. First, note that (2.75) ensures that

$$P\bigg(\limsup_{n\to\infty} |X_n - X_0|_{\mathbb{R}} > 0\bigg) = 0 \tag{2.77}$$

This and Lemma 2.4.3 imply that for all $\varepsilon \in (0, \infty)$ it holds that

$$P\bigg(\limsup_{n\to\infty} |X_n - X_0|_{\mathbb{R}} > \varepsilon\bigg) = 0. \tag{2.78}$$

The fact that P is continuous from above and monotonicity of P hence establish that for all $\varepsilon \in (0, \infty)$ it holds that

$$\begin{aligned}
\limsup_{n\to\infty} P(|X_n - X_0|_{\mathbb{R}} > \varepsilon) &\leq \limsup_{n\to\infty} P\big(\cup_{m\in\mathbb{N}\cap[n,\infty)} \{|X_m - X_0|_{\mathbb{R}} > \varepsilon\}\big) \\
&= P\big(\cap_{n\in\mathbb{N}} \cup_{m\in\mathbb{N}\cap[n,\infty)} \{|X_m - X_0|_{\mathbb{R}} > \varepsilon\}\big) \\
&= P\bigg(\limsup_{n\to\infty} |X_n - X_0|_{\mathbb{R}} > \varepsilon\bigg) = 0.
\end{aligned} \tag{2.79}$$

The proof of Lemma 2.4.4 is thus completed. □

The next theorem, known as *strong law of large numbers*, proves that Monte Carlo approximations are strongly consistent and therefore, in particular, shows that Monte Carlo approximations are consistent (see Lemma 2.4.4 above).

Theorem 2.4.5 (Strong law of large numbers)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $X_n \in \mathcal{L}^1(P; |\cdot|_{\mathbb{R}})$, $n \in \mathbb{N}$, be i.i.d. random variables. Then the sequence $\frac{1}{N}(X_1 + \ldots + X_N)$, $N \in \mathbb{N}$, converges P-almost surely to $\mathbb{E}_P[X_1]$, i.e., it holds that*

$$P\left(\limsup_{N\to\infty} \left| \left[\frac{X_1 + \ldots + X_N}{N}\right] - \mathbb{E}_P\big[X_1\big] \right|_{\mathbb{R}} = 0\right) = 1. \tag{2.80}$$

In other words, the Monte Carlo approximation sequence $\frac{1}{N}(X_1 + \ldots + X_N)$, $N \in \mathbb{N}$, of $\mathbb{E}_P\big[X_1\big]$ is strongly P-consistent for $\mathbb{E}_P\big[X_1\big]$.

Theorem 2.4.5 is, for example, proved as Theorem 5.12 in [Klenke(2008)].

2.4.2 Root mean square error of the Monte Carlo method

Lemma 2.4.6 (Variance of a finite sum of random variables)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probabilty space, let $N \in \mathbb{N}$, and let $X_1, \ldots, X_N \in \mathcal{L}^2(P; |\cdot|_{\mathbb{R}})$. Then*

$$\mathsf{Var}_P\left(\sum_{i=1}^N X_i\right) = \sum_{i,j=1}^N \mathsf{Cov}_P(X_i, X_j) = \sum_{i=1}^N \mathsf{Var}_P(X_i) + \sum_{\substack{i,j \in \{1,\ldots,N\},\\ i \neq j}} \mathsf{Cov}_P(X_i, X_j)\,. \tag{2.81}$$

Proof of Lemma 2.4.6. Note that

$$\begin{aligned} \mathsf{Var}_P\left(\sum_{i=1}^N X_i\right) &= \mathbb{E}_P\left[\left(\sum_{i=1}^N (X_i - \mathbb{E}_P[X_i])\right)^2\right] \\ = \sum_{i,j=1}^N \mathbb{E}_P\Big[\big(X_i - \mathbb{E}_P[X_i]\big)\big(X_j - \mathbb{E}_P[X_j]\big)\Big] &= \sum_{i,j=1}^N \mathsf{Cov}_P(X_i, X_j)\,. \end{aligned} \tag{2.82}$$

This completes the proof of Lemma 2.4.6. □

Corollary 2.4.7 (Variance of a finite sum of uncorrelated random variables)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probabilty space, let $N \in \mathbb{N}$, and let $X_1, \ldots, X_N \in \mathcal{L}^2(P; |\cdot|_{\mathbb{R}})$ be pairwise P-uncorrelated. Then*

$$\mathsf{Var}_P(X_1 + \ldots + X_N) = \mathsf{Var}_P(X_1) + \ldots + \mathsf{Var}_P(X_N)\,. \tag{2.83}$$

Using Corollary 2.4.7, we now analyze the mean square error of Monte Carlo approximations.

Theorem 2.4.8 (Root mean square error of the Monte Carlo method)**.** *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space and let* $X_n \in \mathcal{L}^2(P; |\cdot|_{\mathbb{R}})$*,* $n \in \mathbb{N}$*, be i.i.d. random variables. Then it holds for all* $N \in \mathbb{N}$ *that*

$$\left\| \mathbb{E}_P\big[X_1\big] - \frac{X_1 + \ldots + X_N}{N} \right\|_{\mathcal{L}^2(P;|\cdot|_{\mathbb{R}})} = \frac{\sqrt{\mathsf{Var}_P(X_1)}}{\sqrt{N}}. \tag{2.84}$$

Proof of Theorem 2.4.8. Lemma 2.3.8 and Corollary 2.4.7 imply that for all $N \in \mathbb{N}$ it holds that

$$\begin{aligned} \mathbb{E}_P\left[\left|\mathbb{E}_P\big[X_1\big] - \frac{X_1 + \ldots + X_N}{N}\right|_{\mathbb{R}}^2\right] &= \mathsf{Var}_P\left(\frac{X_1 + \ldots + X_N}{N}\right) \\ = \frac{\mathsf{Var}_P(X_1) + \ldots + \mathsf{Var}_P(X_N)}{N^2} &= \frac{N \cdot \mathsf{Var}_P(X_1)}{N^2} = \frac{\mathsf{Var}_P(X_1)}{N}. \end{aligned} \tag{2.85}$$

This completes the proof of Theorem 2.4.8. ☐

Theorem 2.4.8, in particular, proves that the Monte Carlo approximations $\frac{1}{N}\big(X_1 + \ldots + X_N\big)$, $N \in \mathbb{N}$, of $\mathbb{E}_P\big[X_1\big]$ converge in the root mean square sense with order $\frac{1}{2}$ to the real number $\mathbb{E}_P\big[X_1\big]$ and that the constant appearing in the error estimate (2.84) is the standard deviation $\sqrt{\mathsf{Var}_P(X_1)}$ of X_1.

Remark 2.4.9. *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space and let* $X_n \in \mathcal{L}^0(P; |\cdot|_{\mathbb{R}})$*,* $n \in \mathbb{N}$*, be i.i.d. random variables. Then observe that Theorem 2.4.8 is applicable under the assumption that for all* $n \in \mathbb{N}$ *it holds that*

$$X_n \in \mathcal{L}^2(P; |\cdot|_{\mathbb{R}}) \subseteq \mathcal{L}^1(P; |\cdot|_{\mathbb{R}}) \tag{2.86}$$

while Theorem 2.4.5 is applicable under the assumption that for all $n \in \mathbb{N}$ *it holds that*

$$X_n \in \mathcal{L}^1(P; |\cdot|_{\mathbb{R}}). \tag{2.87}$$

Exercise 2.4.10. *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space, let* $f \in \mathcal{M}(\mathcal{B}(\mathbb{R}), \mathcal{B}(\mathbb{R}))$ *be a bounded function, and let* $U_n \in \mathcal{M}(\mathcal{F}, \mathcal{B}(\mathbb{R}))$*,* $n \in \mathbb{N}$*, be independent* $\mathcal{U}_{(-1,1)}$*-distributed random variables. Prove or disprove the following statement: It holds that*

$$\left(\mathbb{E}_P\left[\left|\frac{f(U_1) + \ldots + f(U_{5000})}{2500} - \int_{-1}^{1} f(x)\,dx\right|_{\mathbb{R}}^2\right]\right)^{1/2} \leq \frac{\sup_{x\in\mathbb{R}} |f(x)|_{\mathbb{R}}}{30}. \tag{2.88}$$

2.4.3 Markov's and Chebyshev's inequality

In this subsection the Markov inequality (see Lemma 2.4.11 below) and the Chebyshev inequality (see Corollary 2.4.12) are presented.

Lemma 2.4.11 (Markov inequality)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, let $\varepsilon \in (0, \infty)$, and let $X \colon \Omega \to [0, \infty)$ be an $\mathcal{F}/\mathcal{B}([0, \infty))$-measurable function. Then*

$$\mu\big(X \geq \varepsilon\big) \leq \frac{\int_\Omega X \, d\mu}{\varepsilon}. \tag{2.89}$$

Proof of Lemma 2.4.11. The fact that $X \geq 0$ proves that

$$\mathbb{1}^\Omega_{\{X \geq \varepsilon\}} = \frac{\varepsilon \cdot \mathbb{1}^\Omega_{\{X \geq \varepsilon\}}}{\varepsilon} \leq \frac{X \cdot \mathbb{1}^\Omega_{\{X \geq \varepsilon\}}}{\varepsilon} \leq \frac{X}{\varepsilon}. \tag{2.90}$$

Integration of (2.90) with respect to μ results in (2.89). This completes the proof of Lemma 2.4.11. □

A direct consequence of the Markov inequality is the Chebyshev inequality which is presented in the next corollary.

Corollary 2.4.12 (Chebyshev inequality)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $X \colon \Omega \to \mathbb{R}$ be an $\mathcal{F}/\mathcal{B}(\mathbb{R})$-measurable function with $\mathbb{E}_P\big[|X|_\mathbb{R}\big] < \infty$, and let $\varepsilon, q \in (0, \infty)$. Then*

$$P\big(\,|X - \mathbb{E}_P[X]|_\mathbb{R} \geq \varepsilon\big) \leq \left(\frac{\mathbb{E}_P\big[|X - \mathbb{E}_P[X]|^q\big]}{\varepsilon^q}\right) \tag{2.91}$$

$$\textit{and} \qquad P\big(\,|X - \mathbb{E}_P[X]|_\mathbb{R} \geq \varepsilon\big) \leq \frac{\mathsf{Var}_P(X)}{\varepsilon^2}. \tag{2.92}$$

An immediate consequence of the Markov inequality is the fact that strong convergence implies convergence in probability. This is the subject of the following lemma.

Lemma 2.4.13 (L^p-convergence implies convergence in probability)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $p \in (0, \infty)$, and let $X_n \in \mathcal{L}^p(P; |\cdot|_\mathbb{R})$, $n \in \mathbb{N}$, satisfy $\limsup_{n\to\infty} \|X_n\|_{\mathcal{L}^p(P;|\cdot|_\mathbb{R})} = 0$. Then $(X_n)_{n\in N}$ converges to zero in probability, i.e., it holds for all $\varepsilon \in (0, \infty)$ that $\limsup_{n\to\infty} P(|X_n|_\mathbb{R} \geq \varepsilon) = 0$.*

Proof of Lemma 2.4.13. First of all, observe that the Markov inequality proves that for all $n \in \mathbb{N}$, $\varepsilon \in (0, \infty)$ it holds that

$$P(|X_n|_\mathbb{R} \geq \varepsilon) = P(|X_n|^p_\mathbb{R} \geq \varepsilon^p) \leq \frac{\mathbb{E}_P\big[|X_n|^p_\mathbb{R}\big]}{\varepsilon^p}. \tag{2.93}$$

The assumption that $\limsup_{n\to\infty} \mathbb{E}_P\big[|X_n|^p_\mathbb{R}\big] = 0$ hence implies that for all $\varepsilon \in (0, \infty)$ it holds that $\limsup_{n\to\infty} P(|X_n|_\mathbb{R} \geq \varepsilon) = 0$. The proof of Lemma 2.4.13 is thus completed. □

2.4.4 Examples and counterexamples

Proposition 2.4.14. *Let $X_n \colon [0,1] \to \mathbb{R}$, $n \in \mathbb{N}$, be the functions which satisfy for all $n \in \mathbb{N}$, $x \in [0,1]$ that*

$$X_n(x) = 2^n \cdot \mathbb{1}^{[0,1]}_{(0,1/n)}(x). \tag{2.94}$$

Then

(i) it holds for all $x \in [0,1]$ that $\limsup_{n\to\infty} |X_n(x)| = 0$,

(ii) it holds that $B_{[0,1]}\big(\limsup_{n\to\infty} |X_n| = 0\big) = 1$,

(iii) it holds for all $\varepsilon \in (0,\infty)$ that $\limsup_{n\to\infty} B_{[0,1]}\big(|X_n| \geq \varepsilon\big) = 0$, and

(iv) it holds for all $p \in (0,\infty)$ that $\liminf_{n\to\infty} \mathbb{E}_{B_{[0,1]}}\big[|X_n|^p\big] = \infty$.

Proof of Proposition 2.4.14. Observe that for all $x \in (0,1]$ and all $n \in \mathbb{N} \cap (\frac{1}{x},\infty)$ it holds that

$$X_n(x) = 0. \tag{2.95}$$

This proves Item (i). Item (ii) is an immediate consequence from Item (i). Item (iii) follows from Item (ii) together with Lemma 2.4.4. Moreover, observe that for all $n \in \mathbb{N}$, $p \in (0,\infty)$ it holds that

$$\mathbb{E}_{B_{[0,1]}}\big[|X_n|^p\big] = 2^{np} \cdot \tfrac{1}{n}. \tag{2.96}$$

This establishes Item (iv). The proof of Proposition 2.4.14 is thus completed. □

Proposition 2.4.15. *Let $X_n \colon [0,1] \to \mathbb{R}$, $n \in \mathbb{N}$, be the functions which satisfy for all $n \in \mathbb{N}$, $l \in \mathbb{N}_0$, $k \in \{1,2,3,\ldots,2^l\}$, $x \in [0,1]$ with $n = 2^l + k - 1$ that*

$$X_n(x) = \mathbb{1}^{[0,1]}_{[\frac{k-1}{2^l},\frac{k}{2^l}]}(x). \tag{2.97}$$

Then

(i) it holds for all $p \in (0,\infty)$ that $\limsup_{n\to\infty} \mathbb{E}_{B_{[0,1]}}\big[|X_n|^p\big] = 0$,

(ii) it holds for all $\varepsilon \in (0,\infty)$ that $\limsup_{n\to\infty} B_{[0,1]}\big(|X_n| \geq \varepsilon\big) = 0$, and

(iii) it holds that $\{x \in [0,1] \colon \limsup_{n\to\infty} |X_n(x)| = 0\} = \emptyset$.

Proof of Proposition 2.4.15. Observe that for all $p \in (0,\infty)$, $n \in \mathbb{N}$, $l \in \mathbb{N}_0$, $k \in \{1,2,3,\ldots,2^l\}$ with $n = 2^l + k - 1$ it holds that

$$\mathbb{E}_{B_{[0,1]}}\big[|X_n|^p\big] = \mathbb{E}_{B_{[0,1]}}\Big[\mathbb{1}^{[0,1]}_{[\frac{k-1}{2^l},\frac{k}{2^l}]}\Big] = \tfrac{1}{2^l}. \tag{2.98}$$

This implies Item (i). Item (ii) is an immediate consequence from Item (i) and Lemma 2.4.13. Item (iii) follows from the fact that for all $x \in [0,1]$ it holds that $\limsup_{n\to\infty} X_n(x) = 1$. The proof of Proposition 2.4.15 is thus completed. □

2.5 Approximating the variance of a random variable

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $X_n \in \mathcal{L}^2(P; |\cdot|_{\mathbb{R}})$ be i.i.d. random variables. Then Lemma 2.4.4, Theorem 2.4.5, and Theorem 2.4.8 prove that the Monte Carlo approximation sequence

$$\frac{(X_1 + \ldots + X_N)}{N}, \quad N \in \mathbb{N}, \tag{2.99}$$

of $\mathbb{E}_P[X_1]$ converges P-almost surely, in probability, and in the root mean square sense to $\mathbb{E}_P[X_1]$. In that sense the random variable $\frac{1}{N}(X_1 + \ldots + X_N)$ is a good approximation of the expectation $\mathbb{E}_P[X_1]$ of the random variable X_1 if $N \in \mathbb{N}$ is sufficiently large. In this subsection we are interested to compute an approximation of the variance

$$\mathsf{Var}_P(X_1) = \mathbb{E}_P\big[(X_1 - \mathbb{E}_P[X_1])^2\big] \tag{2.100}$$

of the random variable X_1. A central reason why we are interested to obtain such an approximation is that the variance of X_1 appears on the right hand side of the root mean square error estimate (2.84) in Theorem 2.4.8. To compute an approximation of (2.100), we consider random variables $\bar{X}_n \in \mathcal{L}^1(P; |\cdot|_{\mathbb{R}})$, $n \in \mathbb{N}$, which satisfy for all $n \in \mathbb{N}$ that

$$\bar{X}_n = (X_n - \mathbb{E}_P[X_1])^2 . \tag{2.101}$$

Then we obtain from Theorem 2.4.5 that the sequence $\frac{1}{N}(\bar{X}_1 + \ldots + \bar{X}_N)$, $N \in \mathbb{N}$, is strongly P-consistent for $\mathbb{E}_P[\bar{X}_1] = \mathsf{Var}_P(X_1)$. This suggests that

$$\frac{\bar{X}_1 + \ldots + \bar{X}_N}{N} = \frac{1}{N}\left(\sum_{n=1}^{N} (X_n - \mathbb{E}_P[X_1])^2\right) \tag{2.102}$$

is a good approximation of $\mathsf{Var}_P(X_1)$ if $N \in \mathbb{N}$ is sufficiently large. However, (2.102) does only help in simulations if one already knows the exact value of the expectation $\mathbb{E}_P[X_1]$ of the random variable X_1. If one does not know the exact value of the expectation $\mathbb{E}_P[X_1]$ of the random variable X_1 (which is often the case), then one can in general not generate realizations of the random variables $\bar{X}_n$, $n \in \mathbb{N}$. Instead we are looking for random variables that are – in an appropriate sense – good approximations of $\mathsf{Var}_P(X_1)$ and from which one can generate realizations without the explicit knowledge of $\mathbb{E}_P[X_1]$. An obvious idea is to replace the expectation $\mathbb{E}_P[X_1]$ in (2.102) by its Monte Carlo approximations. This leads to the random variables

$$\frac{1}{N}\sum_{n=1}^{N}\left(X_n - \frac{X_1 + \ldots + X_N}{N}\right)^2 \tag{2.103}$$

for $N \in \mathbb{N}$ as approximations of $\mathsf{Var}_P(X_1)$. The random variables (2.103) converge, under suitable assumptions, in the mean square sense to $\mathsf{Var}_P(X_1)$ (see Subsection 2.5.3 below). However, we will note that the random variables (2.103) are P-biased with respect to $\mathsf{Var}_P(X_1)$. This is the subject of the next subsection.

2.5.1 On P-biased and P-unbiased variance approximations and Bessel's correction

The next result, Proposition 2.5.1, in particular, proves that the random variables (2.103) are P-biased with respect to $\mathsf{Var}_P(X_1)$.

Proposition 2.5.1 (Biasedness and unbiasedness of approximations for the variance of a random variable)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $X_n \in \mathcal{L}^2(P; |\cdot|_{\mathbb{R}})$, $n \in \mathbb{N}$, be i.i.d. random variables. Then*

(i) it holds for all $N \in \mathbb{N}$ that

$$\mathbb{E}_P\left[\sum_{n=1}^{N}\left(X_n - \frac{X_1 + \ldots + X_N}{N}\right)^2\right] = (N-1)\cdot \mathsf{Var}_P(X_1), \tag{2.104}$$

(ii) it holds for every $N \in \mathbb{N}$ with $\mathsf{Var}_P(X_1) > 0$ that the $\mathcal{F}/\mathcal{B}(\mathbb{R})$-measurable function $\frac{1}{N}\sum_{n=1}^{N}\left(X_n - \frac{X_1+\ldots+X_N}{N}\right)^2$ is P-biased with respect to $\mathsf{Var}_P(X_1)$, and

(iii) it holds for every $N \in \{2, 3, \ldots\}$ that the $\mathcal{F}/\mathcal{B}(\mathbb{R})$-measurable function $\frac{1}{(N-1)}\sum_{n=1}^{N}\left(X_n - \frac{X_1+\ldots+X_N}{N}\right)^2$ is P-unbiased with respect to $\mathsf{Var}_P(X_1)$.

Proof of Proposition 2.5.1. Lemma 2.4.6 shows that for all $N \in \mathbb{N}$ it holds that

$$\begin{aligned}
&\mathbb{E}_P\left[\sum_{n=1}^{N}\left(X_n - \frac{X_1 + \ldots + X_N}{N}\right)^2\right] = \sum_{n=1}^{N}\mathbb{E}_P\left[\left(X_n - \frac{X_1 + \ldots + X_N}{N}\right)^2\right] \\
&= \sum_{n=1}^{N}\mathsf{Var}_P\left(X_n - \frac{X_1 + \ldots + X_N}{N}\right) \\
&= \sum_{n=1}^{N}\left((N-1)\,\mathsf{Var}_P\big(\tfrac{X_1}{N}\big) + \mathsf{Var}_P\Big(\tfrac{(N-1)}{N}X_1\Big)\right) \\
&= N\left(\tfrac{(N-1)}{N^2}\,\mathsf{Var}_P(X_1) + \tfrac{(N-1)^2}{N^2}\,\mathsf{Var}_P(X_1)\right) = \left(\tfrac{(N-1)}{N} + \tfrac{(N-1)^2}{N}\right)\mathsf{Var}_P(X_1) \\
&= (N-1)\,\mathsf{Var}_P(X_1)\,.
\end{aligned} \tag{2.105}$$

This completes the proof of Proposition 2.5.1. □

Proposition 2.5.1 shows that the random variables (2.103) are P-biased with respect to $\mathsf{Var}_P(X_1)$ and Proposition 2.5.1 also shows that the random variables

$$\frac{1}{(N-1)}\sum_{n=1}^{N}\left(X_n - \frac{X_1 + \ldots + X_N}{N}\right)^2 \tag{2.106}$$

for $N \in \{2, 3, \ldots\}$ are P-unbiased with respect to $\mathsf{Var}_P(X_1)$. These random variables also converge in the root mean square sense to $\mathsf{Var}_P(X_1)$ if it holds for all $n \in \mathbb{N}$ that $X_n \in \mathcal{L}^4(P; |\cdot|_{\mathbb{R}})$. This is the subject of the next subsection.

2.5.2 Root mean square error of a P-unbiased variance approximation (unbiased/corrected sample variance)

In Theorem 2.5.3 below we prove, in particular, that the random variables (2.106) converge in the mean square sense to $\mathsf{Var}_P(X_1)$ if it holds for all $n \in \mathbb{N}$ that $X_n \in \mathcal{L}^4(P; |\cdot|_{\mathbb{R}})$. The proof of Theorem 2.5.3 uses the following elementary identity.

Lemma 2.5.2. *Let $N \in \mathbb{N}$, $x_1, x_2, \ldots, x_N \in \mathbb{R}$. Then*

$$\sum_{n=1}^{N} \left(x_n - \frac{x_1 + \ldots + x_N}{N} \right)^2 = \left(\sum_{n=1}^{N} (x_n)^2 \right) - \frac{1}{N} \left(\sum_{n=1}^{N} x_n \right)^2. \tag{2.107}$$

Proof of Lemma 2.5.2. Note that

$$\begin{aligned} &\sum_{n=1}^{N} \left(x_n - \frac{x_1 + \ldots + x_N}{N} \right)^2 \\ &= \left[\sum_{n=1}^{N} (x_n)^2 \right] - 2 \left[\sum_{n=1}^{N} x_n \left(\frac{x_1 + \ldots + x_N}{N} \right) \right] + \left[\sum_{n=1}^{N} \left(\frac{x_1 + \ldots + x_N}{N} \right)^2 \right] \\ &= \left(\sum_{n=1}^{N} (x_n)^2 \right) - \frac{1}{N} \left(\sum_{n=1}^{N} x_n \right)^2. \end{aligned} \tag{2.108}$$

The proof of Lemma 2.5.2 is thus completed. □

Theorem 2.5.3. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $X_n \in \mathcal{L}^4(P; |\cdot|_{\mathbb{R}})$, $n \in \mathbb{N}$, be i.i.d. random variables. Then*

(i) it holds for all $N \in \{2, 3, \ldots\}$ that

$$\begin{aligned} &\left\| \mathsf{Var}_P(X_1) - \frac{1}{(N-1)} \sum_{n=1}^{N} \left(X_n - \frac{X_1 + \ldots + X_N}{N} \right)^2 \right\|_{\mathcal{L}^2(P;|\cdot|_{\mathbb{R}})} \\ &= \frac{\sqrt{(N-1)^2 \, \mathbb{E}_P[(X_1 - \mathbb{E}_P[X_1])^4] + (4N - N^2 - 3) \, |\mathsf{Var}_P(X_1)|_{\mathbb{R}}^2}}{\sqrt{N} \, (N-1)} \end{aligned} \tag{2.109}$$

and

(ii) it holds for all $N \in \{3, 4, \ldots\}$ that

$$\left\| \mathsf{Var}_P(X_1) - \tfrac{1}{(N-1)} \sum_{n=1}^{N} \left(X_n - \frac{X_1 + \ldots + X_N}{N} \right)^2 \right\|_{\mathcal{L}^2(P;|\cdot|_{\mathbb{R}})} \leq \frac{\sqrt{\mathbb{E}_P\left[(X_1 - \mathbb{E}_P[X_1])^4\right]}}{\sqrt{N}}. \tag{2.110}$$

Proof of Theorem 2.5.3. Let $Y_n \in \mathcal{L}^4(P; |\cdot|_{\mathbb{R}})$, $n \in \mathbb{N}$, be random variables which satisfy for all $n \in \mathbb{N}$ that

$$Y_n = X_n - \mathbb{E}_P\big[X_n\big]. \tag{2.111}$$

Next observe that Proposition 2.5.1 implies that

$$\begin{aligned}
&\mathbb{E}_P\left[\left|\mathsf{Var}_P(X_1) - \frac{1}{(N-1)}\sum_{n=1}^{N}\left(X_n - \frac{X_1 + \ldots + X_N}{N}\right)^2\right|_{\mathbb{R}}^2\right] \\
&= \mathsf{Var}_P\left(\frac{1}{(N-1)}\sum_{n=1}^{N}\left(X_n - \frac{X_1 + \ldots + X_N}{N}\right)^2\right) \\
&= \frac{1}{(N-1)^2}\,\mathsf{Var}_P\left(\sum_{n=1}^{N}\left(X_n - \frac{X_1 + \ldots + X_N}{N}\right)^2\right) \\
&= \frac{1}{(N-1)^2}\,\mathsf{Var}_P\left(\sum_{n=1}^{N}\left(Y_n - \frac{Y_1 + \ldots + Y_N}{N}\right)^2\right).
\end{aligned} \tag{2.112}$$

Lemma 2.5.2 and Lemma 2.4.6 and the fact that $\forall\, i \in \{1, \ldots, N\}\colon \mathbb{E}_P\big[Y_i\big] = 0$ hence show that for all $N \in \{2, 3, \ldots\}$ it holds that

$$\begin{aligned}
&\mathbb{E}_P\left[\left|\mathsf{Var}_P(X_1) - \frac{1}{(N-1)}\sum_{n=1}^{N}\left(X_n - \frac{X_1 + \ldots + X_N}{N}\right)^2\right|_{\mathbb{R}}^2\right] \\
&= \frac{1}{(N-1)^2}\,\mathsf{Var}_P\left(\sum_{n=1}^{N}(Y_n)^2 - \frac{\left(\sum_{n=1}^{N} Y_n\right)^2}{N}\right) \\
&= \tfrac{1}{(N-1)^2}\left\{\mathsf{Var}_P\left(\sum_{n=1}^{N}(Y_n)^2\right) - 2\,\mathsf{Cov}_P\left(\sum_{n=1}^{N}(Y_n)^2, \tfrac{\left(\sum_{n=1}^{N} Y_n\right)^2}{N}\right) + \mathsf{Var}_P\left(\tfrac{\left(\sum_{n=1}^{N} Y_n\right)^2}{N}\right)\right\} \\
&= \tfrac{1}{(N-1)^2}\left\{\sum_{n=1}^{N}\mathsf{Var}_P\big((Y_n)^2\big) - \frac{2}{N}\sum_{n,i,j=1}^{N}\mathsf{Cov}_P\big((Y_n)^2, Y_iY_j\big) + \tfrac{\mathsf{Var}\left(\left(\sum_{n=1}^{N} Y_n\right)^2\right)}{N^2}\right\}.
\end{aligned} \tag{2.113}$$

Again Lemma 2.4.6 therefore shows that for all $N \in \{2, 3, \dots\}$ it holds that

$$
\begin{aligned}
&\mathbb{E}_P\left[\left|\mathsf{Var}_P(X_1) - \frac{1}{(N-1)}\sum_{n=1}^{N}\left(X_n - \frac{X_1+\ldots+X_N}{N}\right)^2\right|_{\mathbb{R}}^2\right] \\
&= \tfrac{1}{(N-1)^2}\left\{N\,\mathsf{Var}_P\big((Y_1)^2\big) - \frac{2}{N}\sum_{n=1}^{N}\mathsf{Var}_P\big((Y_n)^2\big) + \tfrac{\mathbb{E}_P\left[\left(\sum_{n=1}^{N} Y_n\right)^4\right] - \left|\mathbb{E}_P\left[\left(\sum_{n=1}^{N} Y_n\right)^2\right]\right|_{\mathbb{R}}^2}{N^2}\right\} \\
&= \tfrac{1}{(N-1)^2}\left\{(N-2)\,\mathsf{Var}_P\big((Y_1)^2\big) + \tfrac{\sum_{n_1,n_2,n_3,n_4=1}^{N}\mathbb{E}_P\left[Y_{n_1}Y_{n_2}Y_{n_3}Y_{n_4}\right] - \left|\mathsf{Var}\left(\sum_{n=1}^{N} Y_n\right)\right|_{\mathbb{R}}^2}{N^2}\right\} \\
&= \tfrac{(N-2)\,\mathsf{Var}\left((Y_1)^2\right)}{(N-1)^2} + \tfrac{\sum_{n=1}^{N}\mathbb{E}_P\left[(Y_n)^4\right] + 3\sum_{n,m\in\{1,\ldots,N\},n\neq m}\mathbb{E}_P\left[(Y_n)^2(Y_m)^2\right] - |N\cdot\mathsf{Var}(Y_1)|_{\mathbb{R}}^2}{N^2(N-1)^2} \\
&= \frac{(N-2)}{(N-1)^2}\,\mathsf{Var}_P\big((Y_1)^2\big) + \frac{N\,\mathbb{E}_P\big[(Y_1)^4\big] + 3N(N-1)\,|\mathbb{E}_P[(Y_1)^2]|_{\mathbb{R}}^2 - N^2\,|\mathbb{E}_P[(Y_1)^2]|_{\mathbb{R}}^2}{N^2\,(N-1)^2} \\
&= \frac{N\,(N-2)\left\{\mathbb{E}_P[(Y_1)^4] - |\mathbb{E}_P[(Y_1)^2]|_{\mathbb{R}}^2\right\}}{N\,(N-1)^2} + \frac{\mathbb{E}_P\big[(Y_1)^4\big] + (2N-3)\,|\mathbb{E}_P[(Y_1)^2]|_{\mathbb{R}}^2}{N\,(N-1)^2} \\
&= \frac{(N-1)^2\,\mathbb{E}_P[(Y_1)^4] + (4N-N^2-3)\,|\mathbb{E}_P[(Y_1)^2]|_{\mathbb{R}}^2}{N\,(N-1)^2}.
\end{aligned}
\tag{2.114}
$$

This proves equation (2.109). Moreover, combining (2.114) with the fact that $\forall\, N \in \{3, 4, \dots\}\colon 4N - N^2 - 3 \leq 0$ proves that for all $N \in \{3, 4, \dots\}$ it holds that

$$
\begin{aligned}
&\mathbb{E}_P\left[\left|\mathsf{Var}_P(X_1) - \frac{1}{(N-1)}\sum_{n=1}^{N}\left(X_n - \frac{X_1+\ldots+X_N}{N}\right)^2\right|_{\mathbb{R}}^2\right] \\
&\leq \frac{(N-1)^2\,\mathbb{E}_P[(Y_1)^4]}{N\,(N-1)^2} = \frac{\mathbb{E}_P[(Y_1)^4]}{N}.
\end{aligned}
\tag{2.115}
$$

This completes the proof of Theorem 2.5.3. □

2.5.3 Root mean square error of a P-biased variance approximation (sample variance)

The random variables in (2.103) also converge in the mean square sense to $\mathsf{Var}_P(X_1)$ if it holds for all $n \in \mathbb{N}$ that $X_n \in \mathcal{L}^4(P; |\cdot|_{\mathbb{R}})$. This is formulated in the next corollary.

Corollary 2.5.4. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $X_n \in \mathcal{L}^4(P; |\cdot|_{\mathbb{R}})$, $n \in \mathbb{N}$, be i.i.d. random variables. Then it holds for all $N \in \mathbb{N}$ that*

$$\begin{aligned}
&\left\| \mathsf{Var}_P(X_1) - \frac{1}{N} \sum_{n=1}^{N} \left(X_n - \frac{X_1 + \ldots + X_N}{N} \right)^2 \right\|_{\mathcal{L}^2(P;|\cdot|_{\mathbb{R}})} \\
&= \frac{\sqrt{(N-1)^2 \, \mathbb{E}_P[(X_1 - \mathbb{E}_P[X_1])^4] + (5N - N^2 - 3) \, |\mathsf{Var}_P(X_1)|^2_{\mathbb{R}}}}{N^{3/2}} \\
&\leq \frac{\sqrt{\mathbb{E}_P[(X_1 - \mathbb{E}_P[X_1])^4]}}{\sqrt{N}}.
\end{aligned} \tag{2.116}$$

Proof of Corollary 2.5.4. First of all, observe that Proposition 2.5.1 implies that for all $N \in \{2, 3, \ldots\}$ it holds that

$$\begin{aligned}
&\left\| \mathsf{Var}_P(X_1) - \frac{1}{N} \sum_{n=1}^{N} \left(X_n - \frac{X_1 + \ldots + X_N}{N} \right)^2 \right\|^2_{\mathcal{L}^2(P;|\cdot|_{\mathbb{R}})} \\
&= \left| \mathsf{Var}_P(X_1) - \tfrac{(N-1)\,\mathsf{Var}_P(X_1)}{N} \right|^2_{\mathbb{R}} + \mathsf{Var}_P\!\left(\frac{1}{N} \sum_{n=1}^{N} \left(X_n - \frac{X_1 + \ldots + X_N}{N} \right)^2 \right) \\
&= \frac{|\mathsf{Var}_P(X_1)|^2_{\mathbb{R}}}{N^2} + \frac{(N-1)^2}{N^2} \mathsf{Var}_P\!\left(\frac{1}{(N-1)} \sum_{n=1}^{N} \left(X_n - \frac{X_1 + \ldots + X_N}{N} \right)^2 \right).
\end{aligned} \tag{2.117}$$

Theorem 2.5.3 hence shows that for all $N \in \{2, 3, \ldots\}$ it holds that

$$\begin{aligned}
&\left\| \mathsf{Var}_P(X_1) - \frac{1}{N} \sum_{n=1}^{N} \left(X_n - \frac{X_1 + \ldots + X_N}{N} \right)^2 \right\|^2_{\mathcal{L}^2(P;|\cdot|_{\mathbb{R}})} \\
&= \frac{|\mathsf{Var}_P(X_1)|^2_{\mathbb{R}}}{N^2} + \frac{(N-1)^2 \, \mathbb{E}_P[(X_1 - \mathbb{E}_P[X_1])^4] + (4N - N^2 - 3) \, |\mathsf{Var}_P(X_1)|^2_{\mathbb{R}}}{N^3} \\
&= \frac{(N-1)^2 \, \mathbb{E}_P[(X_1 - \mathbb{E}_P[X_1])^4] + (5N - N^2 - 3) \, |\mathsf{Var}_P(X_1)|^2_{\mathbb{R}}}{N^3} \\
&= \frac{\mathbb{E}_P[(X_1 - \mathbb{E}_P[X_1])^4]}{N} + \frac{(1 - 2N) \, \mathbb{E}_P[(X_1 - \mathbb{E}_P[X_1])^4] + (5N - N^2 - 3) \, |\mathsf{Var}_P(X_1)|^2_{\mathbb{R}}}{N^3}
\end{aligned} \tag{2.118}$$

The estimate $\mathbb{E}_P[(X_1 - \mathbb{E}_P[X_1])^4] \geq |\mathsf{Var}_P(X_1)|^2_{\mathbb{R}}$ therefore proves that for all $N \in \{2, 3, \ldots\}$ it holds that

$$\begin{aligned}
&\left\| \mathsf{Var}_P(X_1) - \frac{1}{N} \sum_{n=1}^{N} \left(X_n - \frac{X_1 + \ldots + X_N}{N} \right)^2 \right\|^2_{\mathcal{L}^2(P;|\cdot|_{\mathbb{R}})} \\
&\leq \frac{\mathbb{E}_P[(X_1 - \mathbb{E}_P[X_1])^4]}{N} + \frac{(3N - N^2 - 2) \, |\mathsf{Var}_P(X_1)|^2_{\mathbb{R}}}{N^3} \leq \frac{\mathbb{E}_P[(X_1 - \mathbb{E}_P[X_1])^4]}{N}.
\end{aligned} \tag{2.119}$$

Combining (2.118) and (2.119) proves (2.116) in the case $N \in \{2, 3, \dots\}$. Next note that it is clear that (2.116) holds in the case $N = 1$. The proof of Corollary 2.5.4 is thus completed. □

2.5.4 Comparison of the root mean square errors of an P-unbiased and a P-biased variance approximation

In this subsection we compare the strong root mean square errors obtained in Corollary 2.5.4 and Theorem 2.5.3.

Proposition 2.5.5. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $X \in \mathcal{L}^4(P; |\cdot|_{\mathbb{R}})$ satisfy $\mathsf{Var}_P(X) > 0$. Then*

(i) it holds for all $N \in \{2, 3, \dots\}$ that

$$\begin{aligned} &\tfrac{(N-1)^2\,\mathbb{E}_P[(X-\mathbb{E}_P[X])^4]+(4N-N^2-3)|\mathsf{Var}_P(X)|^2_{\mathbb{R}}}{N(N-1)^2} - \tfrac{(N-1)^2\,\mathbb{E}_P[(X-\mathbb{E}_P[X])^4]+(5N-N^2-3)|\mathsf{Var}_P(X)|^2_{\mathbb{R}}}{N^3} \\ &= \frac{\left[2 - \frac{3}{N} + \frac{1}{N^2}\right]\mathbb{E}_P\!\left[(X-\mathbb{E}_P[X])^4\right] - \left[3 - \frac{8}{N} + \frac{3}{N^2}\right]|\mathsf{Var}_P(X)|^2_{\mathbb{R}}}{N\,(N-1)} \end{aligned} \tag{2.120}$$

and

(ii) it holds that

$$\begin{aligned} &\limsup_{N\to\infty}\left(N\,(N-1)\left[\begin{array}{c}\frac{(N-1)^2\,\mathbb{E}_P[(X-\mathbb{E}_P[X])^4]+(4N-N^2-3)|\mathsf{Var}_P(X)|^2_{\mathbb{R}}}{N(N-1)^2} \\ -\ \frac{(N-1)^2\,\mathbb{E}_P[(X-\mathbb{E}_P[X])^4]+(5N-N^2-3)|\mathsf{Var}_P(X)|^2_{\mathbb{R}}}{N^3}\end{array}\right]\right) \\ &= \liminf_{N\to\infty}\left(N\,(N-1)\left[\begin{array}{c}\frac{(N-1)^2\,\mathbb{E}_P[(X-\mathbb{E}_P[X])^4]+(4N-N^2-3)|\mathsf{Var}_P(X)|^2_{\mathbb{R}}}{N(N-1)^2} \\ -\ \frac{(N-1)^2\,\mathbb{E}_P[(X-\mathbb{E}_P[X])^4]+(5N-N^2-3)|\mathsf{Var}_P(X)|^2_{\mathbb{R}}}{N^3}\end{array}\right]\right) \\ &= 2\,\mathbb{E}_P\!\left[(X-\mathbb{E}_P[X])^4\right] - 3\,|\mathsf{Var}_P(X)|^2_{\mathbb{R}} \in \mathbb{R}. \end{aligned} \tag{2.121}$$

Proof. Note that for all $N \in \{2, 3, \dots\}$ it holds that

$$\begin{aligned}
&\frac{(N-1)^2\,\mathbb{E}_P[(X-\mathbb{E}_P[X])^4] + (4N - N^2 - 3)\,|\mathsf{Var}_P(X)|_{\mathbb{R}}^2}{N\,(N-1)^2} \\
&- \frac{(N-1)^2\,\mathbb{E}_P[(X-\mathbb{E}_P[X])^4] + (5N - N^2 - 3)\,|\mathsf{Var}_P(X)|_{\mathbb{R}}^2}{N^3} \\
&= \left[\frac{1}{N} - \frac{(N^2 - 2N + 1)}{N^3}\right] \mathbb{E}_P\big[(X-\mathbb{E}_P[X])^4\big] \\
&+ \left[\frac{(3-N)}{N\,(N-1)} - \frac{(5N - N^2 - 3)}{N^3}\right] |\mathsf{Var}_P(X)|_{\mathbb{R}}^2 \\
&= \left[\frac{(2N-1)}{N^3}\right] \mathbb{E}_P\big[(X-\mathbb{E}_P[X])^4\big] + \left[\tfrac{N^2(3-N) - \left(5N-N^2-3\right)(N-1)}{N^3(N-1)}\right] |\mathsf{Var}_P(X)|_{\mathbb{R}}^2 .
\end{aligned} \tag{2.122}$$

This implies that for all $N \in \{2, 3, \dots\}$ it holds that

$$\begin{aligned}
&\frac{(N-1)^2\,\mathbb{E}_P[(X-\mathbb{E}_P[X])^4] + (4N - N^2 - 3)\,|\mathsf{Var}_P(X)|_{\mathbb{R}}^2}{N\,(N-1)^2} \\
&- \frac{(N-1)^2\,\mathbb{E}_P[(X-\mathbb{E}_P[X])^4] + (5N - N^2 - 3)\,|\mathsf{Var}_P(X)|_{\mathbb{R}}^2}{N^3} \\
&= \left[\frac{(2N-1)}{N^3}\right] \mathbb{E}_P\big[(X-\mathbb{E}_P[X])^4\big] + \left[\tfrac{\left(3N^2-N^3-5N^2+N^3+3N+5N-N^2-3\right)}{N^3(N-1)}\right] |\mathsf{Var}_P(X)|_{\mathbb{R}}^2 \\
&= \left[\frac{(2N-1)}{N^3}\right] \mathbb{E}_P\big[(X-\mathbb{E}_P[X])^4\big] - \left[\frac{(3N^2 - 8N + 3)}{N^3\,(N-1)}\right] |\mathsf{Var}_P(X)|_{\mathbb{R}}^2 .
\end{aligned} \tag{2.123}$$

This shows that for all $N \in \{2, 3, \dots\}$ it holds that

$$\begin{aligned}
&\tfrac{(N-1)^2\,\mathbb{E}_P[(X-\mathbb{E}_P[X])^4] + (4N - N^2 - 3)|\mathsf{Var}_P(X)|_{\mathbb{R}}^2}{N(N-1)^2} - \tfrac{(N-1)^2\,\mathbb{E}_P[(X-\mathbb{E}_P[X])^4] + (5N - N^2 - 3)|\mathsf{Var}_P(X)|_{\mathbb{R}}^2}{N^3} \\
&= \frac{[2N^2 - 3N + 1]\,\mathbb{E}_P\big[(X-\mathbb{E}_P[X])^4\big] - [3N^2 - 8N + 3]\,|\mathsf{Var}_P(X)|_{\mathbb{R}}^2}{N^3\,(N-1)} \\
&= \frac{\left[2 - \frac{3}{N} + \frac{1}{N^2}\right] \mathbb{E}_P\big[(X-\mathbb{E}_P[X])^4\big] - \left[3 - \frac{8}{N} + \frac{3}{N^2}\right] |\mathsf{Var}_P(X)|_{\mathbb{R}}^2}{N\,(N-1)}.
\end{aligned} \tag{2.124}$$

This completes the proof of Proposition 2.5.5. □

2.6 Confidence intervals

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $X_n \in \mathcal{L}^2(P; |\cdot|_{\mathbb{R}})$, $n \in \mathbb{N}$, be i.i.d. random variables. Then Lemma 2.4.4 and Theorems 2.4.5 and 2.4.8 prove that the Monte Carlo approximations $\frac{1}{N}(X_1 + \ldots + X_N)$, $N \in \mathbb{N}$, converge P-almost surely, in probability, and in the root mean square sense to $\mathbb{E}_P[X_1]$. We thus know that the random variable

$$\frac{(X_1 + \ldots + X_N)}{N} \tag{2.125}$$

is in the sense above close to the real number $\mathbb{E}_P[X_1]$ if $N \in \mathbb{N}$ is large. However, in many situations it holds for all $N \in \mathbb{N}$ that

$$P\bigg(\frac{(X_1 + \ldots + X_N)}{N} = \mathbb{E}_P[X_1]\bigg) = 0. \tag{2.126}$$

For instance, identity (2.126) holds if $X_1(P)_{\mathcal{B}(\mathbb{R})}$ is absolutely continuous since in that case it holds for all $N \in \mathbb{N}$, $c \in \mathbb{R}$ that

$$P\bigg(\frac{(X_1 + \ldots + X_N)}{N} = c\bigg) = 0. \tag{2.127}$$

Equation (2.126) is somehow plausible since the random variable in (2.125) is close to $\mathbb{E}_P[X_1]$ if $N \in \mathbb{N}$ is large (according to Lemma 2.4.4 and Theorems 2.4.5 and 2.4.8 above) but it is very unlikely that the random variable in (2.125) is precisely equal to $\mathbb{E}_P[X_1]$. In this subsection we roughly speaking intend to quantify in a suitable sense how close $\mathbb{E}_P[X_1]$ and the random variable in (2.125) are. For this the notion of a *confidence interval* is crucial. To introduce this notion, a few technical issues are presented first.

Definition 2.6.1 (Measurable space of nonempty compact intervals)**.** *Define the set* $\mathrm{CI}_{\mathbb{R}} := \{[a, b] \subseteq \mathbb{R} \colon (a, b \in \mathbb{R} \text{ and } a \leq b)\}$ *of all nonempty compact intervals of real numbers and define the sigma-algebra*

$$\begin{aligned} \mathcal{I}_{\mathbb{R}} &:= \sigma_{\mathrm{CI}_{\mathbb{R}}}\Big(\big\{\{A \in \mathrm{CI}_{\mathbb{R}} \colon \inf(A) < c\} \colon c \in \mathbb{R}\big\} \cup \big\{\{A \in \mathrm{CI}_{\mathbb{R}} \colon \sup(A) < c\} \colon c \in \mathbb{R}\big\}\Big) \\ &= \sigma_{\mathrm{CI}_{\mathbb{R}}}\Big(\mathrm{CI}_{\mathbb{R}} \ni A \mapsto \inf(A) \in \mathbb{R}, \mathrm{CI}_{\mathbb{R}} \ni A \mapsto \sup(A) \in \mathbb{R}\Big) \end{aligned} \tag{2.128}$$

on $\mathrm{CI}_{\mathbb{R}}$*.*

The next lemma provides a describtion of the sigma-algebra $\mathcal{I}_{\mathbb{R}}$ on the set of nonempty compact intervals $\mathrm{CI}_{\mathbb{R}}$.

Lemma 2.6.2 (Sigma-algebra on the set of nonempty compact intervals)**.** *It holds for all* $x \in \mathbb{R}$ *that* $\{A \in \mathrm{CI}_{\mathbb{R}} \colon x \in A\} \in \mathcal{I}_{\mathbb{R}}$*.*

Proof of Lemma 2.6.2. Observe that

$$\{A \in \mathrm{CI}_{\mathbb{R}} \colon x \in A\} = \underbrace{\{A \in \mathrm{CI}_{\mathbb{R}} \colon \inf(A) \leq x\}}_{\in \mathcal{I}_{\mathbb{R}}} \cap \underbrace{\{A \in \mathrm{CI}_{\mathbb{R}} \colon \sup(A) \geq x\}}_{\in \mathcal{I}_{\mathbb{R}}} \in \mathcal{I}_{\mathbb{R}} \tag{2.129}$$

for all $x \in \mathbb{R}$. This completes the proof of Lemma 2.6.2. □

Definition 2.6.3 (random interval)**.** *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space. Then an* $\mathcal{F}/\mathcal{I}_{\mathbb{R}}$*-measurable function* $X\colon \Omega \to \mathrm{CI}_{\mathbb{R}}$ *is called* (nonempty compact) random interval*.*

Using Lemma 2.6.2 and Definition 2.6.3, we now introduce the notion of a confidence interval.

Definition 2.6.4 (Confidence interval)**.** *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space, let* $c \in \mathbb{R}$ *and* $\alpha \in [0,1]$ *be real numbers and let* $A\colon \Omega \to \mathrm{CI}_{\mathbb{R}}$ *be a random interval. Then we say that* A *is an* α-confidence interval *for* c *if* $P(c \in A) \geq \alpha$*.*

If A is an α-confidence interval for c in Definition 2.6.4, then the parameter $\alpha \in [0,1]$ is also referred as *confidence level* of the confidence interval A. Let us collect a few properties of confidence intervals.

Lemma 2.6.5 (Supersets of confidence intervals)**.** *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space, let* $\alpha \in [0,1]$ *and* $c \in \mathbb{R}$ *be real numbers, let* $A_1\colon \Omega \to \mathrm{CI}_{\mathbb{R}}$ *be an* α*-confidence interval for* c *and let* $A_2\colon \Omega \to \mathrm{CI}_{\mathbb{R}}$ *be a random interval with* $A_1(\omega) \subseteq A_2(\omega)$ *for all* $\omega \in \Omega$*. Then* A_2 *is an* α*-confidence interval for* c*.*

Proof of Lemma 2.6.5. The assumption $A_1(\omega) \subseteq A_2(\omega)$ for all $\omega \in \Omega$ implies that

$$\{c \in A_1\} = \{\omega \in \Omega\colon c \in A_1(\omega)\} \subseteq \{\omega \in \Omega\colon c \in A_2(\omega)\} = \{c \in A_2\} \tag{2.130}$$

and the assumption that A_1 is an α-confidence interval for c and the monotonicity of the probability measure P hence show that

$$\alpha \leq P(c \in A_1) \leq P(c \in A_2)\,. \tag{2.131}$$

This finishes the proof of Lemma 2.6.5. □

Lemma 2.6.6 (Enlargement of independent confidence intervals)**.** *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space, let* $\alpha_1, \alpha_2 \in [0,1]$ *and* $c \in \mathbb{R}$ *be real numbers, let* $A_1\colon \Omega \to \mathrm{CI}_{\mathbb{R}}$ *be an* α_1*-confidence interval for* c*, let* $A_2\colon \Omega \to \mathrm{CI}_{\mathbb{R}}$ *be an* α_2*-confidence interval for* c *and assume that* A_1, A_2 *are independent. Then the random interval* $[\inf(A_1 \cup A_2), \sup(A_1 \cup A_2)]\colon \Omega \to \mathrm{CI}_{\mathbb{R}}$ *is an* $(1-(1-\alpha_1)(1-\alpha_2))$*-confidence interval for* c*.*

Proof of Lemma 2.6.6. Observe that

$$\begin{split} &P\Big(c \in \big[\inf(A_1 \cup A_2), \sup(A_1 \cup A_2)\big]\Big) \\ &\geq P\big(\{c \in A_1\} \cup \{c \in A_2\}\big) = 1 - P\big(\{c \notin A_1\} \cap \{c \notin A_2\}\big) \\ &= 1 - P\big(c \notin A_1\big) \cdot P\big(c \notin A_2\big) \geq 1 - (1-\alpha_1)\,(1-\alpha_2) = \alpha_1 + \alpha_2 - \alpha_1\alpha_2. \end{split} \tag{2.132}$$

This finishes the proof of Lemma 2.6.6. □

2.6.1 (Asymptotically valid) Confidence intervals based on the Chebyshev inequality

In this subsection confidence intervals based on Monte Carlo approximations and the *Chebyshev inequality* are derived.

Corollary 2.6.7 (Confidence intervals based on Monte Carlo approximations and the Chebyshev inequality I)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $\alpha \in [0,1)$ be a real number and let $X_n \in \mathcal{L}^2(P; |\cdot|_{\mathbb{R}})$, $n \in \mathbb{N}$, be i.i.d. random variables. Then*

$$\begin{aligned} &P\Bigg(\mathbb{E}_P[X_1] \in \Bigg[\tfrac{X_1+\ldots+X_N}{N} - \tfrac{\sqrt{\mathsf{Var}_P(X_1)}}{\sqrt{(1-\alpha)N}}, \tfrac{X_1+\ldots+X_N}{N} + \tfrac{\sqrt{\mathsf{Var}_P(X_1)}}{\sqrt{(1-\alpha)N}}\Bigg]\Bigg) \\ &= P\left(\left|\mathbb{E}_P[X_1] - \frac{X_1+\ldots+X_N}{N}\right|_{\mathbb{R}} \leq \frac{\sqrt{\mathsf{Var}_P(X_1)}}{\sqrt{(1-\alpha)N}}\right) \geq \alpha. \end{aligned} \tag{2.133}$$

In other words, for every $N \in \mathbb{N}$ the random interval

$$\Bigg[\tfrac{X_1+\ldots+X_N}{N} - \tfrac{\sqrt{\mathsf{Var}_P(X_1)}}{\sqrt{(1-\alpha)N}}, \tfrac{X_1+\ldots+X_N}{N} + \tfrac{\sqrt{\mathsf{Var}_P(X_1)}}{\sqrt{(1-\alpha)N}}\Bigg] \tag{2.134}$$

is an α-confidence interval for $\mathbb{E}_P[X_1]$.

Proof of Corollary 2.6.7. If $\mathsf{Var}_P(X_1) > 0$, then the Chebyshev inequality in (2.92) in Corollary 2.4.12, Lemma 2.3.8, and Theorem 2.4.8 imply that

$$\begin{aligned} &P\left(\left|\mathbb{E}_P[X_1] - \frac{X_1+\ldots+X_N}{N}\right|_{\mathbb{R}} \geq \frac{\sqrt{\mathsf{Var}_P(X_1)}}{\sqrt{(1-\alpha)N}}\right) \leq \frac{\mathsf{Var}_P\big(\frac{X_1+\ldots+X_N}{N}\big)}{\left(\frac{\sqrt{\mathsf{Var}_P(X_1)}}{\sqrt{(1-\alpha)N}}\right)^2} \\ &= \frac{(1-\alpha)\,N}{\mathsf{Var}_P(X_1)} \cdot \mathsf{Var}_P\left(\frac{X_1+\ldots+X_N}{N}\right) = \frac{(1-\alpha)\,N}{\mathsf{Var}_P(X_1)} \cdot \frac{\mathsf{Var}_P(X_1)}{N} = 1-\alpha. \end{aligned} \tag{2.135}$$

This shows in the case $\mathsf{Var}_P(X_1) > 0$ that

$$\begin{aligned} &P\left(\left|\mathbb{E}_P[X_1] - \frac{X_1+\ldots+X_N}{N}\right|_{\mathbb{R}} < \frac{\sqrt{\mathsf{Var}_P(X_1)}}{\sqrt{(1-\alpha)N}}\right) \\ &= 1 - P\left(\left|\mathbb{E}_P[X_1] - \frac{X_1+\ldots+X_N}{N}\right|_{\mathbb{R}} \geq \frac{\sqrt{\mathsf{Var}_P(X_1)}}{\sqrt{(1-\alpha)N}}\right) \geq \alpha \end{aligned} \tag{2.136}$$

and this proves (2.133) in the case $\mathsf{Var}_P(X_1) > 0$. Moreover, note that (2.133) is clear in the case $\mathsf{Var}_P(X_1) = 0$. The proof of Corollary 2.6.7 is thus completed. □

In many situations the variance $\mathsf{Var}_P(X_1)$ appearing in the confidence interval (2.134) is not explicitly known and therefore, the confidence interval (2.134) can not be calculated in that case. However, in many situations at least an upper bound for $\mathsf{Var}_P(X_1)$ in Corollary 2.6.7 is known and then confidence intervals can be calculcated. This is illustrated in the following corollary.

Corollary 2.6.8 (Confidence intervals based on Monte Carlo approximations and the Chebyshev inequality II)**.** *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space, let* $\alpha \in [0,1)$ *and* $c \in (0,\infty)$*, let* $X_n \in \mathcal{L}^2(P; |\cdot|_{\mathbb{R}})$*,* $n \in \mathbb{N}$*, be i.i.d. random variables with* $\sqrt{\mathsf{Var}_P(X_1)} \leq c$ *and define* $E_N \in \mathcal{L}^2(P; |\cdot|_{\mathbb{R}})$*,* $N \in \mathbb{N}$*, through* $E_N := \frac{X_1 + \ldots + X_N}{N}$ *for all* $N \in \mathbb{N}$*. Then* $P\Big(\mathbb{E}_P[X_1] \in \big[E_N - \frac{c}{\sqrt{(1-\alpha)N}}, E_N + \frac{c}{\sqrt{(1-\alpha)N}}\big]\Big) \geq \alpha$ *for all* $N \in \mathbb{N}$*. In other words, for every* $N \in \mathbb{N}$ *the random interval* $\big[E_N - \frac{c}{\sqrt{(1-\alpha)N}}, E_N + \frac{c}{\sqrt{(1-\alpha)N}}\big]$ *is an* α*-confidence interval for* $\mathbb{E}_P[X_1]$*.*

Proof of Corollary 2.6.8. Corollary 2.6.8 follows immediately from Corollary 2.6.7 and Lemma 2.6.5. □

If the variance $\mathsf{Var}_P(X_1)$ appearing in the confidence interval (2.134) in Corollary 2.6.7 is not explicitly known and if there is also no upper bound for $\mathsf{Var}_P(X_1)$ available (cf. Corollary 2.6.8), then it is not clear how to derive suitable confidence intervals for $\mathbb{E}_P[X_1]$. However, a certain weaker statement can be derived. This is the subject of the next definition (cf. Appendix A in [Glasserman(2004)]).

Definition 2.6.9 (Asymptotically valid confidence intervals)**.** *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space, let* $m \in \mathbb{N}$*,* $\alpha \in [0,1)$*,* $c \in \mathbb{R}$ *and let* $A_n \colon \Omega \to \mathrm{CI}_{\mathbb{R}}$*,* $n \in \{m, m+1, \ldots\}$*, be random intervals. Then we say that* $(A_n)_{n \in \{m,m+1,\ldots\}}$ *are* asymptotically valid α-confidence intervals for c *if*

$$\liminf_{n\to\infty} P(c \in A_n) \geq \alpha. \tag{2.137}$$

In Corollary 2.6.10 below we derive asymptotically valid confidence intervals for the expectation of a random variable.

Corollary 2.6.10 (Asymptotically valid confidence intervals based on Monte Carlo approximations, the Chebyshev inequality and variance approximations)**.** *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space, let* $\alpha \in [0,1)$ *be a real number, let* $X_n \in \mathcal{L}^4(P; |\cdot|_{\mathbb{R}})$*,* $n \in \mathbb{N}$*, be i.i.d. random variables and let* $E_N \in \mathcal{L}^4(P; |\cdot|_{\mathbb{R}})$*,* $N \in \mathbb{N}$*, and* $V_N \in \mathcal{L}^2(P; |\cdot|_{\mathbb{R}})$*,* $N \in \{2,3,\ldots\}$*, be defined through* $E_N := \frac{X_1 + \ldots + X_N}{N}$ *for all* $N \in \mathbb{N}$ *and through* $V_N := \frac{1}{(N-1)} \sum_{n=1}^{N} (X_n - E_N)^2$ *for all* $N \in \{2,3,\ldots\}$*. Then*

$$\liminf_{N\to\infty} P\left(\mathbb{E}_P[X_1] \in \left[E_N - \frac{\sqrt{V_N}}{\sqrt{(1-\alpha)N}}, E_N + \frac{\sqrt{V_N}}{\sqrt{(1-\alpha)N}}\right]\right) \geq \alpha. \tag{2.138}$$

In other words, the random intervals $\big[E_N - \frac{\sqrt{V_N}}{\sqrt{(1-\alpha)N}}, E_N + \frac{\sqrt{V_N}}{\sqrt{(1-\alpha)N}}\big]$ *for* $N \in \{2,3,\ldots\}$ *are asymptotically valid* α*-confidence intervals for* $\mathbb{E}_P[X_1]$*.*

Proof of Corollary 2.6.10. Observe that

$$\begin{aligned}
&P\left(|\mathbb{E}_P[X_1]-E_N|_{\mathbb{R}} \leq \frac{\sqrt{V_N}}{\sqrt{(1-\alpha)N}}\right)\\
&\geq P\left(\left\{|\mathbb{E}_P[X_1]-E_N|_{\mathbb{R}} \leq \frac{\sqrt{V_N}}{\sqrt{(1-\alpha)N}}\right\} \cap \left\{|V_N - \mathsf{Var}_P(X_1)|_{\mathbb{R}} \leq \varepsilon\right\}\right)\\
&\geq P\left(\left\{|\mathbb{E}_P[X_1]-E_N|_{\mathbb{R}} \leq \frac{\sqrt{(\mathsf{Var}_P(X_1)-\varepsilon)}}{\sqrt{(1-\alpha)N}}\right\} \cap \left\{|V_N - \mathsf{Var}_P(X_1)|_{\mathbb{R}} \leq \varepsilon\right\}\right) \qquad (2.139)\\
&= P\left(|\mathbb{E}_P[X_1]-E_N|_{\mathbb{R}} \leq \frac{\sqrt{(\mathsf{Var}_P(X_1)-\varepsilon)}}{\sqrt{(1-\alpha)N}}\right)\\
&- P\left(\left\{|\mathbb{E}_P[X_1]-E_N|_{\mathbb{R}} \leq \frac{\sqrt{(\mathsf{Var}_P(X_1)-\varepsilon)}}{\sqrt{(1-\alpha)N}}\right\} \cap \left\{|V_N - \mathsf{Var}_P(X_1)|_{\mathbb{R}} > \varepsilon\right\}\right)
\end{aligned}$$

and the Chebyshev inequality in (2.92) in Corollary 2.4.12 therefore shows that

$$\begin{aligned}
&P\left(|\mathbb{E}_P[X_1]-E_N|_{\mathbb{R}} \leq \frac{\sqrt{V_N}}{\sqrt{(1-\alpha)N}}\right)\\
&\geq P\left(|\mathbb{E}_P[X_1]-E_N|_{\mathbb{R}} \leq \frac{\sqrt{(\mathsf{Var}_P(X_1)-\varepsilon)}}{\sqrt{(1-\alpha)N}}\right) - P(|V_N - \mathsf{Var}_P(X_1)|_{\mathbb{R}} > \varepsilon)\\
&\geq 1 - P\left(|\mathbb{E}_P[X_1]-E_N|_{\mathbb{R}} > \frac{\sqrt{(\mathsf{Var}_P(X_1)-\varepsilon)}}{\sqrt{(1-\alpha)N}}\right) - P(|V_N - \mathsf{Var}_P(X_1)|_{\mathbb{R}} \geq \varepsilon) \qquad (2.140)\\
&\geq 1 - \frac{\mathsf{Var}_P(E_N)}{\left(\frac{\mathsf{Var}_P(X_1)-\varepsilon}{(1-\alpha)N}\right)} - P(|V_N - \mathsf{Var}_P(X_1)|_{\mathbb{R}} \geq \varepsilon)\\
&= 1 - \tfrac{(1-\alpha)\,\mathsf{Var}_P(X_1)}{(\mathsf{Var}_P(X_1)-\varepsilon)} - P(|V_N - \mathsf{Var}_P(X_1)|_{\mathbb{R}} \geq \varepsilon)
\end{aligned}$$

for all $N \in \{2, 3, \dots\}$ and all $\varepsilon \in (0, \mathsf{Var}_P(X_1))$ where we used Theorem 2.4.8 and Lemma 2.3.8 in the last step. Combining this with Theorem 2.5.3 and Lemma 2.4.13 proves that

$$\begin{aligned}
&\liminf_{N\to\infty} P\left(|\mathbb{E}_P[X_1]-E_N|_{\mathbb{R}} \leq \frac{\sqrt{V_N}}{\sqrt{(1-\alpha)N}}\right)\\
&\geq 1 - \frac{(1-\alpha)\,\mathsf{Var}_P(X_1)}{(\mathsf{Var}_P(X_1)-\varepsilon)} - \lim_{N\to\infty} P(|V_N - \mathsf{Var}_P(X_1)|_{\mathbb{R}} \geq \varepsilon) = 1 - \frac{(1-\alpha)\,\mathsf{Var}_P(X_1)}{(\mathsf{Var}_P(X_1)-\varepsilon)} \qquad (2.141)
\end{aligned}$$

for all $\varepsilon \in (0, \mathsf{Var}_P(X_1))$. Letting $\varepsilon \searrow 0$ in (2.141) results in (2.138) and this completes the proof of Corollary 2.6.10. □

2.6.2 Asymptotically valid confidence intervals based on the Central limit theorem

Another way to derive asymptotically valid confidence intervals is to use the central limit theorem in Theorem 1.3.1. This is the subject of the next theorem.

Theorem 2.6.11 (Asymptotically valid confidence intervals based on Monte Carlo approximations and the central limit theorem). *Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $\beta, \gamma \in [0, \infty)$ be real numbers, let $X_n \in \mathcal{L}^2(P; |\cdot|_{\mathbb{R}})$, $n \in \mathbb{N}$, be i.i.d. random variables and define $E_N \in \mathcal{L}^2(P; |\cdot|_{\mathbb{R}})$, $N \in \mathbb{N}$, through $E_N := \frac{X_1 + \ldots + X_N}{N}$ for all $N \in \mathbb{N}$. Then*

$$\lim_{N \to \infty} P\left(\mathbb{E}_P[X_1] \in \left[E_N - \frac{\beta\sqrt{\mathsf{Var}_P(X_1)}}{\sqrt{N}}, E_N + \frac{\gamma\sqrt{\mathsf{Var}_P(X_1)}}{\sqrt{N}}\right]\right) = \frac{1}{\sqrt{2\pi}} \int_{-\beta}^{\gamma} e^{\frac{-x^2}{2}}\, dx. \tag{2.142}$$

In particular, the random intervals $\left[E_N - \frac{\beta\sqrt{\mathsf{Var}_P(X_1)}}{\sqrt{N}}, E_N + \frac{\gamma\sqrt{\mathsf{Var}_P(X_1)}}{\sqrt{N}}\right]$ for $N \in \mathbb{N}$ are asymptotically valid $\left(\frac{1}{\sqrt{2\pi}} \int_{-\beta}^{\gamma} e^{\frac{-x^2}{2}}\, dx\right)$-confidence intervals for $\mathbb{E}_P[X_1]$.

Proof of Theorem 2.6.11. Observe that

$$\begin{aligned} &P\left(\mathbb{E}_P[X_1] \in \left[E_N - \frac{\beta\sqrt{\mathsf{Var}_P(X_1)}}{\sqrt{N}}, E_N + \frac{\gamma\sqrt{\mathsf{Var}_P(X_1)}}{\sqrt{N}}\right]\right) \\ &= P\left(E_N - \frac{\beta\sqrt{\mathsf{Var}_P(X_1)}}{\sqrt{N}} \leq \mathbb{E}_P[X_1] \leq E_N + \frac{\gamma\sqrt{\mathsf{Var}_P(X_1)}}{\sqrt{N}}\right) \\ &= P\left(\frac{-\gamma\sqrt{\mathsf{Var}_P(X_1)}}{\sqrt{N}} \leq E_N - \mathbb{E}_P[X_1] \leq \frac{\beta\sqrt{\mathsf{Var}_P(X_1)}}{\sqrt{N}}\right) \\ &= P\left(-\gamma \leq \frac{X_1 + \ldots + X_N - N \cdot \mathbb{E}_P[X_1]}{\sqrt{N\, \mathsf{Var}_P(X_1)}} \leq \beta\right) \end{aligned} \tag{2.143}$$

for all $N \in \mathbb{N}$. The central limit theorem (see Theorem 1.3.1) hence shows that

$$\begin{aligned} &\lim_{N \to \infty} P\left(\mathbb{E}_P[X_1] \in \left[E_N - \frac{\beta\sqrt{\mathsf{Var}_P(X_1)}}{\sqrt{N}}, E_N + \frac{\gamma\sqrt{\mathsf{Var}_P(X_1)}}{\sqrt{N}}\right]\right) \\ &= \lim_{N \to \infty} P\left(-\gamma \leq \frac{X_1 + \ldots + X_N - N \cdot \mathbb{E}_P[X_1]}{\sqrt{N\, \mathsf{Var}_P(X_1)}} \leq \beta\right) = \frac{1}{\sqrt{2\pi}} \int_{-\gamma}^{\beta} e^{\frac{-x^2}{2}}\, dx = \frac{1}{\sqrt{2\pi}} \int_{-\beta}^{\gamma} e^{\frac{-x^2}{2}}\, dx. \end{aligned} \tag{2.144}$$

This completes the proof of Theorem 2.6.11. □

If the variance $\mathsf{Var}_P(X_1)$ in the asymptotically valid confidence intervals (2.134) is unknown, then it can be approximated by variance approximations (see Section 2.5) as in Corollary 2.6.10. This is the subject of the following corollary.

Corollary 2.6.12 (Asymptotically valid confidence intervals based on Monte Carlo approximations, the central limit theorem and variance approximations)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $\beta, \gamma \in [0, \infty)$ be real numbers, let $X_n \in \mathcal{L}^4(P; |\cdot|_{\mathbb{R}})$, $n \in \mathbb{N}$, be i.i.d. random variables and define $E_N \in \mathcal{L}^4(P; |\cdot|_{\mathbb{R}})$, $N \in \mathbb{N}$, and $V_N \in \mathcal{L}^2(P; |\cdot|_{\mathbb{R}})$, $N \in \{2, 3, \dots\}$, through $E_N := \frac{X_1 + \ldots + X_N}{N}$ for all $N \in \mathbb{N}$ and through $V_N := \frac{1}{(N-1)} \sum_{n=1}^{N} (X_n - E_N)^2$ for all $N \in \{2, 3, \dots\}$. Then the random intervals $\big[E_N - \frac{\beta\sqrt{V_N}}{\sqrt{N}}, E_N + \frac{\gamma\sqrt{V_N}}{\sqrt{N}}\big]$ for $N \in \{2, 3, \dots\}$ are asymptotically valid $\big(\frac{1}{\sqrt{2\pi}} \int_{-\beta}^{\gamma} e^{-\frac{1}{2}x^2} dx\big)$-confidence intervals for $\mathbb{E}_P[X_1]$.*

Proof of Corollary 2.6.12. W.l.o.g. we assume that $\mathsf{Var}_P(X_1) > 0$. Then we obtain that

$$\begin{aligned}
&P\Big(\mathbb{E}_P[X_1] \in \Big[E_N - \tfrac{\beta\sqrt{V_N}}{\sqrt{N}}, E_N + \tfrac{\gamma\sqrt{V_N}}{\sqrt{N}}\Big]\Big) \\
&\geq P\Big(\Big\{\mathbb{E}_P[X_1] \in \Big[E_N - \tfrac{\beta\sqrt{V_N}}{\sqrt{N}}, E_N + \tfrac{\gamma\sqrt{V_N}}{\sqrt{N}}\Big]\Big\} \cap \Big\{V_N \geq \tfrac{\mathsf{Var}_P(X_1)}{\rho}\Big\}\Big) \\
&\geq P\bigg(\bigg\{\mathbb{E}_P[X_1] \in \bigg[E_N - \tfrac{\beta\sqrt{\mathsf{Var}_P(X_1)}}{\sqrt{\rho N}}, E_N + \tfrac{\gamma\sqrt{\mathsf{Var}_P(X_1)}}{\sqrt{\rho N}}\bigg]\bigg\} \cap \Big\{V_N \geq \tfrac{\mathsf{Var}_P(X_1)}{\rho}\Big\}\bigg) \\
&\geq P\bigg(\mathbb{E}_P[X_1] \in \bigg[E_N - \tfrac{\beta\sqrt{\mathsf{Var}_P(X_1)}}{\sqrt{\rho N}}, E_N + \tfrac{\gamma\sqrt{\mathsf{Var}_P(X_1)}}{\sqrt{\rho N}}\bigg]\bigg) - P\Big(V_N < \tfrac{\mathsf{Var}_P(X_1)}{\rho}\Big)
\end{aligned} \tag{2.145}$$

for all $N \in \{2, 3, \dots\}$ and all $\rho \in (1, \infty)$. Theorem 2.6.11 hence gives that

$$\begin{aligned}
&\liminf_{N\to\infty} P\Big(\mathbb{E}_P[X_1] \in \Big[E_N - \tfrac{\beta\sqrt{V_N}}{\sqrt{N}}, E_N + \tfrac{\gamma\sqrt{V_N}}{\sqrt{N}}\Big]\Big) \\
&\geq \lim_{N\to\infty} P\bigg(\mathbb{E}_P[X_1] \in \bigg[E_N - \tfrac{\beta\sqrt{\mathsf{Var}_P(X_1)}}{\sqrt{\rho N}}, E_N + \tfrac{\gamma\sqrt{\mathsf{Var}_P(X_1)}}{\sqrt{\rho N}}\bigg]\bigg) \\
&\quad - \lim_{N\to\infty} P\Big(V_N - \mathsf{Var}_P(X_1) < \tfrac{\mathsf{Var}_P(X_1)(1-\rho)}{\rho}\Big) \\
&\geq \frac{1}{\sqrt{2\pi}} \int_{\frac{-\beta}{\sqrt{\rho}}}^{\frac{\gamma}{\sqrt{\rho}}} e^{\frac{-x^2}{2}}\, dx - \lim_{N\to\infty} P\Big(|V_N - \mathsf{Var}_P(X_1)|_{\mathbb{R}} \geq \tfrac{\mathsf{Var}_P(X_1)(\rho-1)}{\rho}\Big) \\
&= \frac{1}{\sqrt{2\pi}} \int_{\frac{-\beta}{\sqrt{\rho}}}^{\frac{\gamma}{\sqrt{\rho}}} e^{\frac{-x^2}{2}}\, dx
\end{aligned} \tag{2.146}$$

for all $\rho \in (1, \infty)$ where we used Theorem 2.5.3 and Lemma 2.4.13 in the last step. Letting $\rho \searrow 1$ in (2.146) finishes the proof of Corollary 2.6.12. □

2.6.3 Summary

The confidence intervals and the asymptotically valid confidence intervals derived above are summarized in the following corollary. It follows immediately from Corollaries 2.6.7, 2.6.8 and 2.6.10 ((asymptotically valid) confidence intervals based on the Chebyshev inequality) and Theorem 2.6.11 and Corollary 2.6.12 (asymptotically valid confidence intervals based on the central limit theorem) and from Lemma 2.6.5.

Corollary 2.6.13 (Summary for (asymptotically valid) confidence intervals)**.** *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space, let* $\alpha \in [0, 1)$*,* $c, \beta, \gamma \in [0, \infty)$ *be real numbers with* $\frac{1}{\sqrt{2\pi}} \int_{-\beta}^{\gamma} e^{-x^2/2}\, dx \geq \alpha$ *and* $\sqrt{\mathsf{Var}_P(X_1)} \leq c$*, let* $X_n \in \mathcal{L}^2(P; |\cdot|_{\mathbb{R}})$*,* $n \in \mathbb{N}$*, be i.i.d. random variables and define* $E_N \in \mathcal{L}^2(P; |\cdot|_{\mathbb{R}})$*,* $N \in \mathbb{N}$*, through* $E_N := \frac{X_1 + \ldots + X_N}{N}$ *for all* $N \in \mathbb{N}$*. Then*

$$\left[E_N - \frac{\sqrt{\mathsf{Var}_P(X_1)}}{\sqrt{(1-\alpha)N}}, E_N + \frac{\sqrt{\mathsf{Var}_P(X_1)}}{\sqrt{(1-\alpha)N}}\right], \ N \in \mathbb{N}, \ \text{and}$$
$$\left[E_N - \frac{c}{\sqrt{(1-\alpha)N}}, E_N + \frac{c}{\sqrt{(1-\alpha)N}}\right], \ N \in \mathbb{N}, \qquad \text{(Chebyshev inequality)}$$

are α*-confidence intervals and*

$$\left[E_N - \frac{\beta\sqrt{\mathsf{Var}_P(X_1)}}{\sqrt{N}}, E_N + \frac{\gamma\sqrt{\mathsf{Var}_P(X_1)}}{\sqrt{N}}\right], \ N \in \mathbb{N}, \ \text{and}$$
$$\left[E_N - \tfrac{\beta c}{\sqrt{N}}, E_N + \tfrac{\gamma c}{\sqrt{N}}\right], \ N \in \mathbb{N}, \qquad \text{(Central limit theorem)}$$

are asymptotically valid α*-confidence intervals. Moreover, if* $X_n \in \mathcal{L}^4(P; |\cdot|_{\mathbb{R}})$ *for all* $n \in \mathbb{N}$ *and if* $V_N \in \mathcal{L}^2(P; |\cdot|_{\mathbb{R}})$*,* $N \in \{2, 3, \ldots\}$*, are defined through* $V_N := \frac{1}{(N-1)} \sum_{n=1}^{N-1} (X_n - E_N)^2$ *for all* $N \in \{2, 3, \ldots\}$ *(variance approximations) in addition to the above assumptions, then*

$$\left[E_N - \frac{\sqrt{V_N}}{\sqrt{(1-\alpha)N}}, E_N + \frac{\sqrt{V_N}}{\sqrt{(1-\alpha)N}}\right], \ N \in \{2, 3, \ldots\}, \ \text{and} \qquad \text{(Chebys. \& variance appr.)}$$
$$\left[E_N - \tfrac{\beta\sqrt{V_N}}{\sqrt{N}}, E_N + \tfrac{\gamma\sqrt{V_N}}{\sqrt{N}}\right], \ N \in \{2, 3, \ldots\}, \qquad \text{(C.l.t. \& variance appr.)}$$

are asymptotically valid α*-confidence intervals.*

Let us briefly compare the (asymptotically valid) confidence intervals presented in (Chebyshev inequality) in Corollary 2.6.13 and (Central limit theorem) in Corollary 2.6.13 respectively. For this the following simple fact (see Lemma 2.22 in Klenke [Klenke(2008)]) is used.

Lemma 2.6.14 (Tails for the normal distribution)**.** *For all* $x \in (0, \infty)$ *it holds that*

$$\mathcal{N}_{0,1}([x, \infty)) = \int_x^\infty \frac{e^{-\frac{1}{2}y^2}}{\sqrt{2\pi}}\, dy < \frac{e^{-\frac{1}{2}x^2}}{x\sqrt{2\pi}}. \tag{2.147}$$

Proof of Lemma 2.6.14. Integration by parts implies that for all $x \in (0, \infty)$ it holds that

$$\begin{aligned} \int_x^\infty \frac{e^{-\frac{1}{2}y^2}}{\sqrt{2\pi}}\, dy &= \int_x^\infty \frac{1}{y} \cdot \frac{y\, e^{-\frac{1}{2}y^2}}{\sqrt{2\pi}}\, dy = \left[\frac{1}{y} \cdot \frac{-e^{-\frac{1}{2}y^2}}{\sqrt{2\pi}}\right]_{y=x}^{y=\infty} + \int_x^\infty \frac{1}{y^2} \cdot \frac{-e^{-\frac{1}{2}y^2}}{\sqrt{2\pi}}\, dy \\ &= \frac{e^{-\frac{1}{2}x^2}}{x\sqrt{2\pi}} - \int_x^\infty \frac{1}{y^2} \cdot \frac{e^{-\frac{1}{2}y^2}}{\sqrt{2\pi}}\, dy < \frac{e^{-\frac{1}{2}x^2}}{x\sqrt{2\pi}}. \end{aligned} \tag{2.148}$$

This completes the proof of Lemma 2.6.14. □

In the next step we use Lemma 2.6.14 to establish the following lemma.

Lemma 2.6.15. *For all $\alpha \in [0, 1)$ it holds that*

$$\frac{1}{\sqrt{2\pi}} \int_{\frac{-1}{\sqrt{(1-\alpha)}}}^{\frac{1}{\sqrt{(1-\alpha)}}} e^{-\frac{1}{2}x^2} dx > \alpha. \tag{2.149}$$

Proof of Lemma 2.6.15. First of all, we observe that the fact $\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{1}{2}x^2} dx = 1$ and Lemma 2.6.14 imply that for all $\beta \in [1, \infty)$ it holds that

$$\begin{aligned} \frac{1}{\sqrt{2\pi}} \int_{-\beta}^{\beta} e^{-\frac{1}{2}x^2} dx &= 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-\beta} e^{-\frac{1}{2}x^2} dx - \frac{1}{\sqrt{2\pi}} \int_{\beta}^{\infty} e^{-\frac{1}{2}x^2} dx \\ &= 1 - \frac{2}{\sqrt{2\pi}} \int_{\beta}^{\infty} e^{-\frac{1}{2}x^2} dx > 1 - \frac{2\, e^{-\frac{1}{2}\beta^2}}{\beta\sqrt{2\pi}} \geq 1 - \frac{\sqrt{2}\, e^{-\frac{1}{2}\beta^2}}{\sqrt{\pi}} > 1 - \frac{1}{e^{\frac{1}{2}\beta^2}}. \end{aligned} \tag{2.150}$$

The fact that $\forall\, y \in [0, \infty) \colon 1 + y + \frac{y^2}{2} \leq e^y$ hence proves that for all $\beta \in [1, \infty)$ it holds that

$$\frac{1}{\sqrt{2\pi}} \int_{-\beta}^{\beta} e^{-\frac{1}{2}x^2} dx > 1 - \frac{1}{\left(1 + \frac{\beta^2}{2} + \frac{\beta^4}{8}\right)}. \tag{2.151}$$

The estimate $\forall\, y \in \mathbb{R} \colon y \leq 1 + \frac{y}{2} + \frac{y^2}{8}$ therefore shows that for all $\beta \in [1, \infty)$ it holds that

$$\frac{1}{\sqrt{2\pi}} \int_{-\beta}^{\beta} e^{-\frac{1}{2}x^2} dx > 1 - \frac{1}{\left(1 + \frac{\beta^2}{2} + \frac{\beta^4}{8}\right)} \geq 1 - \frac{1}{\beta^2}. \tag{2.152}$$

This implies that for all $\alpha \in [0, 1)$ it holds that

$$\frac{1}{\sqrt{2\pi}} \int_{\frac{-1}{\sqrt{(1-\alpha)}}}^{\frac{1}{\sqrt{(1-\alpha)}}} e^{-\frac{1}{2}x^2} dx > 1 - (1 - \alpha) = \alpha. \tag{2.153}$$

The proof of Lemma 2.6.15 is thus completed. □

Lemma 2.6.15 proves that if $\mathsf{Var}_P(X_1) > 0$, if $\beta = \gamma$ and if $\frac{1}{\sqrt{2\pi}} \int_{-\beta}^{\gamma} e^{\frac{-x^2}{2}} dx = \alpha$ in Corollary 2.6.13, then the asymptotically valid α-confidence intervals in (Central limit theorem) in Corollary 2.6.13 are smaller than the α-confidence intervals in (Chebyshev inequality) in Corollary 2.6.13. However, observe that the concept of asymptotically valid confidence intervals (see Definition 2.6.9) is a much weaker concept than the concept of an confidence interval (see Definition 2.6.4).

2.7 Monte Carlo algorithms for numerical integration

Let $d \in \mathbb{N}$, let $A \in \mathcal{B}(\mathbb{R}^d)$ be a bounded and Borel measurable set and let $f \in \mathcal{L}^1(B_A; |\cdot|_{\mathbb{R}})$. Suppose in this section that we want to approximate the real number

$$\int_A f(x)\,dx \tag{2.154}$$

by Monte Carlo approximations. For this let $(\Omega, \mathcal{F}, P)$ be a probability space and let $Y_n \colon \Omega \to A$, $n \in \mathbb{N}$, be independent $\mathcal{U}_A$-distributed random variables. In view of (2.58), we rewrite (2.154) by

$$\begin{aligned}\int_A f(x)\,dx &= \lambda_{\mathbb{R}^d}(A) \int_A \frac{f(x)}{\lambda_{\mathbb{R}^d}(A)}\,dx = \lambda_{\mathbb{R}^d}(A) \left[\int_A f(x)\,\mathcal{U}_A(dx)\right] \\ &= \mathbb{E}\big[\lambda_{\mathbb{R}^d}(A) \cdot f(Y_1)\big].\end{aligned} \tag{2.155}$$

Then the random variables

$$\frac{\lambda_{\mathbb{R}^d}(A)}{N}\,(f(Y_1) + \ldots + f(Y_N)) \tag{2.156}$$

for $N \in \mathbb{N}$ are Monte Carlo approximations of $\int_A f(x)\,dx$. If $N \in \mathbb{N}$ is a given natural number, then the following algorithm computes realizations of (2.156).

Monte Carlo approximations I

Output: Realization x of $X \sim P_{\frac{\lambda_{\mathbb{R}^d}(A)}{N}(f(Y_1)+\ldots+f(Y_N))} \approx \int_A f(x)\,dx$

$s = 0$
for $n = 1 \to N$ **do**
  Generate realization y of $Y_n \sim \mathcal{U}_A$
  $s = s + f(y)$
end for
$x = \frac{\lambda_{\mathbb{R}^d}(A)\cdot s}{N}$

The above algorithm requires the knowledge of $\lambda_{\mathbb{R}^d}(A)$ and also requires realizations from $\mathcal{U}_A$-distributed random variables (cf. Lemma 1.2.28). To avoid these requirements, one can choose real numbers $a_1, \ldots, a_d, b_1, \ldots, b_d \in \mathbb{R}$ with $a_1 \le b_1, \ldots, a_d \le b_d$ and $A \subseteq [a_1, b_1] \times \ldots \times [a_d, b_d]$. (Such real numbers exist since A is assumed to be bounded.) Next define a function $\tilde{f} \colon [a_1, b_1] \times \ldots \times [a_d, b_d] \to \mathbb{R}$ by $\tilde{f}(x) := f(x)$ for all $x \in A$ and by $\tilde{f}(x) := 0$ for all $x \in \big([a_1, b_1] \times \ldots \times [a_d, b_d]\big)\backslash A$ and assume that $U_n \colon \Omega \to [a_1, b_1] \times \ldots \times [a_d, b_d]$, $n \in \mathbb{N}$, are independent $\mathcal{U}_{[a_1,b_1]\times\ldots\times[a_d,b_d]}$-distributed random variables. Then we obtain that

$$\begin{aligned}\int_A f(x)\,dx &= \int_{[a_1,b_1]\times\ldots\times[a_d,b_d]} \tilde{f}(x)\,dx \\ &= \lambda_{\mathbb{R}^d}\big([a_1, b_1] \times \ldots \times [a_d, b_d]\big) \int_{[a_1,b_1]\times\ldots\times[a_d,b_d]} \tilde{f}(x)\,\mathcal{U}_{[a_1,b_1]\times\ldots\times[a_d,b_d]}(dx) \\ &= \mathbb{E}_P\left[\left(\prod_{i=1}^{d}(b_i - a_i)\right) \cdot \tilde{f}(U_1)\right].\end{aligned} \tag{2.157}$$

The following algorithm computes realizations of the Monte Carlo approximation

$$\frac{\left(\prod_{i=1}^{d}(b_i - a_i)\right)}{N}\left(\tilde{f}(U_1) + \ldots + \tilde{f}(U_N)\right) \tag{2.158}$$

of $\int_A f(x)\,dx$ where $N \in \mathbb{N}$ is a given natural number.

Monte Carlo approximations II

Output: Realization x of $X \sim P_{\frac{(b_1-a_1)\cdot\ldots\cdot(b_d-a_d)}{N}\left(\tilde{f}(U_1)+\ldots+\tilde{f}(U_N)\right)} \approx \int_A f(x)\,dx$

$s = 0$
for $n = 1 \to N$ **do**
  Generate realization u of $U_n \sim \mathcal{U}_{[a_1,b_1]\times\cdots\times[a_d,b_d]}$
  if $u \in A$ **then**
    $s = s + f(u)$
  end if
end for
$x = \frac{(b_1-a_1)\cdot\ldots\cdot(b_d-a_d)\cdot s}{N}$

Next we are interested in an algorithm that returns a confidence interval. For this let $\alpha \in (0,1)$ and $c \in [0,\infty)$ be real numbers and assume that

$$\left(\prod_{i=1}^{d}(b_i - a_i)\right)\sqrt{\mathsf{Var}_P(\tilde{f}(U_1)} \leq c. \tag{2.159}$$

For instance, if f is bounded, then

$$\begin{aligned}
&\left(\prod_{i=1}^{d}(b_i - a_i)\right)\sqrt{\mathsf{Var}\left(\tilde{f}(U_1)\right)} \leq \left(\prod_{i=1}^{d}(b_i - a_i)\right)\left\|\tilde{f}(U_1)\right\|_{\mathcal{L}^2(P;|\cdot|_{\mathbb{R}})} \\
&= \sqrt{\left(\prod_{i=1}^{d}(b_i - a_i)\right)\int_{[a_1,b_1]\times\ldots\times[a_d,b_d]}\left|\tilde{f}(x)\right|^2 dx} \\
&= \sqrt{\left(\prod_{i=1}^{d}(b_i - a_i)\right)\int_A |f(x)|^2\,dx} \leq \left(\prod_{i=1}^{d}(b_i - a_i)\right)\left[\sup_{x\in A}|f(x)|\right] < \infty.
\end{aligned} \tag{2.160}$$

Next define random variables $X_N^1, X_N^2 \colon \Omega \to \mathbb{R}$, $N \in \mathbb{N}$, by

$$\begin{aligned}
X_N^1 &:= \frac{1}{N}\left(\prod_{i=1}^{d}(b_i - a_i)\right)\left(\tilde{f}(U_1) + \ldots + \tilde{f}(U_N)\right) - \frac{c}{\sqrt{(1-\alpha)N}}, \\
X_N^2 &:= \frac{1}{N}\left(\prod_{i=1}^{d}(b_i - a_i)\right)\left(\tilde{f}(U_1) + \ldots + \tilde{f}(U_N)\right) + \frac{c}{\sqrt{(1-\alpha)N}}
\end{aligned} \tag{2.161}$$

for all $N \in \mathbb{N}$. Corollary 2.6.8 then shows that

$$P\left(\int_A f(x)\,dx \in \left[X_N^1, X_N^2\right]\right) \geq \alpha \tag{2.162}$$

for all $N \in \mathbb{N}$. The following algorithm returns realizations of X_N^1 and X_N^2 where $N \in \mathbb{N}$ is a given natural number.

Confidence interval

Output: Realization (x_1, x_2) of (X_N^1, X_N^2)
$s = 0$
for $n = 1 \to N$ **do**
Generate realization u of $U_n \sim \mathcal{U}_{[a_1,b_1]\times\cdots\times[a_d,b_d]}$
if $u \in A$ **then**
$s = s + f(u)$
end if
end for
$x_1 = \frac{(b_1-a_1)\cdot\ldots\cdot(b_d-a_d)\cdot s}{N} - \frac{c}{\sqrt{(1-\alpha)N}}$
$x_2 = \frac{(b_1-a_1)\cdot\ldots\cdot(b_d-a_d)\cdot s}{N} + \frac{c}{\sqrt{(1-\alpha)N}}$

The convidence level $\alpha \in (0,1)$ can be increased to $1-(1-\alpha)^2 \in (\alpha, 1)$ if we rerun the above algorithm and then enlarge the confidence interval in the sense of Lemma 2.6.6. The above algorithm can be used if the real number $c \in [0,\infty)$ in (2.159) is explicitly known. We now consider the case where $c \in [0,\infty)$ in (2.159) is not explicitly known. In that case we additionally assume that $\int_A |f(x)|_{\mathbb{R}}^4 \, dx < \infty$. Next define a real number $\beta \in (0,\infty)$ through

$$\frac{1}{\sqrt{2\pi}} \int_{-\beta}^{\beta} e^{-\frac{1}{2}x^2} \, dx = \alpha \tag{2.163}$$

and define random variables $E_N, V_N, Z_N^1, Z_N^2 \colon \Omega \to \mathbb{R}$, $N \in \{2,3,\ldots\}$, through

$$\begin{aligned} E_N &:= \frac{1}{N} \left(\prod_{i=1}^{d} (b_i - a_i) \right) \left(\tilde{f}(U_1) + \ldots + \tilde{f}(U_N) \right), \\ V_N &:= \frac{1}{(N-1)} \sum_{n=1}^{N} \left(\left(\prod_{i=1}^{d} (b_i - a_i) \right) \tilde{f}(U_n) - E_N \right)^2, \\ Z_N^1 &:= E_N - \frac{\beta\sqrt{V_N}}{\sqrt{N}} \qquad \text{and} \qquad Z_N^2 := E_N + \frac{\beta\sqrt{V_N}}{\sqrt{N}}, \end{aligned} \tag{2.164}$$

for all $N \in \{2,3,\ldots\}$. The value of β can, e.g., be computed approximatively with the Matlab function "*erfinv*". To be more specific, note that

$$\alpha = \frac{1}{\sqrt{2\pi}} \int_{-\beta}^{\beta} e^{-\frac{1}{2}x^2} \, dx = \frac{1}{\sqrt{\pi}} \int_{-\beta/\sqrt{2}}^{\beta/\sqrt{2}} e^{-x^2} \, dx = \frac{2}{\sqrt{\pi}} \int_{0}^{\beta/\sqrt{2}} e^{-x^2} \, dx \approx \text{“}\mathit{erf(\beta/sqrt(2))}\text{”} \tag{2.165}$$

and this illustrates that the Matlab command "*erfinv(α)*sqrt(2)*" returns an approximation of $\beta \in (0,\infty)$. In the next step we observe that Corollary 2.6.12 proves that

$$\liminf_{N\to\infty} P\left(\int_A f(x) \, dx \in \left[Z_N^1, Z_N^2 \right] \right) \geq \alpha. \tag{2.166}$$

In addition, note that Lemma 2.5.2 proves that for all $N \in \{2, 3, \dots\}$ and all $k \in \{1, 2\}$ it holds that

$$\begin{aligned} Z_N^k &= E_N + \frac{(-1)^k \beta \sqrt{V_N}}{\sqrt{N}} \\ &= E_N + \frac{(-1)^k \beta}{\sqrt{N}} \left[\frac{\sum_{n=1}^N |\tilde{f}(U_n) \prod_{i=1}^d (b_i - a_i) - E_N|^2}{(N-1)} \right]^{1/2} \\ &= E_N + \frac{(-1)^k \beta \prod_{i=1}^d (b_i - a_i)}{\sqrt{N}} \left[\frac{[\sum_{n=1}^N |\tilde{f}(U_n)|^2] - \frac{1}{N} |\sum_{n=1}^N \tilde{f}(U_n)|^2}{(N-1)} \right]^{1/2} \\ &= \frac{\prod_{i=1}^d (b_i - a_i)}{\sqrt{N}} \left[\frac{\tilde{f}(U_1) + \ldots + \tilde{f}(U_N)}{\sqrt{N}} + \beta \, (-1)^k \left[\frac{[\sum_{n=1}^N |\tilde{f}(U_n)|^2] - \frac{1}{N} |\sum_{n=1}^N \tilde{f}(U_n)|^2}{(N-1)} \right]^{1/2} \right]. \end{aligned} \tag{2.167}$$

The following algorithm uses (2.167) and returns realizations of Z_N^1 and Z_N^2 where $N \in \{2, 3, \dots\}$ is a given natural number.

Asymptotically valid confidence intervals

Output: Realization (z_1, z_2) of (Z_N^1, Z_N^2)
$s = 0$
$r = 0$
for $n = 1 \to N$ **do**
  Generate realization u of $U_n \sim \mathcal{U}_{[a_1,b_1] \times \cdots \times [a_d,b_d]}$
  if $u \in A$ **then**
    $x = f(u)$
    $s = s + x$
    $r = r + x^2$
  end if
end for
$q = \sqrt{N}$
$r = \beta \sqrt{\frac{1}{(N-1)} (r - \frac{s^2}{N})}$
$\nu = \frac{(b_1 - a_1) \cdot \ldots \cdot (b_d - a_d)}{q}$
$z_1 = \nu \left(\frac{s}{q} - r \right)$
$z_2 = \nu \left(\frac{s}{q} + r \right)$

If $N \in \{2, 3, \dots\}$ is large, then the third last command in the above algorithm may result in a roundoff error that can not be neglected. There are a few algorithms in the literature that significantly reduce these roundoff errors (see, e.g., [Knuth(1998)] and the references therein). We close this secton with the following remark on possibly unbounded domains of integration.

Remark 2.7.1. *Let $d \in \mathbb{N}$, let $A \in \mathcal{B}(\mathbb{R}^d)$ be a Borel measurable set and let $f \in \mathcal{L}^1(B_A; \mathbb{R})$. Then $\int_A f(x)\,dx$ can also be approximated by Monte Carlo approximations although A is not assumed to be bounded. For this one just needs to rewrite $\int_A f(x)\,dx$ as the expectation of an random variable from which one can generate realizations. For instance, define $\tilde{f}\colon \mathbb{R}^d \to \mathbb{R}$ through $\tilde{f}(x) := \sqrt{2\pi} \cdot f(x) \cdot e^{\frac{1}{2}\|x\|^2_{\mathbb{R}^d}}$ for all $x \in A$ and through $\tilde{f}(x) := 0$ for all $x \in \mathbb{R}^d \backslash A$. Moreover, let $(\Omega, \mathcal{F}, P)$ be a probability space and let $Y_n\colon \Omega \to \mathbb{R}^d$, $n \in \mathbb{N}$, be independent $\mathcal{N}_{0,I}$-distributed random variables. Then*

$$\int_A f(x)\,dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}^d} \tilde{f}(x) \cdot e^{-\frac{1}{2}\|x\|^2_{\mathbb{R}^d}}\,dx = \mathbb{E}_P\Big[\tilde{f}(Y_1)\Big] \tag{2.168}$$

and the random variables

$$\frac{\tilde{f}(Y_1) + \ldots + \tilde{f}(Y_N)}{N} \tag{2.169}$$

for $N \in \mathbb{N}$ are thus Monte Carlo approximations of $\int_A f(x)\,dx$.

3 Stochastic processes and Itô stochastic calculus

This chapter reviews basic ideas, concepts, and facts from the literature on stochastic processes and Itô stochastic calculus to the extent that they are necessary for the analysis of numerical approximation schemes for stochastic ordinary differential equations (SODEs) presented in the subsequent chapters.

Disclaimer: These notes are not intended as a mathematical introduction to stochastic analysis or stochastic processes; they merely recapitulate main definitions, fix notation to be used in the sequel, and collect key results from stochastic analysis which will be used in the chapters ahead for the numerical analysis of SODEs. We therefore often give results without proof but indicate where these can be found.

For a self-contained introduction to stochastic processes and stochastic analysis, we refer to the course "Stochastic Processes and Stochastic Analysis" and, in particular, to the Lecture Notes by M. Schweizer. We also recommend [Kuo(2006), Øksendal(2003), Ikeda and Watanabe(1989), Karatzas and Shreve(1988), Revuz and Yor(1999), Jacod and Shiryaev(2003)] as references for the content of this chapter. For an introduction to SODEs with more general semimartingale integrators, we refer, e.g., to [Métivier(1982), Protter(2004), Kühn(2004)].

3.1 Stochastic processes

Let us begin with the definition of a *stochastic process*.

Definition 3.1.1 (Stochastic process)**.** *We say that X is an $\mathbf{S}$-valued stochastic process with time set $\mathbb{T}$ on $\mathbf{\Omega}$ (we say that X is an $\mathbf{S}$-valued stochastic process with time set $\mathbb{T}$, we say that X is a stochastic process on $\mathbf{\Omega}$, we say that X is a stochastic process) if and only if there exist S, $\mathcal{S}$, Ω, $\mathcal{F}$, P such that it holds*

(i) that $\mathbb{T} \subseteq \mathbb{R}$,

(ii) that $\mathbf{\Omega} = (\Omega, \mathcal{F}, P)$ is a probability space,

(iii) that $\mathbf{S} = (S, \mathcal{S})$ is a measurable space,

(iv) that $X \in \mathbb{M}(\mathbb{T} \times \Omega, S)$, and

(v) that for every $t \in \mathbb{T}$ it holds that

$$\Omega \ni \omega \mapsto X(t, \omega) \in S \tag{3.1}$$

is $\mathcal{F}/\mathcal{S}$-measurable.

Definition 3.1.2 (Marginals)**.** *Let $\mathbb{T} \subseteq \mathbb{R}$ be a subset of the real numbers, let $t \in \mathbb{T}$, let $(\Omega, \mathcal{F}, P)$ be a probability space, let $(S, \mathcal{S})$ be a measurable space, and let $X : \mathbb{T} \times \Omega \to S$ be a function. Then we denote by $X_t : \Omega \to S$ the function which satisfies for all $\omega \in \Omega$ that*

$$X_t(\omega) = X(t, \omega). \tag{3.2}$$

Definition 3.1.3 (State space)**.** *We say that S is the state space of the stochastic process X (we say that S is the state space of X) if and only if there exist $\mathcal{S}$, $\mathbb{T}$, $\mathbf{\Omega}$ such that X is an $(S, \mathcal{S})$-valued stochastic process with time set $\mathbb{T}$ on $\mathbf{\Omega}$.*

In many examples of stochastic processes it holds that there exist $T \in (0, \infty)$ and $N \in \mathbb{N}$ such that the time interval set $\mathbb{T}$ appearing in Definition 3.1.1 is equal to $[0, \infty)$, $[0, T]$, $\mathbb{N}_0$, or $\{0, 1, \ldots, N\}$. Moreover, in the case of many examples of stochastic processes it holds that there exists a natural number $d \in \mathbb{N}$ such that the measurable space $(S, \mathcal{S})$ in Definition 3.1.1 is equal to $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

Example 3.1.4 (A simple stochastic process)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $Y_n \colon \Omega \to \{-1, 1\}$, $n \in \mathbb{N}$, be P-independent* $\mathrm{Unif}_{\{-1,1\}}$*-distributed random variables, and let $X \colon \mathbb{N}_0 \times \Omega \to \mathbb{R}$ be the function which satisfies for all $n \in \mathbb{N}_0$ that*

$$X_n = \sum_{k=1}^{n} Y_k. \tag{3.3}$$

Then it holds that X is an $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$-valued stochastic process with time set $\mathbb{N}_0$ on $(\Omega, \mathcal{F}, P)$.

Remark 3.1.5. *Let $\mathbb{T} \subseteq \mathbb{R}$ be a subset of the real numbers, let $(\Omega, \mathcal{F}, P)$ be a probability space, let $(S, \mathcal{S})$ be a measurable space, and let $X \colon \mathbb{T} \times \Omega \to S$ be a stochastic process. There are several ways how a stochastic process $X \colon \mathbb{T} \times \Omega \to S$ can be interpreted:*

- *as a two-parameter function $X \colon \mathbb{T} \times \Omega \to S$ from the cartesian product $\mathbb{T} \times \Omega$ to S,*
- *as a one-parameter family $X_t \colon \Omega \to S$, $t \in \mathbb{T}$, of random variables from Ω to S with the index set $\mathbb{T} \subseteq \mathbb{R}$ and*
- *as a family $\mathbb{T} \ni t \mapsto X_t(\omega) \in S$, $\omega \in \Omega$, of functions from $\mathbb{T}$ to S with the index set Ω (family of "sample paths").*

Example 3.1.6 (Deterministic functions)**.** *Let $T \in (0, \infty)$ be a real number, let $f \colon [0, T] \to \mathbb{R}$ be an arbitrary not necessarily $\mathcal{B}([0, T])/\mathcal{B}(\mathbb{R})$-measurable function, let $(\Omega, \mathcal{F}, P)$ be the probability space given by*

$$(\Omega, \mathcal{F}, P) = \big(\{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \delta_\emptyset^{\{\emptyset\}}\big), \tag{3.4}$$

and let $X \colon [0, T] \times \Omega \to \mathbb{R}$ be the function which satisfies for all $t \in [0, T]$ that

$$X_t(\emptyset) = f(t). \tag{3.5}$$

Then it holds that X is an $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$-valued stochastic process with time set $[0, T]$ on $(\Omega, \mathcal{F}, P)$. This example, in particular, illustrates that the sample paths of a stochastic process do not necessarily need to be Borel measurable functions.

We are often interested in stochastic processes whose sample paths enjoy certain regularity properties (cf. Example 3.1.6) such as, for instance, stochastic processes whose sample paths are continuous functions. This is the subject of the next notion.

Definition 3.1.7 (Stochastic process with continuous sample paths). *We say that X is an $\mathbf{E}$-valued stochastic process with continuous sample paths and time set $\mathbb{T}$ on $\mathbf{\Omega}$ (we say that X is an $\mathbf{E}$-valued stochastic process with continuous sample paths, we say that X is a stochastic process with continuous sample paths) if and only if there exist E, $\mathcal{E}$, Ω, $\mathcal{F}$, P such that it holds*

(i) that $\mathbf{E} = (E, \mathcal{E})$ is a topological space,

(ii) that X is an $(E, \mathcal{B}(E))$-valued stochastic process with time set $\mathbb{T}$ on $\mathbf{\Omega}$,

(iii) that $\mathbf{\Omega} = (\Omega, \mathcal{F}, P)$, and

(iv) that for every $\omega \in \Omega$ it holds that $\mathbb{T} \ni t \mapsto X_t(\omega) \in E$ is a continuous function.

Next we address two notions that somehow describe when two stochastic processes are "equal up to sets of measure zero".

Definition 3.1.8 (Modifications). *Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $(S, \mathcal{S})$ be a measurable space, let $\mathbb{T} \subseteq \mathbb{R}$ be a set, and let $X, Y : \mathbb{T} \times \Omega \to S$ be stochastic processes. Then we say that X and Y are $(S, \mathcal{S})$-valued modifications of each other on $(\Omega, \mathcal{F}, P)$ (we say that X and Y are modifications of each other, we say that X is a modification of Y, we say that Y is a modification of X) if and only if for every $t \in \mathbb{T}$ it holds that there exists an event $A \in \mathcal{F}$ with $P(A) = 1$ and*

$$A \subseteq \{X_t = Y_t\}. \tag{3.6}$$

Exercise 3.1.9. *Prove or disprove the following statement: For all measurable spaces $(\Omega, \mathcal{F})$ it holds that $\{(\omega, \omega) \in \Omega^2 : \omega \in \Omega\} \in \mathcal{F} \otimes \mathcal{F}$.*

Exercise 3.1.10. *Specify explicitly measurable spaces $(\Omega, \mathcal{F})$ and $(S, \mathcal{S})$ and $\mathcal{F}/\mathcal{S}$-measurable functions $X, Y : \Omega \to S$ such that*

$$\{X = Y\} = \{\omega \in \Omega : X(\omega) = Y(\omega)\} \notin \mathcal{F}. \tag{3.7}$$

Prove that your result is correct.

Definition 3.1.11 (Indistinguishablility). *Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $(S, \mathcal{S})$ be a measurable space, let $\mathbb{T} \subseteq \mathbb{R}$ be a set, and let $X, Y : \mathbb{T} \times \Omega \to S$ be stochastic processes. Then we say that X and Y are indistinguishable from each other with values in $(S, \mathcal{S})$ on $(\Omega, \mathcal{F}, P)$ (we say that X and Y are indistinguishable from each other, we say that X is indistinguishable from Y, we say that Y is indistinguishable from X) if and only if there exists an event $A \in \mathcal{F}$ with $P(A) = 1$ and*

$$A \subseteq \left(\cap_{t \in \mathbb{T}} \{X_t = Y_t\}\right). \tag{3.8}$$

Let us illustrate Definitions 3.1.8 and 3.1.11 through a simple example (see, e.g., [Kühn(2004)]).

Example 3.1.12. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $U\colon \Omega \to \mathbb{R}$ be an $\mathcal{U}_{[0,1]}$-distributed random variable with $U(\Omega) \subseteq [0,1]$, and let $X, Y\colon [0,1] \times \Omega \to \mathbb{R}$ be the functions which satisfy for all $\omega \in \Omega$, $t \in [0,1]$ that*

$$X_t(\omega) = 0 \qquad \text{and} \qquad Y_t(\omega) = \begin{cases} 1 & : t = U(\omega) \\ 0 & : t \neq U(\omega) \end{cases}. \tag{3.9}$$

Then

(i) it holds that

$$\Big\{\omega \in \Omega\colon \big(\forall\, t \in [0,T]\colon X_t(\omega) = Y_t(\omega)\big)\Big\} = \Big\{\omega \in \Omega\colon \big(\forall\, t \in [0,T]\colon Y_t(\omega) = 0\big)\Big\} = \emptyset \tag{3.10}$$

and

(ii) it holds for all $t \in [0,T]$ that

$$P(X_t = Y_t) = P(Y_t = 0) = P(U \neq t) = 1. \tag{3.11}$$

This shows that X and Y are modification of each other but X and Y are not indistinguishable from each other.

Lemma 3.1.13 (Modifications with continuous sample paths)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $T \in [0,\infty)$, $m \in \mathbb{N}$, let $X, Y\colon [0,T] \times \Omega \to \mathbb{R}^m$ be stochastic processes with continuous sample paths which satisfy for all $t \in [0,T]$ that $P(X_t = Y_t) = 1$. Then it holds that X and Y are indistinguishable from each other.*

Proof of Lemma 3.1.13. First of all, observe that for all $t \in [0,T]$ it holds that

$$\{X_t = Y_t\} \in \mathcal{F}. \tag{3.12}$$

This implies that

$$\cap_{t\in[0,T]\cap\mathbb{Q}} \{X_t = Y_t\} \in \mathcal{F}. \tag{3.13}$$

The continuity of the sample path of X and Y hence proves that

$$\begin{aligned} \cap_{t\in[0,T]} \{X_t = Y_t\} &= \{\omega \in \Omega\colon \forall\, t \in [0,T]\colon X_t(\omega) = Y_t(\omega)\} \\ &= \{\omega \in \Omega\colon \forall\, t \in [0,T] \cap \mathbb{Q}\colon X_t(\omega) = Y_t(\omega)\} \\ &= \cap_{t\in[0,T]\cap\mathbb{Q}} \{X_t = Y_t\} \in \mathcal{F}. \end{aligned} \tag{3.14}$$

This proves that

$$\begin{aligned} &P\big(\cap_{t\in[0,T]} \{X_t = Y_t\}\big) = P\big(\cap_{t\in[0,T]\cap\mathbb{Q}} \{X_t = Y_t\}\big) \\ &= 1 - P\big(\Omega \backslash \big[\cap_{t\in[0,T]\cap\mathbb{Q}} \{X_t = Y_t\}\big]\big) = 1 - P\big(\cup_{t\in[0,T]\cap\mathbb{Q}} \{X_t \neq Y_t\}\big) \\ &\geq 1 - \sum_{t\in[0,T]\cap\mathbb{Q}} P(X_t \neq Y_t) = 1 - \sum_{t\in[0,T]\cap\mathbb{Q}} 0 = 1. \end{aligned} \tag{3.15}$$

The proof of Lemma 3.1.13 is thus completed. □

3.2 Measurability properites of stochastic processes

3.2.1 Filtrations

The following definition is crucial for the investigation of measurability properites of stochastic processes.

Definition 3.2.1 (Filtration). *We say that $\mathbb{F}$ is a filtration on $\mathbf{\Omega}$ if and only if there exist Ω and $\mathcal{F}$ such that it holds*

(i) that $\mathbf{\Omega} = (\Omega, \mathcal{F})$ *is a measurable space,*

(ii) that $\mathbb{F}$ *is a mapping,*

(iii) that $\operatorname{domain}(\mathbb{F}) \subseteq [-\infty, \infty]$,

(iv) that $\operatorname{codomain}(\mathbb{F}) = \mathcal{P}(\mathcal{P}(\Omega))$, *and*

(v) that for all $t_1, t_2 \in \operatorname{domain}(\mathbb{F})$ *with* $t_1 \leq t_2$ *it holds that* $\sigma_\Omega(\mathbb{F}_{t_1}) = \mathbb{F}_{t_1} \subseteq \mathbb{F}_{t_2} \subseteq \mathcal{F}$.

3.2.1.1 Continuity properties of filtrations

Definition 3.2.2 (Filtrations associated to a filtration). *Let* $(\Omega, \mathcal{F})$ *be a measurable space and let* $\mathbb{F}$ *be a filtration on* $(\Omega, \mathcal{F})$. *Then we denote by* $\mathbb{F}^- \in \mathbb{M}(\operatorname{domain}(\mathbb{F}), \mathcal{P}(\mathcal{P}(\Omega)))$ *and* $\mathbb{F}^+ \in \mathbb{M}(\operatorname{domain}(\mathbb{F}), \mathcal{P}(\mathcal{P}(\Omega)))$ *the filtrations on* $(\Omega, \mathcal{F})$ *which satisfy for all* $t \in \operatorname{domain}(\mathbb{F})$ *that*

$$\mathbb{F}_t^- = \begin{cases} \sigma_\Omega\big(\cup_{s \in \mathbb{T} \cap [-\infty, t)} \mathbb{F}_s\big) & : t > \inf(\operatorname{domain}(\mathbb{F})) \\ \mathbb{F}_t & : t = \inf(\operatorname{domain}(\mathbb{F})) \end{cases} \tag{3.16}$$

and

$$\mathbb{F}_t^+ = \begin{cases} \cap_{s \in \mathbb{T} \cap (t, \infty]} \mathbb{F}_s & : t < \sup(\operatorname{domain}(\mathbb{F})) \\ \mathbb{F}_t & : t = \sup(\operatorname{domain}(\mathbb{F})) \end{cases}. \tag{3.17}$$

Lemma 3.2.3 (Properties of the filtrations associated to a filtration). *Let* $(\Omega, \mathcal{F})$ *be a measurable space and let* $\mathbb{F}$ *be a filtration on* $(\Omega, \mathcal{F})$. *Then*

(i) it holds for all $t \in \operatorname{domain}(\mathbb{F})$ *that* $\mathbb{F}_t^- \subseteq \mathbb{F}_t \subseteq \mathbb{F}_t^+$,

(ii) it holds for all $s, t \in \operatorname{domain}(\mathbb{F})$ *with* $s < t$ *that* $\mathbb{F}_s \subseteq \mathbb{F}_t^- \subseteq \mathbb{F}_t$, *and*

(iii) it holds for all $s, t \in \operatorname{domain}(\mathbb{F})$ *with* $s > t$ *that* $\mathbb{F}_t \subseteq \mathbb{F}_t^+ \subseteq \mathbb{F}_s$.

Proof of Lemma 3.2.3. Items (i)–(iii) are an immediate consequence of Definition 3.2.1, (3.16), and (3.17). The proof of Lemma 3.2.3 is thus completed. □

Lemma 3.2.4 (Further properties of the filtrations associated to a filtration)**.** *Let* $(\Omega, \mathcal{F})$ *be a measurable space, let* $a \in [-\infty, \infty]$*,* $b \in [a, \infty]$*, and let* $(\mathbb{F}_t)_{t\in[a,b]}$ *be a filtration on* $(\Omega, \mathcal{F})$*. Then it holds for all* $t \in [a, b]$ *that*

$$\big((\mathbb{F}_s^-)_{s\in[a,b]}\big)_t^- = \mathbb{F}_t^- \qquad \textit{and} \qquad \big((\mathbb{F}_s^+)_{s\in[a,b]}\big)_t^+ = \mathbb{F}_t^+. \tag{3.18}$$

Proof of Lemma 3.2.4. Throughout this proof assume w.l.o.g. that $a < b$. Next note that Item (ii) of Lemma 3.2.3 ensures that for all $t \in (a, b]$, $r \in [a, t) = [a, b] \cap [-\infty, t)$ it holds that

$$\mathbb{F}_r \subseteq \big(\cup_{s\in[a,b]\cap[-\infty,t)} \mathbb{F}_s^-\big). \tag{3.19}$$

This implies for all $t \in (a, b]$ that

$$\big(\cup_{s\in[a,b]\cap[-\infty,t)} \mathbb{F}_s\big) \subseteq \big(\cup_{s\in[a,b]\cap[-\infty,t)} \mathbb{F}_s^-\big). \tag{3.20}$$

Hence, we obtain for all $t \in (a, b]$ that

$$\mathbb{F}_t^- = \sigma_\Omega\big(\cup_{s\in[a,b]\cap[-\infty,t)} \mathbb{F}_s\big) = \sigma_\Omega\big(\cup_{s\in[a,b]\cap[-\infty,t)} \mathbb{F}_s^-\big) = \big((\mathbb{F}_s^-)_{s\in[a,b]}\big)_t^-. \tag{3.21}$$

Next observe that Item (iii) in Lemma 3.2.3 shows that for all $t \in [a, b)$, $r \in (t, b] = [a, b] \cap (t, \infty]$ it holds that

$$\big(\cap_{s\in[a,b]\cap(t,\infty]} \mathbb{F}_s^+\big) \subseteq \mathbb{F}_r. \tag{3.22}$$

This implies for all $t \in [a, b)$ that

$$\big((\mathbb{F}_s^+)_{s\in[a,b]}\big)_t^+ = \big(\cap_{s\in[a,b]\cap(t,\infty]} \mathbb{F}_s^+\big) = \big(\cap_{s\in[a,b]\cap(t,\infty]} \mathbb{F}_s\big) = \mathbb{F}_t^+. \tag{3.23}$$

Combining (3.21) and (3.23) completes the proof of Lemma 3.2.4. ☐

Definition 3.2.5 (Left-continuity of a filtration)**.** *We say that* $\mathbb{F}$ *is a left-continuous filtration on* $\boldsymbol{\Omega}$ *if and only if it holds that* $\mathbb{F}$ *is a filtration on* $\boldsymbol{\Omega}$ *which satisfies for all* $t \in \text{domain}(\mathbb{F})$ *that*

$$\mathbb{F}_t = \mathbb{F}_t^-. \tag{3.24}$$

Definition 3.2.6 (Right-continuity of a filtration)**.** *We say that* $\mathbb{F}$ *is a right-continuous filtration on* $\boldsymbol{\Omega}$ *if and only if it holds that* $\mathbb{F}$ *is a filtration on* $\boldsymbol{\Omega}$ *which satisfies for all* $t \in \text{domain}(\mathbb{F})$ *that*

$$\mathbb{F}_t = \mathbb{F}_t^+. \tag{3.25}$$

Let $(\Omega, \mathcal{F})$, let $\mathbb{T} \subseteq [-\infty, \infty]$ be a set, and let $(\mathbb{F}_t)_{t\in\mathbb{T}}$ be a filtration on $(\Omega, \mathcal{F})$. Then, in general, it does not hold that for all $t \in \mathbb{T}$ with $t > \inf(\mathbb{T})$ it holds that

$$\mathbb{F}_t^- = \cup_{s\in\mathbb{T}\cap(-\infty,t)}\mathbb{F}_s \tag{3.26}$$

because, in general, it does not hold that for all $t \in \mathbb{T}$ with $t > \inf(\mathbb{T})$ it holds that $\cup_{s\in\mathbb{T}\cap(-\infty,t)}\mathbb{F}_s$ is a sigma-algebra on Ω. This is illustrated in the next example.

Example 3.2.7. *Let $(\Omega, \mathcal{F})$ be the measurable space given by $\Omega = \mathbb{N}_0 = \{0, 1, 2, \dots\}$ and $\mathcal{F} = \mathcal{P}(\Omega)$, let $\mathbb{T} \subseteq \mathbb{R}$ be the set given by $\mathbb{T} = [0, 1]$, and let $\mathbb{F}_t \subseteq \mathcal{P}(\Omega)$, $t \in \mathbb{T}$, be the sets which satisfy for all $n \in \mathbb{N}_0$, $t \in [1 - 1/2^n, 1 - 1/2^{(n+1)})$ that $\mathbb{F}_1 = \mathcal{P}(\Omega)$ and*

$$\mathbb{F}_t = \sigma_\Omega\big(\{\{0\}, \{1\}, \{2\}, \dots, \{n\}\}\big). \tag{3.27}$$

Then

(i) observe

- *that for all $t \in [0, 1/2)$ it holds that $\mathbb{F}_t = \sigma_\Omega\big(\{\{0\}\}\big)$,*
- *that for all $t \in [1/2, 3/4)$ it holds that $\mathbb{F}_t = \sigma_\Omega\big(\{\{0\}, \{1\}\}\big)$,*
- *that for all $t \in [3/4, 7/8)$ it holds that $\mathbb{F}_t = \sigma_\Omega\big(\{\{0\}, \{1\}, \{2\}\}\big)$,*
- *. . . ,*

(ii) observe that $(\mathbb{F}_t)_{t\in\mathbb{T}}$ is a right-continuous filtration on $(\Omega, \mathcal{F})$,

(iii) observe that $(\mathbb{F}_t)_{t\in\mathbb{T}}$ is not a left-continuous filtration on $(\Omega, \mathcal{F})$,

(iv) observe that for all $n \in \Omega$ it holds that

$$\{n\} \in \cup_{s\in\mathbb{T}\cap(-\infty,1)}\mathbb{F}_s = \cup_{s\in[0,1)}\mathbb{F}_s, \tag{3.28}$$

(v) observe that

$$\cup_{n\in\{0,2,4,6,\dots\}} \{n\} = \{0, 2, 4, 6, \dots\} \notin \cup_{s\in\mathbb{T}\cap(-\infty,1)}\mathbb{F}_s = \cup_{s\in[0,1)}\mathbb{F}_s, \tag{3.29}$$

and

(vi) observe that $\cup_{s\in\mathbb{T}\cap(-\infty,1)}\mathbb{F}_s = \cup_{s\in[0,1)}\mathbb{F}_s$ is not a sigma-algebra.

Class exercise 3.2.8. *Let $(\Omega, \mathcal{F})$ be a measurable space, let $\mathbb{T} \subseteq [-\infty, \infty]$ be a set, and let $\mathbb{F}_t \subseteq \mathcal{P}(\Omega)$, $t \in \mathbb{T}$, be a filtration on $(\Omega, \mathcal{F})$.*

(i) Is $(\mathbb{F}_t^-)_{t\in\mathbb{T}}$ a left-continuous filtration on $(\Omega, \mathcal{F})$?

(ii) Is $(\mathbb{F}_t^+)_{t\in\mathbb{T}}$ a right-continuous filtration on $(\Omega, \mathcal{F})$?

3.2.1.2 Filtered probability spaces

Definition 3.2.9. *We say that $\mathbf{\Omega}$ is a filtered probability space if and only if it holds that there exist Ω, $\mathcal{F}$, P, $\mathbb{F}$ such that it holds*

(i) that $(\Omega, \mathcal{F}, P)$ is a probability space,

(ii) that $\mathbb{F}$ is a filtration on $(\Omega, \mathcal{F})$, and

(iii) that $\mathbf{\Omega} = (\Omega, \mathcal{F}, P, \mathbb{F})$.

3.2.1.3 Stochastic bases

Next we present the notions of a normal filtration (cf., e.g., Definition 2.1.11 in [Prévôt and Röckner(2007)]) and of a stochastic basis (cf. Appendix E in [Prévôt and Röckner(2007)]).

Definition 3.2.10 (Normal filtration)**.** *We say that* $\mathbb{F}$ *is a normal filtration on* $\mathbf{\Omega}$ *if and only if there exist* Ω, $\mathcal{F}$, P *such that it holds*

(i) that $\mathbf{\Omega} = (\Omega, \mathcal{F}, P)$ *is a probability space,*

(ii) that $\mathbb{F}$ *is a right-continuous filtration on* $(\Omega, \mathcal{F})$*, and*

(iii) that $\{A \in \mathcal{F} \colon P(A) = 0\} \subseteq (\cap_{t \in \mathrm{domain}(\mathbb{F})} \mathbb{F}_t)$.

Definition 3.2.11 (Stochastic basis)**.** *We say that* $\mathbf{\Omega}$ *is a stochastic basis if and only if there exist* Ω, $\mathcal{F}$, P, $\mathbb{F}$ *such that it holds*

(i) that $\mathbf{\Omega} = (\Omega, \mathcal{F}, P, \mathbb{F})$ *is a filtered probability space and*

(ii) that $\mathbb{F}$ *is a normal filtration on* $(\Omega, \mathcal{F}, P)$.

Let $a \in [-\infty, \infty]$, $b \in [a, \infty]$, let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $(\mathbb{F}_t)_{t \in [a,b]}$ be a filtration on $(\Omega, \mathcal{F})$. Then sometimes the quadrupel $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t \in [a,b]})$ is called a stochastic basis in the literature although $(\mathbb{F}_t)_{t \in [a,b]}$ is not necessarily normal.

Proposition 3.2.12 (Construction of a stochastic basis)**.** *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space, let* $a \in [-\infty, \infty]$, $b \in [a, \infty]$, *let* $(\mathbb{F}_t)_{t \in [a,b]}$ *be a filtration on* $(\Omega, \mathcal{F})$*, and let* $\mathbb{G}_t \subseteq \mathcal{P}(\Omega)$, $t \in [a, b]$, *be the function which satisfies for all* $t \in [a, b]$ *that*

$$\mathbb{G}_t = \sigma_\Omega(\mathbb{F}_t \cup \{A \in \mathcal{F} \colon P(A) = 0\}) \,. \tag{3.30}$$

Then

(i) it holds that $(\Omega, \mathcal{F}, P, (\mathbb{G}_t^+)_{t \in [a,b]})$ *is a stochastic basis and*

(ii) it holds for all normal filtrations $(\mathbb{H}_t)_{t \in [a,b]}$ *on* $(\Omega, \mathcal{F}, P)$ *with* $\forall\, t \in [a, b] \colon \mathbb{F}_t \subseteq \mathbb{H}_t$ *that* $\forall\, t \in [a, b] \colon \mathbb{G}_t^+ \subseteq \mathbb{H}_t$.

Proof of Proposition 3.2.12. First, observe that (3.30) ensures that for all $t \in [a, b]$ it holds that

$$\{A \in \mathcal{F} \colon P(A) = 0\} \subseteq \mathbb{G}_t. \tag{3.31}$$

Item (i) in Lemma 3.2.3 hence ensures that for all $t \in [a, b]$ it holds that

$$\{A \in \mathcal{F} \colon P(A) = 0\} \subseteq \mathbb{G}_t^+. \tag{3.32}$$

Combining this with Lemma 3.2.4 establishes that $(\mathbb{G}_t^+)_{t \in [a,b]}$ is a normal filtration on $(\Omega, \mathcal{F}, P)$. This proves (i). Next observe that (3.30) ensures that for all normal filtrations

$(\mathbb{H}_t)_{t\in[a,b]}$ on $(\Omega, \mathcal{F}, P)$ with $\forall\, t \in [a, b] \colon \mathbb{F}_t \subseteq \mathbb{H}_t$ it holds that $\forall\, t \in [a, b] \colon \mathbb{G}_t \subseteq \mathbb{H}_t$. This implies that for all normal filtrations $(\mathbb{H}_t)_{t\in[a,b]}$ on $(\Omega, \mathcal{F}, P)$ with $\forall\, t \in [a, b] \colon \mathbb{F}_t \subseteq \mathbb{H}_t$ it holds that

$$\forall\, t \in [a, b] \colon \mathbb{G}_t^+ \subseteq \mathbb{H}_t^+ = \mathbb{H}_t. \tag{3.33}$$

This establishes (ii). The proof of Proposition 3.2.12 is thus completed. □

3.2.2 Adaptivity

Every stochastic process induces a filtration. This is the subject of the next definition.

Definition 3.2.13. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $(S, \mathcal{S})$ be a measurable space, let $\mathbb{T} \subseteq \mathbb{R}$ be a set, and let $X \colon \mathbb{T} \times \Omega \to S$ be a stochastic process. Then we denote by $\mathbb{F}^X = (\mathbb{F}_t^X)_{t\in\mathbb{T}} \in \mathbb{M}(\mathbb{T}, \mathcal{P}(\mathcal{P}(\Omega)))$ the function which satisfies for all $t \in \mathbb{T}$ that*

$$\mathbb{F}_t^X = \sigma_\Omega\big((X_s)_{s\in\mathbb{T}\cap(-\infty,t]}\big) \tag{3.34}$$

and we call $\mathbb{F}^X$ the filtration on $(\Omega, \mathcal{F})$ generated by the $(S, \mathcal{S})$-valued stochastic process X (we call $\mathbb{F}^X$ the filtration generated by X).

Note that $(\mathbb{F}_t^X)_{t\in\mathbb{T}}$ in Definition 3.2.13 is indeed a filtration on $(\Omega, \mathcal{F})$. The next definition relates the notion of a filtration with the notion of a stochastic process and is fundamental in the theory of stochastic integration (which we will treat in Section 3.4 below).

Definition 3.2.14 (Adaptivity)**.** *We say that X is $\mathbb{F}/\mathcal{S}$-adapted (we say that X is $\mathbb{F}$-adapted, we say that X is an $\mathbb{F}/\mathcal{S}$-adapted stochastic process, we say that X is an $\mathbb{F}$-adapted stochastic process) if and only if there exist Ω, $\mathcal{F}$, P, S such that it holds*

(i) that $\mathbb{F}$ is a filtration on $(\Omega, \mathcal{F})$,

(ii) that X is an $(S, \mathcal{S})$-valued stochastic process with time set domain$(\mathbb{F})$ *on $(\Omega, \mathcal{F}, P)$, and*

(iii) that $\forall\, t \in \text{domain}(\mathbb{F}) \colon X_t \in \mathcal{M}(\mathbb{F}_t, \mathcal{S})$.

Class exercise 3.2.15. *Let $\mathbb{T} \subseteq \mathbb{R}$ be a set, let $(\Omega, \mathcal{F}, P)$ be a probability space, let $(S, \mathcal{S})$ be a measurable space, and let $X \colon \mathbb{T} \times \Omega \to S$ be a stochastic process. Does it holds that X is $\mathbb{F}^X/\mathcal{S}$-adapted?*

Lemma 3.2.16 (A characterization for adaptivity)**.** *Let $\mathbb{T} \subseteq \mathbb{R}$ be a set, let $(\Omega, \mathcal{F}, P)$ be a probability space with a filtration $(\mathbb{F}_t)_{t\in\mathbb{T}}$, let $(S, \mathcal{S})$ be a measurable space, and let $X \colon \mathbb{T}\times\Omega \to S$ be a stochastic process. Then the following two statements are equivalent:*

(i) It holds that X is $(\mathbb{F}_t)_{t\in\mathbb{T}}$-adapted.

(ii) It holds for every $t \in \mathbb{T}$ that $\mathbb{F}_t^X \subseteq \mathbb{F}_t$.

Lemma 3.2.16 is an immediate consequence from Definition 3.2.13 and Definition 3.2.14. Let us illustrate the notions presented above through a simple example; cf. Example 3.1.4.

Example 3.2.17. *Let $N \in \{3, 4, \dots\}$, let $(\Omega, \mathcal{F}, P)$ be the probability space given by*

$$(\Omega, \mathcal{F}, P) = \big(\{-1, 1\}^N, \mathcal{P}(\{-1, 1\}^N), \mathrm{Unif}_\Omega\big), \tag{3.35}$$

let $U_n\colon \Omega \to \{-1, 1\}$, $n \in \{1, 2, \dots, N\}$, be the functions which satisfy for all $\omega = (\omega_1, \omega_2, \dots, \omega_N) \in \Omega$, $n \in \{1, 2, \dots, N\}$ that

$$U_n(\omega) = \omega_n, \tag{3.36}$$

let $\mathbb{T} \subseteq \mathbb{R}$ be the set given by $\mathbb{T} = \{0, 1, \dots, N\}$, let $X, Y\colon \mathbb{T} \times \Omega \to \mathbb{R}$ be the functions which satisfy for all $n \in \mathbb{T}$ that

$$X_n = \sum_{k=1}^{n} U_k \qquad \text{and} \qquad Y_n = U_3, \tag{3.37}$$

and let $\mathbb{F}_t$, $t \in \mathbb{T}$, be the sigma-algebras which satisfy for all $t \in \mathbb{T}$ that

$$\mathbb{F}_t = \begin{cases} \{\emptyset, \Omega\} & : t = 0 \\ \sigma_\Omega(U_1) = \left\{ \begin{array}{c} \emptyset, \Omega, \{(-1, i_2, \dots, i_N)\colon i_2, \dots, i_N \in \{-1, 1\}\}, \\ \{(1, i_2, \dots, i_N)\colon i_2, \dots, i_N \in \{-1, 1\}\} \end{array} \right\} & : t = 1 \\ \sigma_\Omega(U_1, U_2) = \sigma_\Omega\left(\begin{array}{c} \{(-1, -1, i_3, \dots, i_N)\colon i_3, \dots, i_N \in \{-1, 1\}\}, \\ \{(-1, 1, i_3, \dots, i_N)\colon i_3, \dots, i_N \in \{-1, 1\}\} \\ \{(1, -1, i_3, \dots, i_N)\colon i_3, \dots, i_N \in \{-1, 1\}\} \\ \{(1, 1, i_3, \dots, i_N)\colon i_3, \dots, i_N \in \{-1, 1\}\} \end{array} \right) & : t = 2 \\ \vdots & \vdots \\ \sigma_\Omega(U_1, \dots, U_N) = \mathcal{P}(\Omega) & : t = N \end{cases}. \tag{3.38}$$

Then observe

(i) that U_1, U_2, $\dots$, U_N are P-independent Unif$_{\{-1,1\}}$-distributed random variables,

(ii) that $(\mathbb{F}_t)_{t \in \mathbb{T}}$ is a filtration on $(\Omega, \mathcal{F})$,

(iii) that $(\mathbb{F}_t)_{t \in \mathbb{T}} = (\mathbb{F}_t^X)_{t \in \mathbb{T}}$, that is, $\forall\, t \in \mathbb{T}\colon \mathbb{F}_t = \mathbb{F}_t^X$,

(iv) that X and Y are $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$-valued stochastic processes with time set $\mathbb{T}$ on $(\Omega, \mathcal{F}, P)$,

(v) that X is $(\mathbb{F}_t)_{t \in \mathbb{T}}/\mathcal{B}(\mathbb{R})$-adapted, but

(vi) that Y is not $(\mathbb{F}_t)_{t \in \mathbb{T}}/\mathcal{B}(\mathbb{R})$-adapted as $Y_0 = U_3$ is not a $\mathbb{F}_0/\mathcal{B}(\mathbb{R})$-measurable function.

3.2.3 Predictability

Definition 3.2.18 (Predictable sigma-algebra). *Let $(\Omega, \mathcal{F})$ be a measurable space, let $\mathbb{F}$ be a filtration on $(\Omega, \mathcal{F})$, and let $\mathbb{T}$ be the set given by $\mathbb{T} = \operatorname{domain}(\mathbb{F})$. Then we denote by $\operatorname{Pred}(\mathbb{F})$ the sigma-algebra given by*

$$\operatorname{Pred}(\mathbb{F}) = \sigma_{\mathbb{T}\times\Omega}\Big(\Big\{((s,t]\cap\mathbb{T})\times A\colon A\in\mathbb{F}_s,\ s\in\mathbb{T}\cap[-\infty,t),\ t\in\mathbb{T}\Big\} \cup \Big\{(\{\inf(\mathbb{T})\}\cap\mathbb{T})\times A)\colon A\in\cap_{t\in\mathbb{T}}\mathbb{F}_t\Big\}\Big) \tag{3.39}$$

and we call $\operatorname{Pred}(\mathbb{F})$ the predictable sigma-algebra of $\mathbb{F}$.

Let $T \in (0,\infty)$ be a real number and let $(\Omega, \mathcal{F})$ be a measurable space with a filtration $(\mathbb{F}_t)_{t\in[0,T]}$. Then note that the definition of $\operatorname{Pred}\big((\mathbb{F}_t)_{t\in[0,T]}\big)$ depends on the specific choice of the filtration $(\mathbb{F}_t)_{t\in[0,T]}$.

Definition 3.2.19 (Predictability). *We say that X is $\mathbb{F}/\mathcal{S}$-predictable (we say that X is $\mathbb{F}$-predictable, we say that X is an $\mathbb{F}$-predictable function, we say that X is an $\mathbb{F}$-predictable stochastic process, we say that X is an $\mathbb{F}/\mathcal{S}$-predictable function, we say that X is an $\mathbb{F}/\mathcal{S}$-predictable stochastic process) if and only if there exist Ω and $\mathcal{F}$ such that it holds*

(i) that $\mathbb{F}$ is a filtration on $(\Omega, \mathcal{F})$ and

(ii) that X is a $\operatorname{Pred}(\mathbb{F})/\mathcal{S}$-measurable function.

Let $T \in (0,\infty)$, let $(\Omega, \mathcal{F}, P)$ be a probability space with a filtration $(\mathbb{F}_t)_{t\in[0,T]}$, let $(S, \mathcal{S})$ be a measurable space, and let $X\colon [0,T]\times\Omega \to S$ be $(\mathbb{F}_t)_{t\in[0,T]}/\mathcal{S}$-predictable. Then it holds that X is a stochastic process, that is, it holds for every $t\in[0,T]$ that X_t is an $\mathcal{F}/\mathcal{S}$-measurable function (see Definition 3.1.1 above for the definition of a stochastic process). This property is an immediate consequence of Corollary 3.2.21 below. Corollay 3.2.21, in turn, is a special case of the following lemma, Lemma 3.2.20. Lemma 3.2.20 also helps us to better understand the notion of a predictable stochastic process (cf., e.g., [Kühn(2004)]).

Lemma 3.2.20 ($(\mathbb{F}_t)_{t\in[0,T]}$-Predictability implies $(\mathbb{F}_t^-)_{t\in[0,T]}$-Adaptivity). *Let $T\in(0,\infty)$ be a real number, let $(\Omega, \mathcal{F}, P)$ be a probability space with a filtration $(\mathbb{F}_t)_{t\in[0,T]}$, let $(S,\mathcal{S})$ be a measurable space, and let $X\colon[0,T]\times\Omega\to S$ be an $(\mathbb{F}_t)_{t\in[0,T]}/\mathcal{S}$-predictable function. Then it holds that $X\colon[0,T]\times\Omega\to S$ is $(\mathbb{F}_t^-)_{t\in[0,T]}/\mathcal{S}$-adapted.*

Proof of Lemma 3.2.20. Lemma 0.2.11 implies that for every $t_0\in[0,T]$, $A_0\in\mathcal{S}$ it holds

that

$$
\begin{aligned}
&\{t_0\} \times X_{t_0}^{-1}(A_0) = (\{t_0\} \times \Omega) \cap \underbrace{X^{-1}(A_0)}_{\in \mathrm{Pred}((\mathbb{F}_t)_{t\in[0,T]})} \\
&\in (\{t_0\} \times \Omega) \Cap \underbrace{\sigma_{[0,T]\times\Omega}\left(\begin{array}{c}\{(s,t] \times A \colon A \in \mathbb{F}_s \text{ and } s,t \in [0,T] \text{ with } s < t\} \\ \cup \{\{0\} \times A \colon A \in \mathbb{F}_0\}\end{array}\right)}_{=\mathrm{Pred}((\mathbb{F}_t)_{t\in[0,T]})} \\
&= \sigma_{\{t_0\}\times\Omega}\left((\{t_0\} \times \Omega) \Cap \left(\begin{array}{c}\{(s,t] \times A \colon A \in \mathbb{F}_s \text{ and } s,t \in [0,T] \text{ with } s < t\} \\ \cup \{\{0\} \times A \colon A \in \mathbb{F}_0\}\end{array}\right)\right) \\
&= \sigma_{\{t_0\}\times\Omega}\left(\begin{array}{c}\{\{t_0\} \times A \colon A \in \mathbb{F}_s \text{ and } s,t \in [0,T] \text{ with } s < t_0 \le t\} \\ \cup \{(\{t_0\} \cap \{0\}) \times A \colon A \in \mathbb{F}_0\}\end{array}\right) \\
&= \sigma_{\{t_0\}\times\Omega}\left(\begin{array}{c}\left\{\{t_0\} \times A \colon A \in \cup_{s\in[0,t_0)}\mathbb{F}_s\right\} \\ \bigcup \left\{\left(\{t_0\} \cap \{0\}\right) \times A \colon A \in \mathbb{F}_0\right\}\end{array}\right) \\
&\subseteq \sigma_{\{t_0\}\times\Omega}\left(\left\{\{t_0\} \times A \colon A \in \mathbb{F}_{t_0}^-\right\}\right) = \left\{\{t_0\} \times A \colon A \in \mathbb{F}_{t_0}^-\right\}.
\end{aligned} \tag{3.40}
$$

This shows that for every $t_0 \in [0,T]$, $A_0 \in \mathcal{S}$ it holds that

$$
X_{t_0}^{-1}(A_0) \in \mathbb{F}_{t_0}^-. \tag{3.41}
$$

The proof of Lemma 3.2.20 is thus completed. □

We now present the promised Corollary 3.2.21, which is a special case of Lemma 3.2.20 above.

Corollary 3.2.21 (Predictability implies adaptivity)**.** *Let $T \in (0,\infty)$ be a real number, let $(\Omega, \mathcal{F}, P)$ be a probability space with a filtration $(\mathbb{F}_t)_{t\in[0,T]}$, let $(S,\mathcal{S})$ be a measurable space, and let $X\colon [0,T]\times\Omega \to S$ be an $(\mathbb{F}_t)_{t\in[0,T]}/\mathcal{S}$-predictable function. Then $X\colon [0,T]\times\Omega \to S$ is $(\mathbb{F}_t)_{t\in[0,T]}/\mathcal{S}$-adapted.*

Lemma 3.2.22 (Adaptivity together with continuous sample paths implies predictabilitiy)**.** *Let $T \in (0,\infty)$ be a real number, let $(\Omega, \mathcal{F}, P)$ be a probability space with a filtration $(\mathbb{F}_t)_{t\in[0,T]}$, let (E, d_E) be a metric space, and let $X\colon [0,T]\times\Omega \to E$ be an $(\mathbb{F}_t)_{t\in[0,T]}/\mathcal{B}(E)$-adapted stochastic process with continuous sample paths. Then X is $(\mathbb{F}_t)_{t\in[0,T]}/\mathcal{B}(E)$-predictable.*

The proof of Lemma 3.2.22 is omitted and can, e.g., be found in Lemma 2.5.1 in [Kallenberg(2002)].

Exercise 3.2.23 (Product measurable random fields)**.** *Let $(I,\mathcal{I})$, $(\Omega,\mathcal{F})$, and $(S,\mathcal{S})$ be measurable spaces and let $X\colon I\times\Omega \to S$ be an $(\mathcal{I}\otimes\mathcal{F})/\mathcal{S}$-measurable function. Prove then that for all $\omega \in \Omega$ it holds that $I \ni i \mapsto X(i,\omega) \in S$ is an $\mathcal{I}/\mathcal{S}$-measurable function.*

3.3 Standard Brownian motions

This subsection introduces a class of stochastic processes known as standard Brownian motions. These stochastic processes are fundamental objects of probability theory (see, e.g., Section 21 in [Klenke(2008)]). A detailed analysis of various aspects of standard Brownian motions can be found in the book [Mörters and Peres(2010)] on standard Brownian motions.

Definition 3.3.1 (Standard Brownian motion with respect to a filtration)**.** *We say that W is an m-dimensional standard $\mathbf{\Omega}$-Brownian motion (we say that W is a standard $\mathbf{\Omega}$-Brownian motion) if and only if there exist T, Ω, $\mathcal{F}$, P, $\mathbb{F}$ such that it holds*

(i) that $T \in (0, \infty)$, $m \in \mathbb{N}$,

(ii) that $\mathbf{\Omega} = (\Omega, \mathcal{F}, P, \mathbb{F})$ is a filtered probability space,

(iii) that W is an $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$-valued stochastic process with continuous sample paths and time set $[0, T]$ on $(\Omega, \mathcal{F}, P)$,

(iv) that W is an $\mathbb{F}/\mathcal{B}(\mathbb{R}^m)$-adapted stochastic process,

(v) that $W_0 = 0 \in \mathbb{R}^m$,

(vi) that for every $t_1, t_2 \in [0, T]$ with $t_1 \leq t_2$ it holds that $W_{t_2} - W_{t_1}$ is $\mathcal{N}_{0,(t_2-t_1)I_{\mathbb{R}^m}}$-distributed, and

(vii) that for every $t_1, t_2 \in [0, T]$ with $t_1 \leq t_2$ it holds that $\sigma_\Omega(W_{t_2} - W_{t_1})$ and $\mathbb{F}_{t_1}$ are P-independent.

Observe that the filtration $\mathbb{F} = (\mathbb{F}_t)_{t\in[0,T]}$ in Definition 3.3.1 is a substantial ingredient of the definition of a standard $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$-Brownian motion. In some situations, the particular filtration is not of importance. For this the following notion is used.

Definition 3.3.2 (Standard Brownian motion)**.** *We say that W is an m-dimensional P-standard Brownian motion (we say that W is an m-dimensional standard Brownian motion, we say that W is a P-standard Brownian motion, we say that W is a standard Brownian motion) if and only if there exist Ω, $\mathcal{F}$, $\mathbb{F}$ such that it holds that W is an m-dimensional standard $(\Omega, \mathcal{F}, P, \mathbb{F})$-Brownian motion.*

3.3.1 Elementary properties of standard Brownian motions

In the next step a few elementary properties of standard Brownian motions are collected.

Proposition 3.3.3 (Properties of standard Brownian motions)**.** *Let $T \in (0, \infty)$, $m \in \mathbb{N}$, let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $W \colon [0, T] \times \Omega \to \mathbb{R}^m$ be a standard Brownian motion. Then*

(i) it holds for all $t \in [0, T]$ that $\mathbb{E}_P\big[W_t\big] = 0 \in \mathbb{R}^m$ and $\mathsf{Cov}_P(W_t) = t \cdot I_{\mathbb{R}^m} \in \mathbb{R}^{m \times m}$,

(ii) it holds for all $t \in [0, T]$, $s \in (0, T]$ that $W_t = \frac{\sqrt{t}}{\sqrt{s}} W_s$ in distribution on $\mathcal{B}(\mathbb{R}^m)$,

(iii) it holds that W has P-independent increments, i.e., it holds for every $n \in \{3, 4, \dots\}$, $t_1, \dots, t_n \in [0, T]$ with $t_1 \leq \ldots \leq t_n$ that the random variables $W_{t_2} - W_{t_1}$, $\dots$, $W_{t_n} - W_{t_{n-1}}$ are P-independent, and

(iv) it holds that W has stationary increments, i.e., it holds for every $n \in \{2, 3, \dots\}$, $h, t_1, \dots, t_n \in [0, T]$ with $t_1 \leq \ldots \leq t_n \leq t_n + h \leq T$ that

$$\big(W_{t_2} - W_{t_1}, \dots, W_{t_n} - W_{t_{n-1}}\big) = \big(W_{t_2+h} - W_{t_1+h}, \dots, W_{t_n+h} - W_{t_{n-1}+h}\big) \tag{3.42}$$

in distribution on $\mathcal{B}(\mathbb{R}^{n \cdot m})$.

Proof of Proposition 3.3.3. Items (i) and (ii) follow immediately from Definition 3.3.2. In the next step we show Item (iii). For this we observe that the fact that W is $(\mathbb{F}^W_t)_{t \in [0,T]}$-adapted (see Lemma 3.2.16) proves that for all $n \in \{3, 4, 5, \dots\}$, $t_1, \dots, t_n \in [0, T]$ with $t_1 \leq \ldots \leq t_n$ it holds that

$$\sigma_\Omega\big(\big(W_{t_2} - W_{t_1}, \dots, W_{t_{n-1}} - W_{t_{n-2}}\big)\big) \subseteq \mathbb{F}^W_{t_{n-1}}. \tag{3.43}$$

Item (vii) in Definition 3.3.1 hence implies that for all $n \in \{3, 4, 5, \dots\}$, $t_1, \dots, t_n \in [0, T]$ with $t_1 \leq \ldots \leq t_n$ it holds that

$$\sigma_\Omega\big(\big(W_{t_2} - W_{t_1}, \dots, W_{t_{n-1}} - W_{t_{n-2}}\big)\big) \qquad \text{and} \qquad \sigma_\Omega\big(W_{t_n} - W_{t_{n-1}}\big) \tag{3.44}$$

are P-independent. Therefore, we get that for all $n \in \{3, 4, 5, \dots\}$, $t_1, \dots, t_n \in [0, T]$, $A_1, \dots, A_{n-1} \in \mathcal{B}(\mathbb{R}^m)$ with $t_1 \leq \ldots \leq t_n$ it holds that

$$\begin{aligned} &P\big(\{W_{t_2} - W_{t_1} \in A_1\} \cap \dots \cap \big\{W_{t_n} - W_{t_{n-1}} \in A_{n-1}\big\}\big) \\ &= P\big(\{W_{t_2} - W_{t_1} \in A_1\} \cap \dots \cap \big\{W_{t_{n-1}} - W_{t_{n-2}} \in A_{n-2}\big\}\big) \cdot P\big(W_{t_n} - W_{t_{n-1}} \in A_{n-1}\big). \end{aligned} \tag{3.45}$$

Iterating (3.45), in turn, ensures that for all $n \in \{3, 4, 5, \dots\}$, $t_1, \dots, t_n \in [0, T]$, $A_1, \dots, A_{n-1} \in \mathcal{B}(\mathbb{R}^m)$ with $t_1 \leq \ldots \leq t_n$ it holds that

$$\begin{aligned} &P\big(\{W_{t_2} - W_{t_1} \in A_1\} \cap \dots \cap \big\{W_{t_n} - W_{t_{n-1}} \in A_{n-1}\big\}\big) \\ &= P(W_{t_2} - W_{t_1} \in A_1) \cdot \ldots \cdot P\big(W_{t_n} - W_{t_{n-1}} \in A_{n-1}\big). \end{aligned} \tag{3.46}$$

This proves Item (iii). It thus remains to show Item (iv). For this we observe that Item (vi) in Definition 3.3.1 implies that for every $t_1, t_2 \in [0, T]$, $h \in [0, T - t_2]$ with $t_1 \leq t_2$ it holds that

$$W_{t_2+h} - W_{t_1+h} \tag{3.47}$$

is $\mathcal{N}_{0,(t_2-t_1)I_{\mathbb{R}^m}}$-distributed. Combining this and Item (iii) proves Item (iv). The proof of Proposition 3.3.3 is thus completed. □

3.3.2 Gaussian stochastic processes

We first present a simple result for independency of jointly normal distributed random variables is formulated.

Lemma 3.3.4 (Uncorrelated normally distributed random variables are independent)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $n \in \mathbb{N}$, $k_1, \ldots, k_n \in \mathbb{N}$, and let $(Y_{1,1}, \ldots, Y_{1,k_1}, \ldots, Y_{n,1}, \ldots, Y_{n,k_n}) \colon \Omega \to \mathbb{R}^{(k_1+\ldots+k_n)}$ be a normally distributed random variable with the property that for all $i, \hat{i} \in \{1, \ldots, n\}$, $l \in \{1, \ldots, k_i\}$, $\hat{l} \in \{1, \ldots, k_{\hat{i}}\}$ with $i \neq \hat{i}$ it holds that*

$$\mathsf{Cov}\big(Y_{i,l}, Y_{\hat{i},\hat{l}}\big) = 0. \tag{3.48}$$

Then the random variables $(Y_{1,1}, \ldots, Y_{1,k_1}) \colon \Omega \to \mathbb{R}^{k_1}$, $\ldots$, $(Y_{n,1}, \ldots, Y_{n,k_n}) \colon \Omega \to \mathbb{R}^{k_n}$ are P-independent.

Lemma 3.3.4 is a straightforward consequence of Proposition 0.4.15 and Corollary 0.4.16. The next result, Corollary 3.3.5, relates Definition 3.3.2 above to Exercise 1.3.11 in Chapter 1. Combining Proposition 0.4.15, Lemma 3.3.4, and Proposition 3.3.3 results in the next corollary.

Corollary 3.3.5 (Covariances for one-dimensional standard Brownian motions)**.** *Let $T \in (0, \infty)$, let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $W \colon [0,T] \times \Omega \to \mathbb{R}$ be a stochastic process with continuous sample paths and with $W_0 = 0$. Then W is a one-dimensional standard Brownian motion if and only if it holds for every $n \in \mathbb{N}$, $t_1, \ldots, t_n \in [0,T]$ that the random variable $(W_{t_1}, \ldots, W_{t_n}) \colon \Omega \to \mathbb{R}^n$ is $\mathcal{N}_{0,(\min\{t_i,t_j\})_{(i,j)\in\{1,\ldots,n\}^2}}$-distributed.*

Proof of Corollary 3.3.5. If W is a one-dimensional standard Brownian motion, then Proposition 3.3.3 implies that for all $i, j \in \{1, 2, \ldots, n\}$ it holds that

$$\begin{aligned}
&\mathbb{E}_P\big[W_{t_i} W_{t_j}\big] = \mathbb{E}_P\big[W_{\max\{t_i,t_j\}} W_{\min\{t_i,t_j\}}\big] \\
&= \underbrace{\mathbb{E}_P\big[\big(W_{\max\{t_i,t_j\}} - W_{\min\{t_i,t_j\}}\big)\,\big(W_{\min\{t_i,t_j\}} - W_0\big)\big]}_{=\mathbb{E}_P\left[W_{\max\{t_i,t_j\}} - W_{\min\{t_i,t_j\}}\right]\mathbb{E}_P\left[W_{\min\{t_i,t_j\}} - W_0\right] = 0} + \mathbb{E}_P\Big[\big|W_{\min\{t_i,t_j\}}\big|^2\Big] \\
&= \min\{t_i, t_j\}.
\end{aligned} \tag{3.49}$$

Moreover, if W is a one-dimensional standard Brownian motion, then Proposition 3.3.3 and Proposition 0.4.15 ensure that $(W_{t_1}, W_{t_2}, \ldots, W_{t_n}) \colon \Omega \to \mathbb{R}^n$ is normally distributed. This and (3.49) show the "⇒" direction in Corollary 3.3.5. Let us now assume that for every $n \in \mathbb{N}$, $t_1, \ldots, t_n \in [0,T]$ it holds that $(W_{t_1}, \ldots, W_{t_n})$ is $\mathcal{N}_{0,(\min(t_i,t_j))_{(i,j)\in\{1,2,\ldots,n\}^2}}$-distributed. Proposition 0.4.15 then implies that for every $n \in \mathbb{N}$, $s_1, s_2, \ldots, s_n, t_1, t_2 \in [0,T]$ with $s_1 \leq \cdots \leq s_n \leq t_1 \leq t_2$ it holds that $(W_{s_1}, \ldots, W_{s_n}, W_{t_2} - W_{t_1})$ is normally distributed with the property that for all $i \in \{1, 2, \ldots, n\}$ it holds that

$$\mathbb{E}_P\big[\,(W_{t_2} - W_{t_1})\,W_{s_i}\big] = \min\{t_2, s_i\} - \min\{t_1, s_i\} = 0 \tag{3.50}$$

and

$$\begin{aligned} \mathbb{E}_P\big[(W_{t_2} - W_{t_1})^2\big] &= \mathbb{E}_P\big[(W_{t_2})^2\big] + \mathbb{E}_P\big[(W_{t_1})^2\big] - 2\,\mathbb{E}_P[W_{t_1} W_{t_2}] \\ &= t_2 + t_1 - 2t_1 = t_2 - t_1. \end{aligned} \tag{3.51}$$

Lemma 3.3.4 hence proves the "$\Leftarrow$" direction and this completes the proof of Corollary 3.3.5. □

Based on Definition 0.4.17, we now introduce the notion of a *Gaussian* stochastic process.

Definition 3.3.6 (Gaussian stochastic process)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $m \in \mathbb{N}$, let $\mathbb{T} \subseteq \mathbb{R}$ be a set, and let $X = (X^1, \ldots, X^m)\colon \mathbb{T} \times \Omega \to \mathbb{R}^m$ be a stochastic process with the property that for every $n \in \mathbb{N}$, $t_1, \ldots, t_n \in \mathbb{T}$ it holds that*

$$\Omega \ni \omega \mapsto \big(X^1_{t_1}(\omega), \ldots, X^m_{t_1}(\omega), X^1_{t_2}(\omega), \ldots, X^m_{t_2}(\omega), \ldots, X^1_{t_n}(\omega), \ldots, X^m_{t_n}(\omega)\big) \in \mathbb{R}^{n\cdot m} \tag{3.52}$$

is Gaussian distributed. Then we say that X is a Gaussian stochastic process*.*

See, e.g., [Ash(2013)] and [Jentzen(2014)] for the next result.

Proposition 3.3.7 (Marginally normally distributed versus jointly normally distributed)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $X\colon \Omega \to \mathbb{R}$ be a standard normal random variable, let $Z\colon \Omega \to \mathbb{R}$ be an $\big(\frac{1}{2}\,\delta^{\mathbb{R}}_{-1}|_{\mathcal{B}(\mathbb{R})} + \frac{1}{2}\,\delta^{\mathbb{R}}_{1}|_{\mathcal{B}(\mathbb{R})}\big)$-distributed random variable with the property that X and Z are independent, and let $Y\colon \Omega \to \mathbb{R}$ be given by $Y = ZX$. Then*

- *it holds that X and Y are standard normal random variables,*
- *it holds that X and Y are uncorrelated, i.e., $\mathsf{Cov}(X, Y) = \mathbb{E}_P\big[XY\big] = 0$, but*
- *it does not hold that $\Omega \ni \omega \mapsto (X(\omega), Y(\omega)) \in \mathbb{R}^2$ is normally distributed.*

Proof of Proposition 3.3.7. First of all, observe that the definition of Y and the assumption that X and Z are independent ensures that for all $x \in \mathbb{R}$ it holds that

$$\begin{aligned} P(Y \leq x) &= P(\{Y \leq x\} \cap \{Z = 1\}) + P(\{Y \leq x\} \cap \{Z = -1\}) \\ &= P(\{X \leq x\} \cap \{Z = 1\}) + P(\{-X \leq x\} \cap \{Z = -1\}) \\ &= \tfrac{1}{2} \cdot P(X \leq x) + \tfrac{1}{2} \cdot P(-X \leq x)\,. \end{aligned} \tag{3.53}$$

Next we note that the assumption that $X(P)_{\mathcal{B}(\mathbb{R})} = \mathcal{N}_{0,I_{\mathbb{R}}}$ implies that $X(P)_{\mathcal{B}(\mathbb{R})} = (-X)(P)_{\mathcal{B}(\mathbb{R})} = \mathcal{N}_{0,I_{\mathbb{R}}}$. This and (3.53) imply that for all $x \in \mathbb{R}$ it holds that

$$P(Y \leq x) = \tfrac{1}{2} \cdot \mathcal{N}_{0,I_{\mathbb{R}}}((-\infty, x]) + \tfrac{1}{2} \cdot \mathcal{N}_{0,I_{\mathbb{R}}}((-\infty, x]) = \mathcal{N}_{0,I_{\mathbb{R}}}((-\infty, x]). \tag{3.54}$$

Moreover, we observe that

$$\mathbb{E}_P\big[XY\big] = \mathbb{E}_P\big[X(ZX)\big] = \mathbb{E}_P\big[X^2 Z\big] = \mathbb{E}_P\big[X^2\big]\,\mathbb{E}_P\big[Z\big] = 1 \cdot 0 = 0. \tag{3.55}$$

Furthermore, we note that

$$\begin{aligned} P(X+Y=0) &= P(X+ZX=0) = P\big((1+Z)X=0\big) \\ &= P(1+Z=0) = P(Z=-1) = \tfrac{1}{2}. \end{aligned} \tag{3.56}$$

This proves that

$$\Omega \ni \omega \mapsto X(\omega) + Y(\omega) \in \mathbb{R} \tag{3.57}$$

is not normal distributed. Proposition 0.4.15 hence proves that

$$\Omega \ni \omega \mapsto (X(\omega), Y(\omega)) \in \mathbb{R}^2 \tag{3.58}$$

is not normal distributed. The proof of Proposition 3.3.7 is thus completed. □

3.3.3 Approximative simulation of sample paths of standard Brownian motions

We now present a short Matlab code for the approximative generation of sample paths of 1-dimensional standard Brownian motions (cf. Exercise 1.3.11). More formally, let $(\Omega, \mathcal{F}, P)$ be a probability space, let $T \in (0,\infty)$, $N \in \mathbb{N}$, let $W\colon [0,T] \times \Omega \to \mathbb{R}$ be a 1-dimensional P-standard Brownian motion, and let $\tilde{W}\colon [0,T] \times \Omega \to \mathbb{R}$ be the function which satisfies for all $n \in \{0,1,\dots,N-1\}$, $t \in [\frac{nT}{N}, \frac{(n+1)T}{N}]$ that

$$\tilde{W}_t = \big(n+1-\tfrac{tN}{T}\big)\, W_{\frac{nT}{N}} + \big(\tfrac{tN}{T} - n\big)\, W_{\frac{(n+1)T}{N}}. \tag{3.59}$$

Observe that if $N \in \mathbb{N}$ is large, then $(\tilde{W}_t)_{t\in[0,T]}$ is in a suitable sense a good approximation of $(W_t)_{t\in[0,T]}$. The following Matlab code plots a realization of an pseudo $\tilde{W}(P)_{\otimes_{t\in[0,T]}\mathcal{B}(\mathbb{R})}$-distributed random variable in the case $T = 1$ and $N = 1000$.

```
T = 1;
N = 1000;
BM = cumsum( [0,randn(1,N)]*sqrt(T/N) );
plot( (0:T/N:T), BM );
```

Matlab code 3.1: A Matlab code which plots a realization of an pseudo $\tilde{W}(P)_{\otimes_{t\in[0,T]}\mathcal{B}(\mathbb{R})}$-distributed random variable in the case $T = 1$ and $N = 1000$ in (3.59).

Definition 3.3.8 (Positive part and negative part)**.** *Let $a \in \mathbb{R}$. Then we denote by $a^+ \in \mathbb{R}$ and $a^- \in \mathbb{R}$ the real numbers given by*

$$a^+ = \max\{a,0\} \qquad \textit{and} \qquad a^- = \max\{-a,0\} = -\min\{a,0\}, \tag{3.60}$$

we call a^+ the positive part of a, and we call a^- the negative part of a.

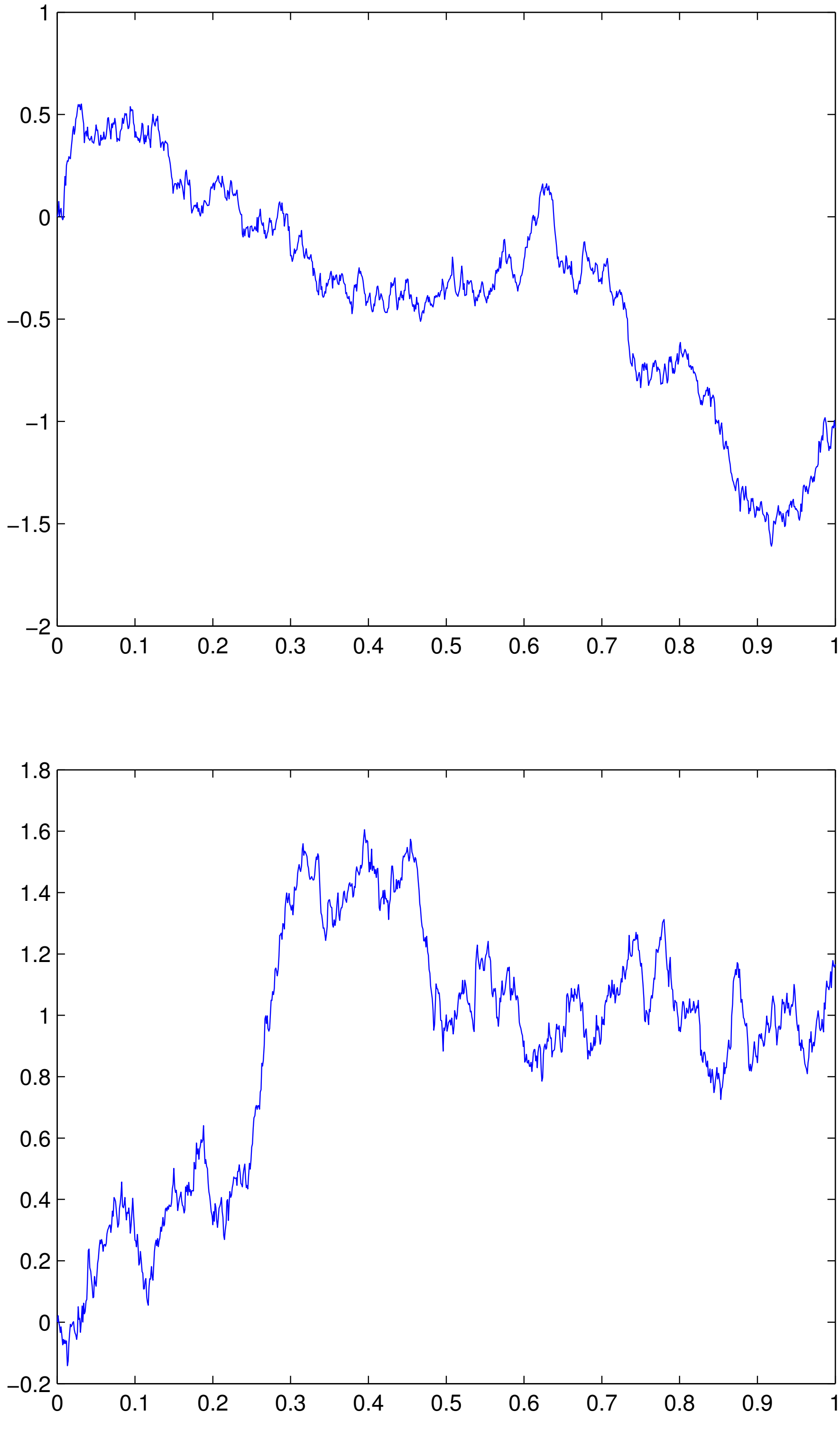


Figure 3.1: Results of two calls of the Matlab code 3.1.

Exercise 3.3.9 (Geometric Brownian motion)**.** *In this exercise we do not distinguish between pseudo random numbers and actual random numbers. Let* $T, x_0 \in (0,\infty)$, $\alpha, \beta \in \mathbb{R}$, *let* $(\Omega, \mathcal{F}, P)$ *be a probability space, let* $W\colon [0,T]\times\Omega \to \mathbb{R}$ *be a standard Brownian motion, let* $X\colon [0,T]\times\Omega\to\mathbb{R}$ *be the stochastic process which satisfies for all* $t \in [0,T]$ *that*

$$X_t = e^{(\alpha t + \beta W_t)}\, x_0, \tag{3.61}$$

and let $f \in \mathcal{L}^1(X_T(P)_{\mathcal{B}(\mathbb{R})}; |\cdot|_{\mathbb{R}})$. *The stochastic process* X *is known as* geometric Brownian motion *in the literature. Write a* Matlab *function* `MonteCarloGBM(T,α,β,x0,f,N)` *with input* $T \in (0,\infty)$, $\alpha, \beta \in \mathbb{R}$, $x_0 \in (0,\infty)$, $f \in \mathcal{L}^1(X_T(P)_{\mathcal{B}(\mathbb{R})}; |\cdot|_{\mathbb{R}})$, $N \in \mathbb{N}$ *and output a Monte Carlo approximation of*

$$\mathbb{E}\big[f(X_T)\big] \tag{3.62}$$

based on $N \in \mathbb{N}$ *samples. Call your* Matlab *function* `MonteCarloGBM(T, α,β,x0,f,N)` *in the case* $T = 1$, $\beta = \frac{1}{10}$, $\alpha = \ln(1.06) - \frac{\beta^2}{2}$, $x_0 = 92$, $f = \mathbb{R} \ni x \mapsto [x-100]^+ \in \mathbb{R}$, $N = 10^4$.

Exercise 3.3.10. *In this exercise we do not distinguish between pseudo random numbers and actual random numbers. Let* $(\Omega, \mathcal{F}, P)$ *be a probability space, let* $m, N \in \mathbb{N}$, *let* $W\colon [0,T]\times\Omega\to\mathbb{R}^m$ *be a standard Brownian motion, and let* $\tilde{W}^M\colon [0,T]\times\Omega\to\mathbb{R}^m$, $M \in \mathbb{N}$, *be the functions which satisfy for all* $M \in \mathbb{N}$, $n \in \{0,1,\ldots,M-1\}$, $t \in \left[\frac{nT}{M}, \frac{(n+1)T}{M}\right]$ *that*

$$\tilde{W}^M_t = \big(n + 1 - \tfrac{tM}{T}\big)\, W_{\frac{nT}{M}} + \big(\tfrac{tM}{T} - n\big)\, W_{\frac{(n+1)T}{M}}. \tag{3.63}$$

- *(i) Write a* Matlab *function* `BrownianMotion(T,m,N)` *with input* $T \in (0,\infty)$, $m, N \in \mathbb{N}$ *and output a realization of an* $\big(W_0, W_{\frac{T}{N}}, W_{\frac{2T}{N}}, \ldots, W_{\frac{(N-1)T}{N}}, W_T\big)(P)_{\mathcal{B}(\mathbb{R}^{m\times(N+1)})}$*-distributed random variable.*
- *(ii) Assume that* $T = 1$ *and that* $m = 2$. *Write a* Matlab *function* `BrownianMotion2DPlot()` *which uses your* Matlab *function* `BrownianMotion(T, m, N)` *to plot one realization of an* $(\tilde{W}^{1000})(P)_{\otimes_{t\in[0,T]}\mathcal{B}(\mathbb{R}^2)}$*-distributed random variable in a three-dimensional coordinate system.*

Exercise 3.3.11 (Geometric Brownian motion revisited)**.** *Let* $T, x_0, \beta \in (0,\infty)$, $\alpha \in \mathbb{R}$, *let* $(\Omega, \mathcal{F}, P)$ *be a probability space, let* $W\colon [0,T]\times\Omega\to\mathbb{R}$ *be a standard Brownian motion, let* $X\colon [0,T]\times\Omega\to\mathbb{R}$ *be the function which satisfies for all* $t \in [0,T]$ *that*

$$X_t = e^{(\alpha t + \beta W_t)}\, x_0 \tag{3.64}$$

(cf. Exercise 3.3.9), and let $\Phi\colon \mathbb{R}\to\mathbb{R}$ *be the function which satisfies for all* $y \in \mathbb{R}$ *that* $\Phi(y) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2} dx$.

- *(i) Show that for all* $K \in \mathbb{R}$ *it holds that*

$$\begin{aligned} &\mathbb{E}\big[\max\{X_T - K, 0\}\big] \\ &= \begin{cases} e^{\left(\alpha + \frac{1}{2}\beta^2\right)T}\, x_0 - K & : K \leq 0 \\ e^{\left(\alpha + \frac{1}{2}\beta^2\right)T}\, x_0\, \Phi\Big(\frac{\alpha T + \ln(\frac{x_0}{K})}{\beta\sqrt{T}} + \beta\sqrt{T}\Big) - K\, \Phi\Big(\frac{\alpha T + \ln(\frac{x_0}{K})}{\beta\sqrt{T}}\Big) & : K > 0 \end{cases}. \end{aligned} \tag{3.65}$$

(ii) *Use Item* (i) *and the built-in* Matlab *function* `erf(...)` *to calculate* $\mathbb{E}\big[\max\{X_T - K, 0\}\big]$ *approximatively in the case* $T = 1$, $\beta = \frac{1}{10}$, $\alpha = \ln(1.06) - \frac{\beta^2}{2}$, $x_0 = 92$, $K = 100$. *Compare this result with the result of Exercise 3.3.9.*

3.3.4 Temporal regularity of standard Brownian motions

This section investigates temporal regularity properties of standard Brownian motions. To do so, a few notions are introduced.

Lemma 3.3.12 (Absolute moments of standard normal random variables)**.** *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space and let* $Y\colon \Omega \to \mathbb{R}$ *be an* $\mathcal{N}_{0,1}$*-distributed random variable. Then*

(i) *it holds for all* $p \in [2, \infty)$ *that*

$$\mathbb{E}_P\big[|Y|_{\mathbb{R}}^p\big] = (p-1) \cdot \mathbb{E}_P\big[|Y|_{\mathbb{R}}^{(p-2)}\big], \tag{3.66}$$

(ii) *it holds for all* $p \in \{2, 3, 4, \dots\}$ *that*

$$\begin{aligned} \mathbb{E}_P\big[|Y|_{\mathbb{R}}^p\big] &= \begin{cases} (p-1)\cdot(p-3)\cdot\ldots\cdot 5\cdot 3\cdot 1 & : p \text{ even} \\ (p-1)\cdot(p-3)\cdot\ldots\cdot 4\cdot 2\cdot \mathbb{E}_P\big[|Y|_{\mathbb{R}}\big] & : p \text{ odd}, \end{cases} \\ &= \begin{cases} (p-1)\cdot(p-3)\cdot\ldots\cdot 5\cdot 3\cdot 1 & : p \text{ even} \\ (p-1)\cdot(p-3)\cdot\ldots\cdot 4\cdot 2\cdot \sqrt{\frac{2}{\pi}} & : p \text{ odd}, \end{cases} \end{aligned} \tag{3.67}$$

and

(iii) *it holds for all* $p \in [1, \infty)$ *that* $\|Y\|_{\mathcal{L}^p(P;|\cdot|_{\mathbb{R}})} \leq \sqrt{p}$.

Proof of Lemma 3.3.12. Note that integration by parts proves that for all $p \in (2, \infty)$ it holds that

$$\begin{aligned} \mathbb{E}_P\big[|Y|_{\mathbb{R}}^p\big] &= \int_{\mathbb{R}} |y|_{\mathbb{R}}^p \tfrac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}y^2}\, dy = \tfrac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \Big[y\,|y|_{\mathbb{R}}^{(p-2)}\Big]\Big[y\, e^{-\frac{1}{2}y^2}\Big]\, dy \\ &= \tfrac{-1}{\sqrt{2\pi}} \Big[y\,|y|_{\mathbb{R}}^{(p-2)}\, e^{-\frac{1}{2}y^2}\Big]_{y=-\infty}^{y=\infty} + \tfrac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \Big[(p-1)\,|y|_{\mathbb{R}}^{(p-2)}\Big]\Big[e^{-\frac{1}{2}y^2}\Big]\, dy \\ &= (p-1) \int_{\mathbb{R}} |y|_{\mathbb{R}}^{(p-2)} \tfrac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}y^2}\, dy = (p-1)\, \mathbb{E}_P\big[|Y|_{\mathbb{R}}^{(p-2)}\big]. \end{aligned} \tag{3.68}$$

This proves (3.66). In addition, note that

$$\begin{aligned} \mathbb{E}_P\big[|Y|_{\mathbb{R}}\big] &= \int_{\mathbb{R}} |y|_{\mathbb{R}} \tfrac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}y^2}\, dy = 2 \int_0^\infty y\, \tfrac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}y^2}\, dy \\ &= \sqrt{\tfrac{2}{\pi}} \int_0^\infty y\, e^{-\frac{1}{2}y^2}\, dy = -\sqrt{\tfrac{2}{\pi}} \Big[e^{-\frac{1}{2}y^2}\Big]_{y=0}^{y=\infty} = \sqrt{\tfrac{2}{\pi}}. \end{aligned} \tag{3.69}$$

Combining this and (3.66) establishes (3.67). Next we note that Jensen's inequality, Definition 1.2.11, and (3.67) imply that for all $p \in (1, \infty)$ it holds that

$$\|Y\|_{\mathcal{L}^p(P;|\cdot|_{\mathbb{R}})} \leq \|Y\|_{\mathcal{L}^{\lceil p \rceil_1}(P;|\cdot|_{\mathbb{R}})} \leq \Big((\lceil p \rceil_1 - 1)^{\frac{(\lceil p \rceil_1 - 1)}{2}} \Big)^{\frac{1}{\lceil p \rceil_1}} \leq (\lceil p \rceil_1 - 1)^{\frac{1}{2}} \leq \sqrt{p}. \quad (3.70)$$

The proof of Lemma 3.3.12 is thus completed. ☐

Theorem 3.3.13 (Kolmogorov-Chentsov theorem)**.** *Let $T, p \in (0, \infty)$, $\alpha \in (0,1)$, $m \in \mathbb{N}$ with $\alpha > \frac{1}{p}$, let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $X \colon [0,T] \times \Omega \to \mathbb{R}^m$ be a stochastic process with $(X_t)_{t \in [0,T]} \in \mathcal{C}^\alpha([0,T], L^p(P; \|\cdot\|_{\mathbb{R}^m}))$, i.e., with*

$$\sup_{\substack{t_1, t_2 \in [0,T] \\ t_1 \neq t_2}} \frac{\|X_{t_2} - X_{t_1}\|_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^m})}}{|t_2 - t_1|^\alpha} < \infty. \quad (3.71)$$

Then there exists a stochastic process $Y \colon [0,T] \times \Omega \to \mathbb{R}^m$ which is a modification of X (i.e., which fulfills that for all $t \in [0,T]$ it holds that $P(X_t = Y_t) = 1$) and which satisfies that for all $\omega \in \Omega$, $\beta \in (0, \alpha - \frac{1}{p})$ it holds that $Y(\omega) \in \mathcal{C}^\beta([0,T], \mathbb{R}^m)$.

The proof of Theorem 3.3.13 is omitted. It can, e.g., be found in Theorem 21.6 in [Klenke(2008)].

Proposition 3.3.14 (Temporal regularity of standard Brownian motions)**.** *Let $T \in (0, \infty)$, $m \in \mathbb{N}$, let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $W = \big(W^{(1)}, \dots, W^{(m)}\big) \colon [0,T] \times \Omega \to \mathbb{R}^m$ be an m-dimensional standard Brownian motion. Then*

(i) it holds for all $p \in [1, \infty)$, $t_1, t_2 \in [0,T]$ with $t_1 \leq t_2$ that

$$\|W_{t_2} - W_{t_1}\|_{\mathcal{L}^2(P;\|\cdot\|_{\mathbb{R}^m})} = \sqrt{m}\,(t_2 - t_1)^{1/2}, \quad (3.72)$$

$$\|W_{t_2} - W_{t_1}\|_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^m})} \leq m\sqrt{p}\,(t_2 - t_1)^{1/2}, \quad (3.73)$$

(ii) it holds for all $\alpha \in (0, \frac{1}{2}]$, $p \in [1, \infty)$ that $(W_t)_{t \in [0,T]} \in \mathcal{C}^\alpha([0,T], L^p(P; \|\cdot\|_{\mathbb{R}^m}))$ and $\|W\|_{\mathcal{C}^{1/2}([0,T], L^p(P;\|\cdot\|_{\mathbb{R}^m}))} \leq m\sqrt{p}$, and

(iii) it holds for all $\alpha \in (0,1)$ that

$$P\big(W \in \mathcal{C}^\alpha([0,T], \mathbb{R}^m)\big) = \begin{cases} 1 & : \alpha \in (0, \frac{1}{2}) \\ 0 & : \alpha \in [\frac{1}{2}, 1) \end{cases}. \quad (3.74)$$

Proof of Proposition 3.3.14. Observe that for all $t_1, t_2 \in [0,T]$ with $t_1 \le t_2$ it holds that

$$\begin{aligned} \mathbb{E}_P\big[\|W_{t_2} - W_{t_1}\|^2_{\mathbb{R}^m}\big] &= \mathbb{E}_P\big[\|W_{(t_2-t_1)}\|^2_{\mathbb{R}^m}\big] = \mathbb{E}_P\left[\sum_{j=1}^m |W^{(j)}_{(t_2-t_1)}|^2\right] \\ &= \sum_{j=1}^m \mathbb{E}_P\Big[|W^{(j)}_{(t_2-t_1)}|^2\Big] = m\,\mathbb{E}_P\Big[|W^{(1)}_{(t_2-t_1)}|^2\Big] = m\,(t_2 - t_1)\,. \end{aligned} \tag{3.75}$$

This implies (3.72). Next note that the estimate

$$\forall\, x = (x_1, \ldots, x_n) \in \mathbb{R}^m\colon \qquad \|x\|_{\mathbb{R}^m} = \left[\sum_{i=1}^m |x_i|^2\right]^{1/2} \le \sum_{i=1}^m |x_i| \tag{3.76}$$

and Lemma 3.3.12 imply that for all $p \in [1,\infty)$, $t_1, t_2 \in [0,T]$ with $t_1 < t_2$ it holds that

$$\begin{aligned} &\|W_{t_2} - W_{t_1}\|_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^m})} = \|W_{(t_2-t_1)}\|_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^m})} = \big\|\|W_{(t_2-t_1)}\|_{\mathbb{R}^m}\big\|_{\mathcal{L}^p(P;|\cdot|_{\mathbb{R}})} \\ &\le \left\|\sum_{j=1}^m |W^{(j)}_{(t_2-t_1)}|\right\|_{\mathcal{L}^p(P;|\cdot|_{\mathbb{R}})} \le \sum_{j=1}^m \left\|W^{(j)}_{(t_2-t_1)}\right\|_{\mathcal{L}^p(P;|\cdot|_{\mathbb{R}})} = m\,\big\|W^{(1)}_{(t_2-t_1)}\big\|_{\mathcal{L}^p(P;|\cdot|_{\mathbb{R}})} \\ &= m\,(t_2-t_1)^{1/2}\,\big\|\tfrac{1}{(t_2-t_1)^{1/2}} W^{(1)}_{(t_2-t_1)}\big\|_{\mathcal{L}^p(P;|\cdot|_{\mathbb{R}})} \le m\sqrt{p}\,(t_2-t_1)^{1/2}\,. \end{aligned} \tag{3.77}$$

This shows (3.73). In particular, we obtain that

$$\forall\, p \in [1,\infty)\colon \qquad \|W\|_{\mathcal{C}^{1/2}([0,T],L^p(P;\|\cdot\|_{\mathbb{R}^m}))} \le m\sqrt{p}. \tag{3.78}$$

Clearly, this implies that

$$\forall\, \alpha \in (0,\tfrac{1}{2}], p \in [1,\infty)\colon \qquad (W_t)_{t\in[0,T]} \in \mathcal{C}^\alpha([0,T], L^p(P;\|\cdot\|_{\mathbb{R}^m}))\,. \tag{3.79}$$

Theorem 3.3.13 and Lemma 3.1.13 hence show that

$$\forall\, \alpha \in (0,\tfrac{1}{2})\colon \qquad P\big(W \in \mathcal{C}^\alpha([0,T],\mathbb{R}^m)\big) = 1. \tag{3.80}$$

For the proof of the fact that

$$\forall\, \alpha \in [\tfrac{1}{2},1)\colon \qquad P\big(W \in \mathcal{C}^\alpha([0,T],\mathbb{R}^m)\big) = 0, \tag{3.81}$$

the reader is referred to [Mörters and Peres(2010)] (see Remark 1.21 in [Mörters and Peres(2010)]). This completes the proof of Proposition 3.3.14. □

Exercise 3.3.15 (Quadratic variation of standard Brownian motions)**.** *Let $T \in (0,\infty)$, $N \in \mathbb{N}$, $0 = t_0 < t_1 < \cdots < t_N = T$, let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $W\colon [0,T] \times \Omega \to \mathbb{R}$ be a standard Brownian motion. Prove that*

$$\left\|T - \sum_{n=0}^{N-1} \big(W_{t_{n+1}} - W_{t_n}\big)^2\right\|_{\mathcal{L}^2(P;|\cdot|_{\mathbb{R}})} \le \sqrt{2T}\left[\max_{n\in\{0,1,\ldots,N-1\}} |t_{n+1} - t_n|\right]^{1/2}. \tag{3.82}$$

3.3.5 Construction of standard Brownian motions

The purpose of this subsection is to show for every $T \in (0, \infty)$ and every $m \in \mathbb{N}$ that there exists a probability space $(\Omega, \mathcal{F}, P)$ on which a standard Brownian motion $W\colon [0,T]\times\Omega \to \mathbb{R}^m$ is defined. To show this result we follow here a method that is known as *Paul Lévy's construction of Brownian motion* in the literature (see, e.g., Section 1.1 in [Mörters and Peres(2010)]). This method is based on Lemma 3.3.19 below. Before Lemma 3.3.19 and the construction of a standard Brownian motion based on it are presented, several preparations are presented first.

Definition 3.3.16. *Let Ω be a set and let $A_n \subseteq \Omega$, $n \in \mathbb{N}$, be sets. Then define the* limes superior

$$\limsup_{n\to\infty} A_n := \cap_{n=1}^{\infty} \left(\cup_{m=n}^{\infty} A_m \right) = \{a \in \Omega\colon a \in A_n \textit{ for infinitely many } n \in \mathbb{N}\} \tag{3.83}$$

and the limes inferior

$$\liminf_{n\to\infty} A_n := \cup_{n=1}^{\infty} \left(\cap_{m=n}^{\infty} A_m \right) = \{a \in \Omega\colon a \in A_n \textit{ for almost all } n \in \mathbb{N}\} \tag{3.84}$$

of the sets $A_n \subseteq \Omega$, $n \in \mathbb{N}$.

In the setting of Definition 3.3.16, observe that $\liminf_{n\to\infty} A_n \subseteq \limsup_{n\to\infty} A_n$ and that $\Omega\backslash\left(\liminf_{n\to\infty} A_n\right) = \limsup_{n\to\infty} (\Omega\backslash A_n)$.

Lemma 3.3.17 (Borel-Cantelli lemma)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $A_n \in \mathcal{F}$, $n \in \mathbb{N}$, be events with $\sum_{n=1}^{\infty} P((A_n)^c) < \infty$. Then $P(\liminf_{n\to\infty} A_n) = 1$.*

Proof of Lemma 3.3.17. Note that the assumption $\sum_{m=1}^{\infty} P((A_m)^c) < \infty$ implies that

$$\begin{aligned} P\Big(\big(\liminf_{n\to\infty} A_n \big)^c \Big) &= P\Big(\limsup_{n\to\infty} (A_n)^c \Big) = P\big(\cap_{n\in\mathbb{N}} \left(\cup_{m=n}^{\infty} (A_m)^c \right)\big) \\ &= \lim_{n\to\infty} P\big(\cup_{m=n}^{\infty} (A_m)^c \big) \leq \lim_{n\to\infty} \left(\sum_{m=n}^{\infty} P\big((A_m)^c \big) \right) = 0 \end{aligned} \tag{3.85}$$

This completes the proof of Lemma 3.3.17. □

We also use the following tool for limits of jointly normally distributed random variables.

Lemma 3.3.18 (The limit of centered jointly normally distributed random variables)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $k \in \mathbb{N}$, let $Q_n \in \mathbb{R}^{k\times k}$, $n \in \mathbb{N}$, be a convergent sequence of $(k \times k)$-matrices, let $X\colon \Omega \to \mathbb{R}^k$ and $X^{(n)}\colon \Omega \to \mathbb{R}^k$, $n \in \mathbb{N}$, be random variables such that for every $n \in \mathbb{N}$ it holds that $X^{(n)}$ is $\mathcal{N}_{0,Q_n}$-distributed and such that for every continuous and bounded function $\varphi\colon \mathbb{R}^k \to \mathbb{R}$ it holds that $\lim_{n\to\infty} \mathbb{E}_P[\varphi(X^{(n)})] = \mathbb{E}_P[\varphi(X)]$. Then X is $\mathcal{N}_{0,\lim_{n\to\infty} Q_n}$-distributed.*

Lemma 3.3.18 can, e.g., be proved by using characteristic functions (see, for example, Theorem 15.8 in [Klenke(2008)]). We now present the promised lemma on which *Paul Lévy's construction of Brownian motion* is based on. Its proof makes use of Proposition 0.4.15.

Lemma 3.3.19 (Lemma for Levy's construction of standard Brownian motions)**.** *Let $t_1, t_2, t_3 \in [0,\infty)$ be real numbers with $t_1 \leq t_3$ and $t_2 = \frac{t_1+t_3}{2}$, let $(\Omega, \mathcal{F}, P)$ be a probability space, let $(Z_1, Z_3)\colon \Omega \to \mathbb{R}^2$ be an $\mathcal{N}_{0,(\min(t_i,t_j))_{(i,j)\in\{1,3\}^2}}$-distributed random variable, let $\Delta\colon \Omega \to \mathbb{R}$ be an $\mathcal{N}_{0,(t_3-t_1)/4}$-distributed random variable and assume that (Z_1, Z_3) and Δ are independent. Then the random variable $(Z_1, \frac{Z_1+Z_3}{2} + \Delta, Z_3)\colon \Omega \to \mathbb{R}^3$ is $\mathcal{N}_{0,(\min(t_i,t_j))_{(i,j)\in\{1,2,3\}^2}}$-distributed.*

The proof of Lemma 3.3.19 is the subject of Exercise 3.5.6 below. In our following construction of a standard Brownian motion we follow the presentation in [Wakolbinger(2004)] and now use Lemma 3.3.19 to construct a standard Brownian motion.

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $Y_{n,k}\colon \Omega \to \mathbb{R}$, $n \in \mathbb{N}_0$, $k \in \mathbb{N}$, be independent standard normal distributed random variables. Recall that such a probability space does indeed exist. (It can, for instance, be constructed as the product space $\big(\times_{(n,k)\in\mathbb{N}_0\times\mathbb{N}} \mathbb{R}, \otimes_{(n,k)\in\mathbb{N}_0\times\mathbb{N}} \mathcal{B}(\mathbb{R}), \otimes_{(n,k)\in\mathbb{N}_0\times\mathbb{N}} \mathcal{N}_{0,1}\big)$; see, e.g., Corollary 14.33 in [Klenke(2008)] for details.) Then we define a family $t_i^{(n)} \in \mathbb{R}$, $i \in \{0,1,\ldots,2^n\}$, $n \in \mathbb{N}_0$, of real numbers by

$$t_i^{(n)} := \frac{i}{2^n} \tag{3.86}$$

for all $i \in \{0,1,\ldots,2^n\}$ and all $n \in \mathbb{N}_0$. In the next step we define *recursively* a family $W^{(n)}\colon [0,1]\times\Omega \to \mathbb{R}$, $n \in \mathbb{N}$, of stochastic processes with continuous sample paths by

$$W_t^{(0)} := t \cdot Y_{0,1} \tag{3.87}$$

for all $t \in [0,1]$ and by

$$W_t^{(n)} := W_t^{(n-1)} + \left(1 - \frac{\left|t - t_i^{(n)}\right|}{\left(t_i^{(n)} - t_{i-1}^{(n)}\right)}\right) Y_{n,i} \sqrt{\frac{\left(t_{i+1}^{(n)} - t_{i-1}^{(n)}\right)}{4}} \tag{3.88}$$

for all $t \in \left[t_{i-1}^{(n)}, t_{i+1}^{(n)}\right]$, $i \in \{1,3,5,\ldots,2^n-3,2^n-1\}$ and all $n \in \mathbb{N}$. We now investigate a few properties of the stochastic processes $W^{(n)}$, $n \in \mathbb{N}_0$.

First, we claim for every $n \in \mathbb{N}_0$ that

$$\left(W^{(n)}_{t_0^{(n)}}, W^{(n)}_{t_1^{(n)}}, \ldots, W^{(n)}_{t_{2^n}^{(n)}}\right) \quad \text{is} \quad \mathcal{N}_{0,(\min(t_i^{(n)},t_j^{(n)}))_{(i,j)\in\{0,1,2,\ldots,2^n\}^2}}\text{-distributed.} \tag{3.89}$$

We prove (3.89) by induction on $n \in \mathbb{N}_0$. Observe that $(W_0^{(0)}, W_1^{(0)}) = (0, Y_{0,1})$ is $\mathcal{N}_{0,(\min(i,j))_{(i,j)\in\{0,1\}^2}}$-distributed and this shows (3.89) in the base case $n = 0$. For the induction step $\mathbb{N}_0 \ni n-1 \to n \in \mathbb{N}$ note for every $n \in \mathbb{N}$ that

$$\begin{aligned}&\left(W^{(n)}_{t^{(n)}_0}, W^{(n)}_{t^{(n)}_1}, \dots, W^{(n)}_{t^{(n)}_{2^n}}\right)\\&= \Bigg(W^{(n-1)}_{t^{(n)}_0}, \frac{W^{(n-1)}_{t^{(n)}_0}+W^{(n-1)}_{t^{(n)}_2}}{2} + Y_{n,1}\sqrt{\frac{\left(t^{(n)}_2-t^{(n)}_0\right)}{4}},\\&\quad W^{(n-1)}_{t^{(n)}_2}, \frac{W^{(n-1)}_{t^{(n)}_2}+W^{(n-1)}_{t^{(n)}_4}}{2} + Y_{n,3}\sqrt{\frac{\left(t^{(n)}_4-t^{(n)}_2\right)}{4}},\\&\quad\vdots \qquad\qquad \vdots\\&\quad W^{(n-1)}_{t^{(n)}_{2^n-1}}, \frac{W^{(n-1)}_{t^{(n)}_{2^n-2}}+W^{(n-1)}_{t^{(n)}_{2^n}}}{2} + Y_{n,2^n-1}\sqrt{\frac{\left(t^{(n)}_{2^n}-t^{(n)}_{2^n-2}\right)}{4}}, W^{(n-1)}_{t^{(n)}_{2^n}}\Bigg)\end{aligned} \tag{3.90}$$

for all $n \in \mathbb{N}$. Combining this with Proposition 0.4.15 and Lemma 3.3.19 then shows that the induction step $\mathbb{N}_0 \ni n-1 \to n \in \mathbb{N}$. This completes the proof of (3.89).

In the next step for every $n \in \mathbb{N}$ the distance between $W^{(n)}$ and $W^{(n-1)}$ is estimated. Observe that

$$\sup_{t\in[0,1]}\left|W^{(n)}_t - W^{(n-1)}_t\right| \leq \sqrt{\frac{\left(t^{(n)}_2-t^{(n)}_0\right)}{4}}\left(\max_{i\in\{1,3,5,\dots,2^n-1\}}|Y_{n,i}|\right) \leq \frac{\max_{i\in\{1,2,\dots,2^n\}}|Y_{n,i}|}{2^{(n+1)/2}} \tag{3.91}$$

for all $n \in \mathbb{N}$ and therefore

$$\begin{aligned}&P\left(\sup_{t\in[0,1]}\left|W^{(n)}_t - W^{(n-1)}_t\right| > \frac{n}{2^{(n+1)/2}}\right) \leq P\left(\frac{\max_{i\in\{1,2,\dots,2^n\}}|Y_{n,i}|}{2^{(n+1)/2}} > \frac{n}{2^{(n+1)/2}}\right)\\&= P\left(\max_{i\in\{1,2,\dots,2^n\}}|Y_{n,i}| > n\right) \leq \sum_{i=1}^{2^n} P(|Y_{n,i}|>n) = 2^n\mathcal{N}_{0,1}((n,\infty)) \leq \frac{2^n e^{-\frac{1}{2}n^2}}{n\sqrt{2\pi}} \leq \frac{2^n}{e^{\frac{1}{2}n^2}}\end{aligned} \tag{3.92}$$

for all $n \in \mathbb{N}$ where we used Lemma 2.6.14 in the last step. Combining the fact that $\sum_{n=1}^{\infty}\frac{2^n}{e^{\frac{1}{2}n^2}} < \infty$ with the Borel-Cantelli lemma (see Lemma 3.3.17 above) therefore shows that

$$\begin{aligned}1 &= P\left(\liminf_{n\to\infty}\left\{\sup_{t\in[0,1]}\left|W^{(n)}_t - W^{(n-1)}_t\right| \leq \frac{n}{2^{(n+1)/2}}\right\}\right)\\&= P\left(\exists\, n_0 \in \mathbb{N}\colon \forall\, n \in \{n_0, n_0+1, \dots\}\colon \sup_{t\in[0,1]}\left|W^{(n)}_t - W^{(n-1)}_t\right| \leq \frac{n}{2^{(n+1)/2}}\right)\\&\leq P\big((W^{(n)})_{n\in\mathbb{N}} \text{ is a Cauchy sequence in } C([0,1],\mathbb{R})\big)\\&= P\big((W^{(n)})_{n\in\mathbb{N}} \text{ is convergence in } C([0,1],\mathbb{R})\big)\end{aligned} \tag{3.93}$$

where we used that the space $C([0,1],\mathbb{R})$ of continuous functions from $[0,1]$ to $\mathbb{R}$ is *complete* in the last step. Next we define a stochastic process $W\colon [0,1]\times\Omega\to\mathbb{R}$ through

$$W_t(\omega) := \begin{cases} \lim_{n\to\infty} W_t^{(n)}(\omega) & : (W^{(n)}(\omega))_{n\in\mathbb{N}} \text{ is convergent in } C([0,1],\mathbb{R}) \\ 0 & : \text{else} \end{cases} \tag{3.94}$$

for all $t\in[0,1]$ and all $\omega\in\Omega$. By construction we have that $P\big(\lim_{n\to\infty}\sup_{t\in[0,T]}|W_t - W_t^{(n)}| = 0\big) = 1$ and that for every $n\in\mathbb{N}$ and every $t_1,\dots,t_n \in \cup_{m\in\mathbb{N}}\cup_{l=0}^{2^m}\{\frac{l}{2^m}\}$ it holds that $(W_{t_1},W_{t_2},\dots,W_{t_n})$ is $\mathcal{N}_{0,(\min(t_i,t_j))_{(i,j)\in\{1,2,\dots,n\}^2}}$-distributed. From Lemma 3.3.18 we therefore get for every $n\in\mathbb{N}$ and every $t_1,\dots,t_n\in[0,1]$ that $(W_{t_1},W_{t_2},\dots,W_{t_n})$ is $\mathcal{N}_{0,(\min(t_i,t_j))_{(i,j)\in\{1,2,\dots,n\}^2}}$-distributed. Combining this with Corollary 3.3.5 results in the following theorem.

Theorem 3.3.20 (Existence of a one-dimensional standard Brownian motion)**.** *There exists a probability space $(\Omega,\mathcal{F},P)$ on which a one-dimensional standard Brownian motion $W\colon [0,1]\times\Omega\to\mathbb{R}$ is defined.*

We have thus constructed a one-dimensional standard Brownian motion on the interval $[0,1]$. The following elementary transformation can be used to generalize this result to arbitrary time intervals $[0,T]$ where $T\in(0,\infty)$.

Lemma 3.3.21 (Transformation of standard Brownian motions)**.** *Let $T,\hat{T}\in(0,\infty)$, let $(\Omega,\mathcal{F},P)$ be a probability space and let $W\colon[0,T]\times\Omega\to\mathbb{R}$ be a one-dimension standard Brownian motion. Then the stochastic process $\hat{W}\colon[0,\hat{T}]\times\Omega\to\mathbb{R}$ defined through*

$$\hat{W}_t := \sqrt{\tfrac{\hat{T}}{T}}\cdot W_{\frac{tT}{\hat{T}}} \tag{3.95}$$

for all $t\in[0,\hat{T}]$ is a one-dimension standard Brownian motion.

The proof of Lemma 3.3.21 is an easy exercise. The following generalization of Theorem 3.3.20 follows from Lemma 3.3.21.

Corollary 3.3.22 (Existence of standard Brownian motions)**.** *Let $T\in(0,\infty)$, $m\in\mathbb{N}$. Then there exists a probability space $(\Omega,\mathcal{F},P)$ and a mapping $W\colon[0,T]\times\Omega\to\mathbb{R}^m$ such that $W\colon[0,T]\times\Omega\to\mathbb{R}^m$ is a standard Brownian motion.*

Proof of Corollary 3.3.22. By Theorem 3.3.20, there exists a probability space $(\Omega,\mathcal{F},P)$ and a one-dimensional standard Brownian motion $W\colon[0,1]\times\Omega\to\mathbb{R}$. Next define $\hat{W}\colon[0,T]\times\Omega\to\mathbb{R}$ through $\hat{W}_t := \sqrt{T}\cdot W_{\frac{t}{T}}$ for all $t\in[0,T]$. Lemma 3.3.21 shows that $\hat{W}$ is a one-dimensional standard Brownian motion. In the next step define the product space $\tilde{\Omega} := \Omega^m$, the product sigma-algebra $\tilde{\mathcal{F}} := \mathcal{F}^{\otimes m} = \mathcal{F}\otimes\mathcal{F}\otimes\ldots\otimes\mathcal{F}$ and the product probability measure $\tilde{P} := P^{\otimes m} = P\otimes P\otimes\ldots\otimes P$. Moreover, define $\tilde{W}\colon[0,T]\times\tilde{\Omega}\to\mathbb{R}^m$ through $\tilde{W}_t(\omega_1,\dots,\omega_m) := (\hat{W}_t(\omega_1),\dots,\hat{W}_t(\omega_m))$ for all $\omega=(\omega_1,\dots,\omega_m)\in\tilde{\Omega}$ and all $t\in[0,T]$. By construction, $(\tilde{\Omega},\tilde{\mathcal{F}},\tilde{P})$ is a probability space and $\tilde{W}$ is an m-dimensional standard Brownian motion on it. □

Theorem 3.3.23 (Existence of stochastic bases and standard Brownian motions)**.** *Let $T \in (0, \infty)$, $m \in \mathbb{N}$. Then there exists a stochastic basis $\big(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]}\big)$ and a function*

$$W\colon [0,T] \times \Omega \to \mathbb{R}^m \tag{3.96}$$

such that W is an m-dimensional standard $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$-Brownian motion.

3.4 Stochastic Integration with respect to standard Brownian motions

In this section the stochastic integral with respect to a standard Brownian motion is defined and some of its properties are formulated. We follow the presentations in [Jentzen(2014)].

3.4.1 Norms on matrices

3.4.1.1 Operator norm induced by the Euclidean norm

Definition 3.4.1 (Operator norm)**.** *Let $d, m \in \mathbb{N}$. Then we denote by $\|\cdot\|_{\mathbb{R}^{d\times m}} : \mathbb{R}^{d\times m} \to [0, \infty)$ the function which satisfies for all $A \in \mathbb{R}^{d\times m}$ that*

$$\|A\|_{\mathbb{R}^{d\times m}} = \sup_{v\in\mathbb{R}^m\setminus\{0\}} \left[\frac{\|Av\|_{\mathbb{R}^d}}{\|v\|_{\mathbb{R}^m}}\right] \tag{3.97}$$

and we call $\|\cdot\|_{\mathbb{R}^{d\times m}}$ the operator norm on $\mathbb{R}^{d\times m}$.

Note that for every $d, m \in \mathbb{N}$ it holds that $(\mathbb{R}^{d\times m}, \|\cdot\|_{\mathbb{R}^{d\times m}})$ is a normed $\mathbb{R}$-vector space (it is even an $\mathbb{R}$-Banach space).

3.4.1.2 Hilbert-Schmidt norm

Definition 3.4.2 (Hilbert-Schmidt norm)**.** *Let $m, d \in \mathbb{N}$. Then we denote by $\|\cdot\|_{HS(\mathbb{R}^m,\mathbb{R}^d)} : \mathbb{R}^{d\times m} \to [0, \infty)$ the function which satisfies for all $A = (A_{i,j})_{i\in\{1,\ldots,d\},j\in\{1,\ldots,m\}} \in \mathbb{R}^{d\times m}$ that*

$$\|A\|_{HS(\mathbb{R}^m,\mathbb{R}^d)} = \sqrt{\sum_{i=1}^{d}\sum_{j=1}^{m} |A_{i,j}|^2} \tag{3.98}$$

and we call $\|\cdot\|_{HS(\mathbb{R}^m,\mathbb{R}^d)}$ the Hilbert-Schmidt norm on $\mathbb{R}^{d\times m}$ (we call $\|\cdot\|_{HS(\mathbb{R}^m,\mathbb{R}^d)}$ the Frobenius norm on $\mathbb{R}^{d\times m}$).

Observe that for every $d, m \in \mathbb{N}$ it holds that $(\mathbb{R}^{d\times m}, \|\cdot\|_{HS(\mathbb{R}^m,\mathbb{R}^d)})$ is a normed $\mathbb{R}$-vector space (it is even an $\mathbb{R}$-Hilbert space).

Example 3.4.3. *Let $A = (A_{i,j})_{i\in\{1,2\},j\in\{1,2\}} \in \mathbb{R}^{2\times 2}$ be the 2×2-matrix given by*

$$A = I_{\mathbb{R}^2} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{3.99}$$

Then

(i) it holds that

$$\|A\|_{\mathbb{R}^{2\times 2}} = \sup_{v\in\mathbb{R}^2\setminus\{0\}} \left[\frac{\|Av\|_{\mathbb{R}^2}}{\|v\|_{\mathbb{R}^2}}\right] = \sup_{v\in\mathbb{R}^2\setminus\{0\}} \left[\frac{\|v\|_{\mathbb{R}^2}}{\|v\|_{\mathbb{R}^2}}\right] = 1 \tag{3.100}$$

and

(ii) it holds that

$$\|A\|_{HS(\mathbb{R}^2,\mathbb{R}^2)} = \sqrt{\sum_{i,j=1}^{2} |A_{i,j}|^2} = \sqrt{2} > 1 = \|A\|_{\mathbb{R}^{2\times 2}}. \tag{3.101}$$

Lemma 3.4.4 (Comparison between operator norm and Hilbert-Schmidt)**.** *Let $d, m \in \mathbb{N}$ and let $A = (A_{i,j})_{i\in\{1,\dots,d\},j\in\{1,\dots,m\}} \in \mathbb{R}^{d\times m}$. Then*

$$\|A\|_{\mathbb{R}^{d\times m}} \leq \|A\|_{HS(\mathbb{R}^m,\mathbb{R}^d)}. \tag{3.102}$$

Proof of Lemma 3.4.4. Note that the triangle inequality and the Cauchy-Schwarz inequality prove that for all $v = (v_1, \dots, v_m) \in \mathbb{R}^m$ it holds that

$$\begin{aligned}
\|Av\|_{\mathbb{R}^d}^2 &= \left\| v_1 \begin{pmatrix} A_{1,1} \\ \vdots \\ A_{d,1} \end{pmatrix} + \ldots + v_m \begin{pmatrix} A_{1,m} \\ \vdots \\ A_{d,m} \end{pmatrix} \right\|_{\mathbb{R}^d}^2 = \left\| \sum_{k=1}^{m} v_k \begin{pmatrix} A_{1,k} \\ \vdots \\ A_{d,k} \end{pmatrix} \right\|_{\mathbb{R}^d}^2 \\
&\leq \left[\sum_{k=1}^{m} |v_k| \left\| \begin{pmatrix} A_{1,k} \\ \vdots \\ A_{d,k} \end{pmatrix} \right\|_{\mathbb{R}^d} \right]^2 \leq \left[\sum_{k=1}^{m} |v_k|^2 \right] \left[\sum_{k=1}^{m} \left\| \begin{pmatrix} A_{1,k} \\ \vdots \\ A_{d,k} \end{pmatrix} \right\|_{\mathbb{R}^d}^2 \right] \\
&= \|v\|_{\mathbb{R}^m}^2 \, \|A\|_{HS(\mathbb{R}^m,\mathbb{R}^d)}^2.
\end{aligned} \tag{3.103}$$

This completes the proof of Lemma 3.4.4. ∎

3.4.2 Product measure on the predictable sigma-algebra

Remark 3.4.5. *Let $T \in (0, \infty)$ and let $(\Omega, \mathcal{F})$ be a measurable space with a filtration $\mathbb{F}_t \in \mathcal{P}(\mathcal{P}(\Omega))$, $t \in [0, T]$. Then observe that*

$$\operatorname{Pred}((\mathbb{F}_t)_{t\in[0,T]}) \subseteq \sigma_{[0,T]\times\Omega}\big(\{B \times A \colon B \in \mathcal{B}([0, T]), A \in \mathbb{F}_T\}\big) = \mathcal{B}([0, T]) \otimes \mathbb{F}_T. \tag{3.104}$$

This fact is used in the next definition.

Definition 3.4.6 (Product measure on the predictable sigma-algebra)**.** *Let $T \in (0, \infty)$ and let $(\Omega, \mathcal{F}, P)$ be a probability space with a filtration $(\mathbb{F}_t)_{t\in[0,T]}$. Then we denote by*

$$\mathcal{P}_{P,(\mathbb{F}_t)_{t\in[0,T]}} \colon \operatorname{Pred}((\mathbb{F}_t)_{t\in[0,T]}) \to [0, \infty] \tag{3.105}$$

the measure given by

$$\mathcal{P}_{P,(\mathbb{F}_t)_{t\in[0,T]}} = (B_{[0,T]} \otimes P)|_{\operatorname{Pred}((\mathbb{F}_t)_{t\in[0,T]})}. \tag{3.106}$$

Let $T \in (0, \infty)$ and let $(\Omega, \mathcal{F}, P)$ be a probability space with a filtration $(\mathbb{F}_t)_{t\in[0,T]}$. Then note that for all $t_1, t_2 \in [0, T]$, $A \in \mathbb{F}_{t_1}$ with $t_1 < t_2$ it holds that

$$\mathcal{P}_{P,(\mathbb{F}_t)_{t\in[0,T]}}((t_1, t_2] \times A) = (t_2 - t_1) \cdot P(A). \tag{3.107}$$

3.4.3 Vector spaces of equivalence classes of predictable stochastic processes

Let $T \in (0, \infty)$, $d, m \in \mathbb{N}$, let $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$ be a stochastic basis, and let $W \colon [0, T] \times \Omega \to \mathbb{R}^m$ be a standard $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$-Brownian motion. Then observe that for all $p \in [0, \infty)$ it holds that

$$\begin{split} &L^p\big(\mathcal{P}_{P,(\mathbb{F}_t)_{t\in[0,T]}}; \|\cdot\|_{HS(\mathbb{R}^m,\mathbb{R}^d)}\big) \\ &= \left\{ \left\{ \begin{array}{c} Y \colon [0, T] \times \Omega \to \mathbb{R}^{d\times m} \colon \\ Y \text{ is } (\mathbb{F}_t)_{t\in[0,T]}\text{-predictable and} \\ \int_0^T \mathbb{E}_P\big[\|X_s - Y_s\|_{HS(\mathbb{R}^m,\mathbb{R}^d)}\big]\, ds = 0 \end{array} \right\} \colon \left(\begin{array}{c} X \colon [0, T] \times \Omega \to \mathbb{R}^{d\times m} \\ \text{is } (\mathbb{F}_t)_{t\in[0,T]}\text{-predictable and} \\ \int_0^T \mathbb{E}_P\big[\|X_s\|^p_{HS(\mathbb{R}^m,\mathbb{R}^d)}\big]\, ds < \infty \end{array} \right) \right\} \\ &= L^p\big(\mathcal{P}_{P,(\mathbb{F}_t)_{t\in[0,T]}}; \|\cdot\|_{\mathbb{R}^{d\times m}}\big). \end{split} \tag{3.108}$$

Furthermore, note that for all $p \in (0, \infty)$, $X \in \mathcal{L}^p(\mathcal{P}_{P,(\mathbb{F}_t)_{t\in[0,T]}}; \|\cdot\|_{\mathbb{R}^{d\times m}})$ it holds that

$$\begin{split} \|X\|_{\mathcal{L}^p(\mathcal{P}_{P,(\mathbb{F}_t)_{t\in[0,T]}}; \|\cdot\|_{HS(\mathbb{R}^m,\mathbb{R}^d)})} &= \left(\int_0^T \mathbb{E}_P\Big[\|X_s\|^p_{HS(\mathbb{R}^m,\mathbb{R}^d)}\Big]\, ds \right)^{\frac{1}{p}}, \\ \|X\|_{\mathcal{L}^p(\mathcal{P}|_{(\mathbb{F}_t)_{t\in[0,T]}}; \|\cdot\|_{\mathbb{R}^{d\times m}})} &= \left(\int_0^T \mathbb{E}_P\Big[\|X_s\|^p_{\mathbb{R}^{d\times m}}\Big]\, ds \right)^{\frac{1}{p}}. \end{split} \tag{3.109}$$

3.4.4 Metrics for convergence in probability

Exercise 3.4.7 (Metrization of convergence in probability)**.** *Let $d \in \mathbb{N}$ and let $d_{p,c}\colon L^0(P; \|\cdot\|_{\mathbb{R}^d}) \times L^0(P; \|\cdot\|_{\mathbb{R}^d}) \to [0,\infty)$, $p \in [1,\infty)$, $c \in (0,\infty)$, be the mappings with the property that for all $p \in [1,\infty)$, $c \in (0,\infty)$, $X, Y \in L^0(P; \|\cdot\|_{\mathbb{R}^d})$ it holds that*

$$d_{p,c}(X,Y) = \|\min\{c, \|X-Y\|_{\mathbb{R}^d}\}\|_{L^p(P;|\cdot|_{\mathbb{R}})} = \left|\mathbb{E}_P\big[\min\{c^p, \|X-Y\|_{\mathbb{R}^d}^p\}\big]\right|^{1/p}. \tag{3.110}$$

Prove then that

(i) for all $p \in [1,\infty)$, $c \in (0,\infty)$ it holds that $(L^0(P; \|\cdot\|_{\mathbb{R}^d}), d_{p,c})$ is a metric space

(ii) and for all $p \in [1,\infty)$, $c \in (0,\infty)$, $(X_n)_{n\in\mathbb{N}_0} \subseteq L^0(P; \|\cdot\|_{\mathbb{R}^d})$ it holds that $\limsup_{n\to\infty} d_{p,c}(X_n, X_0) = 0$ if and only if $\forall\, \varepsilon \in (0,\infty)\colon \limsup_{n\to\infty} P(\|X_n - X_0\|_{\mathbb{R}^d} \geq \varepsilon) = 0$

Exercise 3.4.8. *Let $d, m \in \mathbb{N}$, $T \in (0,\infty)$ and let*

$$d_{p,c}\colon \Big\{X \in L^0(\mathcal{P}_{P,(\mathbb{F}_t)_{t\in[0,T]}}; \|\cdot\|_{\mathbb{R}^{d\times m}})\colon P\big(\textstyle\int_0^T \|X_s\|^2_{\mathbb{R}^{d\times m}}\, ds < \infty\big) = 1\Big\}^2 \to [0,\infty), \tag{3.111}$$

$p \in [1,\infty)$, $c \in (0,\infty)$, be the mappings with the property that for all $p \in [1,\infty)$, $c \in (0,\infty)$, $(X,Y) \in \operatorname{dom}(d_{p,c})$ it holds that

$$d_{p,c}(X,Y) = \left\| \sqrt{\min\left\{c, \int_0^T \|X_s - Y_s\|^2_{\mathbb{R}^{d\times m}}\, ds\right\}} \right\|_{L^p(P;|\cdot|_{\mathbb{R}})}. \tag{3.112}$$

Prove then that

(i) for all $p \in [1,\infty)$, $c \in (0,\infty)$ it holds that $\big(L^0(\mathcal{P}_{P,(\mathbb{F}_t)_{t\in[0,T]}}; \|\cdot\|_{\mathbb{R}^d}), d_{p,c}\big)$ is a metric space

(ii) and for all $p \in [1,\infty)$, $c \in (0,\infty)$, $(X^n)_{n\in\mathbb{N}_0} \subseteq L^0(\mathcal{P}_{P,(\mathbb{F}_t)_{t\in[0,T]}}; \|\cdot\|_{\mathbb{R}^{d\times m}})$ with $\forall\, n \in \mathbb{N}_0\colon P\big(\int_0^T \|X^n_s\|^2_{\mathbb{R}^{d\times m}}\, ds < \infty\big) = 1$ it holds that $\limsup_{n\to\infty} d_{p,c}(X^n, X^0) = 0$ if and only if $\forall\, \varepsilon \in (0,\infty)\colon \limsup_{n\to\infty} P\big(\int_0^T \|X^n_s - X^0_s\|^2_{\mathbb{R}^{d\times m}}\, ds \geq \varepsilon\big) = 0$.

3.4.5 Simple processes

Definition 3.4.9 (Simple predictable process)**.** *Let $T \in (0,\infty)$, $d, m \in \mathbb{N}$ and let $(\Omega, \mathcal{F})$ be a measurable space with a filtration $(\mathbb{F}_t)_{t\in[0,T]}$. Then a mapping $X\colon [0,T]\times\Omega \to \mathbb{R}^{d\times m}$ is called $(\mathbb{F}_t)_{t\in[0,T]}$-*simple *(or just* simple*) if there exist $n \in \mathbb{N}$, $0 \leq t_1 < \ldots < t_n \leq T$ and for every $k \in \{1,\ldots,n-1\}$ an $\mathbb{F}_{t_k}/\mathcal{B}(\mathbb{R}^{d\times m})$-measurable mapping $H_k\colon \Omega \to \mathbb{R}^{d\times m}$ such that for all $t \in [0,T]$ it holds that*

$$X_t = \sum_{k=1}^{n-1} H_k \cdot \mathbb{1}_{(t_k,t_{k+1}]}(t)\,. \tag{3.113}$$

Simple processes in the sense of Definition 3.4.9 are predictable. This is an immediate consequence of the next exercise.

Exercise 3.4.10 (Simple processes)**.** *Let $T \in (0, \infty)$, $d, m \in \mathbb{N}$ and let $(\Omega, \mathcal{F})$ be a measurable space with a filtration $(\mathbb{F}_t)_{t\in[0,T]}$. Prove that*

$$\begin{aligned}&\operatorname{Pred}\big((\mathbb{F}_t)_{t\in[0,T]}\big) = \sigma_{[0,T]\times\Omega}\bigg(\{\{0\}\times A\colon A\in\mathbb{F}_0\}\\&\cup\big\{X^{-1}(A)\colon X\colon[0,T]\times\Omega\to\mathbb{R}^{d\times m} \text{ is } (\mathbb{F}_t)_{t\in[0,T]}\text{-simple}, A\in\mathcal{B}(\mathbb{R}^{d\times m})\big\}\bigg).\end{aligned}\tag{3.114}$$

3.4.5.1 Density of simple processes

If $T \in (0, \infty)$, $d, m \in \mathbb{N}$ and if $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$ is a stochastic basis, then simple stochastic processes that are in $L^2\big(\mathcal{P}_{P,(\mathbb{F}_t)_{t\in[0,T]}}; \|\cdot\|_{HS(\mathbb{R}^m,\mathbb{R}^d)}\big)$ are dense in $L^2\big(\mathcal{P}_{P,(\mathbb{F}_t)_{t\in[0,T]}}; \|\cdot\|_{HS(\mathbb{R}^m,\mathbb{R}^d)}\big)$. This is the topic of the next lemma.

Proposition 3.4.11 (Density of simple processes I)**.** *Let $T \in (0, \infty)$, $d, m \in \mathbb{N}$ and let $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$ be a stochastic basis. Then it holds that the set*

$$\Big\{Y \in L^2\big(\mathcal{P}_{P,(\mathbb{F}_t)_{t\in[0,T]}}; \|\cdot\|_{HS(\mathbb{R}^m,\mathbb{R}^d)}\big)\colon Y \text{ is } (\mathbb{F}_t)_{t\in[0,T]}\text{-simple}\Big\}\tag{3.115}$$

is dense in $L^2\big(\mathcal{P}_{P,(\mathbb{F}_t)_{t\in[0,T]}}; \|\cdot\|_{HS(\mathbb{R}^m,\mathbb{R}^d)}\big)$, that is, for every $X \in L^2\big(\mathcal{P}_{P,(\mathbb{F}_t)_{t\in[0,T]}}; \|\cdot\|_{HS(\mathbb{R}^m,\mathbb{R}^d)}\big)$ there exist $(\mathbb{F}_t)_{t\in[0,T]}$-simple $Y^{(n)} \in L^2\big(\mathcal{P}_{P,(\mathbb{F}_t)_{t\in[0,T]}}; \|\cdot\|_{HS(\mathbb{R}^m,\mathbb{R}^d)}\big)$, $n \in \mathbb{N}$, such that

$$\lim_{n\to\infty} \underbrace{\big\|X - Y^{(n)}\big\|_{L^2(\mathcal{P}_{P,(\mathbb{F}_t)_{t\in[0,T]}};\|\cdot\|_{HS(\mathbb{R}^m,\mathbb{R}^d)})}_{=\left(\int_0^T \mathbb{E}_P\left[\|X_s - X_s^{(n)}\|^2_{HS(\mathbb{R}^m,\mathbb{R}^d)}\right] ds\right)^{1/2}} = 0.\tag{3.116}$$

Proposition 3.4.11 is, e.g., proved as a special case of Theorem 25.9 in [Klenke(2008)]. Let $T \in (0, \infty)$, $d, m \in \mathbb{N}$ and let $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$ be a stochastic basis. Then note that for all $X \in L^2\big(\mathcal{P}_{P,(\mathbb{F}_t)_{t\in[0,T]}}; \|\cdot\|_{HS(\mathbb{R}^m,\mathbb{R}^d)}\big)$ it holds that

$$\begin{aligned}&\|X\|^2_{L^2(\mathcal{P}_{P,(\mathbb{F}_t)_{t\in[0,T]}};\|\cdot\|_{HS(\mathbb{R}^m,\mathbb{R}^d)})} = \int_{[0,T]\times\Omega} \|X_s(\omega)\|^2_{HS(\mathbb{R}^m,\mathbb{R}^d)}\, \mathcal{P}_{P,(\mathbb{F}_t)_{t\in[0,T]}}(ds, d\omega)\\&= \int_{[0,T]\times\Omega} \|X_s(\omega)\|^2_{HS(\mathbb{R}^m,\mathbb{R}^d)}\, (B_{[0,T]} \otimes P)(ds, d\omega)\\&= \int_0^T \mathbb{E}_P\Big[\|X_s\|^2_{HS(\mathbb{R}^m,\mathbb{R}^d)}\Big]\, ds < \infty\end{aligned}\tag{3.117}$$

and

$$P\bigg(\int_0^T \|X_s\|^2_{HS(\mathbb{R}^m,\mathbb{R}^d)}\, ds < \infty\bigg) = 1.\tag{3.118}$$

Proposition 3.4.12 (Density of simple processes II)**.** *Let $T \in (0,\infty)$, $d,m \in \mathbb{N}$, let $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$ be a stochastic basis and let $X \in L^0\big(\mathcal{P}_{P,(\mathbb{F}_t)_{t\in[0,T]}}; \|\cdot\|_{HS(\mathbb{R}^m,\mathbb{R}^d)}\big)$ with the property that it hold P-a.s. that $\int_0^T \|X_s\|^2_{\mathbb{R}^{d\times m}}\,ds < \infty$. Then there exist $(\mathbb{F}_t)_{t\in[0,T]}$-simple $Y^{(n)} \in L^0\big(\mathcal{P}_{P,(\mathbb{F}_t)_{t\in[0,T]}}; \|\cdot\|_{HS(\mathbb{R}^m,\mathbb{R}^d)}\big)$, $n \in \mathbb{N}$, such that*

$$\lim_{n\to\infty} \left\| \min\left\{1, \int_0^T \|X_s - Y_s^{(n)}\|^2_{HS(\mathbb{R}^m,\mathbb{R}^d)}\,ds\right\}\right\|_{L^1(P;|\cdot|_{\mathbb{R}})} = 0. \tag{3.119}$$

3.4.6 Lenglart's inequality

Definition 3.4.13 (Random time)**.** *Let $\mathbb{T} \subseteq \mathbb{R}$ be a set and let $(\Omega, \mathcal{F}, P)$ be a probability space. Then a mapping $\tau\colon \Omega \to \mathbb{T}$ is called a* random time *if τ is $\mathcal{F}/\mathcal{B}(\mathbb{T})$-measurable, that is, if for every $t \in \mathbb{T}$ it holds that $\{\tau \le t\} \in \mathcal{F}$.*

Definition 3.4.14 (Stopping time)**.** *Let $\mathbb{T} \subseteq \mathbb{R}$ be a set and let $(\Omega, \mathcal{F}, P)$ be a probability space with a filtration $(\mathbb{F}_t)_{t\in\mathbb{T}}$. Then a mapping $\tau\colon \Omega \to \mathbb{T}$ is called an* $(\mathbb{F}_t)_{t\in\mathbb{T}}$-stopping time *if for all $t \in \mathbb{T}$ it holds that $\{\tau \le t\} \in \mathbb{F}_t$.*

Exercise 3.4.15. *Let $\mathbb{T} \subseteq \mathbb{R}$ be a set, let $(\Omega, \mathcal{F}, P)$ be a probability space with a filtration $(\mathbb{F}_t)_{t\in\mathbb{T}}$, and let $\tau, \rho\colon \Omega \to \mathbb{T}$ be $(\mathbb{F}_t)_{t\in\mathbb{T}}$-stopping times. Prove then that $\min\{\tau,\rho\}$ is an $(\mathbb{F}_t)_{t\in\mathbb{T}}$-stopping time.*

In (3.122) in the following result, Proposition 3.4.16, we prove a very powerful inequality which is known as Lenglart inequality in the literature. Proposition 3.4.16 and its proof are extensions of Problem 1.4.15, Remark 1.4.17 and Solution 4.15 in Section 1.6 in [Karatzas and Shreve(1988)].

Proposition 3.4.16 (Lenglart inequality)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space with a filtration $(\mathbb{F}_t)_{t\in[0,\infty)}$, let $X, Y\colon [0,\infty)\times\Omega \to [0,\infty)$ be stochastic processes with continuous sample paths such that for all bounded $(\mathbb{F}_t)_{t\in[0,\infty)}$-stopping times $\tau\colon \Omega \to [0,\infty)$ it holds that $\mathbb{E}_P\big[X_\tau\big] \le \mathbb{E}_P\big[\sup_{t\in[0,\tau]} Y_t\big]$. Then for all $\varepsilon, \delta \in (0,\infty)$ and all $(\mathbb{F}_t)_{t\in[0,\infty)}$-stopping times $\tau\colon \Omega \to [0,\infty)$ it holds that*

$$P\big(\sup\nolimits_{t\in[0,\tau]} X_t \ge \varepsilon\big) \le \tfrac{1}{\varepsilon}\,\mathbb{E}_P\big[\sup\nolimits_{t\in[0,\tau]} Y_t\big]\,, \tag{3.120}$$

$$P\big(\sup\nolimits_{t\in[0,\tau]} X_t \ge \varepsilon,\, \sup\nolimits_{t\in[0,\tau]} Y_t < \delta\big) \le \tfrac{1}{\varepsilon}\,\mathbb{E}_P\big[\min\big\{\delta, \sup\nolimits_{t\in[0,\tau]} Y_t\big\}\big]\,, \tag{3.121}$$

$$P\big(\sup\nolimits_{t\in[0,\tau]} X_t \ge \varepsilon\big) \le \tfrac{1}{\varepsilon}\,\mathbb{E}_P\big[\min\big\{\delta, \sup\nolimits_{t\in[0,\tau]} Y_t\big\}\big] + P\big(\sup\nolimits_{t\in[0,\tau]} Y_t \ge \delta\big)\,, \tag{3.122}$$

$$\mathbb{E}_P\big[\min\big\{\varepsilon, \sup\nolimits_{t\in[0,\tau]} X_t\big\}\big] \le \Big[2\sqrt{\varepsilon} + \tfrac{\varepsilon}{\sqrt{\delta}}\Big]\,\big|\mathbb{E}_P\big[\min\big\{\delta, \sup\nolimits_{t\in[0,\tau]} Y_t\big\}\big]\big|^{1/2}\,, \tag{3.123}$$

$$\mathbb{E}_P\big[\min\big\{1, \sup\nolimits_{t\in[0,\tau]} X_t\big\}\big] \le 3\,\big|\mathbb{E}_P\big[\min\big\{1, \sup\nolimits_{t\in[0,\tau]} Y_t\big\}\big]\big|^{1/2}. \tag{3.124}$$

Proof of Proposition 3.4.16. Throughout this proof let $\rho_\varepsilon^X \colon \Omega \to [0,\infty]$, $\varepsilon \in [0,\infty)$, and $\rho_\varepsilon^Y \colon \Omega \to [0,\infty]$, $\varepsilon \in [0,\infty)$, be the mappings with the property that for all $\varepsilon \in [0,\infty)$ it holds that

$$\rho_\varepsilon^X = \inf\bigl(\bigl\{t \in [0,\infty) \colon X_t \geq \varepsilon\bigr\} \cup \{\infty\}\bigr)\,, \tag{3.125}$$

$$\rho_\varepsilon^Y = \inf\bigl(\bigl\{t \in [0,\infty) \colon \sup\nolimits_{s\in[0,t]} Y_s \geq \varepsilon\bigr\} \cup \{\infty\}\bigr). \tag{3.126}$$

Then observe that for all $\varepsilon \in [0,\infty)$, $n \in \mathbb{N}$ and all $(\mathbb{F}_t)_{t\in[0,\infty)}$-stopping times $\tau \colon \Omega \to [0,\infty)$ it holds that

$$\begin{aligned}
&\varepsilon\, P\bigl(\sup\nolimits_{t\in[0,\min\{\tau,n\}]} X_t \geq \varepsilon\bigr) = \varepsilon\, P\Bigl(\exists\, t \in [0,\min\{\tau,n\}] \colon X_t \geq \varepsilon\Bigr)\\
&= \varepsilon\, P\Bigl(\Bigl\{\exists\, t \in [0,\min\{\tau,n\}] \colon X_t \geq \varepsilon\Bigr\} \cap \Bigl\{\rho_\varepsilon^X \leq \min\{\tau,n\}\Bigr\}\Bigr)\\
&= \varepsilon\, P\Bigl(\Bigl\{\exists\, t \in [0,\min\{\tau,n\}] \colon X_t \geq \varepsilon\Bigr\} \cap \Bigl\{\rho_\varepsilon^X \leq \min\{\tau,n\}\Bigr\} \cap \Bigl\{X_{\min\{\tau,n,\rho_\varepsilon^X\}} \geq \varepsilon\Bigr\}\Bigr)\\
&\leq \varepsilon\, P\Bigl(X_{\min\{\tau,n,\rho_\varepsilon^X\}} \geq \varepsilon\Bigr) = \mathbb{E}_P\Bigl[\varepsilon\, \mathbb{1}_{\{X_{\min\{\tau,n,\rho_\varepsilon^X\}} \geq \varepsilon\}}\Bigr]\\
&\leq \mathbb{E}_P\Bigl[X_{\min\{\tau,n,\rho_\varepsilon^X\}}\, \mathbb{1}_{\{X_{\min\{\tau,n,\rho_\varepsilon^X\}} \geq \varepsilon\}}\Bigr] \leq \mathbb{E}_P\bigl[X_{\min\{\tau,n,\rho_\varepsilon^X\}}\bigr]\,.
\end{aligned} \tag{3.127}$$

Combining this with the fact for all $\varepsilon \in [0,\infty)$, $n \in \mathbb{N}$ and all $(\mathbb{F}_t)_{t\in[0,\infty)}$-stopping times $\tau \colon \Omega \to [0,\infty)$ it holds that $\min\{\tau,n,\rho_\varepsilon^X\}$ is a bounded $(\mathbb{F}_t)_{t\in[0,\infty)}$-stopping time (see Exercise 3.4.15) ensures that for all $\varepsilon \in [0,\infty)$, $n \in \mathbb{N}$ and all $(\mathbb{F}_t)_{t\in[0,\infty)}$-stopping times $\tau \colon \Omega \to [0,\infty)$ it holds that

$$\begin{aligned}
\varepsilon\, P\bigl(\sup\nolimits_{t\in[0,\min\{\tau,n\}]} X_t \geq \varepsilon\bigr) &\leq \mathbb{E}_P\bigl[X_{\min\{\tau,n,\rho_\varepsilon^X\}}\bigr] \leq \mathbb{E}_P\bigl[\sup\nolimits_{t\in[0,\min\{\tau,n,\rho_\varepsilon^X\}]} Y_t\bigr]\\
&\leq \mathbb{E}_P\bigl[\sup\nolimits_{t\in[0,\tau]} Y_t\bigr]\,.
\end{aligned} \tag{3.128}$$

Hence, we obtain that for all $\varepsilon \in [0,\infty)$ and all $(\mathbb{F}_t)_{t\in[0,\infty)}$-stopping times $\tau \colon \Omega \to [0,\infty)$ it holds that

$$\begin{aligned}
&\varepsilon\, P\bigl(\sup\nolimits_{t\in[0,\tau]} X_t \geq \varepsilon\bigr) = \varepsilon\, P\bigl(\cup_{n\in\mathbb{N}} \bigl\{\sup\nolimits_{t\in[0,\min\{\tau,n\}]} X_t \geq \varepsilon\bigr\}\bigr)\\
&= \varepsilon \lim_{n\to\infty} P\bigl(\sup\nolimits_{t\in[0,\min\{\tau,n\}]} X_t \geq \varepsilon\bigr) \leq \mathbb{E}_P\bigl[\sup\nolimits_{t\in[0,\tau]} Y_t\bigr]\,.
\end{aligned} \tag{3.129}$$

This proves (3.120). In the next step we observe that (3.120) ensures that for all $\varepsilon, \delta \in (0,\infty)$ and all $(\mathbb{F}_t)_{t\in[0,\infty)}$-stopping times $\tau \colon \Omega \to (0,\infty)$ it holds that

$$\begin{aligned}
&P\bigl(\sup\nolimits_{t\in[0,\tau]} X_t \geq \varepsilon,\, \sup\nolimits_{t\in[0,\tau]} Y_t < \delta\bigr)\\
&= P\bigl(\sup\nolimits_{t\in[0,\tau]} X_t \geq \varepsilon,\, \rho_\delta^Y > \tau,\, \sup\nolimits_{t\in[0,\tau]} Y_t < \delta\bigr)\\
&= P\Bigl(\sup\nolimits_{t\in[0,\min\{\tau,\rho_\delta^Y\}]} X_t \geq \varepsilon,\, \rho_\delta^Y > \tau,\, \sup\nolimits_{t\in[0,\tau]} Y_t < \delta\Bigr)\\
&\leq P\Bigl(\sup\nolimits_{t\in[0,\min\{\tau,\rho_\delta^Y\}]} X_t \geq \varepsilon\Bigr) \leq \tfrac{1}{\varepsilon}\, \mathbb{E}_P\Bigl[\sup\nolimits_{t\in[0,\min\{\tau,\rho_\delta^Y\}]} Y_t\Bigr]\\
&\leq \tfrac{1}{\varepsilon}\, \mathbb{E}_P\Bigl[\min\bigl\{\delta, \sup\nolimits_{t\in[0,\min\{\tau,\rho_\delta^Y\}]} Y_t\bigr\}\Bigr] \leq \tfrac{1}{\varepsilon}\, \mathbb{E}_P\bigl[\min\bigl\{\delta, \sup\nolimits_{t\in[0,\tau]} Y_t\bigr\}\bigr]\,.
\end{aligned} \tag{3.130}$$

This proves (3.121). Furthermore, we observe that (3.121) shows that for all $\varepsilon, \delta \in (0, \infty)$ and all $(\mathbb{F}_t)_{t\in[0,\infty)}$-stopping times $\tau\colon \Omega \to (0, \infty)$ it holds that

$$\begin{aligned} &P\big(\sup\nolimits_{t\in[0,\tau]} X_t \geq \varepsilon\big) \\ &\leq P\big(\sup\nolimits_{t\in[0,\tau]} X_t \geq \varepsilon,\, \sup\nolimits_{t\in[0,\tau]} Y_t < \delta\big) + P\big(\sup\nolimits_{t\in[0,\tau]} Y_t \geq \delta\big) \\ &\leq \tfrac{1}{\varepsilon}\, \mathbb{E}_P\big[\min\big\{\delta, \sup\nolimits_{t\in[0,\tau]} Y_t\big\}\big] + P\big(\sup\nolimits_{t\in[0,\tau]} Y_t \geq \delta\big)\,. \end{aligned} \tag{3.131}$$

This proves (3.122). Next we note that (3.122) and the Markov inequality (see Lemma 2.4.11) show that for all $r, \delta \in (0, \infty)$ and all $(\mathbb{F}_t)_{t\in[0,\infty)}$-stopping times $\tau\colon \Omega \to (0, \infty)$ it holds that

$$\begin{aligned} &P\big(\sup\nolimits_{t\in[0,\tau]} X_t \geq r\big) \\ &\leq \tfrac{1}{r}\, \mathbb{E}_P\big[\min\big\{\delta, \sup\nolimits_{t\in[0,\tau]} Y_t\big\}\big] + P\big(\min\big\{\delta, \sup\nolimits_{t\in[0,\tau]} Y_t\big\} \geq \delta\big) \\ &\leq \big[\tfrac{1}{r} + \tfrac{1}{\delta}\big]\, \mathbb{E}_P\big[\min\big\{\delta, \sup\nolimits_{t\in[0,\tau]} Y_t\big\}\big]\,. \end{aligned} \tag{3.132}$$

This implies that for all $\varepsilon, \delta, r \in (0, \infty)$ and all $(\mathbb{F}_t)_{t\in[0,\infty)}$-stopping times $\tau\colon \Omega \to (0, \infty)$ it holds that

$$\begin{aligned} &\mathbb{E}_P\big[\min\big\{\varepsilon, \sup\nolimits_{t\in[0,\tau]} X_t\big\}\big] \\ &= \mathbb{E}_P\Big[\min\big\{\varepsilon, \sup\nolimits_{t\in[0,\tau]} X_t\big\}\, \mathbb{1}_{\{\sup_{t\in[0,\tau]} X_t < r\}}\Big] \\ &+ \mathbb{E}_P\Big[\min\big\{\varepsilon, \sup\nolimits_{t\in[0,\tau]} X_t\big\}\, \mathbb{1}_{\{\sup_{t\in[0,\tau]} X_t \geq r\}}\Big] \\ &\leq \min\{\varepsilon, r\} + \varepsilon\, P\big(\sup\nolimits_{t\in[0,\tau]} X_t \geq r\big) \\ &\leq \min\{\varepsilon, r\} + \varepsilon\, \big[\tfrac{1}{r} + \tfrac{1}{\delta}\big]\, \mathbb{E}_P\big[\min\big\{\delta, \sup\nolimits_{t\in[0,\tau]} Y_t\big\}\big] \\ &\leq r + \varepsilon\, \big[\tfrac{1}{r} + \tfrac{1}{\delta}\big]\, \mathbb{E}_P\big[\min\big\{\delta, \sup\nolimits_{t\in[0,\tau]} Y_t\big\}\big]\,. \end{aligned} \tag{3.133}$$

Hence, we obtain that for all $\varepsilon, \delta \in (0, \infty)$ and all $(\mathbb{F}_t)_{t\in[0,\infty)}$-stopping times $\tau\colon \Omega \to (0, \infty)$ it holds that

$$\begin{aligned} &\mathbb{E}_P\big[\min\big\{\varepsilon, \sup\nolimits_{t\in[0,\tau]} X_t\big\}\big] \\ &\leq \inf_{r\in(0,\infty)} \big(r + \tfrac{\varepsilon}{r}\, \mathbb{E}_P\big[\min\big\{\delta, \sup\nolimits_{t\in[0,\tau]} Y_t\big\}\big] + \tfrac{\varepsilon}{\delta}\, \mathbb{E}_P\big[\min\big\{\delta, \sup\nolimits_{t\in[0,\tau]} Y_t\big\}\big]\big) \\ &\leq \big|\varepsilon\, \mathbb{E}_P\big[\min\big\{\delta, \sup\nolimits_{t\in[0,\tau]} Y_t\big\}\big]\big|^{1/2} \\ &+ \sqrt{\varepsilon}\, \big|\mathbb{E}_P\big[\min\big\{\delta, \sup\nolimits_{t\in[0,\tau]} Y_t\big\}\big]\big|^{1/2} + \tfrac{\varepsilon}{\delta}\, \mathbb{E}_P\big[\min\big\{\delta, \sup\nolimits_{t\in[0,\tau]} Y_t\big\}\big]\,. \end{aligned} \tag{3.134}$$

This proves (3.123). Moreover, we note that (3.124) is an immediate consequence of (3.123). The proof of Proposition 3.4.16 is thus completed. □

3.4.7 Construction of the stochastic integral

In the next result, Theorem 3.4.17, the existence and uniqueness of the stochastic integral is established (cf., e.g., Proposition 2.26 in [Karatzas and Shreve(1988)]).

Theorem 3.4.17 (Existence and uniqueness of the stochastic integral)**.** *Let $T \in (0, \infty)$, $a \in [0, T]$, $b \in [a, T]$, $d, m \in \mathbb{N}$, let $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t \in [0,T]})$ be a stochastic basis, and let $W\colon [0, T] \times \Omega \to \mathbb{R}^m$ be an m-dimensional standard $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t \in [0,T]})$-Brownian motion. Then there exists a unique linear function*

$$I\colon \left\{ X \in \mathbb{M}([0,T] \times \Omega, \mathbb{R}^{d \times m})\colon \begin{array}{c} X \text{ is } (\mathbb{F}_t)_{t \in [0,T]}\text{-predictable with} \\ P\big(\int_a^b \|X_s\|^2_{\mathbb{R}^{d \times m}} \, ds < \infty \big) = 1 \end{array} \right\} \to L^0(P; \|\cdot\|_{\mathbb{R}^d}) \tag{3.135}$$

which satisfies

(i) that for all $X^n \in \operatorname{dom}(I)$, $n \in \mathbb{N}$, with $\limsup_{n \to \infty} \mathbb{E}_P\big[\min\{1, \int_a^b \|X^n_s\|^2_{\mathbb{R}^{d \times m}} \, ds\}\big] = 0$ it holds that $\limsup_{n \to \infty} \mathbb{E}_P\big[\min\{1, \|I(X^n)\|_{\mathbb{R}^d}\}\big] = 0$ (continuity) and

(ii) that for all $s \in [0, T]$, $t \in (s, T]$, and all $\mathbb{F}_s/\mathcal{B}(\mathbb{R}^{d \times m})$-measurable functions $X\colon \Omega \to \mathbb{R}^{d \times m}$ it holds that

$$I(X \mathbb{1}_{(s,t]}) = \big[X \left(W_{\min\{t,b\}} - W_{\min\{\max\{s,a\},t,b\}}\right)\big]_{P, \mathcal{B}(\mathbb{R}^d)} \tag{3.136}$$

(stochastic integration of simple processes).

Definition 3.4.18 (Stochastic integration)**.** *Let $T \in (0, \infty)$, $a \in [0, T]$, $b \in [a, T]$, $d, m \in \mathbb{N}$, let $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t \in [0,T]})$ be a stochastic basis, and let $W\colon [0, T] \times \Omega \to \mathbb{R}^m$ be an m-dimensional standard $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t \in [0,T]})$-Brownian motion. Then we denote by*

$$I^W_{a,b}\colon \left\{ X \in \mathbb{M}([0,T] \times \Omega, \mathbb{R}^{d \times m})\colon \begin{array}{c} X \text{ is } (\mathbb{F}_t)_{t \in [0,T]}\text{-predictable with} \\ P\big(\int_a^b \|X_s\|^2_{\mathbb{R}^{d \times m}} \, ds < \infty \big) = 1 \end{array} \right\} \to L^0(P; \|\cdot\|_{\mathbb{R}^d}) \tag{3.137}$$

the linear function which satisfies

(i) that for all $X^n \in \operatorname{dom}(I^W_{a,b})$, $n \in \mathbb{N}$, with $\limsup_{n \to \infty} \mathbb{E}_P\big[\min\{1, \int_a^b \|X^n_s\|^2_{\mathbb{R}^{d \times m}} \, ds\}\big] = 0$ it holds that $\limsup_{n \to \infty} \mathbb{E}_P\left[\min\{1, \|I^W_{a,b}(X^n)\|_{\mathbb{R}^d}\}\right] = 0$ (continuity) and

(ii) that for all $s \in [0, T]$, $t \in (s, T]$, and all $\mathbb{F}_s/\mathcal{B}(\mathbb{R}^{d \times m})$-measurable functions $X\colon \Omega \to \mathbb{R}^{d \times m}$ it holds that

$$I^W_{a,b}(X \mathbb{1}_{(s,t]}) = \big[X \left(W_{\min\{t,b\}} - W_{\min\{\max\{s,a\},t,b\}}\right)\big]_{P, \mathcal{B}(\mathbb{R}^d)} \tag{3.138}$$

(stochastic integration of simple processes).

Definition 3.4.19 (Stochastic integral). *Let $T \in (0,\infty)$, $a \in [0,T]$, $b \in [a,T]$, $d,m \in \mathbb{N}$, let $(\Omega,\mathcal{F},P,(\mathbb{F}_t)_{t\in[0,T]})$ be a stochastic basis, let $W\colon [0,T]\times\Omega \to \mathbb{R}^m$ be an m-dimensional standard $(\Omega,\mathcal{F},P,(\mathbb{F}_t)_{t\in[0,T]})$-Brownian motion, and let $X\colon [0,T]\times\Omega \to \mathbb{R}^{d\times m}$ be an $(\mathbb{F}_t)_{t\in[0,T]}/\mathcal{B}(\mathbb{R}^{d\times m})$-predictable stochastic process which satisfies $P(\int_a^b \|X_s\|^2_{\mathbb{R}^{d\times m}}\,ds < \infty) = 1$. Then we denote by $\int_a^b X_s\,dW_s \in L^0(P;\|\cdot\|_{\mathbb{R}^d})$ the set given by*

$$\int_a^b X_s\,dW_s = I^W_{a,b}(X) \tag{3.139}$$

and we call $\int_a^b X_s\,dW_s$ the stochastic integral of X from a to b with respect to W on $(\Omega,\mathcal{F},P,(\mathbb{F}_t)_{t\in[0,T]})$ (we call $\int_a^b X_s\,dW_s$ the stochastic integral of X from a to b).

3.4.8 Properties of the stochastic integral

Exercise 3.4.20. *Let $T \in [0,\infty)$, $t \in [0,T]$, let $(\Omega,\mathcal{F},P,(\mathbb{F}_s)_{s\in[0,T]})$ be a stochastic basis, let $(S,\mathcal{S})$ be a measurable space, let $X\colon \Omega \to S$ be an $\mathcal{F}/\mathcal{S}$-measurable function, let $Y\colon \Omega \to S$ be an $\mathbb{F}_t/\mathcal{S}$-measurable function, and let $A \in \mathcal{F}$ satisfy $P(A) = 1$ and*

$$A \subseteq \{\omega \in \Omega\colon X(\omega) = Y(\omega)\}. \tag{3.140}$$

Prove that X is an $\mathbb{F}_t/\mathcal{S}$-measurable function.

Let $d \in \mathbb{N}$, $T \in (0,\infty)$ and let $(\Omega,\mathcal{F},P,(\mathbb{F}_t)_{t\in[0,T]})$ be a stochastic basis. Then Exercise 3.4.20, in particular, shows that for all $t_1,t_2 \in [0,T]$ with $t_1 \le t_2$ it holds that

$$L^0\big(P|_{\mathbb{F}_0};\|\cdot\|_{\mathbb{R}^d}\big) \subseteq L^0\big(P|_{\mathbb{F}_{t_1}};\|\cdot\|_{\mathbb{R}^d}\big) \subseteq L^0\big(P|_{\mathbb{F}_{t_2}};\|\cdot\|_{\mathbb{R}^d}\big) \subseteq L^0\big(P;\|\cdot\|_{\mathbb{R}^d}\big). \tag{3.141}$$

In Exercise 3.4.20 it is crucial that the filtration is *normal*.

Theorem 3.4.21 (Properties of the stochastic integral). *Let $T \in (0,\infty)$, $d,m \in \mathbb{N}$, $a,b \in [0,T]$ with $a \le b$, let $(\Omega,\mathcal{F},P,(\mathbb{F}_t)_{t\in[0,T]})$ be a stochastic basis, let $W\colon [0,T]\times\Omega \to \mathbb{R}^m$ be an m-dimensional standard $(\Omega,\mathcal{F},P,(\mathbb{F}_t)_{t\in[0,T]})$-Brownian motion, and let $X\colon [0,T]\times\Omega \to \mathbb{R}^{d\times m}$ be an $(\mathbb{F}_t)_{t\in[0,T]}/\mathcal{B}(\mathbb{R}^{d\times m})$-predictable function which satisfies $P(\int_a^b \|X_s\|^2_{\mathbb{R}^{d\times m}}\,ds < \infty) = 1$. Then*

(i) it holds that $\int_a^b X_s\,dW_s \in L^0\big(P|_{\mathbb{F}_b};\|\cdot\|_{\mathbb{R}^d}\big)$,

(ii) it holds that $(\int_a^t X_s\,dW_s)_{t\in[a,b]}$ is an $(\mathbb{F}_t)_{t\in[a,b]}/\mathcal{B}(\mathbb{R}^d)$-adapted stochastic process,

(iii) it holds for all $\alpha,\beta \in \mathbb{R}$ and all $(\mathbb{F}_t)_{t\in[0,T]}/\mathcal{B}(\mathbb{R}^{d\times m})$-predictable stochastic processes $Y,Z\colon [0,T]\times\Omega \to \mathbb{R}^{d\times m}$ with $P\big(\int_a^b \|Y_s\|^2_{HS(\mathbb{R}^m,\mathbb{R}^d)} + \|Z_s\|^2_{HS(\mathbb{R}^m,\mathbb{R}^d)}ds < \infty\big) = 1$ that

$$\int_a^b [\alpha Y_s + \beta Z_s]\,dW_s = \alpha\int_a^b Y_s\,dW_s + \beta\int_a^b Z_s\,dW_s, \tag{3.142}$$

(iv) it holds for all $(\mathbb{F}_t)_{t\in[0,T]}/\mathcal{B}(\mathbb{R}^{d\times m})$-predictable stochastic processes $Y\colon [0,T]\times\Omega\to\mathbb{R}^{d\times m}$ with $\int_a^b \mathbb{E}_P\big[\|Y_s\|^2_{HS(\mathbb{R}^m,\mathbb{R}^d)}\big]\,ds<\infty$ that

$$\mathbb{E}_P\left[\left\|\int_a^b Y_s\,dW_s\right\|^2_{\mathbb{R}^d}\right]=\int_a^b \mathbb{E}_P\Big[\|Y_s\|^2_{HS(\mathbb{R}^m,\mathbb{R}^d)}\Big]\,ds, \qquad \text{(Itô's isometry)}$$

$$\left\|\int_a^b Y_s\,dW_s\right\|_{L^2(P;\|\cdot\|_{\mathbb{R}^d})}=\left(\int_a^b \|Y_s\|^2_{\mathcal{L}^2(P;\|\cdot\|_{HS(\mathbb{R}^m,\mathbb{R}^d)})}\,ds\right)^{\frac12}, \tag{3.143}$$

$$\mathbb{E}_P\left[\int_a^b Y_s\,dW_s\right]=0, \tag{3.144}$$

(v) it holds for all $p\in[2,\infty)$ that

$$\left\|\int_a^b X_s\,dW_s\right\|_{L^p(P;\|\cdot\|_{\mathbb{R}^d})}\le\sqrt{\frac{p\,(p-1)}{2}}\left(\int_a^b \|X_s\|^2_{\mathcal{L}^p(P;\|\cdot\|_{HS(\mathbb{R}^m,\mathbb{R}^d)})}\,ds\right)^{\frac12}$$

$$\left(\mathbb{E}_P\left[\left\|\int_a^b X_s\,dW_s\right\|^p_{\mathbb{R}^d}\right]\right)^{\frac1p}\le\sqrt{\frac{p\,(p-1)}{2}}\left(\int_a^b\left(\mathbb{E}_P\Big[\|X_s\|^p_{HS(\mathbb{R}^m,\mathbb{R}^d)}\Big]\right)^{\frac2p}ds\right)^{\frac12},$$

(Burkholder-Davis-Gundy inequality I)

(vi) there exists an up to indistinguishability unique $(\mathbb{F}_t)_{t\in[a,b]}/\mathcal{B}(\mathbb{R}^d)$-adapted stochastic process $V\colon[a,b]\times\Omega\to\mathbb{R}^d$ with continuous sample paths which satisfies for all $t\in[a,b]$ that $[V_t]_{P,\mathcal{B}(\mathbb{R}^d)}=\int_a^t X_s\,dW_s$ (V is called a continuous modification of $(\int_a^t X_s\,dW_s)_{t\in[a,b]}$), and

(vii) it holds for all $p\in[2,\infty)$ and all continuous modifications $V\colon[a,b]\times\Omega\to\mathbb{R}^d$ of $(\int_a^t X_s\,dW_s)_{t\in[a,b]}$ that

$$\left\|\sup_{s\in[a,b]}\|V_s\|_{\mathbb{R}^d}\right\|_{\mathcal{L}^p(P;|\cdot|_{\mathbb{R}})}\le p\left(\int_a^b \|X_s\|^2_{\mathcal{L}^p(P;\|\cdot\|_{HS(\mathbb{R}^m,\mathbb{R}^d)})}\,ds\right)^{\frac12},$$

$$\left(\mathbb{E}_P\left[\sup_{s\in[a,b]}\|V_s\|^p_{\mathbb{R}^d}\right]\right)^{\frac1p}\le p\left(\int_a^b\left(\mathbb{E}_P\Big[\|X_s\|^p_{HS(\mathbb{R}^m,\mathbb{R}^d)}\Big]\right)^{\frac2p}ds\right)^{\frac12}.$$

(Burkholder-Davis-Gundy inequality II)

The statements of Theorem 3.4.21 and their proofs can, for example, be found in [Klenke(2008)] and [Da Prato and Zabczyk(1992)].

Exercise 3.4.22 (Stochastic integration of L^2-continuous stochastic processes)**.** *Let $T\in(0,\infty)$, $d,m\in\mathbb{N}$, let $(\Omega,\mathcal{F},P,(\mathbb{F}_t)_{t\in[0,T]})$ be a stochastic basis, let $W\colon[0,T]\times\Omega\to\mathbb{R}^m$ be an m-dimensional standard $(\Omega,\mathcal{F},P,(\mathbb{F}_t)_{t\in[0,T]})$-Brownian motion, let $a\in[0,T]$, $b\in$*

$[a,T]$, and let $X\colon [0,T]\times\Omega\to\mathbb{R}^{d\times m}$ be an $(\mathbb{F}_t)_{t\in[0,T]}/\mathcal{B}(\mathbb{R}^{d\times m})$-predictable function with $X\in C([0,T],L^2(P;\|\cdot\|_{\mathbb{R}^{d\times m}}))$. Prove that

$$\limsup_{n\to\infty}\left\|\int_a^b X_s\,dW_s-\left[\sum_{k=0}^{n-1}X_{(a+\frac{k(b-a)}{n})}\left(W_{a+\frac{(k+1)(b-a)}{n}}-W_{a+\frac{k(b-a)}{n}}\right)\right]\right\|_{L^2(P;\|\cdot\|_{\mathbb{R}^d})}=0. \tag{3.145}$$

3.5 Itô stochastic calculus

3.5.1 Itô processes and Itô's formula

This section presents the *Itô formula* from *Itô stochastic calculus* (*Itô calculus*), which is the stochastic analogue of the fundamental theorem of calculus and the chain rule respectively. For this we first briefly review basic properites from deterministic calculus.

Let $T\in(0,\infty)$ be a real number. Then the fundamental theorem of calculus proves that a function (a process) $x\colon[0,T]\to\mathbb{R}$ is continuously differentiable if and only if there exists a continuous function $y\colon[0,T]\to\mathbb{R}$ such that for all $t\in[0,T]$ it holds that

$$x(t)=x(0)+\int_0^t y(s)\,ds \tag{3.146}$$

and in that case it holds for all $t\in[0,T]$ that $x'(t)=y(t)$. Functions of the form (3.146) are crucial in deterministic calculus. The chain rule proves that if $f\colon\mathbb{R}\to\mathbb{R}$ and $x\colon[0,T]\to\mathbb{R}$ are continuously differentiable functions, then so is the composition function $[0,T]\ni t\mapsto f(x(t))\in\mathbb{R}$ and in that case it holds for all $t\in[0,T]$ that

$$\frac{d}{dt}f(x(t))=f'(x(t))\,x'(t). \tag{3.147}$$

Combining (3.146) and (3.147) proves that if $f\colon\mathbb{R}\to\mathbb{R}$ and $x\colon[0,T]\to\mathbb{R}$ are continuously differentiable and if $y\colon[0,T]\to\mathbb{R}$ is a continuous function satisfying (3.146), then it holds for all $t\in[0,T]$ that

$$f(x(t))=f(x(0))+\int_0^t f'(x(s))\,y(s)\,ds. \tag{3.148}$$

Formula (3.148) is the combination of the fundamental theorem of calculus and of the chain rule. In the following we present the stochastic generalization of (3.148) which is known as *Itô's formula* in the literature (see Theorem 3.5.5 below). Recall that the sample paths of a standard Brownian motion are with probability one not continuously differentiable; see (3.74) in Proposition 3.3.14 for details.

Definition 3.5.1 (Itô process – stochastic analogue of a continuously differentiable function/process)**.** *We say that X is an O-valued Itô process on $\mathbf{\Omega}$ with drift Y, diffusion Z, and standard Brownian motion W (we say that X is an Itô process on $\mathbf{\Omega}$ with drift Y, diffusion Z, and standard Brownian motion W, we say that X is an Itô process with drift Y, diffusion Z, and standard Brownian motion W, we say that X is an Itô process) if and only if there exist T, Ω, $\mathcal{F}$, P, $\mathbb{F}$, d, m such that it holds*

(i) that $T \in (0, \infty)$, $d, m \in \mathbb{N}$, $O \in \mathcal{B}(\mathbb{R}^d)$,

(ii) that $\mathbf{\Omega} = (\Omega, \mathcal{F}, P, \mathbb{F})$ is a stochastic basis,

(iii) that $X \in \mathbb{M}([0, T] \times \Omega, O)$ is an $\mathbb{F}/\mathcal{B}(O)$-adapted stochastic process with continuous sample paths,

(iv) that $Y \in \mathbb{M}([0, T] \times \Omega, \mathbb{R}^d)$ is an $\mathbb{F}/\mathcal{B}(\mathbb{R}^d)$-predictable stochastic process,

(v) that $Z \in \mathbb{M}([0, T] \times \Omega, \mathbb{R}^{d \times m})$ is an $\mathbb{F}/\mathcal{B}(\mathbb{R}^{d \times m})$-predictable stochastic process,

(vi) that $P\big(\int_0^T \|Y_s\|_{\mathbb{R}^d} + \|Z_s\|^2_{\mathbb{R}^{d \times m}} \, ds < \infty\big) = 1$,

(vii) that for all $t \in [0, T]$ it holds P-a.s. that

$$X_t = X_0 + \int_0^t Y_s \, ds + \int_0^t Z_s \, dW_s. \tag{3.149}$$

Let $T \in (0, \infty)$, let $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t \in [0,T]})$ be a stochastic basis, and let X be an O-valued Itô process on $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t \in [0,T]})$ with drift Y, diffusion Z, and standard Brownian motion W. Then one often writes

$$dX_t = Y_t \, dt + Z_t \, dW_t \tag{3.150}$$

or

$$dX_t = Y_t \, dt + \sum\nolimits_{i=1}^m Z_t^{(*,i)} \, dW_t^{(i)} \tag{3.151}$$

for $t \in [0, T]$ as an abbreviation for (3.149). Moreover, observe that for all $t_0, t \in [0, T]$ with $t_0 \leq t$ it holds P-a.s. that

$$X_t = X_{t_0} + \int_{t_0}^t Y_s \, ds + \int_{t_0}^t Z_s \, dW_s. \tag{3.152}$$

Theorem 3.4.21 hence shows that for all $t_1, t_2 \in [0, T]$, $p \in [2, \infty)$ with $t_1 < t_2$ it holds

that

$$\begin{aligned}\|X_{t_2}-X_{t_1}\|_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})} &\leq \int_{t_1}^{t_2}\|Y_s\|_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})}\,ds + p\left[\int_{t_1}^{t_2}\|Z_s\|^2_{\mathcal{L}^p(P;\|\cdot\|_{HS(\mathbb{R}^m,\mathbb{R}^d)})}\,ds\right]^{\frac{1}{2}}\\ &\leq \left[\sup_{s\in[0,T]}\|Y_s\|_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})}\right](t_2-t_1)+p\left[\sup_{s\in[0,T]}\|Z_s\|_{\mathcal{L}^p(P;\|\cdot\|_{HS(\mathbb{R}^m,\mathbb{R}^d)})}\right](t_2-t_1)^{\frac{1}{2}}\\ &\leq \left[\sqrt{T}\cdot\sup_{s\in[0,T]}\|Y_s\|_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})}+p\cdot\sup_{s\in[0,T]}\|Z_s\|_{\mathcal{L}^p(P;\|\cdot\|_{HS(\mathbb{R}^m,\mathbb{R}^d)})}\right](t_2-t_1)^{\frac{1}{2}}.\end{aligned} \tag{3.153}$$

This proves in the setting of Definition 3.5.1 that for all $\alpha\in(0,\frac{1}{2}]$, $p\in[2,\infty)$ with

$$\sup_{s\in[0,T]}\|Y_s\|_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})}+\sup_{s\in[0,T]}\|Z_s\|_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^{d\times m}})}<\infty \tag{3.154}$$

it holds that

$$X\in\mathcal{C}^\alpha\big([0,T],L^p(P;\|\cdot\|_{\mathbb{R}^d})\big) \qquad \text{(Temporal regularity for Itô processes)}$$

(cf. Proposition 3.3.14).

Remark 3.5.2. *Let $T\in(0,\infty)$, $p\in[2,\infty)$, $d,m\in\mathbb{N}$, $O\in\mathcal{B}(\mathbb{R}^d)$, let $(\Omega,\mathcal{F},P,(\mathbb{F}_t)_{t\in[0,T]})$ be a stochastic basis, let $W\colon[0,T]\times\Omega\to\mathbb{R}^m$ be an m-dimensional standard $(\Omega,\mathcal{F},P,(\mathbb{F}_t)_{t\in[0,T]})$-Brownian motion, let $X\colon[0,T]\times\Omega\to O$ be an Itô process on $(\Omega,\mathcal{F},P,(\mathbb{F}_t)_{t\in[0,T]})$ with drift $Y\colon[0,T]\times\Omega\to\mathbb{R}^d$, diffusion $Z\colon[0,T]\times\Omega\to\mathbb{R}^{d\times m}$, and standard Brownian motion $W\colon[0,T]\times\Omega\to\mathbb{R}^m$, and assume that $\sup_{t\in[0,T]}\mathbb{E}_P\big[\|Y_t\|^p_{\mathbb{R}^d}+\|Z_t\|^p_{\mathbb{R}^{d\times m}}\big]<\infty$. Then observe that for all $N\in\mathbb{N}$, $\alpha\in(0,\infty)$ it holds that*

$$\begin{aligned}\sum_{k=0}^{N-1}\big\|X_{(k+1)T/N}-X_{kT/N}\big\|^\alpha_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})} &\leq \sum_{k=0}^{N-1}\left[|X|_{\mathcal{C}^{1/2}([0,T],L^p(P;\|\cdot\|_{\mathbb{R}^d}))}\frac{\sqrt{T}}{\sqrt{N}}\right]^\alpha\\ &= T^{\alpha/2}\,|X|^\alpha_{\mathcal{C}^{1/2}([0,T],L^p(P;\|\cdot\|_{\mathbb{R}^d}))}\,N^{(1-\alpha/2)}.\end{aligned} \tag{3.155}$$

This proves that for all $\alpha\in(2,\infty)$ it holds that

$$\lim_{N\to\infty}\left[\sum_{k=0}^{N-1}\big\|X_{(k+1)T/N}-X_{kT/N}\big\|^\alpha_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})}\right]=0. \tag{3.156}$$

Remark 3.5.3. *Let $T \in (0,\infty)$, $p \in [2,\infty)$, $m \in \mathbb{N}$, let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $W\colon [0,T] \times \Omega \to \mathbb{R}^m$ be an m-dimensional standard Brownian motion. Then observe that for all $N \in \mathbb{N}$, $\alpha \in (0,\infty)$ it holds that*

$$\begin{aligned}
&\sum_{k=0}^{N-1} \left\| W_{(k+1)T/N} - W_{kT/N} \right\|^{\alpha}_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})} \\
&= \left| \frac{\sqrt{T}}{\sqrt{N}} \right|^{\alpha} \sum_{k=0}^{N-1} \left\| \frac{\sqrt{N}}{\sqrt{T}} \left[W_{(k+1)T/N} - W_{kT/N} \right] \right\|^{\alpha}_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})} \\
&= T^{\alpha/2} N^{(1-\alpha/2)} \left\| \frac{\sqrt{N}}{\sqrt{T}} W_{\frac{T}{N}} \right\|^{\alpha}_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})} = T^{\alpha/2} N^{(1-\alpha/2)} \left\| \frac{W_T}{\sqrt{T}} \right\|^{\alpha}_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})} \\
&= N^{(1-\alpha/2)} \left\| W_T \right\|^{\alpha}_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})} .
\end{aligned} \tag{3.157}$$

This proves that for all $\alpha \in (0,\infty)$ it holds that

$$\lim_{N\to\infty} \left[\sum_{k=0}^{N-1} \left\| W_{(k+1)T/N} - W_{kT/N} \right\|^{\alpha}_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})} \right] = \begin{cases} \infty & : \alpha < 2 \\ \|W_T\|^{\alpha}_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})} & : \alpha = 2 \\ 0 & : \alpha > 2 \end{cases} . \tag{3.158}$$

The stochastic version of the fundamental theorem of calculus and the chain rule is given in the next result, Theorem 3.5.5. To formulate Theorem 3.5.5, the following notation is used.

Definition 3.5.4 (Canonical basis)**.** *Let $k \in \mathbb{N}$. Then we denote by $e_1^{(k)}, \dots, e_k^{(k)} \in \mathbb{R}^k$ the vectors given by*

$$e_1^{(k)} = (1,0,\dots,0), \quad e_2^{(k)} = (0,1,0,\dots,0), \quad \dots, \quad e_k^{(k)} = (0,\dots,0,1) \tag{3.159}$$

and we call $\{e_1^{(k)}, \dots, e_k^{(k)}\}$ the canonical basis of the $\mathbb{R}^k$.

Theorem 3.5.5 (Itô's formula – stochastic analogue of the fundamental theorem of calculus and the chain rule)**.** *Let $T \in (0,\infty)$, $t_0 \in [0,T]$, $t \in [t_0,T]$, $d,m,v \in \mathbb{N}$, let $O \subseteq \mathbb{R}^d$ be an open set, let $f \in C^2(O,\mathbb{R}^v)$, let $(\Omega,\mathcal{F},P,(\mathbb{F}_t)_{t\in[0,T]})$ be a stochastic basis, let $W = (W^{(1)},\dots,W^{(m)})\colon [0,T]\times\Omega \to \mathbb{R}^m$ be an m-dimensional standard $(\Omega,\mathcal{F},P,(\mathbb{F}_t)_{t\in[0,T]})$-Brownian motion, and let $X\colon [0,T]\times\Omega \to O$ be an O-valued Itô process on $(\Omega,\mathcal{F},P,(\mathbb{F}_t)_{t\in[0,T]})$ with drift $Y = (Y^{(1)},\dots,Y^{(d)})\colon [0,T]\times\Omega\to\mathbb{R}^d$, diffusion $Z = (Z^{(k,i)})_{k\in\{1,\dots,d\},i\in\{1,\dots,m\}}\colon [0,T]\times\Omega\to\mathbb{R}^{d\times m}$, and standard Brownian motion $W\colon [0,T]\times\Omega\to\mathbb{R}^m$. Then*

(i) it holds P-a.s. that

$$f(X_t) = f(X_{t_0}) + \int_{t_0}^t f'(X_s)\,Y_s\,ds + \int_{t_0}^t f'(X_s)\,Z_s\,dW_s + \frac{1}{2}\sum_{i=1}^m \int_{t_0}^t f''(X_s)\big(Z_s e_i^{(m)}, Z_s e_i^{(m)}\big)\,ds, \qquad \text{(Itô's formula)}$$

i.e.,

(ii) it holds P-a.s. that

$$\begin{aligned} &f(X_t) \\ &= f(X_{t_0}) + \int_{t_0}^t \left[\sum_{k=1}^d \Big(\tfrac{\partial}{\partial x_k} f\Big)(X_s)\cdot Y_s^{(k)} + \frac{1}{2}\sum_{i=1}^m \sum_{k,l=1}^d \Big(\tfrac{\partial^2}{\partial x_k \partial x_l} f\Big)(X_s)\cdot Z_s^{(k,i)}\cdot Z_s^{(l,i)}\right] ds \\ &\quad + \sum_{i=1}^m \sum_{k=1}^d \int_{t_0}^t \Big(\tfrac{\partial}{\partial x_k} f\Big)(X_s)\cdot Z_s^{k,i}\,dW_s^{(i)}, \end{aligned} \tag{3.160}$$

i.e.,

(iii) the stochastic process $f(X_t)$, $t \in [0,T]$, is an $\mathbb{R}^v$-valued Itô process on $(\Omega,\mathcal{F},P,(\mathbb{F}_t)_{t\in[0,T]})$ with drift $f'(X_t)\,Y_t + \frac{1}{2}\sum_{i=1}^m f''(X_t)\big(Z_t e_i^{(m)}, Z_t e_i^{(m)}\big)$, $t\in[0,T]$, diffusion $f'(X_t)\,Z_t$, $t\in[0,T]$, and standard Brownian motion W.

Sketch of the proof of Theorem 3.5.5. We restrict ourself in this sketch of the proof of Theorem 3.5.5 to the case where the derivatives of $f\colon O\to\mathbb{R}^k$ are globally bounded and globally Lipschitz continuous, where $[t_0,t] = [0,T]$, and where $\sup_{t\in[0,T]} \mathbb{E}_P\big[\|Y_t\|_{\mathbb{R}^d}^4 + \|Z_t\|_{\mathbb{R}^{d\times m}}^4\big] < \infty$. The proof of Theorem 3.5.5 is based on discretisations of the time interval $[0,T]$. More precisely, let $(t_k^N)_{k\in\{0,1,\dots,N\}} \subseteq [0,T]$ be real numbers with the property that for all $N\in\mathbb{N}$, $k\in\{0,1,\dots,N\}$ it holds that

$$t_k^N = \frac{kT}{N}. \tag{3.161}$$

Then observe that *Taylor's formula* proves that for all $N \in \mathbb{N}$ it holds that

$$\begin{aligned}
f(X_T) &= f\big(X_{t_N^N}\big) \\
&= f(X_0) + \sum_{k=0}^{N-1} f\big(X_{t_{k+1}^N}\big) - f\big(X_{t_k^N}\big) \\
&= f(X_0) + \sum_{k=0}^{N-1} f'\big(X_{t_k^N}\big)\big(X_{t_{k+1}^N} - X_{t_k^N}\big) + \frac{1}{2}\sum_{k=0}^{N-1} f''\big(X_{t_k^N}\big)\big(X_{t_{k+1}^N} - X_{t_k^N}, X_{t_{k+1}^N} - X_{t_k^N}\big) \\
&+ \underbrace{\sum_{k=0}^{N-1} \int_0^1 \Big[f''\big(X_{t_k^N} + r\big(X_{t_{k+1}^N} - X_{t_k^N}\big)\big) - f''\big(X_{t_k^N}\big)\Big]\big(X_{t_{k+1}^N} - X_{t_k^N}, X_{t_{k+1}^N} - X_{t_k^N}\big)\,(1-r)\,dr}_{\to 0 \text{ in probability as } N\to\infty}.
\end{aligned} \tag{3.162}$$

Next observe that for all $N \in \mathbb{N}$ it holds P-a.s. that

$$\begin{aligned}
&\sum_{k=0}^{N-1} f'\big(X_{t_k^N}\big)\big(X_{t_{k+1}^N} - X_{t_k^N}\big) \\
&= \sum_{k=0}^{N-1} \int_{t_k^N}^{t_{k+1}^N} f'\big(X_{t_k^N}\big)\, Y_s\, ds + \sum_{k=0}^{N-1} \int_{t_k^N}^{t_{k+1}^N} f'\big(X_{t_k^N}\big)\, Z_s\, dW_s \\
&= \underbrace{\int_0^T f'\big(X_{\lfloor s \rfloor_{T/N}}\big)\, Y_s\, ds}_{\substack{\to \int_0^T f'(X_s)\, Y_s\, ds \\ \text{in probability as } N\to\infty}} + \underbrace{\int_0^T f'\big(X_{\lfloor s \rfloor_{T/N}}\big)\, Z_s\, dW_s}_{\substack{\to \int_0^T f'(X_s)\, Z_s\, dW_s \\ \text{in probability as } N\to\infty}}.
\end{aligned} \tag{3.163}$$

Furthermore, note that for all $N \in \mathbb{N}$ it holds P-a.s. that

$$\begin{aligned}
&\sum_{k=0}^{N-1} f''\big(X_{t_k^N}\big)\big(X_{t_{k+1}^N} - X_{t_k^N}, X_{t_{k+1}^N} - X_{t_k^N}\big) \\
&= \underbrace{\sum_{k=0}^{N-1} f''\big(X_{t_k^N}\big)\left(\int_{t_k^N}^{t_{k+1}^N} Y_s\, ds, \int_{t_k^N}^{t_{k+1}^N} Y_s\, ds\right)}_{\to 0 \text{ in probability as } N\to\infty} \\
&+ 2\underbrace{\sum_{k=0}^{N-1} f''\big(X_{t_k^N}\big)\left(\int_{t_k^N}^{t_{k+1}^N} Y_s\, ds, \int_{t_k^N}^{t_{k+1}^N} Z_s\, dW_s\right)}_{\to 0 \text{ in probability as } N\to\infty} \\
&+ \underbrace{\sum_{k=0}^{N-1} f''\big(X_{t_k^N}\big)\left(\int_{t_k^N}^{t_{k+1}^N} Z_s\, dW_s, \int_{t_k^N}^{t_{k+1}^N} Z_s\, dW_s\right)}_{\to \sum_{i=1}^m \int_0^T f''(X_s)(Z_s e_i^{(m)}, Z_s e_i^{(m)})\, ds \text{ in probability as } N\to\infty}.
\end{aligned} \tag{3.164}$$

Combining (3.162)–(3.164) completes the proof of Theorem 3.5.5 in the case where the above formulated additional assumptions are fulfilled. □

The function $f \in C^2(O, \mathbb{R}^v)$ in Theorem 3.5.5 is often referred to as *test* function. Let us illustrate the consequences of Theorem 3.5.5 by two examples.

Example 3.5.6 (Iterated stochastic integrals)**.** *Let $T \in (0,\infty)$, let $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$ be a stochastic basis, and let $W\colon [0,T] \times \Omega \to \mathbb{R}$ be a standard $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$-Brownian motion. Then Itô's formula applied to the function $\mathbb{R} \ni x \mapsto x^2 \in \mathbb{R}$ proves that for all $t \in [0,T]$ it holds P-a.s. that*

$$(W_t)^2 = 2 \int_0^t W_s \, dW_s + t. \tag{3.165}$$

This shows, in particular, that for all $t \in [0,T]$ it holds P-a.s. that

$$\int_0^t \int_0^s dW_u \, dW_s = \int_0^t \int_0^s 1 \, dW_u \, dW_s = \frac{(W_t)^2 - t}{2}. \tag{3.166}$$

Example 3.5.7 (Geometric Brownian motion)**.** *Let $T \in (0,\infty)$, $\alpha, \beta, \xi \in \mathbb{R}$, let $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$ be a stochastic basis, let $W\colon [0,T] \times \Omega \to \mathbb{R}$ is a standard $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$-Brownian motion, and let $X\colon [0,T] \times \Omega \to \mathbb{R}$ be the function which satisfies for all $t \in [0,T]$ that*

$$X_t = \exp\Big(\Big(\alpha - \tfrac{\beta^2}{2}\Big)\, t + \beta \, W_t\Big)\, \xi. \tag{3.167}$$

The process X is referred to as geometric Brownian motion; cf. Exercise 3.3.9, Exercise 3.3.11, and Section 4.7.2. Itô's formula applied to the test function $\mathbb{R} \ni x \mapsto e^x \cdot \xi \in \mathbb{R}$ and the Itô process $(\alpha - \frac{\beta^2}{2})\, t + \beta \, W_t$, $t \in [0,T]$, proves that for all $t \in [0,T]$ it holds P-a.s. that

$$\begin{aligned} X_t = \exp\Big(\Big(\alpha - \tfrac{\beta^2}{2}\Big)\, t + \beta \, W_t\Big)\, \xi &= \xi + \int_0^t e^{\left((\alpha - \beta^2/2)s + \beta \, W_s\right)} \, \xi \left(\alpha - \tfrac{\beta^2}{2}\right) ds \\ &\quad + \int_0^t e^{\left((\alpha - \beta^2/2)s + \beta \, W_s\right)} \, \xi \, \beta \, dW_s + \frac{1}{2} \int_0^t e^{\left((\alpha - \beta^2/2)s + \beta \, W_s\right)} \, \xi \, \beta^2 \, ds \\ &= \xi + \int_0^t e^{\left((\alpha - \beta^2/2)s + \beta \, W_s\right)} \, \xi \, \alpha \, ds + \int_0^t e^{\left((\alpha - \beta^2/2)s + \beta \, W_s\right)} \, \xi \, \beta \, dW_s. \end{aligned} \tag{3.168}$$

Putting (3.167) *into the integrands in* (3.168) *shows that for all $t \in [0,T]$ it holds P-a.s. that*

$$X_t = X_0 + \int_0^t \alpha \, X_s \, ds + \int_0^t \beta \, X_s \, dW_s. \tag{3.169}$$

The process X is thus an Itô process with drift αX, diffusion βX, and standard Brownian motion W (cf. Definition 4.2.1 below).

3.5.2 Itô's formula for time-dependent test functions

In Theorem 3.5.5 Itô's formula is presented for "test" functions $f\colon O \to \mathbb{R}^v$ that are twice continuously differentiable functions from O to $\mathbb{R}^v$. Itô's formula can be extended so that it can be applied to test functions $f\colon [t_0, t] \times O \to \mathbb{R}^v$ that depend on both $s \in [t_0, t]$ and $x \in O$. This is the subject of the following corollary of Theorem 3.5.5.

Corollary 3.5.8 (Itô's formula for time-dependent test functions)**.** *Let $T \in (0,\infty)$, $t_0 \in [0,T)$, $t \in (t_0,T]$ $d,m,v \in \mathbb{N}$, let $(\Omega,\mathcal{F},P,(\mathbb{F}_t)_{t\in[0,T]})$ be a stochastic basis, let $W = (W^{(1)},\dots,W^{(m)})\colon [0,T]\times\Omega \to \mathbb{R}^m$ be a standard $(\mathbb{F}_t)_{t\in[0,T]}$-Brownian motion, let $O \subseteq \mathbb{R}^d$ be an open set, let*

$$f\colon [t_0,t]\times O \ni (s,x) \mapsto f(s,x) \in \mathbb{R}^v \qquad \text{(test function)}$$

be a twice *continuously differentiable function, and let $X\colon [0,T]\times\Omega \to O$ be an Itô process on $(\Omega,\mathcal{F},P,(\mathbb{F}_t)_{t\in[0,T]})$ with drift*

$$Y = (Y^{(1)},\dots,Y^{(d)})\colon [0,T]\times\Omega \to \mathbb{R}^d, \qquad \text{(drift)}$$

diffusion

$$Z = (Z^{(k,i)})_{k\in\{1,\dots,d\},i\in\{1,\dots,m\}}\colon [0,T]\times\Omega \to \mathbb{R}^{d\times m} \qquad \text{(diffusion)}$$

and standard Brownian motion $W = (W^{(1)},\dots,W^{(m)})\colon [0,T]\times\Omega \to \mathbb{R}^m$. Then

(i) it holds P-a.s. that

$$\begin{aligned} f(t,X_t) &= f(t_0,X_{t_0}) \\ &+ \int_{t_0}^t \left[\left(\tfrac{\partial}{\partial s} f\right)(s,X_s) + \left(\tfrac{\partial}{\partial x} f\right)(s,X_s)\, Y_s + \frac{1}{2}\sum_{i=1}^m \left(\tfrac{\partial^2}{\partial x^2} f\right)(s,X_s)\big(Z_s e_i^{(m)}, Z_s e_i^{(m)}\big) \right] ds \\ &+ \int_{t_0}^t \left(\tfrac{\partial}{\partial x} f\right)(s,X_s)\, Z_s\, dW_s, \end{aligned} \qquad \text{(Itô's formula)}$$

i.e.,

(ii) it holds P-a.s. that

$$\begin{aligned} f(t,X_t) = f(t_0,X_{t_0}) &+ \int_{t_0}^t \left[\left(\tfrac{\partial}{\partial s} f\right)(s,X_s) + \sum_{k=1}^d \left(\tfrac{\partial}{\partial x_k} f\right)(s,X_s)\cdot Y_s^{(k)} \right] ds \\ &+ \int_{t_0}^t \left[\frac{1}{2}\sum_{i=1}^m \sum_{k,l=1}^d \left(\tfrac{\partial^2}{\partial x_k \partial x_l} f\right)(s,X_s)\cdot Z_s^{(k,i)}\cdot Z_s^{(l,i)} \right] ds \\ &+ \sum_{i=1}^m \sum_{k=1}^d \int_{t_0}^t \left(\tfrac{\partial}{\partial x_k} f\right)(s,X_s)\cdot Z_s^{k,i}\, dW_s^{(i)}. \end{aligned} \tag{3.170}$$

Proof of Corollary 3.5.8. Throughout this proof let $\bar{X}\colon [0,T]\times\Omega \to \mathbb{R}\times O$, $\bar{Y}\colon [0,T]\times\Omega \to \mathbb{R}^{d+1}$, and $\bar{Z}\colon [0,T]\times\Omega \to \mathbb{R}^{(d+1)\times m}$ be $(\mathbb{F}_t)_{t\in[0,T]}$-predictable stochastic processes with the property that for all $t \in [0,T]$ it holds that

$$\bar{X}_t = \begin{pmatrix} t \\ X_t \end{pmatrix}, \qquad \bar{Y}_t = \begin{pmatrix} 1 \\ Y_t \end{pmatrix}, \qquad \text{and} \qquad \bar{Z}_t = \begin{pmatrix} 0\,0\dots0 \\ Z_t \end{pmatrix} \in \mathbb{R}^{(d+1)\times m} \tag{3.171}$$

and let $\bar{f}\colon \mathbb{R}\times O \to \mathbb{R}^v$ be the twice continuously differentiable function which satisfies for all $(s,x)\in[t_0,t]\times O$ that

$$\bar{f}(s,x) = f(s,x). \tag{3.172}$$

Note that $\bar{X}\colon [0,T]\times\Omega\to\mathbb{R}\times O$ is an Itô process on $(\Omega,\mathcal{F},P,(\mathbb{F}_t)_{t\in[0,T]})$ with drift $\bar{Y}$, diffusion $\bar{Z}$, and standard Brownian motion W. Theorem 3.5.5 then proves that it holds P-a.s. that

$$\begin{aligned}\bar{f}(\bar{X}_t) = \bar{f}(\bar{X}_{t_0}) + \int_{t_0}^{t} \bar{f}'(\bar{X}_s)\,\bar{Y}_s\,ds + \int_{t_0}^{t} \bar{f}'(\bar{X}_s)\,\bar{Z}_s\,dW_s \\ + \frac{1}{2}\int_{t_0}^{t} \bar{f}''(\bar{X}_s)\big(\bar{Z}_s e_i^{(m)}, \bar{Z}_s e_i^{(m)}\big)\,ds.\end{aligned} \tag{3.173}$$

This completes the proof of Corollary 3.5.8. □

3.6 Martingales

To introduce the notion of a *martingale*, we briefly recall the concept of conditional expectations.

Definition 3.6.1 (Conditional expectation). *Let $d\in\mathbb{N}$, let $(\Omega,\mathcal{F},P)$ be a probability space, let $\mathcal{G}\subseteq\mathcal{F}$ be a sigma-algebra on Ω, and let $X\in\mathcal{L}^1(P;\|\cdot\|_{\mathbb{R}^d})$ be an $\mathcal{F}/\mathcal{B}(\mathbb{R}^d)$-measurable mapping. Then an $\mathcal{G}/\mathcal{B}(\mathbb{R}^d)$-measurable mapping $Y\in\mathcal{L}^1(P|_{\mathcal{G}};\|\cdot\|_{\mathbb{R}^d})$ is called a* conditional expectation of X given $\mathcal{G}$ *if it holds for all $A\in\mathcal{G}$ that*

$$\mathbb{E}_P\big[\mathbb{1}_A Y\big] = \mathbb{E}_P\big[\mathbb{1}_A X\big]. \tag{3.174}$$

Conditional expecations exist and are unique up to equivalence. This is subject of the next theorem.

Theorem 3.6.2 (Conditional expectation). *Let $d\in\mathbb{N}$, let $(\Omega,\mathcal{F},P)$ be a probability space and let $\mathcal{G}\subseteq\mathcal{F}$ be a sigma-algebra on Ω. Then there exists a unique mapping*

$$\mathbb{E}_P\big[\,\cdot\,|\mathcal{G}\big]\colon L^1(P;\|\cdot\|_{\mathbb{R}^d}) \to L^1(P|_{\mathcal{G}};\|\cdot\|_{\mathbb{R}^d}) \tag{3.175}$$

which fulfills for every $X\in L^1(P;\|\cdot\|_{\mathbb{R}^d})$ that $\mathbb{E}_P\big[X|\mathcal{G}\big]$ is a conditional expectation of X given $\mathcal{G}$. The function $\mathbb{E}_P\big[\,\cdot\,|\mathcal{G}\big]$ is a linear mapping from $L^1(P;\|\cdot\|_{\mathbb{R}^d})$ to $L^1(P|_{\mathcal{G}};\|\cdot\|_{\mathbb{R}^d})$.

Theorem 3.6.2 is, e.g., proved as Theorem 8.2 in [Klenke(2008)]. We also refer Section 8 in [Klenke(2008)] for further properties of conditional expectations.

Proposition 3.6.3 (Conditional expectation as best approximation)**.** *Let $d \in \mathbb{N}$, let $(\Omega, \mathcal{F}, P)$ be a probability space, let $\mathcal{G} \subseteq \mathcal{F}$ be a sigma-algebra on Ω and let $X \in L^2(P; \|\cdot\|_{\mathbb{R}^d})$. Then it holds for all $Y \in L^2(P|_{\mathcal{G}}; \|\cdot\|_{\mathbb{R}^d})$ that*

$$\mathbb{E}_P\big[\|X - Y\|^2_{\mathbb{R}^d}\big] = \mathbb{E}_P\Big[\big\|X - \mathbb{E}_P[X|\mathcal{G}]\big\|^2_{\mathbb{R}^d}\Big] + \mathbb{E}_P\Big[\big\|\mathbb{E}_P[X|\mathcal{G}] - Y\big\|^2_{\mathbb{R}^d}\Big] \tag{3.176}$$

and, in particular, it holds that

$$\min_{Y \in L^2(P|_{\mathcal{G}}; \|\cdot\|_{\mathbb{R}^d})} \|X - Y\|_{L^2(P; \|\cdot\|_{\mathbb{R}^d})} = \big\|X - \mathbb{E}_P[X|\mathcal{G}]\big\|_{L^2(P; \|\cdot\|_{\mathbb{R}^d})}. \tag{3.177}$$

Proof of Proposition 3.6.3. Note that for all $Y \in L^2(P|_{\mathcal{G}}; \|\cdot\|_{\mathbb{R}^d})$ it holds that

$$\begin{aligned} &\mathbb{E}_P\big[\|X - Y\|^2_{\mathbb{R}^d}\big] = \mathbb{E}_P\Big[\big\|\big(X - \mathbb{E}_P[X|\mathcal{G}]\big) + \big(\mathbb{E}_P[X|\mathcal{G}] - Y\big)\big\|^2_{\mathbb{R}^d}\Big] \\ &= \mathbb{E}_P\Big[\big\|X - \mathbb{E}_P[X|\mathcal{G}]\big\|^2_{\mathbb{R}^d} + 2\big\langle X - \mathbb{E}_P[X|\mathcal{G}], \mathbb{E}_P[X|\mathcal{G}] - Y\big\rangle_{\mathbb{R}^d} + \big\|\mathbb{E}_P[X|\mathcal{G}] - Y\big\|^2_{\mathbb{R}^d}\Big] \\ &= \mathbb{E}_P\Big[\big\|X - \mathbb{E}_P[X|\mathcal{G}]\big\|^2_{\mathbb{R}^d}\Big] + \mathbb{E}_P\Big[\big\|\mathbb{E}_P[X|\mathcal{G}] - Y\big\|^2_{\mathbb{R}^d}\Big] \\ &\quad + 2\,\mathbb{E}_P\big[\big\langle X - \mathbb{E}_P[X|\mathcal{G}], \mathbb{E}_P[X|\mathcal{G}] - Y\big\rangle_{\mathbb{R}^d}\big]. \end{aligned} \tag{3.178}$$

Moreover, observe that the *tower property* of the conditional expectation proves that for all $Y \in L^2(P|_{\mathcal{G}}; \|\cdot\|_{\mathbb{R}^d})$ it holds that

$$\begin{aligned} &\mathbb{E}_P\big[\big\langle X - \mathbb{E}_P[X|\mathcal{G}], \mathbb{E}_P[X|\mathcal{G}] - Y\big\rangle_{\mathbb{R}^d}\big] \\ &= \mathbb{E}_P\Big[\mathbb{E}_P\Big[\big\langle X - \mathbb{E}_P[X|\mathcal{G}], \mathbb{E}_P[X|\mathcal{G}] - Y\big\rangle_{\mathbb{R}^d}\,\Big|\,\mathcal{G}\Big]\Big] \\ &= \mathbb{E}_P\Big[\Big\langle \mathbb{E}_P\Big[X - \mathbb{E}_P[X|\mathcal{G}]\,\Big|\,\mathcal{G}\Big], \mathbb{E}_P[X|\mathcal{G}] - Y\Big\rangle_{\mathbb{R}^d}\Big] \\ &= \mathbb{E}_P\Big[\big\langle \mathbb{E}_P[X|\mathcal{G}] - \mathbb{E}_P[X|\mathcal{G}], \mathbb{E}_P[X|\mathcal{G}] - Y\big\rangle_{\mathbb{R}^d}\Big] = 0. \end{aligned} \tag{3.179}$$

Combining (3.178) and (3.179) completes the proof of Proposition 3.6.3. □

In the case $d = 1$, Proposition 3.6.3 is, e.g., also proved as Corollary 8.16 in [Klenke(2008)].

Definition 3.6.4 (Martingale)**.** *Let $\mathbb{T} \subseteq \mathbb{R}$ be a set, let $d \in \mathbb{N}$, let $(\Omega, \mathcal{F}, P)$ be a probability space with a filtration $(\mathbb{F}_t)_{t \in \mathbb{T}}$ and let $X \colon \mathbb{T} \times \Omega \to \mathbb{R}^d$ be an $(\mathbb{F}_t)_{t \in \mathbb{T}}$-adapted stochastic process with the property that for all $t \in \mathbb{T}$ it holds that $X_t \in \mathcal{L}^1(P; \|\cdot\|_{\mathbb{R}^d})$. Then X is called an $(\mathbb{F}_t)_{t \in \mathbb{T}}$-*martingale *if*

$$\mathbb{E}_P\big[X_{t_2}\,|\,\mathbb{F}_{t_1}\big] = X_{t_1} \tag{3.180}$$

P-a.s. for all $t_1, t_2 \in \mathbb{T}$ with $t_1 \leq t_2$.

Proposition 3.6.5 (The stochastic integral process is a martingale)**.** *Let $T \in (0,\infty)$, $d, m \in \mathbb{N}$, let $(\Omega, \mathcal{F}, P)$ be a probability space with a normal filtration $(\mathbb{F}_t)_{t\in[0,T]}$, let $W\colon [0,T]\times\Omega \to \mathbb{R}^m$ be a standard $(\mathbb{F}_t)_{t\in[0,T]}$-Brownian motion and let $X \in L^2(\mathcal{P}_{P,(\mathbb{F}_t)_{t\in[0,T]}}; \|\cdot\|_{\mathbb{R}^{d\times m}})$. Then the stochastic integral process $(\int_0^t X_s\, dW_s)_{t\in[0,T]}$ is an $(\mathbb{F}_t)_{t\in[0,T]}$-martingale, i.e.,*

$$\mathbb{E}_P\left[\int_0^{t_2} X_s\, dW_s \,\middle|\, \mathbb{F}_{t_1}\right] = \int_0^{t_1} X_s\, dW_s \tag{3.181}$$

P-a.s. for all $t_1, t_2 \in [0,T]$ with $t_1 \le t_2$.

Proof of Proposition 3.6.5. First, observe that

$$\begin{aligned} \mathbb{E}_P\big[W_{t_2}\big|\mathbb{F}_{t_1}\big] &= \mathbb{E}_P\big[W_{t_2}-W_{t_1}\big|\mathbb{F}_{t_1}\big] + \mathbb{E}_P\big[W_{t_1}\big|\mathbb{F}_{t_1}\big] = \mathbb{E}_P\big[W_{t_2}-W_{t_1}\big|\mathbb{F}_{t_1}\big] + W_{t_1} \\ &= \mathbb{E}_P[W_{t_2}-W_{t_1}] + W_{t_1} = W_{t_1} \end{aligned} \tag{3.182}$$

P-a.s. for all $t_1, t_2 \in [0,T]$ with $t_1 \le t_2$. *This proves that W is an $(\mathbb{F}_t)_{t\in[0,T]}$-martingale.* Next observe that if X is $(\mathbb{F}_t)_{t\in[0,T]}$-simple, then there exist $n \in \mathbb{N}$, $s_1, \dots, s_n$ with $s_1 < \dots < s_n$ and for every $k \in \{1, \dots, n-1\}$ an $\mathbb{F}_{s_k}/\mathcal{B}(\mathbb{R}^{d\times m})$-measurable mapping $H_k\colon \Omega \to \mathbb{R}^{d\times m}$ such that

$$X_s = \sum_{k=1}^{n-1} H_k \cdot \mathbb{1}_{(s_k, s_{k+1}]}(s) \tag{3.183}$$

for all $s \in [0,T]$ and in that case, we obtain

$$\int_{t_1}^{t_2} X_s\, dW_s = \sum_{\substack{k\in\{1,\dots,n-1\},\\ t_1 < s_{k+1},\\ s_k < t_2}} H_k \left(W_{\min(s_{k+1},t_2)} - W_{\max(s_k,t_1)}\right) \tag{3.184}$$

P-a.s. for all $t_1, t_2 \in [0,T]$ with $t_1 \le t_2$ and therefore

$$\begin{aligned} \mathbb{E}_P\left[\int_{t_1}^{t_2} X_s\, dW_s \middle| \mathbb{F}_{t_1}\right] &= \sum_{\substack{k\in\{1,\dots,n-1\},\\ t_1 < s_{k+1},\\ s_k < t_2}} \mathbb{E}_P\left[H_k \left(W_{\min(s_{k+1},t_2)} - W_{\max(s_k,t_1)}\right) \middle| \mathbb{F}_{t_1}\right] \\ &= \sum_{\substack{k\in\{1,\dots,n-1\},\\ t_1 < s_{k+1},\\ s_k < t_2}} \mathbb{E}_P\left[\mathbb{E}_P\left[H_k \left(W_{\min(s_{k+1},t_2)} - W_{\max(s_k,t_1)}\right) \middle| \mathbb{F}_{\max(s_k,t_1)}\right] \middle| \mathbb{F}_{t_1}\right] \\ &= \sum_{\substack{k\in\{1,\dots,n-1\},\\ t_1 < s_{k+1},\\ s_k < t_2}} \mathbb{E}_P\left[H_k \underbrace{\mathbb{E}_P\left[W_{\min(s_{k+1},t_2)} - W_{\max(s_k,t_1)} \middle| \mathbb{F}_{\max(s_k,t_1)}\right]}_{=0} \middle| \mathbb{F}_{t_1}\right] = 0 \end{aligned} \tag{3.185}$$

P-a.s. for all $t_1, t_2 \in [0,T]$ with $t_1 \leq t_2$. Finally, if $X^{(n)} \in L^2(\mathcal{P}_{P,(\mathbb{F}_t)_{t\in[0,T]}}; \|\cdot\|_{\mathbb{R}^{d\times m}})$, $n \in \mathbb{N}$, are $(\mathbb{F}_t)_{t\in[0,T]}$-simple with $\lim_{n\to\infty} \|X - X^{(n)}\|_{L^2(\mathcal{P}_{P,(\mathbb{F}_t)_{t\in[0,T]}};\|\cdot\|_{\mathbb{R}^{d\times m}})} = 0$, then

$$\mathbb{E}_P\!\left[\int_{t_1}^{t_2} X_s \, dW_s \,\middle|\, \mathbb{F}_{t_1}\right] = L^1(P; \|\cdot\|_{\mathbb{R}^d})\text{ - }\lim_{n\to\infty} \underbrace{\mathbb{E}_P\!\left[\int_{t_1}^{t_2} X_s^{(n)} \, dW_s \,\middle|\, \mathbb{F}_{t_1}\right]}_{=0} = 0 \tag{3.186}$$

P-a.s. for all $t_1, t_2 \in [0,T]$ with $t_1 \leq t_2$. Combining this with Proposition 3.4.11 completes the proof of Proposition 3.6.5. □

4 Stochastic differential equations (SDEs)

In this chapter we specify what we mean by a stochastic differential equation (SDE for short) and by a solution process of such an equation. The content of this chapter can, e.g., be found in [Kloeden and Platen(1992)], [Øksendal(2003)], and [Kuo(2006)].

4.1 Setting

Throughout this chapter the following setting is frequently used. Let $T \in (0, \infty)$, $d, m \in \mathbb{N}$, let $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$ be a stochastic basis, and let $W = (W^{(1)}, \ldots, W^{(m)})\colon [0,T] \times \Omega \to \mathbb{R}^m$ be an m-dimensional standard $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$-Brownian motion.

4.2 Solution processes of SDEs

The next definition describes what we mean by a solution process of a *stochastic differential equation*.

Definition 4.2.1 (Solution processes of stochastic differential equations (SDEs))**.** *Let $T \in (0,\infty)$, $d,m \in \mathbb{N}$, $O \in \mathcal{B}(\mathbb{R}^d)$, $\mu \in \mathcal{M}(\mathcal{B}(O),\mathcal{B}(\mathbb{R}^d))$, $\sigma \in \mathcal{M}(\mathcal{B}(O),\mathcal{B}(\mathbb{R}^{d\times m}))$, let $(\Omega,\mathcal{F},P,(\mathbb{F}_t)_{t\in[0,T]})$ be a stochastic basis, let $\xi \in \mathcal{M}(\mathbb{F}_0,\mathcal{B}(O))$, and let $W = (W^{(1)},\dots,W^{(m)})\colon [0,T]\times\Omega \to \mathbb{R}^m$ be an m-dimensional standard $(\Omega,\mathcal{F},P,(\mathbb{F}_t)_{t\in[0,T]})$-Brownian motion. Then we say that X is a solution process of the SDE*

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t, \qquad t\in[0,T], \qquad X_0 = \xi \tag{4.1}$$

on $(\Omega,\mathcal{F},P,(\mathbb{F}_t)_{t\in[0,T]})$ (we say that X is a solution process of the SDE

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t, \qquad t\in[0,T], \qquad X_0 = \xi) \tag{4.2}$$

if and only if it holds

(i) that $X\colon [0,T]\times\Omega \to O$ is an $(\mathbb{F}_t)_{t\in[0,T]}/\mathcal{B}(O)$-adapted stochastic process with continuous sample paths,

(ii) that

$$P\bigg(\int_0^T \|\mu(X_s)\|_{\mathbb{R}^d} + \|\sigma(X_s)\|^2_{\mathbb{R}^{d\times m}}\,ds < \infty\bigg) = 1, \tag{4.3}$$

and

(iii) that for all $t\in[0,T]$ it holds P-a.s. that

$$X_t = \xi + \int_0^t \mu(X_s)\,ds + \int_0^t \sigma(X_s)\,dW_s. \tag{4.4}$$

Equation (4.1) is referred to as *stochastic differential equation* (SDE), the function μ is called *drift coefficient (function)* of the SDE (4.1), and the function σ is called *diffusion coefficient (function)* of the SDE (4.1). Note in the setting of Definition 4.2.1 that $X\colon [0,T]\times\Omega \to O$ is an O-valued Itô process on $(\Omega,\mathcal{F},P,(\mathbb{F}_t)_{t\in[0,T]})$ with drift $\mu(X_t)$, $t\in[0,T]$, diffusion $\sigma(X_t)$, $t\in[0,T]$, and standard Brownian motion W.

4.3 Gronwall inequalities

4.3.1 Time continuous Gronwall inequality

The following elementary lemma is crucial to investigate properties of SDEs.

Lemma 4.3.1 (Gronwall lemma). *Let $T, \beta \in [0, \infty)$, $\alpha \in \mathbb{R}$, $f \in \mathcal{L}^1(B_{[0,T]}; |\cdot|_{\mathbb{R}})$ satisfy for all $t \in [0, T]$ that*

$$f(t) \leq \alpha + \beta \int_0^t f(s)\, ds. \tag{4.5}$$

Then it holds for all $t \in [0, T]$ that

$$f(t) \leq \alpha \cdot e^{\beta t}. \tag{4.6}$$

Proof of Lemma 4.3.1. W.l.o.g. let $T > 0$. Next let $u\colon [0, T] \to \mathbb{R}$ be the function which satisfies for all $t \in [0, T]$ that

$$u(t) = \alpha + \beta \int_0^t f(s)\, ds. \tag{4.7}$$

Then note that u is absolutely continuous and observe that inequality (4.5) implies that for $B_{[0,T]}$-almost all $t \in [0, T]$ it holds that

$$u'(t) = \beta \cdot f(t) \leq \beta \cdot u(t). \tag{4.8}$$

This shows that for $B_{[0,T]}$-almost all $t \in [0, T]$ it holds that

$$0 \geq e^{-\beta t} \left(u'(t) - \beta \cdot u(t)\right) = \frac{d}{dt} \left[u(t) \cdot e^{-\beta t}\right]. \tag{4.9}$$

The fundamental theorem of calculus hence gives that for all $t \in [0, T]$ it holds that

$$\underbrace{u(t)}_{\geq f(t)} \cdot e^{-\beta t} - \underbrace{u(0)}_{=\alpha} \leq 0. \tag{4.10}$$

Rearranging (4.10) results in (4.6). This completes the proof of Lemma 4.3.1. □

4.3.2 Time discrete Gronwall inequality

In the numerical analysis of SDEs, we also need the discrete counterpart of Lemma 4.3.1.

Lemma 4.3.2. *Let $N \in \mathbb{N}$, $\beta \in [0, \infty)$, $\alpha \in \mathbb{R}$, $f_0, f_1, \ldots, f_N \in \mathbb{R} \cup \{\infty\}$ satisfy for all $n \in \{0, 1, \ldots, N\}$ that*

$$f_n \leq \alpha + \beta \left(\sum_{k=0}^{n-1} f_k \right). \tag{4.11}$$

Then it holds for all $n \in \{0, 1, \ldots, N\}$ that

$$f_n \leq \alpha \cdot (1 + \beta)^n \leq |\alpha| \cdot e^{\beta n} < \infty. \tag{4.12}$$

Proof of Lemma 4.3.2. First of all, we observe that induction and (4.11) prove that for all $n \in \{0, 1, 2, \ldots, N\}$ it holds that $f_n \in \mathbb{R}$. Next let $u_0, u_1, \ldots, u_N \in \mathbb{R}$ be the real numbers which satisfy for all $n \in \{0, 1, 2, \ldots, N\}$ that

$$u_n = \alpha + \beta \left(\sum_{k=0}^{n-1} u_k \right). \tag{4.13}$$

Equation (4.13) ensures that for all $n \in \{0, 1, \ldots, N-1\}$ it holds that

$$u_{n+1} = \alpha + \beta \left(\sum_{k=0}^{n} u_k \right) = \underbrace{\alpha + \beta \left(\sum_{k=0}^{n-1} u_k \right)}_{=u_n} + \beta u_n = (1 + \beta)\, u_n. \tag{4.14}$$

This implies that for all $n \in \{0, 1, \ldots, N\}$ it holds that

$$u_n = \alpha \cdot (1 + \beta)^n. \tag{4.15}$$

Moreover, observe that induction shows that for all $n \in \{0, 1, \ldots, N\}$ it holds that

$$f_n \leq u_n. \tag{4.16}$$

Combining this with (4.15) completes the proof of Lemma 4.3.2. □

4.4 Uniqueness of solution processes of SDEs

Theorem 4.4.2 below shows that solution processes of SDEs are unique up to indistinguishability if both the drift coefficient function and the diffusion coefficient function of the considered SDE are *locally Lipschitz continuous*.

Remark 4.4.1. *Assume the setting in Section 4.1, let* $O \in \mathcal{B}(\mathbb{R}^d)$, $\xi \in \mathcal{M}(\mathbb{F}_0, \mathcal{B}(O))$, $\mu \in \mathcal{M}\big(\mathcal{B}(O), \mathcal{B}(\mathbb{R}^d)\big)$, $\sigma \in \mathcal{M}\big(\mathcal{B}(O), \mathcal{B}(\mathbb{R}^{d\times m})\big)$, *and let* $X\colon [0,T]\times\Omega \to O$ *be a solution process of the SDE*

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t, \qquad t\in[0,T], \qquad X_0 = \xi \tag{4.17}$$

on $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$. *Then observe that the fact that*

$$\cup_{y\in\mathbb{R}^m} (W_T)^{-1}(\{y\}) = (W_T)^{-1}(\cup_{y\in\mathbb{R}^m}\{y\}) = (W_T)^{-1}(\mathbb{R}^m) = \Omega \neq \emptyset \tag{4.18}$$

proves that there exists a real number $y\in\mathbb{R}^m$ *such that* $(W_T)^{-1}(\{y\}) \neq \emptyset$. *This implies that there exists an* $A \in \mathcal{F}\backslash\{\emptyset\}$ *which satisfies*

$$P(A) = 0. \tag{4.19}$$

In the next step let $X^v\colon [0,T]\times\Omega \to O$, $v\in O$, *be the functions which satisfy for all* $v \in O$, $t\in[0,T]$, $\omega\in\Omega$ *that*

$$X^v_t(\omega) = \mathbb{1}_{\Omega\backslash A}(\omega)\, X_t(\omega) + \mathbb{1}_A(\omega)\, v. \tag{4.20}$$

Then

(i) it holds for all $v\in O$ *that* X^v *is a solution process of the SDE* (4.17) *and*

(ii) it holds for all $v, w \in O$ *with* $v\neq w$ *that* $X^v \neq X^w$.

Solution processes of the SDE (4.17) are thus typically *not unique.* However, under suitable additional assumptions (cf., e.g., (4.21) below), solution processes of the SDE (4.17) are *unique up to indistinguishability.* This is the subject of the next result, Theorem 4.4.2 below.

Theorem 4.4.2. *Assume the setting Section 4.1, let* $O \subseteq \mathbb{R}^d$ *be an open set, let* $\xi\in\mathcal{M}(\mathbb{F}_0, \mathcal{B}(O))$, *let* $\mu\colon O\to\mathbb{R}^d$ *and* $\sigma\colon O \to \mathbb{R}^{d\times m}$ *be locally Lipschitz continuous functions, i.e., assume for all compact subsets* $K\subseteq O$ *of* O *that*

$$\sup\!\left(\left\{\frac{\|\mu(x)-\mu(y)\|_{\mathbb{R}^d} + \|\sigma(x)-\sigma(y)\|_{\mathbb{R}^{d\times m}}}{\|x-y\|_{\mathbb{R}^d}} : x,y\in K, x\neq y\right\}\cup\{0\}\right) < \infty, \tag{4.21}$$

and let $X, Y\colon [0,T]\times\Omega\to O$ *be solution processes of the SDE*

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t, \qquad t\in[0,T], \qquad X_0 = \xi \tag{4.22}$$

on $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$. *Then* X *and* Y *are indistinguishable from each other, i.e.,*

$$P\big(\forall\, t\in[0,T]\colon X_t = Y_t\big) = 1. \tag{4.23}$$

The proof of Theorem 4.4.2 uses Gronwall's lemma (see Lemma 4.3.1) and is omitted.

Remark 4.4.3. *Local Lipschitz continuity of both the drift and the diffusion coefficient function (see inequality* (4.21)*) is a sufficient (see Theorem 4.4.2) but not a necessary condition to ensure that the solution processes of an SDE are unique up to indistinguishability.*

Lemma 4.4.4 (Lebesgue's number lemma)**.** *Let $d \in \mathbb{N}$, let $K \subseteq \mathbb{R}^d$ be a compact (closed and bounded) set, let I be a set, and let $U_i \subseteq \mathbb{R}^d$, $i \in I$, be a family of open sets with*

$$K \subseteq \cup_{i \in I} U_i. \tag{4.24}$$

Then there exists a positive real number $\delta \in (0,\infty)$ such that for every $x \in K$ it holds that there exists an $i \in I$ such that

$$\{y \in K \colon \|x - y\|_{\mathbb{R}^d} \leq \delta\} \subseteq U_i. \tag{4.25}$$

Proof of Lemma 4.4.4. We prove Lemma 4.4.4 by a contradiction. We thus assume that there exists a family $(x_\delta)_{\delta \in (0,\infty)} \subseteq K$ with the property that for every $\delta \in (0,\infty)$ and every $i \in I$ it does *not* hold that

$$\{y \in K \colon \|x_\delta - y\|_{\mathbb{R}^d} \leq \delta\} \subseteq U_i. \tag{4.26}$$

Since $K \subseteq \mathbb{R}^d$ is a compact set, there exists a sequence $(\delta_n)_{n\in\mathbb{N}} \subseteq (0,\infty)$ of positive real numbers with the property that for all $n \in \mathbb{N}$ it holds that $\delta_n \leq \frac{1}{n}$ and with the property that $(x_{\delta_n})_{n\in\mathbb{N}} \subseteq K$ is convergent to a vector $x^* \in \mathbb{R}^d$. Combining (4.24) and the fact that $x^* \in K$ shows that there exists an $i \in I$ such that $x^* \in U_i$. This implies that there exists a real number $\varepsilon \in (0,\infty)$ such that

$$\left\{y \in \mathbb{R}^d \colon \|x^* - y\|_{\mathbb{R}^d} \leq \varepsilon\right\} \subseteq U_i. \tag{4.27}$$

Hence, we obtain that for all $n \in \mathbb{N}$ with $\frac{1}{n} < \frac{\varepsilon}{2}$ and $\|x^* - x_{\delta_n}\|_{\mathbb{R}^d} \leq \frac{\varepsilon}{2}$ it holds that

$$\begin{aligned}&\left\{y \in \mathbb{R}^d \colon \|x_{\delta_n} - y\|_{\mathbb{R}^d} \leq \delta_n\right\} \subseteq \left\{y \in \mathbb{R}^d \colon \|x^* - y\|_{\mathbb{R}^d} \leq \delta_n + \|x^* - x_{\delta_n}\|_{\mathbb{R}^d}\right\} \\ &\subseteq \left\{y \in \mathbb{R}^d \colon \|x^* - y\|_{\mathbb{R}^d} \leq \tfrac{1}{n} + \tfrac{\varepsilon}{2}\right\} \subseteq \left\{y \in \mathbb{R}^d \colon \|x^* - y\|_{\mathbb{R}^d} \leq \varepsilon\right\} \subseteq U_i.\end{aligned} \tag{4.28}$$

This contradicts to (4.26). The proof of Lemma 4.4.4 is thus completed. □

4.5 Existence and uniqueness of solution processes of SDEs

The next theorem shows that if both the drift coefficient function and the diffusion coefficient function of an SDE are globally Lipschitz continuous, then there exists an up to indistinguishability unique solution process of the SDE.

Theorem 4.5.1 (Existence and uniqueness of solution processes of SDEs with globally Lipschitz continuous coefficients)**.** *Assume the setting in Section 4.1, let $p \in [2,\infty)$, $\xi \in \mathcal{L}^p(P|_{\mathbb{F}_0}; \|\cdot\|_{\mathbb{R}^d})$, and let $\mu\colon \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d\times m}$ be globally Lipschitz continuous functions, i.e., assume that there exist a real number $C \in [0,\infty)$ such that for all $x, y \in \mathbb{R}^d$ it holds that*

$$\|\mu(x) - \mu(y)\|_{\mathbb{R}^d} + \|\sigma(x) - \sigma(y)\|_{\mathbb{R}^{d\times m}} \leq C\,\|x - y\|_{\mathbb{R}^d}\,. \tag{4.29}$$

Then

(i) there exists an up to indistinguishability unique solution process $X\colon [0,T]\times\Omega \to \mathbb{R}^d$ of the SDE

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t, \qquad t \in [0,T], \qquad X_0 = \xi \tag{4.30}$$

on $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$,

(ii) it holds that $\sup_{t\in[0,T]} \|X_t\|_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})} < \infty$, and

(iii) it holds for all $\alpha \in (0, 1/2]$ that $X \in \mathcal{C}^\alpha([0,T], L^p(P; \|\cdot\|_{\mathbb{R}^d}))$.

Proof of Theorem 4.5.1. Throughout this proof we use the $\mathbb{R}$-vector space $\mathcal{V}$ given by

$$\mathcal{V} = \left\{ \left\{ \begin{array}{c} Y\colon [0,T]\times\Omega \to \mathbb{R}^d\colon \\ Y \text{ is } (\mathbb{F}_t)_{t\in[0,T]}\text{-predictable} \\ \text{and a modification of } X \end{array} \right\} : \left(\begin{array}{c} X\colon [0,T]\times\Omega \to \mathbb{R}^d \text{ is} \\ (\mathbb{F}_t)_{t\in[0,T]}\text{-predictable and} \\ \sup_{t\in[0,T]} \mathbb{E}_P\big[\|X_t\|^p_{\mathbb{R}^d}\big] < \infty \end{array} \right) \right\}. \tag{4.31}$$

As usual, we do in the following not distinguish between an $(\mathbb{F}_t)_{t\in[0,T]}$-predictable stochastic process $X\colon [0,T]\times\Omega \to \mathbb{R}^d$ with $\sup_{t\in[0,T]} \mathbb{E}_P\big[\|X_t\|^p\big] < \infty$ and its equivalence class in $\mathcal{V}$. In the next step let $\|\cdot\|_{\mathcal{V},\lambda}\colon \mathcal{V} \to [0,\infty)$, $\lambda \in \mathbb{R}$, be the functions with the property that for all $X \in \mathcal{V}$, $\lambda \in \mathbb{R}$ it holds that

$$\|X\|_{\mathcal{V},\lambda} = \sup_{t\in[0,T]} \Big(e^{\lambda t}\, \|X_t\|_{L^p(P;\|\cdot\|_{\mathbb{R}^d})} \Big) = \sup_{t\in[0,T]} \Big(e^{\lambda t} \big(\mathbb{E}_P\big[\|X_t\|^p_{\mathbb{R}^d}\big]\big)^{\frac{1}{p}} \Big)\,. \tag{4.32}$$

It can be shown that for every $\lambda \in \mathbb{R}$ it holds that $\big(\mathcal{V}, \|\cdot\|_{\mathcal{V},\lambda}\big)$ is a complete normed $\mathbb{R}$-vector spaces (i.e., an $\mathbb{R}$-Banach spaces). Next let $\Phi\colon \mathcal{V} \to \mathcal{V}$ be the mapping with the property that for all $X \in \mathcal{V}$, $t \in [0,T]$ it holds P-a.s. that

$$\big(\Phi(X)\big)_t = \xi + \int_0^t \mu(X_s)\,ds + \int_0^t \sigma(X_s)\,dW_s. \tag{4.33}$$

It follows from the linear growth estimates

$$\begin{aligned} \|\mu(x)\|_{\mathbb{R}^d} &\leq (C + \|\mu(0)\|_{\mathbb{R}^d})\,(1 + \|x\|_{\mathbb{R}^d})\,, \\ \|\sigma(x)\|_{HS(\mathbb{R}^m,\mathbb{R}^d)} &\leq \Big(C + \|\sigma(0)\|_{HS(\mathbb{R}^m,\mathbb{R}^d)}\Big)\,(1 + \|x\|_{\mathbb{R}^d}) \end{aligned} \tag{4.34}$$

for all $x \in \mathbb{R}^d$ and from inequality (Burkholder-Davis-Gundy inequality I) in Theorem 3.4.21 that Φ is indeed well-defined. In the next step we note that again inequality (Burkholder-Davis-Gundy inequality I) in Theorem 3.4.21 proves that for all $t \in [0,T]$ and all $X, Y \in \mathcal{V}$ it holds that

$$\begin{aligned}
&\left\| \big(\Phi(X)\big)_t - \big(\Phi(Y)\big)_t \right\|_{L^p(P;\|\cdot\|_{\mathbb{R}^d})} \\
&\leq \left\| \int_0^t (\mu(X_s) - \mu(Y_s))\, ds \right\|_{L^p(P;\|\cdot\|_{\mathbb{R}^d})} + \left\| \int_0^t (\sigma(X_s) - \sigma(Y_s))\, dW_s \right\|_{L^p(P;\|\cdot\|_{\mathbb{R}^d})} \\
&\leq \int_0^t \|\mu(X_s) - \mu(Y_s)\|_{L^p(P;\|\cdot\|_{\mathbb{R}^d})}\, ds + p \left[\int_0^t \|\sigma(X_s) - \sigma(Y_s)\|^2_{L^p(P;\|\cdot\|_{HS(\mathbb{R}^m,\mathbb{R}^d)})}\, ds \right]^{\frac{1}{2}} \\
&\leq C \int_0^t \|X_s - Y_s\|_{L^p(P;\|\cdot\|_{\mathbb{R}^d})}\, ds + pC \left[\int_0^t \|X_s - Y_s\|^2_{L^p(P;\|\cdot\|_{\mathbb{R}^d})}\, ds \right]^{\frac{1}{2}} \\
&\leq C\sqrt{T} \left[\int_0^t \|X_s - Y_s\|^2_{L^p(P;\|\cdot\|_{\mathbb{R}^d})}\, ds \right]^{\frac{1}{2}} + pC \left[\int_0^t \|X_s - Y_s\|^2_{L^p(P;\|\cdot\|_{\mathbb{R}^d})}\, ds \right]^{\frac{1}{2}} \\
&\leq C \left(\sqrt{T} + p \right) \left[\int_0^t \|X_s - Y_s\|^2_{L^p(P;\|\cdot\|_{\mathbb{R}^d})}\, ds \right]^{\frac{1}{2}}.
\end{aligned} \tag{4.35}$$

Therefore, we obtain that for all $\lambda \in (-\infty, 0)$, $t \in [0,T]$ and all $X, Y \in \mathcal{V}$ it holds that

$$\begin{aligned}
&e^{\lambda t} \left\| \big(\Phi(X)\big)_t - \big(\Phi(Y)\big)_t \right\|_{L^p(P;\|\cdot\|_{\mathbb{R}^d})} \\
&\leq C \left(\sqrt{T} + p \right) \left[\int_0^t e^{2\lambda(t-s)} \left[e^{\lambda s} \|X_s - Y_s\|_{L^p(P;\|\cdot\|_{\mathbb{R}^d})} \right]^2 ds \right]^{\frac{1}{2}} \\
&\leq C \left(\sqrt{T} + p \right) \left[\int_0^t e^{2\lambda(t-s)}\, ds \right]^{\frac{1}{2}} \|X - Y\|_{\mathcal{V},\lambda} \\
&\leq C \left(\sqrt{T} + p \right) \underbrace{\left[\int_0^T e^{2\lambda s}\, ds \right]^{\frac{1}{2}}}_{= \frac{\sqrt{1 - e^{2\lambda T}}}{\sqrt{2|\lambda|}} \leq \frac{1}{\sqrt{|\lambda|}}} \|X - Y\|_{\mathcal{V},\lambda}.
\end{aligned} \tag{4.36}$$

This proves that for all $\lambda \in (-\infty, 0)$ and all $X, Y \in \mathcal{V}$ it holds that

$$\|\Phi(X) - \Phi(Y)\|_{\mathcal{V},\lambda} \leq \frac{C\left(\sqrt{T} + p\right)}{\sqrt{|\lambda|}} \|X - Y\|_{\mathcal{V},\lambda}. \tag{4.37}$$

Hence, we obtain that for all $X, Y \in \mathcal{V}$ it holds that

$$\|\Phi(X) - \Phi(Y)\|_{\mathcal{V},4C^2[\sqrt{T}+p]^2} \leq \frac{1}{2} \|X - Y\|_{\mathcal{V},4C^2[\sqrt{T}+p]^2}. \tag{4.38}$$

The mapping Φ is thus a *contraction* from $(\mathcal{V}, \|\cdot\|_{\mathcal{V},4C^2[\sqrt{T}+p]^2})$ to $(\mathcal{V}, \|\cdot\|_{\mathcal{V},4C^2[\sqrt{T}+p]^2})$. The Banach fixed point theorem hence proves that there exists a unique $Y \in \mathcal{V}$ with

$$\Phi(Y) = Y, \tag{4.39}$$

i.e., that

$$Y_t = \xi + \int_0^t \mu(Y_s)\, ds + \int_0^t \sigma(Y_s)\, dW_s \tag{4.40}$$

P-a.s. for all $t \in [0,T]$. Item (vi) in Theorem 3.4.21 then implies that there an $(\mathbb{F}_t)_{t\in[0,T]}$-adapted stochastic process $X\colon [0,T]\times\Omega \to \mathbb{R}^d$ with continuous sample paths which is a modification of Y, i.e, which fulfills that for all $t \in [0,T]$ it holds that

$$P(X_t = Y_t) = 1. \tag{4.41}$$

Combining this with (4.40) proves that the stochastic process X is a *solution process* of the SDE (4.30). It thus remains to prove that for all $\alpha \in (0, \frac{1}{2}]$ it holds that

$$X \in C^\alpha([0,T], L^p(P; \|\cdot\|_{\mathbb{R}^d})). \tag{4.42}$$

This, in turn, follows immediately from (Temporal regularity for Itô processes). The proof of Theorem 4.5.1 is thus completed. □

4.6 Autonomization of SDEs with time-dependent coefficient functions

Assume the setting in Section 4.1, let $p \in [2,\infty)$, $\xi \in \mathcal{L}^p(P|_{\mathbb{F}_0}; \|\cdot\|_{\mathbb{R}^d})$, and let $\mu\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}^d$ and $\sigma\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}^{d\times m}$ be globally Lipschitz continuous functions, i.e., assume that

$$\sup_{\substack{x,y\in\mathbb{R}^d\\ x\neq y}} \left(\frac{\|\mu(t_1,x_1)-\mu(t_2,x_2)\|_{\mathbb{R}^d} + \|\sigma(t_1,x_1)-\sigma(t_2,x_2)\|_{\mathbb{R}^{d\times m}}}{|t_1-t_2| + \|x_1-x_2\|_{\mathbb{R}^d}} \right) < \infty. \tag{4.43}$$

Theorem 4.5.1 then shows that there exists an up to indistinguishability unique $(\mathbb{F}_t)_{t\in[0,T]}$-adapted stochastic process $X\colon [0,T]\times\Omega\to\mathbb{R}^d$ with continuous sample paths such that for all $t\in[0,T]$ it holds P-a.s. that

$$X_t = \xi + \int_0^t \mu(s,X_s)\, ds + \int_0^t \sigma(s,X_s)\, dW_s \tag{4.44}$$

and such that for all $\alpha \in (0,\frac{1}{2}]$ it holds that

$$X \in C^\alpha([0,T], L^p(P; \|\cdot\|_{\mathbb{R}^d})). \tag{4.45}$$

Indeed, define $\tilde{\xi}\colon \Omega \to \mathbb{R}^{d+1}$ through

$$\tilde{\xi}(\omega) := (0, \xi(\omega)) \tag{4.46}$$

for all $\omega \in \Omega$, define $p_T\colon \mathbb{R} \to [0,T]$ through

$$p_T(t) := \max\big\{ \min\{t, T\}, 0\big\} \tag{4.47}$$

for all $t \in \mathbb{R}$ and define $\tilde{\mu}\colon \mathbb{R}^{d+1} \to \mathbb{R}^{d+1}$ and $\tilde{\sigma}\colon \mathbb{R}^{d+1} \to \mathbb{R}^{(d+1)\times m}$ through

$$\tilde{\mu}(t,x) := \big(1, \mu\big(p_T(t), x\big)\big) \qquad \text{and} \qquad \tilde{\sigma}(t,x) := \big(0, \sigma\big(p_T(t), x\big)\big) \tag{4.48}$$

for all $t \in \mathbb{R}$ and all $x \in \mathbb{R}^d$. Then observe that assumption (4.43) ensures that $\tilde{\mu}$ and $\tilde{\sigma}$ are globally Lipschitz continuous, i.e., that

$$\sup_{\substack{x,y\in\mathbb{R}^{d+1}\\ x\neq y}} \left(\frac{\|\tilde{\mu}(x) - \tilde{\mu}(y)\|_{\mathbb{R}^{d+1}} + \|\tilde{\sigma}(x) - \tilde{\sigma}(y)\|_{\mathbb{R}^{(d+1)\times m}}}{\|x-y\|_{\mathbb{R}^{d+1}}} \right) < \infty \tag{4.49}$$

and Theorem 4.5.1 hence proves the existence of an up to indistinguishability unique $(\mathbb{F}_t)_{t\in[0,T]}$-adapted stochastic process $\tilde{X} = (\tilde{X}^{(1)}, \ldots, \tilde{X}^{(d+1)})\colon [0,T]\times\Omega \to \mathbb{R}^{d+1}$ with continuous sample paths which fulfills

$$\tilde{X}_t = \tilde{\xi} + \int_0^t \tilde{\mu}(\tilde{X}_s)\, ds + \int_0^t \tilde{\sigma}(\tilde{X}_s)\, dW_s \tag{4.50}$$

P-a.s. for all $t \in [0,T]$. The process $X\colon [0,T]\times\Omega \to \mathbb{R}^d$ defined through

$$X_t := \big(\tilde{X}_t^{(2)}, \ldots, \tilde{X}_t^{(d+1)}\big) \tag{4.51}$$

for all $t \in [0,T]$ is then an $(\mathbb{F}_t)_{t\in[0,T]}$-adapted stochastic process with continuous sample paths which fulfills (4.44).

4.7 Examples of SDEs

In this section several examples of SDEs from the literature are presented. Most of this section comes from [Hutzenthaler and Jentzen(2012)].

4.7.1 Setting

The following setting is used to formulate the examples. Assume the setting in Section 4.1, let $O \subseteq \mathbb{R}^d$ be an open set, let $\xi = (\xi^{(1)}, \ldots, \xi^{(d)}) \in O$, let $\mu\colon O \to \mathbb{R}^d$ and $\sigma\colon O \to \mathbb{R}^{d\times m}$ be locally Lipschitz continuous functions, and let $X = (X^{(1)}, \ldots, X^{(d)})\colon [0,T]\times\Omega \to O$ be a solution process of the SDE

$$dX_t = \mu(X_t)\, dt + \sigma(X_t)\, dW_t, \qquad t\in[0,T], \qquad X_0 = \xi \tag{4.52}$$

on $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$. In particular, we assume that for all $t \in [0,T]$ it holds P-a.s. that

$$X_t = \xi + \int_0^t \mu(X_s)\, ds + \int_0^t \sigma(X_s)\, dW_s. \tag{4.53}$$

Theorem 4.4.2 shows that the stochastic process X is unique up to indistinguishability.

4.7.2 Geometric Brownian motion

In addition to the assumptions in Subsection 4.7.1, let $\alpha, \beta \in \mathbb{R}$ be real numbers, assume that $d = m = 1$, $O = \mathbb{R}$, and assume for all $x \in \mathbb{R}$ that

$$\mu(x) = \alpha x \qquad \text{and} \qquad \sigma(x) = \beta x. \tag{4.54}$$

The SDE (4.52) reads as

$$dX_t = \alpha X_t\, dt + \beta X_t\, dW_t, \qquad t \in [0,T], \qquad X_0 = \xi. \tag{4.55}$$

Observe that μ and σ are globally Lipschitz continuous. Theorem 4.5.1 hence proves that an up to indistinguishability unique solution process $X\colon [0,T]\times\Omega \to \mathbb{R}$ of (4.55) does indeed exist. Moreover, observe that Theorem 4.5.1 shows that for all $p \in [0,\infty)$ it holds that

$$\sup_{t\in[0,T]} \mathbb{E}_P\big[|X_t|_{\mathbb{R}}^p\big] < \infty. \tag{4.56}$$

The solution process $X\colon [0,T]\times\Omega \to \mathbb{R}$ of (4.55) can be calculated explicitly. More precisely, combining Example 3.5.7 and Theorem 4.5.1 proves that for all $t \in [0,T]$ it holds P-a.s. that

$$X_t = \exp\Big(\Big(\alpha - \tfrac{\beta^2}{2}\Big)\, t + \beta\, W_t\Big)\, \xi. \tag{4.57}$$

The solution processes X of (4.55) is also referred to as *geometric Brownian motion*. Taking expectations on both sides of (4.55) shows that the deterministic expectation process $\mathbb{E}_P\big[X_t\big]$, $t \in [0,T]$, satisfies the ordinary differential equation

$$\tfrac{d}{dt}\, \mathbb{E}_P\big[X_t\big] = \alpha \cdot \mathbb{E}_P\big[X_t\big], \qquad \mathbb{E}_P\big[X_0\big] = \xi. \tag{4.58}$$

Hence, we obtain that for all $t \in [0,T]$ it holds that

$$\mathbb{E}_P\big[X_t\big] = e^{\alpha t}\, \xi. \tag{4.59}$$

4.7.3 Black-Scholes model

In addition to the assumptions in Subsection 4.7.1, let $r, \alpha \in \mathbb{R}$, $\beta \in (0,\infty)$ be real numbers, assume that $d = 2$, $m = 1$, $O = (0,\infty)^2$, and assume for all $x = (x_1, x_2) \in (0,\infty)^2$ that

$$\mu(x) = \begin{pmatrix} r x_1 \\ \alpha x_2 \end{pmatrix} \qquad \text{and} \qquad \sigma(x) = \begin{pmatrix} 0 \\ \beta x_2 \end{pmatrix}. \tag{4.60}$$

The SDE (4.52) then reads as

$$\begin{pmatrix} dX_t^{(1)} = rX_t^{(1)} dt \\ dX_t^{(2)} = \alpha X_t^{(2)} dt + \beta X_t^{(2)} dW_t \end{pmatrix}, \qquad t \in [0, T], \qquad X_0 = \xi. \tag{4.61}$$

Observe that μ and σ are globally Lipschitz continuous. Theorem 4.5.1 hence proves that an up to indistinguishability unique solution process $X: [0, T] \times \Omega \to \mathbb{R}$ of (4.61) does indeed exist. Moreover, note that Theorem 4.5.1 shows that for all $p \in [0, \infty)$ it holds that

$$\sup_{t \in [0,T]} \mathbb{E}_P\big[\|X_t\|_{\mathbb{R}^2}^p\big] < \infty. \tag{4.62}$$

In the next step we observe that Subsection 4.7.2 shows that for all $t \in [0, T]$ it holds P-a.s. that

$$X_t^{(1)} = e^{rt} \xi^{(1)} \qquad \text{and} \qquad X_t^{(2)} = \exp\Big((\alpha - \tfrac{\beta^2}{2})\, t + \beta W_t\Big)\, \xi^{(2)}. \tag{4.63}$$

In the remainder of Subsection 4.7.3 we roughly illustrate a few basic ideas from the theory of option pricing from mathematical finance and its applications in the financial engineering industry. The remainder of Subsection 4.7.3 just intends to roughly illustrate a few basic ideas and contains a number of improper and inaccurate descriptions of the material. For a more proper and accurate presentation of this material the reader is referred to the economics and mathematical finance literature including the references mentioned below.

In the *Black-Scholes model* (see [Black and Scholes(1973), Merton(1973)]) the stochastic process $(X_t^{(1)})_{t \in [0,T]}$ in (4.61) represents the price process of a *"risk-free" bank account* with

- the fixed interest rate $r \in \mathbb{R}$ and
- the initial price $\xi^{(1)} > 0$

and the stochastic process $(X_t^{(2)})_{t \in [0,T]}$ in (4.61) models the price process of an *underlying*, e.g., a stock, a commodity, a currency, or an index (e.g., the Standard & Poor's 500 [S & P 500], the Swiss Market Index [SMI]), with

- the expected interest rate $\alpha \in \mathbb{R}$ (see (4.58) above and Lemma 4.7.1 below),
- the *volatility* $\beta > 0$, and
- the initial price $\xi^{(2)} > 0$.

The Black-Scholes model is a popular model for estimating prices of certain *financial derivatives*. A financial derivative is a suitable financial product that is in a certain way derived from an (or some) underlying(s). Simple examples of financial derivatives are *European call options* and *European put options*.

4.7.3.1 Simple examples of financial derivatives: European put and call options

A *European put option* is a *contract* between two parties, the *writer of the option* and the *holder of the option*, that gives the *holder* of the option the *right* but not the *obligation* to sell a *stipulated underlying* (e.g., a stock or a currency) at the *stipulated time* $T > 0$ for the *stipulated price* $K \in (0, \infty)$ (*strike price*) to the *writer* of the option (cf., e.g., [Higham(2004)]). The holder of the European put option thus has the *option* to sell (*put*) the underlying in the sense above. This explains the words "put" and "option" in the label "European put option". The word "European" refers to the in the European put option contract in advance *fixed* stipulated time $T > 0$ in contrast to *American put options* which can be exercised at any time until $T > 0$ (United States of *America*, "the land of opportunity"/"the country of boundless possibilities").

(i) If $x_T \in [0, \infty)$ is the price of the underlying at time T and if the $x_T < K$, then the holder of the European put option would probably make use of his right and *exercise* the European put option, that is, the holder would sell the underlying for the price K. At the same time the holder of the European put option could (try to) buy the underlying at a market (e.g., at a stock market if the underlying is a stock) for the price x_T and, as $x_T < K$, this would result in a profit of

$$K - x_T \tag{4.64}$$

for the holder of the European put option.

(ii) If $x_T \in [0, \infty)$ is the price of the underlying at time T and if the $x_T \geq K$, then the holder of the European put option would probably do nothing and let his right elapse.

Typically the concrete selling of the underlying at time T is replaced by a *cash settlement* which is stipulated in the European put option contract. In particular, if $x_T \in [0, \infty)$ is the price of the underlying at time $T > 0$ and if a cash settlement is stipulated in the European put option contract, then, in view of (i) and (ii), the holder of the option has at time T the claim

$$\max\{K - x_T, 0\} \tag{4.65}$$

to the writer of the European put option. Analogously, a *European call option* is a contract between two parties, the *writer of the option* and the *holder of the option*, that gives the holder the right but not the obligation to buy a *stipulated underlying* (e.g., a stock or a currency) at the *stipulated time* $T > 0$ for the *stipulated price* $K \in (0, \infty)$ (*strike price*) from the *writer* of the option (cf., e.g., [Higham(2004)]). The holder of the European call option thus has the *option* to buy (*call*) the underlying in the sense above.

(i) If $x_T \in [0, \infty)$ is the price of the underlying at time T and if the $x_T > K$, then the holder of the European call option would probably make use of its right and *exercise* the European call option, that is, the holder would buy the underlying for the price K. At the same time the holder of the European call option could (try to) sell the underlying at a market (e.g., at a stock market if the underlying is a stock) for the price x_T and, as $x_T > K$, this would result in a profit of

$$x_T - K \tag{4.66}$$

for the holder of the European call option.

(ii) If $x_T \in [0, \infty)$ is the price of the underlying at time T and if the $x_T \leq K$, then the holder of the European call option would probably do nothing and let his right elapse.

If $x_T \in [0, \infty)$ is the price of the underlying at time $T > 0$ and if a cash settlement is stipulated in the European call option contract, then, in view of (i) and (ii), the holder of the option has at time T the claim

$$\max\{x_T - K, 0\} \tag{4.67}$$

to the writer of the European call option.

4.7.3.2 Trading of financial derivatives

Financial derivatives (such as European put options) can/are – depending on the specific form of the financial derivative – purchased/concluded

- through an exchange including
  - usual stock exchanges (e.g., the *Frankfurt Stock Exchange*) for *warrants* (options formulated as *security papers*)
  - special *exchanges for financial derivatives* (often called as *futures exchanges*) such as, for example,
    - the *Chicago Mercantile Exchange* (*CME*) (Chicago, USA, `http://www.cmegroup.com`, approximatively $9 \cdot 10^6$ contracts per trading day (see [FrankfurterAllgemeineZeitung(2006)]), October 17th, 2006),
    - the *European Exchange* (*Eurex*) (Eschborn, Germany, `http://www.eurexchange.com`, approximatively $6 \cdot 10^6$ contracts per trading day (see [FrankfurterAllgemeineZeitung(2006)]), October 17th, 2006)
    - ...

  or
- *over-the-counter* (*OTC*) (*off-exchange trading*) through contracts between two parties with no supervision of an exchange.

4.7.3.3 Purposes of financial derivatives

Central reasons why investors are interested in buying/concluding a financial derivative include *hedging* and *speculation*. For example, a possibility to hedge against the risk of falling stock prices, falling foreign exchange rates, and/or falling interest rates is to hold/buy suitable *put* options.

For instance, in 2007 the *pension fund* of the *Radobank* (a multinational bank for clients from the food and agribusiness with the head office in Utrecht (the Netherlands) with approximatively 59000 employees; see `http://en.wikipedia.org/wiki/Rabobank` and `http://www.rabobank.de/`) started to implement a hedging stragey consisting of suitable financial derivatives which, in particular, intends to hedge the risk of both "significant decreased" stock prices as well as "dramatically reduced interest rates" (see `http://www.cardano.com/risk_management_client_case.html?id=9` for further details). At `http://www.tagesschau.de/wirtschaft/banken-strafzahlungen100.html` (November 11th, 2013) it is reported that the Radobank payed approximatively 10^9 US Dollar to prevent further investigations regarding the Libor (London Interbank Offered Rate) scandal.

4.7.3.4 Estimation of prices of financial derivatives

There are much *more complex financial derivatives* than European put options and such financial derivatives result in much more complicated claims. For example, let $f\colon C([0,T],\mathbb{R})\to[0,\infty)$ be an at most polynomially growing Borel measurable function. Then we consider in the following a financial derivative that results at time T in the claim

$$f(x) \tag{4.68}$$

of the holder of the financial derivative to the writer of the financial derivative where we think of $x=(x_t)_{t\in[0,T]}$ as the price process of the underlying which is assumed to be a continuous function on $[0,T]$. In the special case of an European call option with cash settlement, $f\colon C([0,T],\mathbb{R})\to[0,\infty)$ satisfies that there exists a real number $K\in[0,\infty)$ such that for all $x=(x_t)_{t\in[0,T]}\in C([0,T],\mathbb{R})$ it holds that

$$f(x)=\max\bigl\{x_T-K,0\bigr\}. \tag{4.69}$$

The contract between the writer and the holder of the financial derivative is concluded at time $t=0$. At time T the holder of the financial derivative has the claim (4.68) to the writer of the option. The holder of the financial derivative has to compensate the writer for this claim when the contract is concluded, that is, at time $t=0$. This compensation is the price of the financial derivative at time $t=0$. The topic of *option pricing* or, more generally, *derivative pricing*, is to investigate what could in a certain way be a "fair" price for the financial derivative "today", that is, <u>at time $t=0$</u>.

4.7.3.5 Derivative pricing in the Black-Scholes model

Under suitable simplifications and assumptions (which are not met in the "real life" trading of financial derivatives), the Black-Scholes model provides an attempt to an answer to this question. This is what we illustrate in the following. For this assume the setting in the beginning of Subsection 4.7.3 and let $f\colon C([0,T],\mathbb{R}) \to [0,\infty)$ be an at most polynomially growing Borel measurable function. Then there exist an up to indistinguishability unique $(\mathbb{F}_t)_{t\in[0,T]}$-adapted stochastic process $D\colon [0,T]\times\Omega \to \mathbb{R}$ with continuous sample paths such that

$$D_T = f(X^{(2)}) \tag{4.70}$$

and such that the financial market model $(X^{(1)}, X^{(2)}, D)$ is in a certain sense *arbitrage free* (see, e.g., [Kühn(2004)]). The process D_t, $t \in [0,T]$, represents in the Black-Scholes model the price process of the financial derivative with the pay-off $f(X^{(2)})$ at time T. The random variable D_0 can be represented explicitly as the expectation of a suitable random variable. More precisely, if $\tilde{X}\colon [0,T]\times\Omega \to \mathbb{R}$ is a solution process of the SDE

$$d\tilde{X}_t = r\tilde{X}_t\,dt + \beta\tilde{X}_t\,dW_t, \qquad t\in[0,T], \qquad \tilde{X}_0 = X_0^{(2)}, \tag{4.71}$$

then it holds P-a.s. that

$$D_0 = \frac{\mathbb{E}_P\big[f(\tilde{X})\big]}{X_T^{(1)}} = e^{-rT}\,\mathbb{E}_P\big[f(\tilde{X})\big] \tag{4.72}$$

(see, e.g., Subsection 4.1.2 in [Kühn(2004)]). The assertions (4.70) and (4.72) above are a consequence of a powerful result that is known as *fundamental theorem of asset pricing* in the mathematical finance literature (see, e.g., [Kallsen(2009)]). It is interesting to observe that the right hand side of (4.72) is completely independent of the expected interest rate α of the underlying in the Black-Scholes model. Let us illustrate equation (4.72) in the case of a simple example, that is, in the case of an European call option. For this the following well-known lemmas are used.

Lemma 4.7.1. *Let $(\Omega,\mathcal{F},P)$ be a probability space, let $c\in(0,\infty)$, and let $X\colon\Omega\to\mathbb{R}$ be an $\mathcal{N}_{0,1}$-distributed random variable. Then $\mathbb{E}_P\big[e^{cX}\big] = \exp\big(\tfrac{1}{2}c^2\big)$.*

Proof of Lemma 4.7.1. Note that

$$\begin{aligned}
\mathbb{E}_P\big[e^{cX}\big] &= \int_{\mathbb{R}} e^{cx}\,\tfrac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}x^2}\,dx = \int_{\mathbb{R}} \tfrac{1}{\sqrt{2\pi}}\,e^{\left(cx-\frac{1}{2}x^2\right)}\,dx \\
&= \int_{\mathbb{R}} \tfrac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}\left(x^2-2cx+c^2-c^2\right)}\,dx = e^{\frac{1}{2}c^2}\int_{\mathbb{R}} \tfrac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}\left(x^2-2cx+c^2\right)}\,dx \\
&= e^{\frac{1}{2}c^2}\int_{\mathbb{R}} \tfrac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}(x-c)^2}\,dx = e^{\frac{1}{2}c^2}\int_{\mathbb{R}} \tfrac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}x^2}\,dx = e^{\frac{1}{2}c^2}.
\end{aligned} \tag{4.73}$$

The proof of Lemma 4.7.1 is thus completed. □

The statement and the proof of the following lemma can in a slightly different form, e.g., also be found in (4.91) in [Kühn(2004)].

Lemma 4.7.2. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $\alpha \in \mathbb{R}$, $\beta \in (0, \infty)$, let $Y : \Omega \to \mathbb{R}$ be an $\mathcal{N}_{\alpha,\beta^2}$-distributed random variable, and let $\Phi \colon \mathbb{R} \to \mathbb{R}$ be the function which satisfies for all $x \in \mathbb{R}$ that*

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \, e^{-\frac{1}{2}y^2} \, dy. \tag{4.74}$$

Then it holds for all $K \in \mathbb{R}$ that

$$\mathbb{E}_P\big[\max\{e^Y - K, 0\}\big] = \begin{cases} e^{\alpha + \frac{1}{2}\beta^2} \, \Phi\Big(\frac{\alpha - \ln(K)}{\beta} + \beta\Big) - K \, \Phi\Big(\frac{\alpha - \ln(K)}{\beta}\Big) & : K > 0 \\ e^{\alpha + \frac{1}{2}\beta^2} - K & : K \leq 0 \end{cases}. \tag{4.75}$$

Proof of Lemma 4.7.2. First of all, observe that Lemma 4.7.1 implies that for all $K \in (-\infty, 0]$ it holds that

$$\begin{aligned} \mathbb{E}_P\big[\max\big(e^Y - K, 0\big)\big] &= \mathbb{E}_P\big[e^Y - K\big] = \mathbb{E}_P\big[e^Y\big] - K \\ &= e^{\alpha} \int_{\mathbb{R}} e^{\beta y} \, \tfrac{1}{\sqrt{2\pi}} \, e^{-\frac{1}{2}y^2} \, dy - K = e^{\alpha + \frac{1}{2}\beta^2} - K. \end{aligned} \tag{4.76}$$

In addition, note that for all $K \in (0, \infty)$ it holds that

$$\begin{aligned} &\mathbb{E}_P\big[\max\big(e^Y - K, 0\big)\big] = \int_{\mathbb{R}} \max\big(e^{\alpha + \beta y} - K, 0\big) \, \tfrac{1}{\sqrt{2\pi}} \, e^{-\frac{1}{2}y^2} \, dy \\ &= \int_{\{x \in \mathbb{R} \colon \exp(\alpha + \beta x) \geq K\}} \big(e^{\alpha + \beta y} - K\big) \, \tfrac{1}{\sqrt{2\pi}} \, e^{-\frac{1}{2}y^2} \, dy \\ &= \int_{\frac{(\ln(K) - \alpha)}{\beta}}^{\infty} \big(e^{\alpha + \beta y} - K\big) \, \tfrac{1}{\sqrt{2\pi}} \, e^{-\frac{1}{2}y^2} \, dy \\ &= e^{\alpha} \int_{\frac{(\ln(K) - \alpha)}{\beta}}^{\infty} \tfrac{1}{\sqrt{2\pi}} \, e^{\left(\beta y - \frac{1}{2}y^2\right)} \, dy - K \int_{\frac{(\ln(K) - \alpha)}{\beta}}^{\infty} \tfrac{1}{\sqrt{2\pi}} \, e^{-\frac{1}{2}y^2} \, dy. \end{aligned} \tag{4.77}$$

This implies that for all $K \in (0, \infty)$ it holds that

$$\begin{aligned} \mathbb{E}_P\big[\max\big(e^Y - K, 0\big)\big] &= e^{\alpha} \int_{\frac{(\ln(K) - \alpha)}{\beta}}^{\infty} \tfrac{1}{\sqrt{2\pi}} \, e^{-\frac{1}{2}\left(y^2 - 2\beta y + \beta^2 - \beta^2\right)} \, dy - K \, \Phi\Big(\tfrac{\alpha - \ln(K)}{\beta}\Big) \\ &= e^{\alpha + \frac{1}{2}\beta^2} \int_{\frac{(\ln(K) - \alpha)}{\beta}}^{\infty} \tfrac{1}{\sqrt{2\pi}} \, e^{-\frac{1}{2}(y - \beta)^2} \, dy - K \, \Phi\Big(\tfrac{\alpha - \ln(K)}{\beta}\Big) \\ &= e^{\alpha + \frac{1}{2}\beta^2} \int_{\left[\frac{(\ln(K) - \alpha)}{\beta} - \beta\right]}^{\infty} \tfrac{1}{\sqrt{2\pi}} \, e^{-\frac{1}{2}y^2} \, dy - K \, \Phi\Big(\tfrac{\alpha - \ln(K)}{\beta}\Big) \\ &= e^{\alpha + \frac{1}{2}\beta^2} \, \Phi\Big(\beta - \tfrac{(\ln(K) - \alpha)}{\beta}\Big) - K \, \Phi\Big(\tfrac{\alpha - \ln(K)}{\beta}\Big) \\ &= e^{\alpha + \frac{1}{2}\beta^2} \, \Phi\Big(\tfrac{\alpha - \ln(K)}{\beta} + \beta\Big) - K \, \Phi\Big(\tfrac{\alpha - \ln(K)}{\beta}\Big). \end{aligned} \tag{4.78}$$

Combining (4.76) and (4.78) completes the proof of Lemma 4.7.2. □

We now use Lemma 4.7.2 and (4.57) to compute the right hand side of (4.72) in the case of an *European call option*. More precisely, Lemma 4.7.2, (4.57), and (4.72) prove that in the case where there exists a real number $K \in \mathbb{R}$ such that f satisfies (4.69) it holds P-a.s. that

$$\begin{aligned} D_0 &= e^{-rT}\,\mathbb{E}_P\big[f(\tilde{X})\big] = e^{-rT}\,\mathbb{E}_P\big[\max\{\tilde{X}_T - K, 0\}\big] \\ &= e^{-rT}\,\mathbb{E}_P\Big[\max\Big\{e^{(r-\frac{1}{2}\beta^2)T+\ln(X_0^{(2)})+\beta W_T} - K, 0\Big\}\Big] \end{aligned} \tag{4.79}$$

and hence it holds P-a.s. that

$$D_0 = \begin{cases} X_0^{(2)}\,\Phi\Big(\frac{(r+\frac{1}{2}\beta^2)T+\ln(X_0^{(2)}/K)}{\beta\sqrt{T}}\Big) - Ke^{-rT}\Phi\Big(\frac{(r-\frac{1}{2}\beta^2)T+\ln(X_0^{(2)}/K)}{\beta\sqrt{T}}\Big) & : K > 0 \\ X_0^{(2)} - Ke^{-rT} & : K \leq 0 \end{cases}. \tag{4.80}$$

Equation (4.80) is (a special case of) the famous *Black-Scholes formula* for option pricing. Let us close this section with a few comments and concluding remarks.

- It is in some sense completely ridiculous to model the price process of a stock price as a geometric Brownian motion (remark: there are a number of substantially more general models; see, e.g., Subsection 4.7.8 below). It might also be quite questionable to model the price process of a stock price as a *stochastic process*.
- There are also a number of other assumptions that are not fulfilled in the real life trading of financial derivatives (e.g., bid-ask spread/transaction costs, default risk, no arbitrage assumption, etc.; remark: there are also more general models that intend to (partially) take such issues into account).
- The aim in the Black-Scholes model is *not* to predict the expected payoff of the underlying nor the financial derivative at time $T > 0$. The aim in the Black-Scholes model is to estimate the price of the financial derivative *today*, that is, *at time* $t = 0$.
- This price of a financial derivative of the form (4.68) is in the setting of the Black-Scholes model completely independent of the expected interest rate α of the underlying.

To be more concrete, we now mention a simple illustrative example of a structured product (which is somehow a combination of one or more underlyings together with one or more financial derivatives) whose price has been estimated by using Monte Carlo methods. Between November 26th, 2012 and Mai 26th, 2014 the price of the structured product with the International Security Identification Number (ISIN) CH0197477877 has been estimated every trading day, in particular, by the swissQuant Group AG (Zurich, Switzerland; see `http://www.swissquant.com/`) by using Monte Carlo methods. The estimates prizes are important to determine the collaterals that the issuer of the considered structured product has to provide. In the case of a bankruptcy of the issuer of the structured product, the collaterals are taken to (partially) hedge the risk of a bankruptcy of the issuer of the considered structured product. The structured product with the ISIN CH0197477877 (cf., e.g., `http://ts.dp-research.com/131014IK014.pdf`) is a "Multi Barrier Reverse Convertible on Gold, Silver" and it is traded at the exchange Scoach Schweiz AG (an exchange in Zurich (Switzerland) which is specialist in structured products).

4.7.4 Stochastic Ginzburg-Landau equation

In addition to the assumptions in Subsection 4.7.1, let $\alpha, \beta, \bar{\beta} \in \mathbb{R}$, $\delta \in (0, \infty)$ be real numbers, assume that $d = m = 1$, $O = \mathbb{R}$, and assume for all $x \in \mathbb{R}$ that

$$\mu(x) = \alpha x - \delta x^3 \qquad \text{and} \qquad \sigma(x) = \beta x + \bar{\beta}. \tag{4.81}$$

The SDE (4.52) then reduces to the *stochastic Ginzburg-Landau equation*

$$dX_t = \left[\alpha X_t - \delta X_t^3\right] dt + \left[\beta X_t + \bar{\beta}\right] dW_t, \qquad t \in [0, T], \qquad X_0 = \xi. \tag{4.82}$$

Here the drift coefficient function μ is not globally Lipschitz continuous and Theorem 4.5.1 can thus not be applied. Nonetheless, there exists an up to indinstinguishability unique solution process of (4.82) (see, e.g., [Gyöngy and Krylov(1996)]) and, in addition, it holds for all $p \in [0, \infty)$ that

$$\sup_{t \in [0,T]} \mathbb{E}_P\big[|X_t|_{\mathbb{R}}^p\big] < \infty. \tag{4.83}$$

4.7.5 Stochastic Verhulst equation

In addition to the assumptions in Subsection 4.7.1, let $c \in \mathbb{R}$, $\eta, \lambda \in (0, \infty)$ be real numbers, assume that $d = m = 1$, $O = (0, \infty)$, and assume for all $x \in \mathbb{R}$ that

$$\mu(x) = \left(\eta + \tfrac{c^2}{2}\right) x - \lambda x^2 \qquad \text{and} \qquad \sigma(x) = cx. \tag{4.84}$$

The SDE (4.52) then reads as

$$dX_t = \left[\left(\eta + \tfrac{c^2}{2}\right) X_t - \lambda \left(X_t\right)^2\right] dt + c\, X_t\, dW_t, \qquad t \in [0, T], \qquad X_0 = \xi. \tag{4.85}$$

Equation (4.85) is known as *stochastic Verhulst equation* in the literature (see, e.g., Section 4.4 in Kloeden & Platen [Kloeden and Platen(1992)]). Here the drift coefficient function μ is not globally Lipschitz continuous and Theorem 4.5.1 can thus not be applied. Nonetheless, there exists an up to indinstinguishability unique solution process of (4.82) (see, e.g., [Gyöngy and Krylov(1996)]) and, in addition, it holds for all $p \in [0, \infty)$ that

$$\sup_{t \in [0,T]} \mathbb{E}_P\big[|X_t|_{\mathbb{R}}^p\big] < \infty. \tag{4.86}$$

4.7.6 Stochastic predator-prey model

In addition to the assumptions above, let $c_1, c_2 \in \mathbb{R}$, $\alpha, \beta, \gamma, \delta \in (0, \infty)$ be real numbers, assume that $d = m = 2$, $O = (0, \infty)^2$, and assume that for all $x = (x_1, x_2) \in O$ it holds that

$$\mu\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \, (\alpha - \beta x_2) \\ x_2 \, (\gamma x_1 - \delta) \end{pmatrix} \tag{4.87}$$

and

$$\sigma\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} c_1 x_1 & 0 \\ 0 & c_2 x_2 \end{pmatrix}. \tag{4.88}$$

The SDE (4.52) then reduces to the *stochastic predator-prey model*

$$dX_t = \begin{pmatrix} X_t^{(1)} \left(\alpha - \beta \cdot X_t^{(2)}\right) \\ X_t^{(2)} \left(\gamma \cdot X_t^{(1)} - \delta\right) \end{pmatrix} dt + \begin{pmatrix} c_1 \cdot X_t^{(1)} & 0 \\ 0 & c_2 \cdot X_t^{(2)} \end{pmatrix} dW_t, \; t \in [0, T], \quad X_0 = \xi. \tag{4.89}$$

The deterministic case ($c_1 = c_2 = 0$) of this model has been introduced by Lotka [Lotka(1920)] and Volterra [Volterra(1926)]. Here the drift coefficient function μ is not globally Lipschitz continuous and Theorem 4.5.1 can thus not be applied. Nonetheless, there exists an up to indinstinguishability unique solution process of (4.82) (see, e.g., [Gyöngy and Krylov(1996)]) and, in addition, it holds for all $p \in [0, \infty)$ that

$$\sup_{t \in [0,T]} \mathbb{E}_P\big[\|X_t\|_{\mathbb{R}^2}^p\big] < \infty. \tag{4.90}$$

4.7.7 Volatility processes

In addition to the assumptions in Subsection 4.7.1, let $a \in [1, \infty)$, $b \in [\frac{1}{2}, \infty)$, $\alpha \in (0, \infty)$, $\beta, \delta \in [0, \infty)$, $\gamma \in \mathbb{R}$ be real numbers with

$$a + 1 \geq 2b, \qquad \delta > -\mathbb{1}_{[1,\infty)}(b), \qquad \text{and} \qquad \delta \geq \mathbb{1}_{\{\frac{1}{2}\}}(b) \cdot \frac{\beta^2}{2}, \tag{4.91}$$

assume that $d = m = 1$, $O = (0, \infty)$ and assume that for all $x \in (0, \infty)$ it holds that

$$\mu(x) = \delta + \gamma\, x - \alpha\, x^a \qquad \text{and} \qquad \sigma(x) = \beta\, x^b. \tag{4.92}$$

The SDE (4.52) then reads as

$$dX_t = \big[\delta + \gamma X_t - \alpha\,(X_t)^a\,\big]\, dt + \beta\,(X_t)^b\, dW_t, \;\; t \in [0, T], \qquad X_0 = \xi. \tag{4.93}$$

Assumption (4.91) ensures the existence of an up to indistinguishability unique solution process of (4.93). The proof of the existence of a solution process of (4.93) is omitted. Let us consider three more specific examples of the SDE (4.93).

4.7.7.1 Cox-Ingersoll-Ross process

In addition to the assumptions above, assume that $a = 1$, $b = \frac{1}{2}$ and $\gamma = 0$. The SDE (4.93) is then the *Cox-Ingersoll-Ross process*

$$dX_t = \big[\delta - \alpha X_t\big]\, dt + \beta\sqrt{X_t}\, dW_t, \;\; t \in [0,T], \qquad X_0 = \xi \tag{4.94}$$

which has been introduced in [Cox et al.(1985)Cox, Ingersoll, and Ross] as a model for instantaneous interest rates. Later, in [Heston(1993)], this process has been proposed as a model for the squared volatility in a Black-Scholes type market model (see Subsection 4.7.8.1 below).

4.7.7.2 Simplified Ait-Sahalia interest rate model

In addition to the assumptions above, assume that $a = 2$ and $b < \frac{3}{2}$. Under these additional assumptions, the SDE (4.93) reads as

$$dX_t = \big[\delta + \gamma X_t - \alpha\,(X_t)^2\,\big]\, dt + \beta\,(X_t)^b\, dW_t, \;\; t \in [0,T], \qquad X_0 = \xi. \tag{4.95}$$

A more general version hereof has been used in Ait-Sahalia [Ait-Sahalia(1996)] for testing continuous-time models of the spot interest rate. More information on these type of models can be found in the introductory section in [Szpruch et al.(2011)Szpruch, Mao, Higham, and Pan] and in the references mentioned therein.

4.7.7.3 Volatility process in the Lewis stochastic volatility model

In addition to the assumptions above, assume that $a = 2$, $b = \frac{3}{2}$, $\gamma > 0$ and $\delta = 0$. The SDE (4.93) is then the instantaneous variance process (squared volatility) in the *Lewis stochastic volatility model* (see [Lewis(2000)] and Subsection 4.7.8.2 below)

$$dX_t = \big[\gamma X_t - \alpha\,(X_t)^2\,\big]\, dt + \beta\,(X_t)^{\frac{3}{2}}\; dW_t, \;\; t \in [0,T], \qquad X_0 = \xi. \tag{4.96}$$

The stochastic volatility model associated to (4.96) is a.k.a. 3/2-*stochastic volatility model*.

4.7.8 Stochastic volatility models

In addition to the assumptions in Subsection 4.7.1, let $\hat{a} \in [1,\infty)$, $\hat{b} \in [\frac{1}{2},\infty)$, $\alpha, \hat{\alpha} \in (0,\infty)$, $\hat{\beta}, \hat{\delta} \in [0,\infty)$, $\hat{\gamma} \in \mathbb{R}$, $\rho \in [0,1]$ be real numbers with

$$\hat{a} + 1 \geq 2\hat{b}, \qquad \hat{\delta} > -\mathbb{1}_{[1,\infty)}(\hat{b}), \qquad \text{and} \qquad \hat{\delta} \geq \mathbb{1}_{\{\frac{1}{2}\}}(\hat{b}) \cdot \tfrac{(\hat{\beta})^2}{2}, \tag{4.97}$$

assume that $d = m = 2$, $O = (0,\infty)^2$, and assume that for all $x = (x_1, x_2) \in (0,\infty)^2$ it holds that

$$\mu\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \alpha\, x_1 \\ \hat{\delta} + \hat{\gamma} x_2 - \hat{\alpha}\,(x_2)^{\hat{a}} \end{pmatrix}, \quad \sigma\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \sqrt{x_2}\, x_1 & 0 \\ \hat{\beta}\,(x_2)^{\hat{b}}\sqrt{1-\rho^2} & \hat{\beta}\,(x_2)^{\hat{b}}\,\rho \end{pmatrix}. \tag{4.98}$$

The SDE (4.52) then reads as

$$\begin{aligned} dX_t &= \begin{pmatrix} \alpha\, X_t^{(1)} \\ \hat{\delta} + \hat{\gamma} X_t^{(2)} - \hat{\alpha}\,\big(X_t^{(2)}\big)^{\hat{a}} \end{pmatrix} dt \\ &\quad + \begin{pmatrix} \sqrt{X_t^{(2)}}\, X_t^{(1)} & 0 \\ \hat{\beta}\,\big(X_t^{(2)}\big)^{\hat{b}}\sqrt{1-\rho^2} & \hat{\beta}\,\big(X_t^{(2)}\big)^{\hat{b}}\,\rho \end{pmatrix} dW_t, \qquad t \in [0,T], \qquad X_0 = \xi. \end{aligned} \tag{4.99}$$

In the next step let $\hat{W}\colon [0,T]\times\Omega \to \mathbb{R}$ be the mapping with the property that for all $t \in [0,T]$ it holds that

$$\hat{W}_t = \sqrt{1-\rho^2}\, W_t^{(1)} + \rho\, W_t^{(2)}. \tag{4.100}$$

Then we observe that $\hat{W}$ is an one-dimensional standard $(\mathbb{F}_t)_{t\in[0,T]}$-Brownian motion. Note that $W^{(1)}$ and $\hat{W}$ are independent if and only if $\rho = 1$. Using this notation, we obtain that the stochastic process $X^{(2)}\colon [0,T]\times\Omega \to (0,\infty)$ is a solution process of the SDE

$$dX_t^{(2)} = \Big[\hat{\delta} + \hat{\gamma} X_t^{(2)} - \hat{\alpha}\,\big(X_t^{(2)}\big)^{\hat{a}}\Big]\, dt + \hat{\beta}\,\big(X_t^{(2)}\big)^{\hat{b}}\, d\hat{W}_t, \;\; t \in [0,T], \qquad X_0^{(2)} = \xi^{(2)}. \tag{4.101}$$

Assumption (4.91) ensures the existence of an up to indistinguishability unique solution process of (4.99). The proof of the existence of a solution process of (4.99) is omitted. Let us consider two more specific examples of the SDE (4.99).

4.7.8.1 Heston model

In addition to the assumptions above, assume that $\hat{a} = 1$, $\hat{b} = \frac{1}{2}$ and $\hat{\gamma} = 0$. The SDE (4.99) is then the *Heston model*

$$dX_t = \begin{pmatrix} \alpha\, X_t^{(1)} \\ \hat{\delta} - \hat{\alpha}\, X_t^{(2)} \end{pmatrix} dt + \begin{pmatrix} \sqrt{X_t^{(2)}}\, X_t^{(1)} & 0 \\ \hat{\beta}\sqrt{1-\rho^2}\sqrt{X_t^{(2)}} & \hat{\beta}\,\rho\,\sqrt{X_t^{(2)}} \end{pmatrix} dW_t,\; t \in [0,T],\; X_0 = \xi \tag{4.102}$$

(see [Heston(1993)]).

4.7.8.2 Lewis stochastic volatility model

In addition to the assumptions above, assume that $\hat{a} = 2$, $\hat{b} = \frac{3}{2}$, $\hat{\gamma} > 0$ and $\hat{\delta} = 0$. The SDE (4.99) is then the *Lewis stochastic volatility model* (a.k.a. 3/2-stochastic volatility model)

$$dX_t = \begin{pmatrix} \alpha\, X_t^{(1)} \\ \hat{\gamma} X_t - \hat{\alpha}\left(X_t^{(2)}\right)^2 \end{pmatrix} dt + \begin{pmatrix} \sqrt{X_t^{(2)}}\, X_t^{(1)} & 0 \\ \hat{\beta}\,\sqrt{1-\rho^2}\left(X_t^{(2)}\right)^{\frac{3}{2}} & \hat{\beta}\,\rho\left(X_t^{(2)}\right)^{\frac{3}{2}} \end{pmatrix} dW_t,\ t \in [0,T],\ X_0 = \xi \tag{4.103}$$

(see [Lewis(2000)] and, e.g., also [Henry-Labordère(2007), Higham(2011)]).

5 Strong approximations for SDEs

Most of this chapter can, e.g., in a bit different form be found in [Kloeden and Platen(1992)].

5.1 Setting

Throughout this chapter the following setting is frequently used. Let $T \in (0, \infty)$, $d, m \in \mathbb{N}$, let $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$ be a stochastic basis, let $W = (W^{(1)}, \ldots, W^{(m)}) \colon [0, T] \times \Omega \to \mathbb{R}^m$ be an m-dimensional standard $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$-Brownian motion, let $O \subseteq \mathbb{R}^d$ be an open set, let $\xi = (\xi^{(1)}, \ldots, \xi^{(d)}) \in \mathcal{M}(\mathbb{F}_0, \mathcal{B}(O))$, $\bar{\mu} \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{R}^d))$, $\bar{\sigma} \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{R}^{d\times m}))$, let

$$\mu = (\mu_1, \ldots, \mu_d) \colon O \to \mathbb{R}^d, \quad \sigma = (\sigma_j)_{j\in\{1,\ldots,m\}} = (\sigma_{i,j})_{i\in\{1,\ldots,d\},j\in\{1,\ldots,m\}} \colon O \to \mathbb{R}^{d\times m} \tag{5.1}$$

be locally Lipschitz continuous functions, let $X = (X^{(1)}, \ldots, X^{(d)}) \colon [0, T] \times \Omega \to O$ be a solution process of the SDE

$$dX_t = \mu(X_t)\, dt + \sigma(X_t)\, dW_t, \qquad t \in [0, T], \qquad X_0 = \xi, \tag{5.2}$$

and assume for all $x \in O$ that

$$\bar{\mu}(x) = \mu(x) \qquad \text{and} \qquad \bar{\sigma}(x) = \sigma(x). \tag{5.3}$$

Remark 5.1.1. *The functions $\bar{\mu}$ and $\bar{\sigma}$ are thus Borel measurable* extentions *of μ and σ. For instance, the functions*

$$\mathbb{R}^d \ni x \mapsto \left\{ \begin{array}{cl} \mu(x) & : x \in O \\ 0 & : x \notin O \end{array} \right\} \in \mathbb{R}^d, \qquad \mathbb{R}^d \ni x \mapsto \left\{ \begin{array}{cl} \sigma(x) & : x \in O \\ 0 & : x \notin O \end{array} \right\} \in \mathbb{R}^{d\times m} \tag{5.4}$$

are Borel measurable extensions of μ and σ.

Remark 5.1.2. *We observe that for all $i \in \{1, 2, \ldots, d\}$, $j \in \{1, 2, \ldots, m\}$ it holds that $\sigma_{i,j} \colon O \to \mathbb{R}$ and $\sigma_j \colon O \to \mathbb{R}^d$ are functions which satisfy for all $x \in O$ that*

$$\sigma_j(x) = (\sigma_{1,j}(x), \ldots, \sigma_{d,j}(x))\,. \tag{5.5}$$

5.2 Notions of convergence for stochastic processes

5.2.1 Growth properties for functions

Definition 5.2.1. *Let (E, d_E) and (F, d_F) be metric spaces. Then we say that f grows at most linearly from (E, d_E) to (F, d_F) (we say that f grows at most linearly) if and only if $f \in \mathbb{M}(E, F)$ is a function from E to F which satisfies for all $v \in E$, $w \in F$ that there exists a real number $c \in \mathbb{R}$ such that for all $x \in E$ it holds that*

$$d_F(w, f(x)) \leq c\,(1 + d_E(v, x)) . \tag{5.6}$$

Definition 5.2.2. *Let (E, d_E) and (F, d_F) be metric spaces. Then we say that f grows at most quadratically from (E, d_E) to (F, d_F) (we say that f grows at most quadratically) if and only if $f \in \mathbb{M}(E, F)$ is a function from E to F which satisfies for all $v \in E$, $w \in F$ that there exists a real number $c \in \mathbb{R}$ such that for all $x \in E$ it holds that*

$$d_F(w, f(x)) \leq c\,(1 + d_E(v, x))^2 . \tag{5.7}$$

Definition 5.2.3. *Let (E, d_E) and (F, d_F) be metric spaces. Then we say that f grows at most polynomially from (E, d_E) to (F, d_F) (we say that f grows at most polynomially) if and only if $f \in \mathbb{M}(E, F)$ is a function from E to F which satisfies for all $v \in E$, $w \in F$ that there exists a real number $c \in [0, \infty)$ such that for all $x \in E$ it holds that*

$$d_F(w, f(x)) \leq c\,(1 + d_E(v, x))^c . \tag{5.8}$$

Exercise 5.2.4. *Let (E, d_E) and (F, d_F) be metric spaces with $E \neq \emptyset$ and let $f : E \to F$ be a function. Prove that f grows at most polynomially from (E, d_E) to (F, d_F) if and only if there exist $v \in E$, $w \in F$ such that*

$$\limsup_{c \to \infty} \sup_{x \in E} \left[\frac{d_F(w, f(x))}{[1 + d_E(v, x)]^c} \right] < \infty. \tag{5.9}$$

5.2.2 Growth properties for derivatives of functions

Definition 5.2.5 (1-Hölder continuous)**.** *Let (E, d_E) and (F, d_F) be metric spaces. Then we say that f is globally Lipschitz continuous from (E, d_E) to (F, d_F) (we say that f is globally Lipschitz continuous, we say that f is Lipschitz continuous from (E, d_E) to (F, d_F), we say that f is Lipschitz continuous) if and only if f is d_E/d_F-1-Hölder continuous.*

Definition 5.2.6. *Let $k, l, v \in \mathbb{N}$. Then we say that f is v-times continuously differentiable with at most polynomially growing derivatives from $\mathbb{R}^k$ to $\mathbb{R}^l$ (we say that f is v-times continuously differentiable with at most polynomially growing derivatives) if and only if $f \in C^v(\mathbb{R}^k, \mathbb{R}^l)$ is a v-times continuously differentiable function from $\mathbb{R}^k$ to $\mathbb{R}^l$ which satisfies that $f^{(v)}$ grows at most polynomially.*

Exercise 5.2.7. *Let $k, l, v \in \mathbb{N}$ and let $f\colon \mathbb{R}^k \to \mathbb{R}^l$ be a v-times continuously differentiable function with at most polynomially growing derivatives. Prove that for all $w \in \{0, 1, \ldots, v\}$ it holds that $f^{(w)}$ grows at most polynomially.*

Definition 5.2.8. *Let $k, l \in \mathbb{N}$. Then we say that f is infinitely often differentiable with at most polynomially growing derivatives from $\mathbb{R}^k$ to $\mathbb{R}^l$ (we say that f is infinitely often differentiable with at most polynomially growing derivatives) if and only if $f \in C^\infty(\mathbb{R}^k, \mathbb{R}^l)$ is an infinitely often differentiable function from $\mathbb{R}^k$ to $\mathbb{R}^l$ which satisfies that for every $v \in \mathbb{N}$ it holds that $f^{(v)}$ grows at most polynomially.*

Corollary 5.2.9. *Let $k, l \in \mathbb{N}$ and let $f\colon \mathbb{R}^k \to \mathbb{R}^l$ be a function. Then f grows at most polynomially if and only if there exists a real number $c \in [0, \infty)$ such that for all $x \in \mathbb{R}^k$ it holds that*

$$\|f(x)\|_{\mathbb{R}^l} \leq c\,(1 + \|x\|_{\mathbb{R}^k}^c) \tag{5.10}$$

Corollary 5.2.9 is a straightforward consequence of Definition 5.2.3 and Exercise 5.2.4 above.

5.2.3 Strong convergence

Definition 5.2.10. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $d \in \mathbb{N}$, $T, p \in (0, \infty)$, and let $Y^N\colon [0, T] \times \Omega \to \mathbb{R}^d$, $N \in \mathbb{N}_0$, be stochastic processes. Then we say that $(Y^N)_{N\in\mathbb{N}}$ converges at time T in the strong L^p-sense to Y^0 on $(\Omega, \mathcal{F}, P)$ (we say that $(Y^N)_{N\in\mathbb{N}}$ converges at time T in the strong L^p-sense to Y^0) if and only if*

$$\limsup_{N\to\infty} \mathbb{E}_P\big[\|Y_T^0 - Y_T^N\|_{\mathbb{R}^d}^p\big] = 0. \tag{5.11}$$

Definition 5.2.11. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $d \in \mathbb{N}$, $T, p, \alpha \in (0, \infty)$, and let $Y^N\colon [0, T] \times \Omega \to \mathbb{R}^d$, $N \in \mathbb{N}_0$, be stochastic processes. Then we say that $(Y^N)_{N\in\mathbb{N}}$ converges at time T in the strong L^p-sense with order α to Y^0 on $(\Omega, \mathcal{F}, P)$ (we say that $(Y^N)_{N\in\mathbb{N}}$ converges at time T in the strong L^p-sense with order α to Y^0) if and only if there exists a real number $C \in \mathbb{R}$ such that for all $N \in \mathbb{N}$ it holds that*

$$\big\|Y_T^0 - Y_T^N\big\|_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})} \leq C \cdot N^{-\alpha}, \tag{5.12}$$

5.2.4 Almost sure convergence

Definition 5.2.12. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $d \in \mathbb{N}$, $T \in (0, \infty)$, and let $Y^N\colon [0, T] \times \Omega \to \mathbb{R}^d$, $N \in \mathbb{N}_0$, be stochastic processes. Then we say that $(Y^N)_{N\in\mathbb{N}}$ converges at time T P-almost surely to Y^0 if and only if*

$$P\bigg(\limsup_{N\to\infty} \|Y_T^0 - Y_T^N\|_{\mathbb{R}^d} = 0\bigg) = 1. \tag{5.13}$$

Definition 5.2.13. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $d \in \mathbb{N}$, $T, \alpha \in (0, \infty)$, and let $Y^N \colon [0,T] \times \Omega \to \mathbb{R}^d$, $N \in \mathbb{N}_0$, be stochastic processes. Then we say that $(Y^N)_{N \in \mathbb{N}}$ converges at time T P-almost surely with order α to Y^0 if and only if there exists an $\mathcal{F}/\mathcal{B}(\mathbb{R})$-measurable function $C \colon \Omega \to \mathbb{R}$ such that for all $N \in \mathbb{N}$ it holds that*

$$P\big(\|Y_T^0 - Y_T^N\|_{\mathbb{R}^d} \leq C \cdot N^{-\alpha}\big) = 1. \tag{5.14}$$

5.2.5 Convergence in probability

Definition 5.2.14. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $d \in \mathbb{N}$, $T \in (0, \infty)$, and let $Y^N \colon [0,T] \times \Omega \to \mathbb{R}^d$, $N \in \mathbb{N}_0$, be stochastic processes. Then we say that $(Y^N)_{N \in \mathbb{N}}$ converges at time T in probability to Y^0 on $(\Omega, \mathcal{F}, P)$ (we say that $(Y^N)_{N \in \mathbb{N}}$ converges at time T in probability to Y^0) if and only if it holds for all $\varepsilon \in (0, \infty)$ that*

$$\limsup_{N \to \infty} P\big(\|Y_T^0 - Y_T^N\|_{\mathbb{R}^d} \geq \varepsilon\big) = 0. \tag{5.15}$$

5.2.6 Numerically weak convergence

Definition 5.2.15. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $d \in \mathbb{N}$, $T \in (0, \infty)$, and let $Y^N \colon [0,T] \times \Omega \to \mathbb{R}^d$, $N \in \mathbb{N}_0$, be stochastic processes. Then we say that $(Y^N)_{N \in \mathbb{N}}$ converges at time T in the numerically weak sense to Y^0 if and only if it holds for every infinitely often differentiable function $\varphi \colon \mathbb{R}^d \to \mathbb{R}$ with at most polynomially growing derivatives that $\forall\, N \in \mathbb{N}_0 \colon\ \mathbb{E}_P\big[|\varphi(Y_T^N)|_{\mathbb{R}}\big] < \infty$ and*

$$\limsup_{N \to \infty} \big|\mathbb{E}_P\big[\varphi(Y_T^0)\big] - \mathbb{E}_P\big[\varphi(Y_T^N)\big]\big|_{\mathbb{R}} = 0. \tag{5.16}$$

Definition 5.2.16. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $d \in \mathbb{N}$, $T, \alpha \in (0, \infty)$, and let $Y^N \colon [0,T] \times \Omega \to \mathbb{R}^d$, $N \in \mathbb{N}_0$, be stochastic processes. Then we say that $(Y^N)_{N \in \mathbb{N}}$ converges at time T in the numerically weak sense with order α to Y^0 on $(\Omega, \mathcal{F}, P)$ (we say that $(Y^N)_{N \in \mathbb{N}}$ converges at time T in the numerically weak sense with order α to Y^0) if and only if it holds for every infinitely often differentiable function $\varphi \colon \mathbb{R}^d \to \mathbb{R}$ with at most polynomially growing derivatives that there exists a real number $C \in \mathbb{R}$ such that for all $N \in \mathbb{N}$ it holds that $\mathbb{E}_P\big[|\varphi(Y_T^0)|_{\mathbb{R}} + |\varphi(Y_T^N)|_{\mathbb{R}}\big] < \infty$ and*

$$\big|\mathbb{E}_P\big[\varphi(Y_T^0)\big] - \mathbb{E}_P\big[\varphi(Y_T^N)\big]\big|_{\mathbb{R}} \leq C \cdot N^{-\alpha}. \tag{5.17}$$

5.3 Euler-Maruyama scheme

Assume the setting in Section 5.1. Then for all $t \in [0,T]$ it holds P-a.s. that

$$X_t = \xi + \int_0^t \mu(X_s)\, ds + \int_0^t \sigma(X_s)\, dW_s. \tag{5.18}$$

This and (3.152) show that for all $t_0 \in [0,T]$, $t \in [t_0,T]$ with $t - t_0$ "sufficiently small" it holds P-a.s. that

$$\begin{aligned} X_t &= X_{t_0} + \int_{t_0}^t \underbrace{\mu(X_s)}_{\approx \mu(X_{t_0})} \, ds + \int_{t_0}^t \underbrace{\sigma(X_s)}_{\approx \sigma(X_{t_0})} \, dW_s \\ &\approx X_{t_0} + \int_{t_0}^t \mu(X_{t_0}) \, ds + \int_{t_0}^t \sigma(X_{t_0}) \, dW_s \\ &= X_{t_0} + \mu(X_{t_0}) \, (t - t_0) + \sigma(X_{t_0}) \, (W_t - W_{t_0}) \, . \end{aligned} \tag{5.19}$$

The approximation in (5.19) motivates the following definition (see [Maruyama(1953), Maruyama(1955)]).

Definition 5.3.1 (Euler-Maruyama approximation)**.** *Let $T \in (0,\infty)$, $d, m, N \in \mathbb{N}$, $\mu \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{R}^d))$, $\sigma \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{R}^{d \times m}))$, let $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t \in [0,T]})$ be a stochastic basis, let $\xi \in \mathcal{L}^0(P|_{\mathbb{F}_0}; \|\cdot\|_{\mathbb{R}^d})$, and let $W\colon [0,T] \times \Omega \to \mathbb{R}^m$ be an m-dimensional standard $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t \in [0,T]})$-Brownian motion. Then we say that Y is an Euler-Maruyama approximation for the SDE*

$$dX_t = \mu(X_t) \, dt + \sigma(X_t) \, dW_t, \qquad t \in [0,T], \qquad X_0 = \xi \tag{5.20}$$

with time step size T/N on $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t \in [0,T]})$ (we say that Y is an Euler-Maruyama approximation for the SDE

$$dX_t = \mu(X_t) \, dt + \sigma(X_t) \, dW_t, \qquad t \in [0,T], \qquad X_0 = \xi \tag{5.21}$$

with time step size T/N) if and only if $Y \in \mathbb{M}(\{0,1,\dots,N\} \times \Omega, \mathbb{R}^d)$ is the function from $\{0,1,\dots,N\} \times \Omega$ to $\mathbb{R}^d$ which satisfies for all $n \in \{0,1,\dots,N-1\}$ that $Y_0 = \xi$ and

$$Y_{n+1} = Y_n + \mu(Y_n) \tfrac{T}{N} + \sigma(Y_n) \big(W_{\frac{(n+1)T}{N}} - W_{\frac{nT}{N}} \big). \tag{5.22}$$

Sometimes Euler-Maruyama approximations are also simply referred to as Euler approximations in the literature.

Definition 5.3.2 (Linearly-interpolated Euler-Maruyama approximation). *Let* $T \in (0, \infty)$, $d, m, N \in \mathbb{N}$, $\mu \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{R}^d))$, $\sigma \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{R}^{d\times m}))$, *let* $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$ *be a stochastic basis, let* $\xi \in \mathcal{L}^0(P|_{\mathbb{F}_0}; \|\cdot\|_{\mathbb{R}^d})$, *and let* $W\colon [0,T]\times\Omega \to \mathbb{R}^m$ *be an* m*-dimensional standard* $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$*-Brownian motion. Then we say that* Y *is a linearly-interpolated Euler-Maruyama approximation for the SDE*

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t, \qquad t \in [0,T], \qquad X_0 = \xi \tag{5.23}$$

with time step size T/N *on* $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$ *(we say that* Y *is a linearly-interpolated Euler-Maruyama approximation for the SDE*

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t, \qquad t \in [0,T], \qquad X_0 = \xi \tag{5.24}$$

with time step size T/N*) if and only if* $Y \in \mathbb{M}([0,T]\times\Omega, \mathbb{R}^d)$ *is the function from* $[0,T]\times\Omega$ *to* $\mathbb{R}^d$ *which satisfies for all* $n \in \{0, 1, \ldots, N-1\}$, $t \in \left[\frac{nT}{N}, \frac{(n+1)T}{N}\right]$ *that* $Y_0 = \xi$ *and*

$$Y_t = Y_{\frac{nT}{N}} + \left(\tfrac{tN}{T} - n\right)\left[\mu(Y_{\frac{nT}{N}})\tfrac{T}{N} + \sigma(Y_{\frac{nT}{N}})\big(W_{\frac{(n+1)T}{N}} - W_{\frac{nT}{N}}\big)\right]. \tag{5.25}$$

In Definition 5.3.1 we introduce Euler-Maruyama approximations at the discretization points $\{0, \frac{T}{N}, \frac{2T}{N}, \ldots, T\}$, $N \in \mathbb{N}$. In the literature sometimes approximations on more complicated possibly non-equidistant discretization points are investigated/used.

5.3.1 Simulation of sample paths of Euler-Maruyama approximations

In the next step two Matlab codes for the Euler-Maruyama method are presented in the case $d = m = 1$.

```
function Y = EulerMaruyama(mu, sigma ,T, x0 ,N)
  Y = zeros(1 ,N+1);
  Y(1) = x0;
  h = T/N;
  sqrth = sqrt(h);
  for n = 1:N
    Y(n+1) = Y(n) + mu(Y(n))*h + sigma(Y(n))*sqrth*randn;
  end
end
```

Matlab code 5.1: A Matlab function for the Euler-Maruyama method in the case $d = m = 1$.

```
T = 1;
N = 1000;
mu = @(x)-x^3;
```

```
plot( (0:T/N:T), EulerMaruyama(mu, @(x)0, T, 2, N), 'r' );
hold on
plot( (0:T/N:T), EulerMaruyama(mu, @(x)1/10, T, 2, N) );
```

Matlab code 5.2: A Matlab code for the Euler-Maruyama method.

Exercise 5.3.3 (Euler-Maruyama)**.** *In this exercise we do not distinguish between pseudo random numbers and actual random numbers. Let* $T \in (0,\infty)$*,* $d, m, N \in \mathbb{N}$*,* $\xi \in \mathbb{R}^d$*,* $\mu \in \mathcal{M}\big(\mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{R}^d)\big)$*,* $\sigma \in \mathcal{M}\big(\mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{R}^{d\times m})\big)$*, let* $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$ *be a stochastic basis, let* $W\colon [0,T] \times \Omega \to \mathbb{R}^m$ *be an* m*-dimensional standard* $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$*-Brownian motion, and let* $Y\colon \{0,1,\dots,N\} \times \Omega \to \mathbb{R}^d$ *be an Euler-Maruyama approximation for the SDE*

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t, \qquad t \in [0,T], \qquad X_0 = \xi \tag{5.26}$$

with time step size T/N*. Write a* Matlab *function* `EulerMaruyama(`T`,`d`,`m`,`N`,`ξ`,`μ`,`σ`)` *with input* $T \in (0,\infty)$*,* $d, m, N \in \mathbb{N}$*,* $\xi \in \mathbb{R}^d$*,* $\mu \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{R}^d))$*,* $\sigma \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{R}^{d\times m}))$ *and output a realization of an* $Y_N(P)_{\mathcal{B}(\mathbb{R}^d)}$*-distributed random variable.*

Exercise 5.3.4 (Monte Carlo Euler for geometric Brownian motion)**.** *In this exercise we do not distinguish between pseudo random numbers and actual random numbers. Let* $T, \alpha, \beta, \xi, K \in (0,\infty)$*,* $N, M \in \mathbb{N}$*, let* $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$ *be a stochastic basis, let* $W^k\colon [0,T] \times \Omega \to \mathbb{R}$*,* $k \in \mathbb{N}$*, be* P*-independent standard* $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$*-Brownian motions, and for every* $k \in \mathbb{N}$ *let* $Y^k\colon \{0,1,\dots,N\} \times \Omega \to \mathbb{R}$ *be an Euler-Maruyama approximation for the SDE*

$$dX_t = \alpha\, X_t\,dt + \beta\, X_t\,dW_t^k, \qquad t \in [0,T], \qquad X_0 = \xi \tag{5.27}$$

with time step size T/N*. Write a* Matlab *function* `MonteCarloEulerGBM(`T`,`α`,`β`,`ξ`,`K`,`N`,`M`)` *with input* $T, \alpha, \beta, \xi, K \in (0,\infty)$*,* $N, M \in \mathbb{N}$ *and output a realization of an* $\big(\frac{1}{M}\sum_{k=1}^{M} \max\{Y_N^k - K, 0\}\big)(P)_{\mathcal{B}(\mathbb{R})}$*-distributed random variable. Type* `MonteCarloEulerGBM(1,log(1.06)-1/200,1/10,92,100,100,10000)` *to test your implementation. Compare this result with the results of Exercise 3.3.9 and Exercise 3.3.11.*

5.3.2 Strong convergence of the Euler-Maruyama method

In Theorem 5.3.10 below we prove strong convergence of the Euler-Maruyama method under the assumption that the coefficients of the SDE are globally Lipschitz continuous. The proof of Theorem 5.3.10 uses the well-known Hölder inequality. For completeness we now present (a special case of) the Hölder inequality.

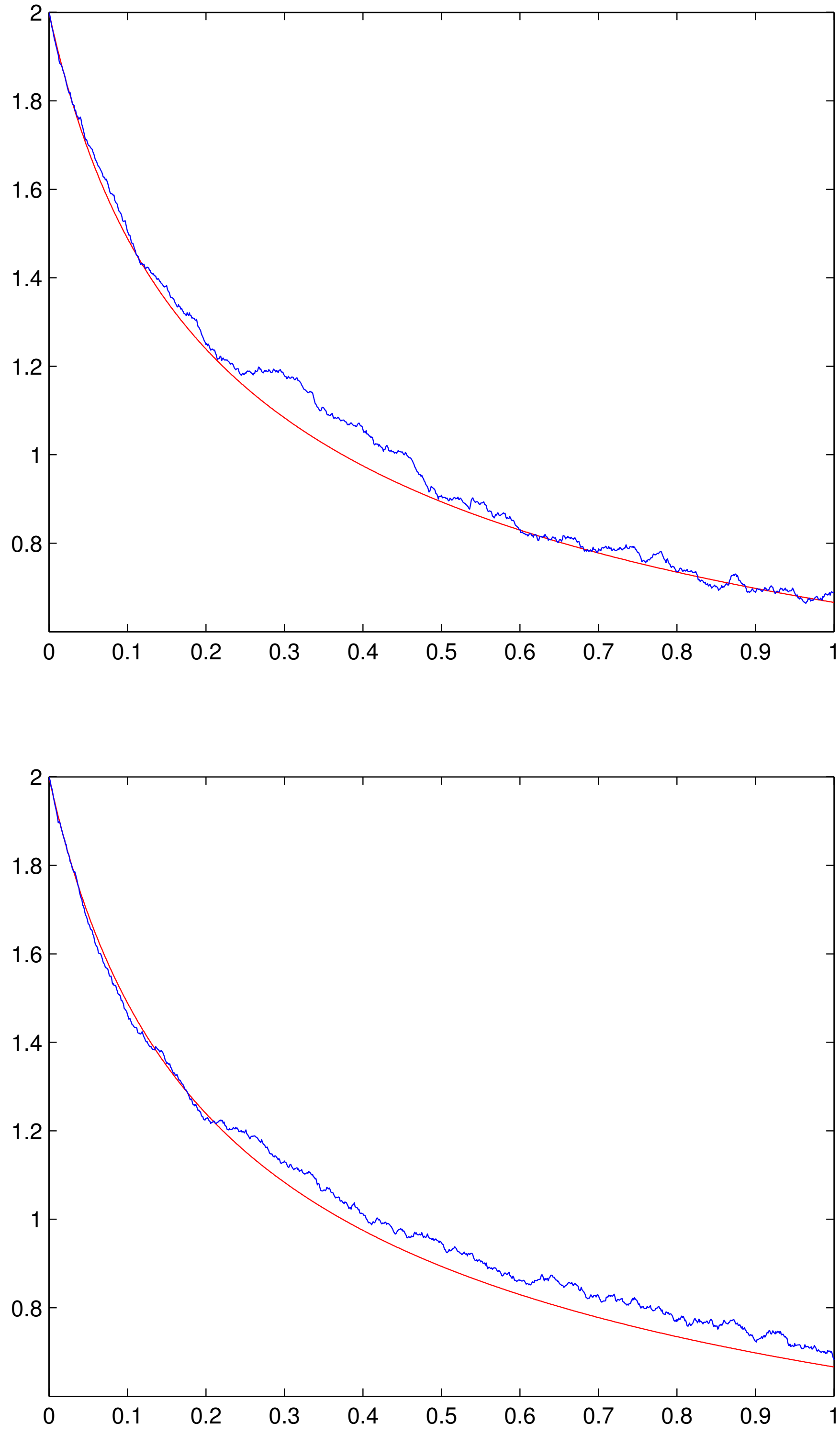


Figure 5.1: Results of two calls of the Matlab code 5.2.

Theorem 5.3.5 (Hölder inequality)**.** *Let* $(\Omega, \mathcal{F}, \mu)$ *be a measure space and let* $r \in (0, \infty)$*,* $f \in \mathcal{L}^{1+r}(\mu; |\cdot|_{\mathbb{R}})$*,* $g \in \mathcal{L}^{1+1/r}(\mu; |\cdot|_{\mathbb{R}})$*. Then*

$$\int_{\Omega} |f(x) \cdot g(x)| \, \mu(dx) = \|f \cdot g\|_{\mathcal{L}^1(\mu; |\cdot|_{\mathbb{R}})} \leq \|f\|_{\mathcal{L}^{1+r}(\mu; |\cdot|_{\mathbb{R}})} \, \|g\|_{\mathcal{L}^{1+1/r}(\mu; |\cdot|_{\mathbb{R}})} . \tag{5.28}$$

Proof of Theorem 5.3.5. Throughout this proof we assume w.l.o.g. that

$$\|f\|_{\mathcal{L}^{1+r}(\mu; |\cdot|_{\mathbb{R}})} > 0 \qquad \text{and} \qquad \|g\|_{\mathcal{L}^{1+1/r}(\mu; |\cdot|_{\mathbb{R}})} > 0 \tag{5.29}$$

(otherwise (5.28) is clear). Next observe that Young's inequality proves that

$$\begin{aligned}
&\int_{\Omega} \left[\frac{|f(x)|}{\|f\|_{\mathcal{L}^{1+r}(\mu; |\cdot|_{\mathbb{R}})}} \right] \cdot \left[\frac{|g(x)|}{\|g\|_{\mathcal{L}^{1+1/r}(\mu; |\cdot|_{\mathbb{R}})}} \right] \mu(dx) \\
&\leq \int_{\Omega} \frac{1}{(1+r)} \left[\frac{|f(x)|}{\|f\|_{\mathcal{L}^{1+r}(\mu; |\cdot|_{\mathbb{R}})}} \right]^{(1+r)} + \frac{1}{(1+\frac{1}{r})} \left[\frac{|g(x)|}{\|g\|_{\mathcal{L}^{1+1/r}(\mu; |\cdot|_{\mathbb{R}})}} \right]^{(1+\frac{1}{r})} \mu(dx) \\
&= \frac{1}{(1+r)} \left[\frac{\int_{\Omega} |f(x)|^{(1+r)} \, \mu(dx)}{\|f\|^{(1+r)}_{\mathcal{L}^{1+r}(\mu; |\cdot|_{\mathbb{R}})}} \right] + \frac{1}{(1+\frac{1}{r})} \left[\frac{\int_{\Omega} |g(x)|^{(1+1/r)} \, \mu(dx)}{\|g\|^{(1+1/r)}_{\mathcal{L}^{1+1/r}(\mu; |\cdot|_{\mathbb{R}})}} \right] \\
&= \frac{1}{(1+r)} + \frac{1}{(1+\frac{1}{r})} = 1.
\end{aligned} \tag{5.30}$$

The proof of Theorem 5.3.5 is thus completed. □

The following inequality is known as Minkowski's integral inequality (see, e.g., [Jentzen and Kloeden(2011)]).

Lemma 5.3.6. *Let* $p \in [1, \infty)$*, let* $(\Omega_1, \mathcal{F}_1, \mu_1)$ *and* $(\Omega_2, \mathcal{F}_2, \mu_2)$ *be finite measure spaces, and let* $f \colon \Omega_1 \times \Omega_2 \to [0, \infty)$ *be a globally bounded* $(\mathcal{F}_1 \otimes \mathcal{F}_2)/\mathcal{B}([0, \infty))$*-measurable function. Then it holds that*

$$\left[\int_{\Omega_1} \left(\int_{\Omega_2} f(x, y) \, \mu_2(dy) \right)^p \mu_1(dx) \right]^{\frac{1}{p}} \leq \int_{\Omega_2} \left(\int_{\Omega_1} |f(x, y)|^p \, \mu_1(dx) \right)^{\frac{1}{p}} \mu_2(dy). \tag{5.31}$$

Proof of Lemma 5.3.6. First, observe that (5.31) follows from Tonelli's theorem in the case $p = 1$. Throughout the rest of this proof assume w.l.o.g. that $p > 1$ and let $q \in (1, \infty)$ be the real number which satisfies that

$$\tfrac{1}{p} + \tfrac{1}{q} = 1. \tag{5.32}$$

Note that Fubini's theorem ensures that

$$\begin{aligned}
&\int_{\Omega_1}\left(\int_{\Omega_2} f(x,y)\,\mu_2(dy)\right)^p \mu_1(dx)\\
&= \int_{\Omega_1}\left(\int_{\Omega_2} f(x,y)\,\mu_2(dy)\right)^{(p-1)}\left(\int_{\Omega_2} f(x,y)\,\mu_2(dy)\right)\mu_1(dx)\\
&= \int_{\Omega_1}\left(\int_{\Omega_2} f(x,y)\,\mu_2(dy)\right)^{(p-1)}\left(\int_{\Omega_2} f(x,u)\,\mu_2(du)\right)\mu_1(dx) \qquad (5.33)\\
&= \int_{\Omega_1}\int_{\Omega_2}\left(\int_{\Omega_2} f(x,y)\,\mu_2(dy)\right)^{(p-1)} f(x,u)\,\mu_2(du)\,\mu_1(dx)\\
&= \int_{\Omega_2}\int_{\Omega_1}\left(\int_{\Omega_2} f(x,y)\,\mu_2(dy)\right)^{(p-1)} f(x,u)\,\mu_1(dx)\,\mu_2(du).
\end{aligned}$$

Hölder's inequality therefore proves that

$$\begin{aligned}
&\int_{\Omega_1}\left(\int_{\Omega_2} f(x,y)\,\mu_2(dy)\right)^p \mu_1(dx)\\
&\leq \int_{\Omega_2}\left[\int_{\Omega_1}\left(\int_{\Omega_2} f(x,y)\,\mu_2(dy)\right)^{(p-1)q}\mu_1(dx)\right]^{\frac{1}{q}}\cdot\left[\int_{\Omega_1}|f(x,u)|^p\,\mu_1(dx)\right]^{\frac{1}{p}}\mu_2(du)\\
&= \left[\int_{\Omega_1}\left(\int_{\Omega_2} f(x,y)\,\mu_2(dy)\right)^p \mu_1(dx)\right]^{(1-\frac{1}{p})}\int_{\Omega_2}\left[\int_{\Omega_1}|f(x,u)|^p\,\mu_1(dx)\right]^{\frac{1}{p}}\mu_2(du).
\end{aligned} \tag{5.34}$$

Next observe that the assumption that f is globally bounded implies that

$$\left[\int_{\Omega_1}\left(\int_{\Omega_2} f(x,y)\,\mu_2(dy)\right)^p \mu_1(dx)\right]^{(1-\frac{1}{p})} < \infty. \tag{5.35}$$

This and (5.34) establish that

$$\left[\int_{\Omega_1}\left(\int_{\Omega_2} f(x,y)\,\mu_2(dy)\right)^p \mu_1(dx)\right]^{\frac{1}{p}} \leq \int_{\Omega_2}\left(\int_{\Omega_1}|f(x,y)|^p\,\mu_1(dx)\right)^{\frac{1}{p}}\mu_2(dy). \tag{5.36}$$

The proof of Lemma 5.3.6 is thus completed. □

Lemma 5.3.7. *Let $p \in [1,\infty)$, let $(\Omega_1, \mathcal{F}_1, \mu_1)$ and $(\Omega_2, \mathcal{F}_2, \mu_2)$ be finite measure spaces, and let $f\colon \Omega_1 \times \Omega_2 \to [0,\infty]$ be an $(\mathcal{F}_1 \otimes \mathcal{F}_2)/\mathcal{B}([0,\infty])$-measurable function. Then it holds that*

$$\left[\int_{\Omega_1}\left(\int_{\Omega_2} f(x,y)\,\mu_2(dy)\right)^p \mu_1(dx)\right]^{\frac{1}{p}} \leq \int_{\Omega_2}\left(\int_{\Omega_1}|f(x,y)|^p\,\mu_1(dx)\right)^{\frac{1}{p}}\mu_2(dy). \tag{5.37}$$

Proof of Lemma 5.3.7. Throughout this proof let $f_N\colon \Omega_1 \times \Omega_2 \to [0,\infty)$, $N \in \mathbb{N}$, be the functions which satisfy for all $N \in \mathbb{N}$, $x \in \Omega_1$, $y \in \Omega_2$ that

$$f_N(x,y) = \min\{f(x,y), N\}. \tag{5.38}$$

Note that for all $N \in \mathbb{N}$ it holds that f_N is globally bounded and $(\mathcal{F}_1 \otimes \mathcal{F}_2)/\mathcal{B}([0,\infty))$-measurable. Lemma 5.3.6 hence establishes that for all $N \in \mathbb{N}$ it holds that

$$\left[\int_{\Omega_1}\left(\int_{\Omega_2} f_N(x,y)\,\mu_2(dy)\right)^p \mu_1(dx)\right]^{\frac{1}{p}} \leq \int_{\Omega_2}\left(\int_{\Omega_1} |f_N(x,y)|^p\,\mu_1(dx)\right)^{\frac{1}{p}} \mu_2(dy). \tag{5.39}$$

This and monotone convergence theorem for Lebesgue integral yield that

$$\begin{aligned}
&\left[\int_{\Omega_1}\left(\int_{\Omega_2} f(x,y)\,\mu_2(dy)\right)^p \mu_1(dx)\right]^{\frac{1}{p}} \\
&= \left[\int_{\Omega_1}\left(\int_{\Omega_2} \lim_{N\to\infty} f_N(x,y)\,\mu_2(dy)\right)^p \mu_1(dx)\right]^{\frac{1}{p}} \\
&= \lim_{N\to\infty}\left[\int_{\Omega_1}\left(\int_{\Omega_2} f_N(x,y)\,\mu_2(dy)\right)^p \mu_1(dx)\right]^{\frac{1}{p}} \\
&\leq \lim_{N\to\infty}\int_{\Omega_2}\left(\int_{\Omega_1} |f_N(x,y)|^p\,\mu_1(dx)\right)^{\frac{1}{p}} \mu_2(dy) \\
&= \int_{\Omega_2}\left(\int_{\Omega_1} \lim_{N\to\infty}|f_N(x,y)|^p\,\mu_1(dx)\right)^{\frac{1}{p}} \mu_2(dy) \\
&= \int_{\Omega_2}\left(\int_{\Omega_1} |f(x,y)|^p\,\mu_1(dx)\right)^{\frac{1}{p}} \mu_2(dy).
\end{aligned} \tag{5.40}$$

The proof of Lemma 5.3.7 is thus completed. □

Proposition 5.3.8. *Let $p \in [1,\infty)$, let $(\Omega_1,\mathcal{F}_1,\mu_1)$ and $(\Omega_2,\mathcal{F}_2,\mu_2)$ be sigma-finite measure spaces, and let $f\colon \Omega_1 \times \Omega_2 \to [0,\infty]$ be an $(\mathcal{F}_1 \otimes \mathcal{F}_2)/\mathcal{B}([0,\infty])$-measurable function. Then it holds that*

$$\left[\int_{\Omega_1}\left(\int_{\Omega_2} f(x,y)\,\mu_2(dy)\right)^p \mu_1(dx)\right]^{\frac{1}{p}} \leq \int_{\Omega_2}\left(\int_{\Omega_1} |f(x,y)|^p\,\mu_1(dx)\right)^{\frac{1}{p}} \mu_2(dy). \tag{5.41}$$

Proof of Proposition 5.3.8. Throughout this proof let $\Omega_i^{(n)} \in \mathcal{F}_i$, $n \in \mathbb{N}$, $i \in \{1,2\}$, be sets which satisfy for all $n \in \mathbb{N}$, $i \in \{1,2\}$ that

$$\mu_i(\Omega_i^{(n)}) < \infty, \quad \Omega_i^{(n)} \subseteq \Omega_i^{(n+1)}, \quad \text{and} \quad \bigcup_{k\in\mathbb{N}} \Omega_i^{(k)} = \Omega_i, \tag{5.42}$$

let $\mu_i^{(n)} \colon \Omega_i^{(n)} \Cap \mathcal{F}_i \to [0,\infty]$, $n \in \mathbb{N}$, $i \in \{1,2\}$, be the measures which satisfy for all $n \in \mathbb{N}$, $i \in \{1,2\}$ that

$$\mu_i^{(n)} = \mu|_{\Omega_i^{(n)} \Cap \mathcal{F}_i}, \tag{5.43}$$

and let $f_n \colon \Omega_1^{(n)} \times \Omega_2^{(n)} \to [0,\infty]$, $n \in \mathbb{N}$, be the functions which satisfy for all $n \in \mathbb{N}$ that

$$f_n = f|_{\Omega_1^{(n)} \times \Omega_2^{(n)}}. \tag{5.44}$$

Observe that (5.42) and (5.43) ensure that for all $n \in \mathbb{N}$ it holds that $\big(\Omega_1^{(n)}, \Omega_1^{(n)} \Cap \mathcal{F}_1, \mu_1^{(n)}\big)$ and $\big(\Omega_2^{(n)}, \Omega_2^{(n)} \Cap \mathcal{F}_2, \mu_2^{(n)}\big)$ are finite measure spaces. The fact that for all $n \in \mathbb{N}$ it holds that f_n is $((\Omega_1^{(n)} \Cap \mathcal{F}_1) \otimes (\Omega_2^{(n)} \Cap \mathcal{F}_2))/\mathcal{B}([0,\infty])$-measurable hence allows us to apply Lemma 5.3.7 to obtain that for all $n \in \mathbb{N}$ it holds that

$$\left[\int_{\Omega_1^{(n)}} \left(\int_{\Omega_2^{(n)}} f_n(x,y)\, \mu_2^{(n)}(dy)\right)^p \mu_1^{(n)}(dx)\right]^{\frac{1}{p}} \leq \int_{\Omega_2^{(n)}} \left(\int_{\Omega_1^{(n)}} |f_n(x,y)|^p\, \mu_1^{(n)}(dx)\right)^{\frac{1}{p}} \mu_2^{(n)}(dy). \tag{5.45}$$

This implies that for all $n \in \mathbb{N}$ it holds that

$$\begin{aligned}
&\left[\int_{\Omega_1} \left(\int_{\Omega_2} f(x,y) \mathbb{1}_{\{\Omega_1^{(n)} \times \Omega_2^{(n)}\}}(x,y)\, \mu_2(dy)\right)^p \mu_1(dx)\right]^{\frac{1}{p}} \\
&= \left[\int_{\Omega_1^{(n)}} \left(\int_{\Omega_2^{(n)}} f_n(x,y)\, \mu_2^{(n)}(dy)\right)^p \mu_1^{(n)}(dx)\right]^{\frac{1}{p}} \\
&\leq \int_{\Omega_2^{(n)}} \left(\int_{\Omega_1^{(n)}} |f_n(x,y)|^p\, \mu_1^{(n)}(dx)\right)^{\frac{1}{p}} \mu_2^{(n)}(dy) \\
&= \int_{\Omega_2} \left(\int_{\Omega_1} \left|f(x,y) \mathbb{1}_{\{\Omega_1^{(n)} \times \Omega_2^{(n)}\}}(x,y)\right|^p \mu_1(dx)\right)^{\frac{1}{p}} \mu_2(dy).
\end{aligned} \tag{5.46}$$

Monotone convergence theorem for Lebesgue integral therefore assures that

$$\begin{aligned}
&\left[\int_{\Omega_1}\left(\int_{\Omega_2} f(x,y)\,\mu_2(dy)\right)^p \mu_1(dx)\right]^{\frac{1}{p}} \\
&= \left[\int_{\Omega_1}\left(\int_{\Omega_2} \lim_{n\to\infty}\left[f(x,y)\mathbb{1}_{\{\Omega_1^{(n)}\times\Omega_2^{(n)}\}}(x,y)\right]\mu_2(dy)\right)^p \mu_1(dx)\right]^{\frac{1}{p}} \\
&= \lim_{n\to\infty}\left[\int_{\Omega_1}\left(\int_{\Omega_2} f(x,y)\mathbb{1}_{\{\Omega_1^{(n)}\times\Omega_2^{(n)}\}}(x,y)\,\mu_2(dy)\right)^p \mu_1(dx)\right]^{\frac{1}{p}} \\
&\leq \lim_{n\to\infty}\int_{\Omega_2}\left(\int_{\Omega_1}\left|f(x,y)\mathbb{1}_{\{\Omega_1^{(n)}\times\Omega_2^{(n)}\}}(x,y)\right|^p \mu_1(dx)\right)^{\frac{1}{p}} \mu_2(dy) \\
&= \int_{\Omega_2}\left(\int_{\Omega_1}\lim_{n\to\infty}\left|f(x,y)\mathbb{1}_{\{\Omega_1^{(n)}\times\Omega_2^{(n)}\}}(x,y)\right|^p \mu_1(dx)\right)^{\frac{1}{p}} \mu_2(dy) \\
&= \int_{\Omega_2}\left(\int_{\Omega_1}|f(x,y)|^p\,\mu_1(dx)\right)^{\frac{1}{p}} \mu_2(dy).
\end{aligned} \tag{5.47}$$

The proof of Proposition 5.3.8 is thus completed. □

The next result, Corollary 5.3.9, follows immediately from Proposition 5.3.8.

Corollary 5.3.9. *Let $T \in (0,\infty)$, $p \in [1,\infty)$, let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, and let $Y\colon [0,T]\times\Omega \to [0,\infty]$ be a $(\mathcal{B}([0,T])\otimes\mathcal{F})/\mathcal{B}([0,\infty])$-measurable function. Then it holds that*

$$\left(\mathbb{E}\left[\left|\int_0^T Y_s\,ds\right|^p\right]\right)^{\frac{1}{p}} \leq \int_0^T \left(\mathbb{E}\big[\,|Y_s|^p\,\big]\right)^{\frac{1}{p}} ds. \tag{5.48}$$

Theorem 5.3.10 (Strong convergence of the Euler-Maruyama method)**.** *Let $T \in (0,\infty)$, L_μ, $L_\sigma \in [0,\infty)$, $p \in [2,\infty)$, $d, m, N \in \mathbb{N}$, let $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$ be a stochastic basis, let $W\colon [0,T] \times \Omega \to \mathbb{R}^m$ be an m-dimensional standard $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$-Brownian motion, let $\xi \in \mathcal{L}^p(P|_{\mathbb{F}_0}; \|\cdot\|_{\mathbb{R}^d})$, let $\mu\colon \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d\times m}$ be functions which satisfy for all $x, y \in \mathbb{R}^d$ that*

$$\|\mu(x) - \mu(y)\|_{\mathbb{R}^d} \leq L_\mu \|x-y\|_{\mathbb{R}^d}, \quad \|\sigma(x) - \sigma(y)\|_{HS(\mathbb{R}^m,\mathbb{R}^d)} \leq L_\sigma \|x-y\|_{\mathbb{R}^d}, \tag{5.49}$$

let $X\colon [0,T] \times \Omega \to \mathbb{R}^d$ be a solution process of the SDE

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t, \qquad t \in [0,T], \qquad X_0 = \xi, \tag{5.50}$$

and let $\bar{Y}\colon [0,T] \times \Omega \to \mathbb{R}^d$ be a linearly interpolated Euler-Maruyama approximation for the SDE (5.50) *with time step size ${}^T\!/\!{}_N$. Then*

$$\begin{aligned} &\sup_{t\in[0,T]} \left(\mathbb{E}_P\big[\|X_t - \bar{Y}_t\|^p_{\mathbb{R}^d}\big]\right)^{1/p} = \sup_{t\in[0,T]} \|X_t - \bar{Y}_t\|_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})} \\ &\leq \underbrace{\left[\exp\!\left(\sqrt{T} + \left(T + T^2\right)\left[L_\mu\sqrt{T} + L_\sigma\sqrt{\tfrac{p\,(p-1)}{2}}\right]^2\right)\|X\|_{\mathcal{C}^{1/2}([0,T],L^p(P;\|\cdot\|_{\mathbb{R}^d}))}\right]}_{<\infty} \frac{1}{\sqrt{N}}. \end{aligned} \tag{5.51}$$

Proof of Theorem 5.3.10. Throughout this proof let $\bar{X}\colon [0,T]\times\Omega \to \mathbb{R}^d$ be the function which satisfies for all $n \in \{0,1,\ldots,N-1\}$, $t \in \left[\frac{nT}{N}, \frac{(n+1)T}{N}\right]$ that

$$\bar{X}_t = \left(\tfrac{tN}{T} - n\right) X_{(n+1)T/N} + \left(n + 1 - \tfrac{tN}{T}\right) X_{nT/N}. \tag{5.52}$$

Next observe that for all $t \in \{0, \frac{T}{N}, \ldots, T\}$ it holds P-a.s. that

$$\bar{Y}_t = \xi + \int_0^t \mu\big(\bar{Y}_{\lfloor s\rfloor_{T/N}}\big)\,ds + \int_0^t \sigma\big(\bar{Y}_{\lfloor s\rfloor_{T/N}}\big)\,dW_s. \tag{5.53}$$

This implies that for all $t \in \{0, \frac{T}{N}, \ldots, T\}$ it holds P-a.s. that

$$\begin{aligned} X_t - \bar{Y}_t &= \int_0^t \mu\big(X_s\big) - \mu\big(\bar{Y}_{\lfloor s\rfloor_{T/N}}\big)\,ds + \int_0^t \sigma\big(X_s\big) - \sigma\big(\bar{Y}_{\lfloor s\rfloor_{T/N}}\big)\,dW_s \\ &= \int_0^t \mu\big(X_s\big) - \mu\big(X_{\lfloor s\rfloor_{T/N}}\big)\,ds + \int_0^t \sigma\big(X_s\big) - \sigma\big(X_{\lfloor s\rfloor_{T/N}}\big)\,dW_s \\ &+ \int_0^t \mu\big(X_{\lfloor s\rfloor_{T/N}}\big) - \mu\big(\bar{Y}_{\lfloor s\rfloor_{T/N}}\big)\,ds + \int_0^t \sigma\big(X_{\lfloor s\rfloor_{T/N}}\big) - \sigma\big(\bar{Y}_{\lfloor s\rfloor_{T/N}}\big)\,dW_s. \end{aligned} \tag{5.54}$$

The triangle inequality hence proves that for all $t \in \{0, \frac{T}{N}, \dots, T\}$ it holds that

$$\begin{aligned}
&\|X_t - \bar{Y}_t\|_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})} \\
&\leq \int_0^t \big\|\mu\big(X_s\big) - \mu\big(X_{\lfloor s \rfloor_{T/N}}\big)\big\|_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})}\, ds \\
&+ \left\| \int_0^t \sigma\big(X_s\big) - \sigma\big(X_{\lfloor s \rfloor_{T/N}}\big)\, dW_s \right\|_{L^p(P;\|\cdot\|_{\mathbb{R}^d})} \\
&+ \int_0^t \big\|\mu\big(X_{\lfloor s \rfloor_{T/N}}\big) - \mu\big(\bar{Y}^N_{\lfloor s \rfloor_{T/N}}\big)\big\|_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})}\, ds \\
&+ \left\| \int_0^t \sigma\big(X_{\lfloor s \rfloor_{T/N}}\big) - \sigma\big(\bar{Y}^N_{\lfloor s \rfloor_{T/N}}\big)\, dW_s \right\|_{L^p(P;\|\cdot\|_{\mathbb{R}^d})}.
\end{aligned} \tag{5.55}$$

The Hölder inequality and inequality (Burkholder-Davis-Gundy inequality I) therefore imply that for all $t \in \{0, \frac{T}{N}, \dots, T\}$ it holds that

$$\begin{aligned}
&\big\|X_t - \bar{Y}_t\big\|_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})} \\
&\leq \left[T \int_0^t \big\|\mu\big(X_s\big) - \mu\big(X_{\lfloor s \rfloor_{T/N}}\big)\big\|^2_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})}\, ds \right]^{1/2} \\
&+ \left[\tfrac{p\,(p-1)}{2} \int_0^t \big\|\sigma\big(X_s\big) - \sigma\big(X_{\lfloor s \rfloor_{T/N}}\big)\big\|^2_{\mathcal{L}^p(P;\|\cdot\|_{HS(\mathbb{R}^m,\mathbb{R}^d)})}\, ds \right]^{1/2} \\
&+ \left[T \int_0^t \big\|\mu\big(X_{\lfloor s \rfloor_{T/N}}\big) - \mu\big(\bar{Y}_{\lfloor s \rfloor_{T/N}}\big)\big\|^2_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})}\, ds \right]^{1/2} \\
&+ \left[\tfrac{p\,(p-1)}{2} \int_0^t \big\|\sigma\big(X_{\lfloor s \rfloor_{T/N}}\big) - \sigma\big(\bar{Y}_{\lfloor s \rfloor_{T/N}}\big)\big\|^2_{\mathcal{L}^p(P;\|\cdot\|_{HS(\mathbb{R}^m,\mathbb{R}^d)})}\, ds \right]^{1/2}.
\end{aligned} \tag{5.56}$$

Assumption (5.49) hence shows that for all $t \in \{0, \frac{T}{N}, \dots, T\}$ it holds that

$$\begin{aligned}
&\big\|X_t - \bar{Y}_t\big\|_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})} \\
&\leq L_\mu \left[T \int_0^t \big\|X_s - X_{\lfloor s \rfloor_{T/N}}\big\|^2_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})}\, ds \right]^{1/2} \\
&+ L_\sigma \left[\tfrac{p\,(p-1)}{2} \int_0^t \big\|X_s - X_{\lfloor s \rfloor_{T/N}}\big\|^2_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})}\, ds \right]^{1/2} \\
&+ L_\mu \left[T \int_0^t \big\|X_{\lfloor s \rfloor_{T/N}} - \bar{Y}_{\lfloor s \rfloor_{T/N}}\big\|^2_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})}\, ds \right]^{1/2} \\
&+ L_\sigma \left[\tfrac{p\,(p-1)}{2} \int_0^t \big\|X_{\lfloor s \rfloor_{T/N}} - \bar{Y}_{\lfloor s \rfloor_{T/N}}\big\|^2_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})}\, ds \right]^{1/2}.
\end{aligned} \tag{5.57}$$

This implies that for all $t \in \{0, \frac{T}{N}, \dots, T\}$ it holds that

$$\begin{aligned} &\|X_t - \bar{Y}_t\|_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})} \\ &\leq \left[L_\mu \sqrt{T} + L_\sigma \sqrt{\tfrac{p\,(p-1)}{2}}\right] \left[\int_0^T \left\|X_s - X_{\lfloor s \rfloor_{T/N}}\right\|^2_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})} ds\right]^{1/2} \\ &+ \left[L_\mu \sqrt{T} + L_\sigma \sqrt{\tfrac{p\,(p-1)}{2}}\right] \left[\int_0^t \left\|X_{\lfloor s \rfloor_{T/N}} - \bar{Y}_{\lfloor s \rfloor_{T/N}}\right\|^2_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})} ds\right]^{1/2}. \end{aligned} \tag{5.58}$$

The well-known fact that for all $n \in \mathbb{N}$, $a_1, \dots, a_n \in \mathbb{R}$ it holds that

$$(a_1 + \ldots + a_n)^2 \leq n\left((a_1)^2 + \ldots + (a_n)^2\right) \tag{5.59}$$

hence proves that for all $t \in \{0, \frac{T}{N}, \dots, T\}$ it holds that

$$\begin{aligned} &\|X_t - \bar{Y}_t\|^2_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})} \\ &\leq 2\left[L_\mu \sqrt{T} + L_\sigma \sqrt{\tfrac{p\,(p-1)}{2}}\right]^2 \int_0^T \left\|X_s - X_{\lfloor s \rfloor_{T/N}}\right\|^2_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})} ds \\ &+ 2\left[L_\mu \sqrt{T} + L_\sigma \sqrt{\tfrac{p\,(p-1)}{2}}\right]^2 \int_0^t \left\|X_{\lfloor s \rfloor_{T/N}} - \bar{Y}_{\lfloor s \rfloor_{T/N}}\right\|^2_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})} ds. \end{aligned} \tag{5.60}$$

The discrete Gronwall lemma (see Lemma 4.3.2) therefore shows that

$$\begin{aligned} &\sup_{t \in \{0, T/N, \dots, T\}} \|X_t - \bar{Y}_t\|^2_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})} \leq \exp\!\left(2\,T\left[L_\mu \sqrt{T} + L_\sigma \sqrt{\tfrac{p\,(p-1)}{2}}\right]^2\right) \\ &\quad \cdot 2\left[L_\mu \sqrt{T} + L_\sigma \sqrt{\tfrac{p\,(p-1)}{2}}\right]^2 \left[\int_0^T \left\|X_s - X_{\lfloor s \rfloor_{T/N}}\right\|^2_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})} ds\right]. \end{aligned} \tag{5.61}$$

This implies that

$$\begin{aligned} &\sup_{t \in \{0, T/N, \dots, T\}} \|X_t - \bar{Y}_t\|_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})} \leq \exp\!\left(T\left[L_\mu \sqrt{T} + L_\sigma \sqrt{\tfrac{p\,(p-1)}{2}}\right]^2\right) \\ &\quad \cdot \sqrt{2}\left[L_\mu \sqrt{T} + L_\sigma \sqrt{\tfrac{p\,(p-1)}{2}}\right] \left[\int_0^T \left\|X_s - X_{\lfloor s \rfloor_{T/N}}\right\|^2_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})} ds\right]^{1/2}. \end{aligned} \tag{5.62}$$

This shows that

$$\begin{aligned} &\sup_{t \in \{0, T/N, \dots, T\}} \|X_t - \bar{Y}_t\|_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})} \leq \exp\!\left(T\left[L_\mu \sqrt{T} + L_\sigma \sqrt{\tfrac{p\,(p-1)}{2}}\right]^2\right) \\ &\quad \cdot \sqrt{2}\left[L_\mu \sqrt{T} + L_\sigma \sqrt{\tfrac{p\,(p-1)}{2}}\right] \|X\|_{\mathcal{C}^{1/2}([0,T], L^p(P;\|\cdot\|_{\mathbb{R}^d}))} \frac{T}{\sqrt{N}} \\ &\leq \exp\!\left(\left(T + T^2\right)\left[L_\mu \sqrt{T} + L_\sigma \sqrt{\tfrac{p\,(p-1)}{2}}\right]^2\right) \frac{\|X\|_{\mathcal{C}^{1/2}([0,T], L^p(P;\|\cdot\|_{\mathbb{R}^d}))}}{\sqrt{N}}. \end{aligned} \tag{5.63}$$

The triangle inequality therefore proves that

$$\begin{aligned}
&\sup_{t\in[0,T]} \big\|X_t - \bar{Y}_t\big\|_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})} \\
&\leq \sup_{t\in[0,T]} \big\|X_t - \bar{X}_t\big\|_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})} + \sup_{t\in[0,T]} \big\|\bar{X}_t - \bar{Y}_t\big\|_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})} \\
&\leq \|X\|_{\mathcal{C}^{1/2}([0,T],L^p(P;\|\cdot\|_{\mathbb{R}^d}))} \frac{\sqrt{T}}{\sqrt{N}} + \sup_{t\in[0,T]} \big\|\bar{X}_t - \bar{Y}_t\big\|_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})} \\
&\leq \left[\sqrt{T} + \exp\!\left(\big(T+T^2\big)\left[L_\mu\sqrt{T} + L_\sigma\sqrt{\tfrac{p\,(p-1)}{2}}\right]^2\right)\right] \frac{\|X\|_{\mathcal{C}^{1/2}([0,T],L^p(P;\|\cdot\|_{\mathbb{R}^d}))}}{\sqrt{N}} \\
&\leq \exp\!\left(\sqrt{T} + \big(T+T^2\big)\left[L_\mu\sqrt{T} + L_\sigma\sqrt{\tfrac{p\,(p-1)}{2}}\right]^2\right) \frac{\|X\|_{\mathcal{C}^{1/2}([0,T],L^p(P;\|\cdot\|_{\mathbb{R}^d}))}}{\sqrt{N}}.
\end{aligned} \tag{5.64}$$

The proof of Theorem 5.3.10 is thus completed. □

5.3.3 Uniform strong convergence of the Euler-Maruyama method

Theorem 5.3.11 (Uniform strong convergence of the Euler-Maruyama method)**.** *Let $T \in (0,\infty)$, $p \in [2,\infty)$, $d, m \in \mathbb{N}$, let $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$ be a stochastic basis, let $W\colon [0,T]\times\Omega \to \mathbb{R}^m$ be an m-dimensional standard $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$-Brownian motion, let $\xi \in \mathcal{L}^p(P|_{\mathbb{F}_0}; \|\cdot\|_{\mathbb{R}^d})$, let $\mu\colon \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d\times m}$ be globally Lipschitz continuous functions, let $X\colon [0,T]\times\Omega \to \mathbb{R}^d$ be a solution process of the SDE*

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t, \qquad t \in [0,T], \qquad X_0 = \xi, \tag{5.65}$$

and for every $N \in \mathbb{N}$ let $\bar{Y}^N\colon [0,T]\times\Omega \to \mathbb{R}^d$ be a linearly interpolated Euler-Maruyama approximation for the SDE (5.65) *with time step size T/N. Then there exists a real number $C \in \mathbb{R}$ such that for all $N \in \mathbb{N}$ it holds that*

$$\left(\mathbb{E}_P\left[\sup_{t\in[0,T]} \|X_t - \bar{Y}^N_t\|^p_{\mathbb{R}^d}\right]\right)^{\frac{1}{p}} = \left\|\sup_{t\in[0,T]} \big\|X_t - \bar{Y}^N_t\big\|_{\mathbb{R}^d}\right\|_{\mathcal{L}^p(P;|\cdot|_{\mathbb{R}})} \leq \frac{C\sqrt{1+\ln(N)}}{\sqrt{N}}. \tag{5.66}$$

Remark 5.3.12. *The proof of Theorem 5.3.11 can be performed similarly as Theorem 5.3.10 but uses inequality* (Burkholder-Davis-Gundy inequality II) *instead of inequality* (Burkholder-Davis-Gundy inequality I) *and also additionally expoits an argument from [Müller-Gronbach(2002)].*

5.4 Strong L^p-convergence with order $\alpha > 1/p$ implies almost sure convergence

In this section a relation between strong L^p-convergence for $p \in (0, \infty)$ and almost sure convergence is presented. To be more precise, Lemma 5.4.1 below shows for every $p, \beta \in (0, \infty)$ with $\beta > 1/p$ that *strong L^p-convergence with order β* implies for every arbitrarily small $\varepsilon \in (0, \beta - 1/p)$ *almost sure convergence with order $\beta - 1/p - \varepsilon$*

Lemma 5.4.1 and its proof are slightly modified versions of Lemma 3.21 and its proof in [Hutzenthaler and Jentzen(2012)] respectively. Lemma 3.21 in [Hutzenthaler and Jentzen(2012)] is a slight generalization of Lemma 2.1 in [Kloeden and Neuenkirch(2007)]. In particular, the last statement in Lemma 5.4.1 (see (5.69)) is precisely the statement of Lemma 2.1 in [Kloeden and Neuenkirch(2007)].

Lemma 5.4.1 (L^p-convergence with order $\alpha \in (1/p, \infty)$ implies almost sure convergence)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $Y_N \colon \Omega \to \mathbb{R}$, $N \in \mathbb{N}$, be $\mathcal{F}/\mathcal{B}(\mathbb{R})$-measurable functions. Then*

(i) it holds for all $p \in (0, \infty)$, $\alpha \in (1/p, \infty)$, $\beta \in (0, \alpha - 1/p)$ that

$$\Big\| \sup_{N \in \mathbb{N}} \big(N^\beta \cdot |Y_N| \big) \Big\|_{\mathcal{L}^p(P; |\cdot|_{\mathbb{R}})} \leq \underbrace{\left[\sum_{N=1}^{\infty} N^{(\beta - \alpha)p} \right]^{1/p}}_{< \infty} \left[\sup_{N \in \mathbb{N}} \big(N^\alpha \cdot \|Y_N\|_{\mathcal{L}^p(P; |\cdot|_{\mathbb{R}})} \big) \right], \tag{5.67}$$

(ii) it holds for all $p \in (0, \infty)$, $\alpha \in (1/p, \infty)$, $\beta \in (0, \alpha - 1/p)$ with $\sup_{N \in \mathbb{N}} (N^\alpha \, \|Y_N\|_{\mathcal{L}^p(P; |\cdot|_{\mathbb{R}})}) < \infty$ that

$$P\Big(\sup_{N \in \mathbb{N}} \big(N^\beta \cdot |Y_N| \big) < \infty \Big) = 1, \tag{5.68}$$

and

(iii) it holds for all $\alpha \in (0, \infty)$, $\beta \in (0, \alpha)$ with $\forall\, p \in (0, \infty) \colon \sup_{N \in \mathbb{N}} (N^\alpha \, \|Y_N\|_{\mathcal{L}^p(P; |\cdot|_{\mathbb{R}})}) < \infty$ that

$$P\Big(\sup_{N \in \mathbb{N}} \big(N^\beta \cdot |Y_N| \big) < \infty \Big) = 1. \tag{5.69}$$

Proof of Lemma 5.4.1. Note that for all $p \in (0, \infty)$, $\alpha \in \mathbb{R}$, $\beta \in (-\infty, \alpha - 1/p)$ it holds

that

$$\begin{aligned}
&\mathbb{E}_P\left[\left\{\sup_{N\in\mathbb{N}}\left(N^{\beta}\cdot|Y_N|\right)\right\}^p\right] = \mathbb{E}_P\left[\sup_{N\in\mathbb{N}}\left(N^{\beta p}\cdot|Y_N|^p\right)\right]\\
&\leq \mathbb{E}_P\left[\sum_{N=1}^{\infty}\left(N^{\beta p}\cdot|Y_N|^p\right)\right] = \sum_{N=1}^{\infty}\left(N^{\beta p}\cdot\mathbb{E}_P\big[\,|Y_N|^p\,\big]\right)\\
&\leq \left(\sum_{N=1}^{\infty}N^{(\beta-\alpha)p}\right)\left(\sup_{N\in\mathbb{N}}N^{\alpha p}\cdot\mathbb{E}_P\big[\,|Y_N|^p\,\big]\right).
\end{aligned} \tag{5.70}$$

This proves inequality (5.67). The assertions in (5.68) and (5.69) follow immediately from inequality (5.67). The proof of Lemma 5.4.1 is thus completed. □

5.4.1 Almost sure convergence of the Euler-Maruyama method

Combining Lemma 5.4.1 with Theorem 5.3.11 results in the following corollary.

Corollary 5.4.2 (Almost sure convergence of the Euler-Maruyama scheme)**.** *Let $T\in(0,\infty)$, $d,m\in\mathbb{N}$, let $(\Omega,\mathcal{F},P,(\mathbb{F}_t)_{t\in[0,T]})$ be a stochastic basis, let $W\colon[0,T]\times\Omega\to\mathbb{R}^m$ be an m-dimensional standard $(\Omega,\mathcal{F},P,(\mathbb{F}_t)_{t\in[0,T]})$-Brownian motion, let $\xi\in\cap_{p\in(0,\infty)}\mathcal{L}^p(P|_{\mathbb{F}_0};\|\cdot\|_{\mathbb{R}^d})$, let $\mu\colon\mathbb{R}^d\to\mathbb{R}^d$ and $\sigma\colon\mathbb{R}^d\to\mathbb{R}^{d\times m}$ be globally Lipschitz continuous functions, let $X\colon[0,T]\times\Omega\to\mathbb{R}^d$ be a solution process of the SDE*

$$dX_t=\mu(X_t)\,dt+\sigma(X_t)\,dW_t,\qquad t\in[0,T],\qquad X_0=\xi \tag{5.71}$$

on $(\Omega,\mathcal{F},P,(\mathbb{F}_t)_{t\in[0,T]})$, and for every $N\in\mathbb{N}$ let $\bar{Y}^N\colon[0,T]\times\Omega\to\mathbb{R}^d$ be a linearly interpolated Euler-Maruyama approximation for the SDE (5.71) with time step size T/N. Then for every $\varepsilon\in(0,1/2)$ there exists an $\mathcal{F}/\mathcal{B}([0,\infty))$-measurable function $C\colon\Omega\to[0,\infty)$ such that for all $N\in\mathbb{N}$ it holds that

$$P\left(\sup_{t\in[0,T]}\big\|X_t-\bar{Y}^N_t\big\|_{\mathbb{R}^d}\leq C\cdot N^{\left(\varepsilon-\frac{1}{2}\right)}\right)=1. \tag{5.72}$$

5.5 Numerical methods for SDEs with non-globally Lipschitz continuous coefficient functions

5.5.1 Almost sure convergence of the Euler-Maryuama method revisited

The assumptions of Corollary 5.4.2 can be significantly relaxed. In particular, the coefficient functions μ and σ of the SDE (5.2) do not need to be globally Lipschitz continuous.

This is subject of the next theorem which is a slightly modified version of Theorem 2.4 in [Gyöngy(1998)].

Theorem 5.5.1 (Almost sure convergence of the Euler-Maruyama scheme)**.** *Assume the setting in Section 5.1 and for every $N \in \mathbb{N}$ let $\bar{Y}^N \colon [0,T] \times \Omega \to \mathbb{R}^d$ be a linearly interpolated Euler-Maruyama approximation for the SDE*

$$dX_t = \bar{\mu}(X_t)\,dt + \bar{\sigma}(X_t)\,dW_t, \qquad t \in [0,T], \qquad X_0 = \xi \tag{5.73}$$

with time step size T/N on $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$. Then for every $\varepsilon \in (0, 1/2)$ there exist an $\mathcal{F}/\mathcal{B}([0,\infty))$-measurable function $C \colon \Omega \to [0,\infty)$ such that for all $N \in \mathbb{N}$ it holds that

$$P\left(\sup_{t\in[0,T]} \left\|X_t - \bar{Y}_t^N\right\|_{\mathbb{R}^d} \le C \cdot N^{\left(\varepsilon - \frac{1}{2}\right)}\right) = 1. \tag{5.74}$$

5.5.2 Strong and numerically weak divergence of the Euler-Maruyama scheme

Theorems 5.3.10 and 5.3.11 can not be generalized to the case where μ and σ are merely locally Lipschitz continuous. This is the subject of Theorem 5.5.2 below. Theorem 5.5.2 is a special case of Theorem 2.1 in [Hutzenthaler et al.(2011b)Hutzenthaler, Jentzen, and Kloeden] (cf. also Theorem 2.1 in [Hutzenthaler et al.(2011a)Hutzenthaler, Jentzen, and Kloeden]).

Theorem 5.5.2 (Strong and weak divergence of the Euler method for SDEs with superlinearly growing coefficients)**.** *Let $T \in (0,\infty)$, let $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$ be a stochastic basis, let $\xi \in \mathcal{M}(\mathbb{F}_0, \mathcal{B}(\mathbb{R}))$, $\mu, \sigma \in \mathcal{M}(\mathcal{B}(\mathbb{R}), \mathcal{B}(\mathbb{R}))$ satisfy $P\big(\sigma(\xi) \neq 0\big) > 0$, let $W \colon [0,T] \times \Omega \to \mathbb{R}$ be a one-dimensional standard $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$-Brownian motion, let $X \colon [0,T] \times \Omega \to \mathbb{R}$ be a solution process of the SDE*

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t, \qquad t \in [0,T], \qquad X_0 = \xi \tag{5.75}$$

on $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$, for every $N \in \mathbb{N}$ let $\bar{Y}^N \colon [0,T] \times \Omega \to \mathbb{R}$ a linearly interpolated Euler-Maruyama approximation for the SDE (5.75) *with time step size T/N, and let $p, \varepsilon \in (0,\infty)$ satisfy $\mathbb{E}_P\big[|X_T|^p\big] < \infty$ and $\forall\, x \in (-\infty, -1/\varepsilon] \cup [1/\varepsilon, \infty)$:*

$$|\mu(x)| + |\sigma(x)| \ge \varepsilon\, |x|^{(1+\varepsilon)}. \tag{5.76}$$

Then it holds for all $q \in (0,p]$ that

$$\lim_{N\to\infty} \mathbb{E}_P\Big[\big|X_T - \bar{Y}_T^N\big|^q\Big] = \infty \qquad \text{and} \qquad \lim_{N\to\infty} \mathbb{E}_P\Big[\big|\bar{Y}_T^N\big|^q\Big] = \infty \neq \mathbb{E}_P\Big[\big|X_T\big|^q\Big]. \tag{5.77}$$

Theorem 5.5.2 applies, for instance, to the *stochastic Ginzburg-Landau equation* in Subsection 4.7.4. Long time divergence results for Euler's method as $T \to \infty$ can be found in Mattingly et al. [Mattingly et al.(2002)Mattingly, Stuart, and Higham] and in the references mentioned therein (see also Milstein & Tretyakov [Milstein and Tretyakov(2004), Milstein and Tretyakov(2005)] for further remarks on this topic). Theorem 5.5.2 is proved as Theorem 2.1 in [Hutzenthaler et al.(2011b)Hutzenthaler, Jentzen, and Kloeden].

5.5.3 Drift-implicit Euler-Maruyama scheme

Definition 5.5.3 (Drift-implicit Euler-Maruyama approximation)**.** *Let $T \in (0,\infty)$, $d, m, N \in \mathbb{N}$, $\mu \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{R}^d))$, $\sigma \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{R}^{d\times m}))$, let $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$ be a stochastic basis, let $\xi \in \mathcal{L}^0(P|_{\mathbb{F}_0}; \|\cdot\|_{\mathbb{R}^d})$, and let $W\colon [0,T]\times\Omega \to \mathbb{R}^m$ be an m-dimensional standard $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$-Brownian motion. Then we say that Y is a drift-implicit Euler-Maruyama approximation for the SDE*

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t, \qquad t \in [0,T], \qquad X_0 = \xi \tag{5.78}$$

with time step size T/N on $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$ (we say that Y is a drift-implicit Euler-Maruyama approximation for the SDE

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t, \qquad t \in [0,T], \qquad X_0 = \xi \tag{5.79}$$

with time step size T/N) if and only if $Y \in \mathbb{M}(\{0,1,\dots,N\}\times\Omega, \mathbb{R}^d)$ is the function from $\{0,1,\dots,N\}\times\Omega$ to $\mathbb{R}^d$ which satisfies for all $n \in \{0,1,\dots,N-1\}$ that $Y_0 = \xi$ and

$$Y_{n+1} = Y_n + \mu\big(Y_{n+1}\big)\tfrac{T}{N} + \sigma\big(Y_n\big)\big(W_{\frac{(n+1)T}{N}} - W_{\frac{nT}{N}}\big). \tag{5.80}$$

Remark 5.5.4. *Sometimes drift-implicit Euler-Maruyama approximations are also referred to as* semi-implicit Euler-Maruyama approximations *or* Backward Euler approximations *in the literature; cf., e.g., [Hu(1996)] and [Higham et al.(2002)Higham, Mao, and Stuart].*

Proposition 5.5.5 (Unique existence of drift-implicit Euler-Maruyama approximations)**.** *Let $d \in \mathbb{N}$, $h \in (0,\infty)$, $L \in \mathbb{R}$ satisfy $Lh < 1$ and let $\mu\colon \mathbb{R}^d \to \mathbb{R}^d$ be a locally Lipschitz continuous function which satisfies for all $x, y \in \mathbb{R}^d$ that*

$$\langle x-y, \mu(x)-\mu(y)\rangle_{\mathbb{R}^d} \leq L\,\|x-y\|^2_{\mathbb{R}^d}\,. \tag{5.81}$$

Then it holds that the function $\mathbb{R}^d \ni x \mapsto x - \mu(x)h \in \mathbb{R}^d$ is bijective, that is, it holds for all $y \in \mathbb{R}^d$ that $\#_{\mathbb{R}^d}\big(\{x \in \mathbb{R}^d\colon x = y + \mu(x)h\}\big) = 1$.

In many situations the drift-implicit Euler approximations converge strongly to the solution process of the SDE (5.2) although the (explicit) Euler-Maruyama approximations fail to converge strongly (see Theorem 5.5.2 above). Assumptions that are sufficient to ensure that the drift-implicit Euler-Maruyama approximations converge strongly to the solution process of the SDE (5.2) can, e.g., be found in [Hu(1996)], [Higham et al.(2002)Higham, Mao, and Stuart] and [Hutzenthaler and Jentzen(2012)].

Exercise 5.5.6. *Let $\mu \in C^1(\mathbb{R}, \mathbb{R})$, $L \in \mathbb{R}$. Prove that $\sup_{x\in\mathbb{R}} \mu'(x) \leq L$ if and only if $\forall\, x, y \in \mathbb{R}$: $(x-y)\cdot(\mu(x)-\mu(y)) \leq L\,(x-y)^2$.*

5.5.4 Increment-tamed Euler-Maruyama scheme

In this subsection increment-tamed Euler-Maryuama approximations are presented; see [Hutzenthaler and Jentzen(2012)].

Definition 5.5.7 (Increment-tamed Euler-Maruyama approximation)**.** *Let $T \in (0,\infty)$, $d, m, N \in \mathbb{N}$, $\mu \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{R}^d))$, $\sigma \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{R}^{d\times m}))$, let $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$ be a stochastic basis, let $\xi \in \mathcal{L}^0(P|_{\mathbb{F}_0}; \|\cdot\|_{\mathbb{R}^d})$, and let $W\colon [0,T]\times\Omega \to \mathbb{R}^m$ be an m-dimensional standard $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$-Brownian motion. Then we say that Y is an increment-tamed Euler-Maruyama approximation for the SDE*

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t, \qquad t \in [0,T], \qquad X_0 = \xi \tag{5.82}$$

with time step size ${}^T\!/_N$ on $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$ (we say that Y is an increment-tamed Euler-Maruyama approximation for the SDE

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t, \qquad t \in [0,T], \qquad X_0 = \xi \tag{5.83}$$

with time step size ${}^T\!/_N$) if and only if $Y \in \mathbb{M}(\{0,1,\dots,N\}\times\Omega, \mathbb{R}^d)$ is the function from $\{0,1,\dots,N\}\times\Omega$ to $\mathbb{R}^d$ which satisfies for all $n \in \{0,1,\dots,N-1\}$ that $Y_0 = \xi$ and

$$Y_{n+1} = Y_n + \frac{\mu(Y_n)\frac{T}{N} + \sigma(Y_n)\,(W_{(n+1)T/N} - W_{nT/N})}{\max\big\{1, \frac{T}{N}\,\big\|\mu(Y_n)\frac{T}{N} + \sigma(Y_n)\,(W_{(n+1)T/N} - W_{nT/N})\big\|_{\mathbb{R}^d}\big\}}. \tag{5.84}$$

Remark 5.5.8 (Implementation of the increment-tamed Euler-Maruyama scheme)**.** *Let $T \in (0,\infty)$, $d, m, N \in \mathbb{N}$, let $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$ be a stochastic basis, let $\xi \in \mathcal{M}(\mathbb{F}_0, \mathcal{B}(\mathbb{R}^d))$, $\mu \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{R}^d))$, $\sigma \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{R}^{d\times m}))$, let $W\colon [0,T]\times\Omega \to \mathbb{R}^m$ be an m-dimensional standard $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$-Brownian motion, and let $Y\colon \{0,1,\dots,N\}\times\Omega \to \mathbb{R}^d$ be an increment-tamed Euler-Maruyama approximation for the SDE*

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t, \qquad t \in [0,T], \qquad X_0 = \xi \tag{5.85}$$

with time step size ${}^T\!/_N$. Then observe that for all $n \in \{0,1,\dots,N-1\}$ it holds that

$$\begin{aligned} Y_{n+1} &= Y_n \\ &+ \mathbb{1}_{\left\{\left\|\mu(Y_n)\frac{T}{N} + \sigma(Y_n)\,(W_{\frac{(n+1)T}{N}} - W_{\frac{nT}{N}})\right\|_{\mathbb{R}^d} \leq \frac{N}{T}\right\}} \left[\mu(Y_n)\tfrac{T}{N} + \sigma(Y_n)\,(W_{\frac{(n+1)T}{N}} - W_{\frac{nT}{N}})\right] \\ &+ \mathbb{1}_{\left\{\left\|\mu(Y_n)\frac{T}{N} + \sigma(Y_n)\,(W_{\frac{(n+1)T}{N}} - W_{\frac{nT}{N}})\right\|_{\mathbb{R}^d} > \frac{N}{T}\right\}} \frac{\frac{N}{T}\left[\mu(Y_n)\frac{T}{N} + \sigma(Y_n)\,(W_{\frac{(n+1)T}{N}} - W_{\frac{nT}{N}})\right]}{\left\|\mu(Y_n)\,\frac{T}{N} + \sigma(Y_n)\,(W_{\frac{(n+1)T}{N}} - W_{\frac{nT}{N}})\right\|_{\mathbb{R}^d}}. \end{aligned} \tag{5.86}$$

The scheme in Definition 5.5.7 is only one possible suggestion of an explicit scheme that overcomes the divergence behaviour of the Euler-Maruyama method in the case of SDEs with superlinearly growing coefficients (see Theorem 5.5.2 above). In particular, the norm $\|\cdot\|_{\mathbb{R}^d}$ in Definition 5.5.7 could be replaced by any other norm on the $\mathbb{R}^d$ and further "suitable tamings" are possible, cf., e.g., [Roberts and Tweedie(1996), Milstein et al.(1998)Milstein, Platen, and Schurz, Hutzenthaler et al.(2012)Hutzenthaler, Jentzen, and Kloeden, Hutzenthaler and Jentzen(2012)] and the references mentioned therein.

In many situations the increment-tamed Euler-Maruyama approximations in (5.84) converge strongly to the solution process of the SDE (5.2) although the (explicit) Euler-Maruyama approximations fail to converge strongly (see Theorem 5.5.2 above). Assumptions that are sufficient to ensure that the increment-tamed Euler-Maruyama approximations converge strongly to the solution process of the SDE (5.2) can be found in [Hutzenthaler and Jentzen(2012)].

Exercise 5.5.9 (Increment-tamed Euler-Maruyama method)**.** *In this exercise we do not distinguish between pseudo random numbers and actual random numbers. Let* $T \in (0,\infty)$, $d, m, N \in \mathbb{N}$, $\xi \in \mathbb{R}^d$, $\mu \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{R}^d))$, $\sigma \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{R}^{d\times m}))$, *let* $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$ *be a stochastic basis, let* $W\colon [0,T]\times\Omega \to \mathbb{R}^m$ *be an* m*-dimensional standard* $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$*-Brownian motion, and let* $Y\colon \{0,1,\ldots,N\}\times\Omega \to \mathbb{R}^d$ *be an increment-tamed Euler-Maruyama approximation for the SDE*

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t, \qquad t \in [0,T], \qquad X_0 = \xi \tag{5.87}$$

with time step size T/N *on* $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$ *(see Definition 5.5.7). Write a* Matlab *function* `IncrementTamed(T,d,m,N,ξ,μ,σ)` *with input* $T \in (0,\infty)$, $d, m, N \in \mathbb{N}$, $\xi \in \mathbb{R}^d$, $\mu \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{R}^d))$, $\sigma \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{R}^{d\times m}))$ *and output a realization of an* $Y_N(P)_{\mathcal{B}(\mathbb{R}^d)}$*-distributed random variable.*

5.6 Stochastic Taylor expansions and higher order numerical methods for SDEs

In Theorem 5.3.10 above it has been shown that the Euler-Maruyama method converges, under suitable assumptions, at time T in the strong L^2-sense with order $1/2$ to the solution process of the SDE under consideration. In this section we will derive other numerical methods that converge, under suitable assumptions, with a higher order than $1/2$ to the solution process of the SDE under consideration. These "higher order schemes" are based on certain stochastic Taylor expansions which will be presented first. For this we assume the setting in Section 5.1 and observe that Itô's formula implies that for all $f = \big(f(x_1,\ldots,x_d)\big)_{(x_1,\ldots,x_d)\in O} \in \cup_{k\in\mathbb{N}} C^2(O,\mathbb{R}^k)$, $t_0, t \in [0,T]$ with $t_0 \le t$ it holds P-a.s.

that

$$\begin{aligned} &f(X_t) \\ &= f(X_{t_0}) + \int_{t_0}^{t} \left[f'(X_s)\,\mu(X_s) + \frac{1}{2}\sum_{i=1}^{m} f''(X_s)\big(\sigma_i(X_s),\sigma_i(X_s)\big)\right] ds \\ &+ \sum_{i=1}^{m}\int_{t_0}^{t} f'(X_s)\,\sigma_i(X_s)\,dW_s^{(i)} \\ &= f(X_{t_0}) + \sum_{i=1}^{m}\int_{t_0}^{t}\left[\sum_{k=1}^{d}\big(\tfrac{\partial}{\partial x_k} f\big)(X_s)\cdot\sigma_{k,i}(X_s)\right] dW_s^{(i)} \\ &+ \int_{t_0}^{t}\left[\sum_{k=1}^{d}\big(\tfrac{\partial}{\partial x_k} f\big)(X_s)\cdot\mu_k(X_s) + \tfrac{1}{2}\sum_{i=1}^{m}\sum_{k,l=1}^{d}\big(\tfrac{\partial^2}{\partial x_k \partial x_l} f\big)(X_s)\cdot\sigma_{k,i}(X_s)\cdot\sigma_{l,i}(X_s)\right] ds. \end{aligned} \tag{5.88}$$

Equation (5.88) is Itô's formula in the special case where the considered Itô process is a solution process of an SDE.

5.6.1 Generator and noise operators associated to an SDE

In this subsection we introduce a few differential operators that allow us to shorten the presentation of identity (5.88) and that play an important role in the analysis of solutions of SDEs.

Definition 5.6.1 (Generator of SDEs)**.** *Let $d, m \in \mathbb{N}$, let $O \subseteq \mathbb{R}^d$ be an open set, and let $\mu\colon O \to \mathbb{R}^d$ and $\sigma = (\sigma_i)_{i\in\{1,2,\dots,m\}} = (\sigma_{k,i})_{k\in\{1,\dots,d\},i\in\{1,\dots,m\}}\colon O \to \mathbb{R}^{d\times m}$ be functions. Then we denote by*

$$L^0_{\mu,\sigma}\colon\ \cup_{k\in\mathbb{N}}\, C^2(O,\mathbb{R}^k) \to \cup_{k\in\mathbb{N}}\,\mathbb{M}(O,\mathbb{R}^k) \tag{5.89}$$

the function which satisfies for all $f \in \cup_{k\in\mathbb{N}} C^2(O,\mathbb{R}^k)$, $x \in O$ that

$$\begin{aligned} \big(L^0_{\mu,\sigma} f\big)(x) &= f'(x)\,\mu(x) + \frac{1}{2}\sum_{i=1}^{m} f''(x)\big(\sigma_i(x),\sigma_i(x)\big) \\ &= \sum_{k=1}^{d}\big(\tfrac{\partial}{\partial x_k} f\big)(x)\cdot\mu_k(x) + \frac{1}{2}\sum_{i=1}^{m}\sum_{k,l=1}^{d}\big(\tfrac{\partial^2}{\partial x_k\partial x_l} f\big)(x)\cdot\sigma_{k,i}(x)\cdot\sigma_{l,i}(x) \end{aligned} \tag{5.90}$$

and we call $L^0_{\mu,\sigma}$ the 0-th noise operator associated to (μ,σ) (we call $L^0_{\mu,\sigma}$ the generator associated to (μ,σ)).

Definition 5.6.2 (Noise operator associated to a stochastic differential equation)**.** *Let $d, m \in \mathbb{N}$, $i \in \{1, 2, \ldots, m\}$, let $O \subseteq \mathbb{R}^d$ be an open set, and let $\mu \colon O \to \mathbb{R}^d$ and $\sigma = (\sigma_i)_{i \in \{1,2,\ldots,m\}} = (\sigma_{k,i})_{k \in \{1,\ldots,d\}, i \in \{1,\ldots,m\}} \colon O \to \mathbb{R}^{d \times m}$ be functions. Then we denote by*

$$L^i_{\mu,\sigma} \colon \cup_{k \in \mathbb{N}} C^1(O, \mathbb{R}^k) \to \cup_{k \in \mathbb{N}} \mathbb{M}(O, \mathbb{R}^k) \tag{5.91}$$

the function which satisfies for all $f \in \cup_{k \in \mathbb{N}} C^1(O, \mathbb{R}^k)$, $x \in O$ that

$$\big(L^i_{\mu,\sigma} f\big)(x) = f'(x)\, \sigma_i(x) = \sum_{k=1}^{d} \big(\tfrac{\partial}{\partial x_k} f\big)(x) \cdot \sigma_{k,i}(x) \tag{5.92}$$

and we call $L^i_{\mu,\sigma}$ the i-th noise operator associated to (μ, σ).

In the next step we employ Definitions 5.6.1 and 5.6.2 to shorten the presentation of identity (5.88). More precisely, assume the setting in Section 5.1 and observe that (5.88) and Definitions 5.6.1 and 5.6.2, ensure that for all $f \in \cup_{k \in \mathbb{N}} C^2(O, \mathbb{R}^k)$, $t_0, t \in [0, T]$ with $t_0 \leq t$ it holds P-a.s. that

$$f(X_t) = f(X_{t_0}) + \int_{t_0}^{t} (L^0_{\mu,\sigma} f)(X_s)\, ds + \sum_{i=1}^{m} \int_{t_0}^{t} (L^i_{\mu,\sigma} f)(X_s)\, dW^{(i)}_s. \tag{5.93}$$

5.6.2 Taylor approximations

Throughout Section 5.6.2 assume the setting in Section 5.1, assume that $\mu \in C^\infty(O, \mathbb{R}^d)$ and $\sigma \in C^\infty(O, \mathbb{R}^{d \times m})$, let $k \in \mathbb{N}$, $f \in C^\infty(O, \mathbb{R}^k)$, $t_0 \in [0, T]$, and let $W^{(0)} \colon [0, T] \to \mathbb{R}$ be the function which satisfies for all $t \in [0, T]$ that $W^{(0)}_t = t$. Equation (5.93) and the fact that $\forall\, t \in [0, T] \colon W^{(0)}_t = t$ show that for all $t \in [t_0, T]$ it holds P-a.s. that

$$f(X_t) = f(X_{t_0}) + \sum_{i=0}^{m} \int_{t_0}^{t} (L^i_{\mu,\sigma} f)(X_s)\, dW^{(i)}_s. \tag{5.94}$$

5.6.2.1 Trivial Taylor approximations

Equation (5.94) assures that for all $s_0 \in [t_0, T]$ it holds P-a.s. that

$$f(X_{s_0}) = \underbrace{f(X_{t_0})}_{\text{Taylor approximation}} + \underbrace{\sum_{\alpha_1=0}^{m} \int_{t_0}^{s_0} (L^{\alpha_1}_{\mu,\sigma} f)(X_{s_1})\, dW^{(\alpha_1)}_{s_1}}_{\text{remainder of the Taylor expansion}}. \tag{5.95}$$

This suggests the approximation

$$f(X_{s_0}) \approx f(X_{t_0}) \tag{5.96}$$

for $s_0 \in [t_0, T]$. In the case where $f = O \ni x \mapsto x \in \mathbb{R}^d$, (5.96) reduces to

$$X_{s_0} \approx X_{t_0} \tag{5.97}$$

for $s_0 \in [t_0, T]$.

5.6.2.2 Taylor approximations corresponding to the Euler-Maruyama scheme

In the next step again equation (5.94) (again Itô's formula) implies that for all $s_1 \in [t_0, T]$, $\alpha_1 \in \{0, 1, \ldots, m\}$ it holds P-a.s. that

$$(L^{\alpha_1}_{\mu,\sigma} f)(X_{s_1}) = (L^{\alpha_1}_{\mu,\sigma} f)(X_{t_0}) + \sum_{\alpha_2=0}^{m} \int_{t_0}^{s_1} \underbrace{\big(L^{\alpha_2}_{\mu,\sigma}(L^{\alpha_1}_{\mu,\sigma} f)\big)(X_{s_2})}_{=(L^{\alpha_2}_{\mu,\sigma} L^{\alpha_1}_{\mu,\sigma} f)(X_{s_2})} \, dW^{(\alpha_2)}_{s_2} \tag{5.98}$$

Putting (5.98) into (5.95) shows that for all $s_0 \in [t_0, T]$ it holds P-a.s. that

$$\begin{aligned} f(X_{s_0}) = {} & \underbrace{f(X_{t_0}) + \sum_{\alpha_1=0}^{m} \int_{t_0}^{s_0} (L^{\alpha_1}_{\mu,\sigma} f)(X_{t_0}) \, dW^{(\alpha_1)}_{s_1}}_{\text{Taylor approximation}} \\ & + \underbrace{\sum_{\alpha_1,\alpha_2=0}^{m} \int_{t_0}^{s_0} \int_{t_0}^{s_1} (L^{\alpha_2}_{\mu,\sigma} L^{\alpha_1}_{\mu,\sigma} f)(X_{s_2}) \, dW^{(\alpha_2)}_{s_2} \, dW^{(\alpha_1)}_{s_1}}_{\text{remainder of the Taylor expansion}} . \end{aligned} \tag{5.99}$$

Identity (5.99) suggests the approximation

$$\begin{aligned} f(X_{s_0}) &\approx f(X_{t_0}) + \sum_{\alpha_1=0}^{m} \int_{t_0}^{s_0} (L^{\alpha_1}_{\mu,\sigma} f)(X_{t_0}) \, dW^{(\alpha_1)}_{s_1} \\ &= f(X_{t_0}) + \sum_{\alpha_1=0}^{m} (L^{\alpha_1}_{\mu,\sigma} f)(X_{t_0}) \cdot \int_{t_0}^{s_0} dW^{(\alpha_1)}_{s_1} \\ &= f(X_{t_0}) + \sum_{\alpha_1=0}^{m} (L^{\alpha_1}_{\mu,\sigma} f)(X_{t_0}) \big(W^{(\alpha_1)}_{s_0} - W^{(\alpha_1)}_{t_0}\big) \end{aligned} \tag{5.100}$$

P-a.s. for $s_0 \in [t_0, T]$. In the case where $f = O \ni x \mapsto x \in \mathbb{R}^d$, (5.100) reads as

$$\begin{aligned} X_{s_0} &\approx X_{t_0} + \sum_{\alpha_1=0}^{m} (L^{\alpha_1}_{\mu,\sigma} f)(X_{t_0}) \big(W^{(\alpha_1)}_{s_0} - W^{(\alpha_1)}_{t_0}\big) \\ &= X_{t_0} + (L^{0}_{\mu,\sigma} f)(X_{t_0}) \big(W^{(0)}_{s_0} - W^{(0)}_{t_0}\big) + \sum_{i=1}^{m} (L^{i}_{\mu,\sigma} f)(X_{t_0}) \big(W^{(i)}_{s_0} - W^{(i)}_{t_0}\big) \\ &= X_{t_0} + \mu(X_{t_0}) \, (s_0 - t_0) + \sum_{i=1}^{m} \sigma_i(X_{t_0}) \big(W^{(i)}_{s_0} - W^{(i)}_{t_0}\big) \\ &= X_{t_0} + \mu(X_{t_0}) \, (s_0 - t_0) + \sigma(X_{t_0}) \big(W_{s_0} - W_{t_0}\big) \end{aligned} \tag{5.101}$$

P-a.s. for $s_0 \in [t_0, T]$.

5.6.2.3 Taylor approximations corresponding to the Milstein scheme

Next note that again equation (5.94) (again Itô's formula) implies that for all $s_2 \in [t_0, T]$, $\alpha_1, \alpha_2 \in \{0, 1, \dots, m\}$ it holds P-a.s. that

$$(L^{\alpha_2}_{\mu,\sigma} L^{\alpha_1}_{\mu,\sigma} f)(X_{s_2}) = (L^{\alpha_2}_{\mu,\sigma} L^{\alpha_1}_{\mu,\sigma} f)(X_{t_0}) + \sum_{\alpha_3=0}^{m} \int_{t_0}^{s_2} (L^{\alpha_3}_{\mu,\sigma} L^{\alpha_2}_{\mu,\sigma} L^{\alpha_1}_{\mu,\sigma} f)(X_{s_3})\, dW^{(\alpha_3)}_{s_3}. \tag{5.102}$$

Putting (5.102) into (5.99) proves that for all $s_0 \in [t_0, T]$ it holds P-a.s. that

$$\begin{aligned} f(X_{s_0}) &= \underbrace{f(X_{t_0}) + \sum_{\alpha_1=0}^{m} (L^{\alpha_1}_{\mu,\sigma} f)(X_{t_0}) \cdot \int_{t_0}^{s_0} dW^{(\alpha_1)}_{s_1}}_{\text{Taylor approximation}} \\ &\quad + \underbrace{\sum_{\alpha_1,\alpha_2=1}^{m} (L^{\alpha_2}_{\mu,\sigma} L^{\alpha_1}_{\mu,\sigma} f)(X_{t_0}) \cdot \int_{t_0}^{s_0} \int_{t_0}^{s_1} dW^{(\alpha_2)}_{s_2}\, dW^{(\alpha_1)}_{s_1}}_{\text{Taylor approximation cont'd}} \\ &\quad + \underbrace{\sum_{\substack{\alpha_1,\alpha_2 \in \{0,1,\dots,m\} \\ \alpha_1 \cdot \alpha_2 = 0}} \int_{t_0}^{s_0} \int_{t_0}^{s_1} (L^{\alpha_2}_{\mu,\sigma} L^{\alpha_1}_{\mu,\sigma} f)(X_{s_2})\, dW^{(\alpha_2)}_{s_2}\, dW^{(\alpha_1)}_{s_1}}_{\text{remainder of the Taylor expansion}} \\ &\quad + \underbrace{\sum_{\alpha_1,\alpha_2=1}^{m} \sum_{\alpha_3=0}^{m} \int_{t_0}^{s_0} \int_{t_0}^{s_1} \int_{t_0}^{s_2} (L^{\alpha_3}_{\mu,\sigma} L^{\alpha_2}_{\mu,\sigma} L^{\alpha_1}_{\mu,\sigma} f)(X_{s_3})\, dW^{(\alpha_3)}_{s_3}\, dW^{(\alpha_2)}_{s_2}\, dW^{(\alpha_1)}_{s_1}}_{\text{remainder of the Taylor expansion cont'd}}. \end{aligned} \tag{5.103}$$

This suggests the approximation

$$\begin{aligned} f(X_{s_0}) &\approx f(X_{t_0}) + \sum_{\alpha_1=0}^{m} (L^{\alpha_1}_{\mu,\sigma} f)(X_{t_0}) \cdot \int_{t_0}^{s_0} dW^{(\alpha_1)}_{s_1} \\ &\qquad + \sum_{\alpha_1,\alpha_2=1}^{m} (L^{\alpha_2}_{\mu,\sigma} L^{\alpha_1}_{\mu,\sigma} f)(X_{t_0}) \cdot \int_{t_0}^{s_0} \int_{t_0}^{s_1} dW^{(\alpha_2)}_{s_2}\, dW^{(\alpha_1)}_{s_1} \\ &= f(X_{t_0}) + \sum_{\alpha_1=0}^{m} (L^{\alpha_1}_{\mu,\sigma} f)(X_{t_0}) \left(W^{(\alpha_1)}_{s_0} - W^{(\alpha_1)}_{t_0}\right) \\ &\qquad + \sum_{\alpha_1,\alpha_2=1}^{m} (L^{\alpha_2}_{\mu,\sigma} L^{\alpha_1}_{\mu,\sigma} f)(X_{t_0}) \cdot \int_{t_0}^{s_0} \int_{t_0}^{s_1} dW^{(\alpha_2)}_{s_2}\, dW^{(\alpha_1)}_{s_1} \end{aligned} \tag{5.104}$$

P-a.s. for $s_0 \in [t_0, T]$. In the case where $f = O \ni x \mapsto x \in \mathbb{R}^d$, (5.104) reads as

$$\begin{aligned} X_{s_0} &\approx X_{t_0} + \sum_{\alpha_1=0}^{m} (L^{\alpha_1}_{\mu,\sigma} f)(X_{t_0}) \left(W^{(\alpha_1)}_{s_0} - W^{(\alpha_1)}_{t_0}\right) \\ &\quad + \sum_{\alpha_1,\alpha_2=1}^{m} (L^{\alpha_2}_{\mu,\sigma} L^{\alpha_1}_{\mu,\sigma} f)(X_{t_0}) \cdot \int_{t_0}^{s_0} \int_{t_0}^{s_1} dW^{(\alpha_2)}_{s_2}\, dW^{(\alpha_1)}_{s_1} \\ &= X_{t_0} + \mu(X_{t_0})\,(s_0 - t_0) + \sigma(X_{t_0})\big(W_{s_0} - W_{t_0}\big) \\ &\quad + \sum_{\alpha_1,\alpha_2=1}^{m} (\sigma_{\alpha_1})'(X_{t_0})\, \sigma_{\alpha_2}(X_{t_0}) \cdot \int_{t_0}^{s_0} \int_{t_0}^{s_1} dW^{(\alpha_2)}_{s_2}\, dW^{(\alpha_1)}_{s_1} \end{aligned} \tag{5.105}$$

P-a.s. for $s_0 \in [t_0, T]$. This *stochastic Taylor approximation* motivates the scheme presented in the following subsection.

5.6.3 Milstein scheme

In this subsection we present the Milstein scheme (see [Milstein(1974)]), which is beside the Euler-Maruyama scheme probably the most known numerical scheme for stochastic differential equations.

Definition 5.6.3 (Milstein approximation)**.** *Let $T \in (0, \infty)$, $d, m, N \in \mathbb{N}$, $\mu \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{R}^d))$, $\sigma = (\sigma_j)_{j\in\{1,\ldots,m\}} \in C^1(\mathbb{R}^d, \mathbb{R}^{d\times m})$, let $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$ be a stochastic basis, let $\xi \in \mathcal{M}(\mathbb{F}_0, \mathcal{B}(\mathbb{R}^d))$, and let $W \colon [0,T]\times\Omega \to \mathbb{R}^m$ be an m-dimensional standard $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$-Brownian motion. Then we say that Y is a Milstein approximation for the SDE*

$$dX_t = \mu(X_t)\, dt + \sigma(X_t)\, dW_t, \qquad t \in [0, T], \qquad X_0 = \xi \tag{5.106}$$

with time step size T/N on $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$ (we say that Y is a Milstein approximation for the SDE

$$dX_t = \mu(X_t)\, dt + \sigma(X_t)\, dW_t, \qquad t \in [0, T], \qquad X_0 = \xi \tag{5.107}$$

with time step size T/N) if and only if $Y \in \mathbb{M}(\{0, 1, \ldots, N\} \times \Omega, \mathbb{R}^d)$ is an $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$-valued stochastic process with time set $\{0, 1, \ldots, N\}$ on $(\Omega, \mathcal{F}, P)$ which satisfies that for all $n \in \{0, 1, \ldots, N-1\}$ it holds P-a.s. that $Y_0 = \xi$ and

$$\begin{aligned} Y_{n+1} &= Y_n + \mu(Y_n)\tfrac{T}{N} + \sigma(Y_n)\big(W_{\frac{(n+1)T}{N}} - W_{\frac{nT}{N}}\big) \\ &\qquad + \sum_{i,j=1}^{m} (\sigma_i)'(Y_n)\, \sigma_j(Y_n) \int_{\frac{nT}{N}}^{\frac{(n+1)T}{N}} \int_{\frac{nT}{N}}^{s} dW^{(j)}_u\, dW^{(i)}_s. \end{aligned} \tag{5.108}$$

Definition 5.6.4 (Linearly-interpolated Milstein approximation). *Let $T \in (0,\infty)$, $d,m,N \in \mathbb{N}$, $\mu \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d),\mathcal{B}(\mathbb{R}^d))$, $\sigma = (\sigma_j)_{j\in\{1,\ldots,m\}} \in C^1(\mathbb{R}^d,\mathbb{R}^{d\times m})$, let $(\Omega,\mathcal{F},P,(\mathbb{F}_t)_{t\in[0,T]})$ be a stochastic basis, let $\xi \in \mathcal{M}(\mathbb{F}_0,\mathcal{B}(\mathbb{R}^d))$, and let $W\colon [0,T]\times\Omega \to \mathbb{R}^m$ be an m-dimensional standard $(\Omega,\mathcal{F},P,(\mathbb{F}_t)_{t\in[0,T]})$-Brownian motion. Then we say that Y is a linearly-interpolated Milstein approximation for the SDE*

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t, \qquad t\in[0,T], \qquad X_0=\xi \tag{5.109}$$

with time step size T/N on $(\Omega,\mathcal{F},P,(\mathbb{F}_t)_{t\in[0,T]})$ (we say that Y is a linearly-interpolated Milstein approximation for the SDE

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t, \qquad t\in[0,T], \qquad X_0=\xi \tag{5.110}$$

with time step size T/N) if and only if $Y \in \mathbb{M}([0,T]\times\Omega,\mathbb{R}^d)$ is an $(\mathbb{R}^d,\mathcal{B}(\mathbb{R}^d))$-valued stochastic process with time set $[0,T]$ on $(\Omega,\mathcal{F},P)$ which satisfies that for all $n\in\{0,1,\ldots,N-1\}$, $t\in\left[\frac{nT}{N},\frac{(n+1)T}{N}\right]$ it holds P-a.s. that

$$\begin{aligned} Y_t = Y_{\frac{nT}{N}} &+ \left(\tfrac{tN}{T}-n\right)\left[\mu\big(Y_{\frac{nT}{N}}\big)\tfrac{T}{N} + \sigma\big(Y_{\frac{nT}{N}}\big)\big(W_{\frac{(n+1)T}{N}} - W_{\frac{nT}{N}}\big)\right] \\ &+ \left(\tfrac{tN}{T}-n\right)\left[\sum_{i,j=1}^{m}(\sigma_i)'\big(Y_{\frac{nT}{N}}\big)\,\sigma_j\big(Y_{\frac{nT}{N}}\big)\int_{\frac{nT}{N}}^{\frac{(n+1)T}{N}}\int_{\frac{nT}{N}}^{s} dW_u^{(j)}\,dW_s^{(i)}\right]. \end{aligned} \tag{5.111}$$

5.6.3.1 Strong convergence of the Milstein scheme

Under suitable assumptions, the Milstein approximations converge at time T in the strong L^p-sense to the solution process $X\colon[0,T]\times\Omega\to D$ of the SDE (5.2). This is the subject of Theorem 5.6.5 below which is a slightly modified version of the convergence results in [Kloeden and Platen(1992)].

Theorem 5.6.5 (Strong convergence of the Milstein method). *Let $T\in(0,\infty)$, $d,m,N\in\mathbb{N}$, $p\in[2,\infty)$, let $(\Omega,\mathcal{F},P,(\mathbb{F}_t)_{t\in[0,T]})$ be a stochastic basis, let $\xi\in\mathcal{L}^{2p}(P|_{\mathbb{F}_0};\|\cdot\|_{\mathbb{R}^d})$, let $W\colon[0,T]\times\Omega\to\mathbb{R}^m$ be an m-dimensional standard $(\Omega,\mathcal{F},P,(\mathbb{F}_t)_{t\in[0,T]})$-Brownian motion, let $\mu\colon\mathbb{R}^d\to\mathbb{R}^d$ and $\sigma\colon\mathbb{R}^d\to\mathbb{R}^{d\times m}$ be twice continuously differentiable functions with globally bounded derivatives, let $X\colon[0,T]\times\Omega\to\mathbb{R}^d$ be a solution process of the SDE*

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t, \qquad t\in[0,T], \qquad X_0=\xi \tag{5.112}$$

on $(\Omega,\mathcal{F},P,(\mathbb{F}_t)_{t\in[0,T]})$, and for every $N\in\mathbb{N}$ let $Y^N\colon[0,T]\times\Omega\to\mathbb{R}^d$ be a linearly-interpolated Milstein approximation for the SDE (5.112) with time step size T/N. Then there exists a real number $C\in\mathbb{R}$ such that for all $N\in\mathbb{N}$ it holds that

$$\left(\mathbb{E}_P\big[\|X_T - Y_T^N\|_{\mathbb{R}^d}^p\big]\right)^{\frac{1}{p}} = \|X_T - Y_T^N\|_{\mathcal{L}^p(P;\|\cdot\|_{\mathbb{R}^d})} \le \frac{C}{N}. \tag{5.113}$$

The proof of Theorem 5.6.5 is omitted. Theorem 5.6.5 shows that under the assumptions of Theorem 5.6.5 it holds that the *linearly-interpolated Milstein approximations* $(Y^N)_{N\in\mathbb{N}}$

converge at time T in the *strong L^p-sense with order* 1 to X.

5.6.3.2 Simulation of sample paths of Milstein approximations

In general, the Milstein scheme is difficult/impossible to simulate/implement. However, under a suitable further assumption (see (Commutative noise) below), the Milstein scheme can be simulated efficiently. This is the subject of the next proposition.

Proposition 5.6.6. *Let $T \in (0,\infty)$, $d, m, N \in \mathbb{N}$, $\mu \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{R}^d))$, $\sigma = (\sigma_j)_{j\in\{1,\dots,m\}} \in C^1(\mathbb{R}^d, \mathbb{R}^{d\times m})$, let $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$ be a stochastic basis, let $\xi \in \mathcal{M}(\mathbb{F}_0, \mathcal{B}(\mathbb{R}^d))$, let $W\colon [0,T] \times \Omega \to \mathbb{R}^m$ be an m-dimensional standard $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$-Brownian motion, let $Y\colon \{0,1,\dots,N\} \times \Omega \to \mathbb{R}^d$ be a Milstein approximation for the SDE*

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t, \qquad t \in [0,T], \qquad X_0 = \xi \tag{5.114}$$

with time step size T/N on $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$, and assume for all $x \in \mathbb{R}^d$, $i, j \in \{1, 2, \dots, m\}$ that

$$(\sigma_i)'(x)\,\sigma_j(x) = (\sigma_j)'(x)\,\sigma_i(x). \tag{Commutative noise}$$

Then for all $n \in \{0, 1, \dots, N-1\}$ it holds P-a.s. that

$$\begin{aligned}
&Y_{n+1} = Y_n + \mu(Y_n)\,\tfrac{T}{N} + \sigma(Y_n)\big(W_{\frac{(n+1)T}{N}} - W_{\frac{nT}{N}}\big)\\
&+ \frac{1}{2}\,\sigma'(Y_n)\Big(\sigma(Y_n)\big(W_{\frac{(n+1)T}{N}} - W_{\frac{nT}{N}}\big)\Big)\big(W_{\frac{(n+1)T}{N}} - W_{\frac{nT}{N}}\big) - \frac{T}{2N}\sum_{i=1}^m (\sigma_i)'(Y_n)\,\sigma_i(Y_n)\\
&= Y_n + \mu(Y_n)\,\tfrac{T}{N} + \sum_{j=1}^m \sigma_j(Y_n)\big(W^{(j)}_{\frac{(n+1)T}{N}} - W^{(j)}_{\frac{nT}{N}}\big) - \frac{T}{2N}\sum_{i=1}^m (\sigma_i)'(Y_n)\,\sigma_i(Y_n)\\
&+ \frac{1}{2}\sum_{i,j=1}^m (\sigma_i)'(Y_n)\,\sigma_j(Y_n)\,\big(W^{(i)}_{\frac{(n+1)T}{N}} - W^{(i)}_{\frac{nT}{N}}\big)\,\big(W^{(j)}_{\frac{(n+1)T}{N}} - W^{(j)}_{\frac{nT}{N}}\big).
\end{aligned} \tag{5.115}$$

Proof of Proposition 5.6.6. Observe that Itô's formula (cf. Example 3.5.6) shows that for all $n \in \{0, 1, \ldots, N-1\}$ it holds P-a.s. that

$$\begin{aligned}
&\sum_{i,j=1}^{m} (\sigma_i)'(Y_n)\, \sigma_j(Y_n) \cdot \int_{\frac{nT}{N}}^{\frac{(n+1)T}{N}} \int_{\frac{nT}{N}}^{s} dW_u^{(j)}\, dW_s^{(i)} \\
&= \sum_{i=1}^{m} (\sigma_i)'(Y_n)\, \sigma_i(Y_n) \cdot \int_{\frac{nT}{N}}^{\frac{(n+1)T}{N}} \int_{\frac{nT}{N}}^{s} dW_u^{(i)}\, dW_s^{(i)} \\
&+ \frac{1}{2} \sum_{\substack{i,j \in \{1,2,\ldots,m\},\\ i \neq j}} \Bigg[(\sigma_i)'(Y_n)\, \sigma_j(Y_n) \cdot \int_{\frac{nT}{N}}^{\frac{(n+1)T}{N}} \int_{\frac{nT}{N}}^{s} dW_u^{(j)}\, dW_s^{(i)} \\
&+ (\sigma_j)'(Y_n)\, \sigma_i(Y_n) \cdot \int_{\frac{nT}{N}}^{\frac{(n+1)T}{N}} \int_{\frac{nT}{N}}^{s} dW_u^{(i)}\, dW_s^{(j)} \Bigg] \\
&= \frac{1}{2} \sum_{i=1}^{m} (\sigma_i)'(Y_n)\, \sigma_i(Y_n) \left[\left[W^{(i)}_{\frac{(n+1)T}{N}} - W^{(i)}_{\frac{nT}{N}} \right]^2 - \frac{T}{N} \right] \\
&+ \tfrac{1}{2} \sum_{\substack{i,j \in \{1,2,\\ \ldots,m\}, i \neq j}} (\sigma_i)'(Y_n)\, \sigma_j(Y_n) \left[\int_{\frac{nT}{N}}^{\frac{(n+1)T}{N}} \int_{\frac{nT}{N}}^{s} dW_u^{(j)}\, dW_s^{(i)} + \int_{\frac{nT}{N}}^{\frac{(n+1)T}{N}} \int_{\frac{nT}{N}}^{s} dW_u^{(i)}\, dW_s^{(j)} \right].
\end{aligned} \tag{5.116}$$

Again Itô's formula hence gives that for all $n \in \{0, 1, \ldots, N-1\}$ it holds P-a.s. that

$$\begin{aligned}
&\sum_{i,j=1}^{m} (\sigma_i)'(Y_n)\, \sigma_j(Y_n) \cdot \int_{\frac{nT}{N}}^{\frac{(n+1)T}{N}} \int_{\frac{nT}{N}}^{s} dW_u^{(j)}\, dW_s^{(i)} \\
&= \frac{1}{2} \sum_{i=1}^{m} (\sigma_i)'(Y_n)\, \sigma_i(Y_n) \left[\left[W^{(i)}_{\frac{(n+1)T}{N}} - W^{(i)}_{\frac{nT}{N}} \right]^2 - \frac{T}{N} \right] \\
&+ \frac{1}{2} \sum_{\substack{i,j \in \{1,2,\ldots,m\},\\ i \neq j}} (\sigma_i)'(Y_n)\, \sigma_j(Y_n) \left[W^{(i)}_{\frac{(n+1)T}{N}} - W^{(i)}_{\frac{nT}{N}} \right] \left[W^{(j)}_{\frac{(n+1)T}{N}} - W^{(j)}_{\frac{nT}{N}} \right] \\
&= \frac{1}{2} \sum_{i,j \in \{1,2,\ldots,m\}} (\sigma_i)'(Y_n) \Big(\sigma_j(Y_n) \big(W^{(j)}_{\frac{(n+1)T}{N}} - W^{(j)}_{\frac{nT}{N}} \big) \Big) \big(W^{(i)}_{\frac{(n+1)T}{N}} - W^{(i)}_{\frac{nT}{N}} \big) \\
&- \frac{T}{2N} \sum_{i=1}^{m} (\sigma_i)'(Y_n)\, \sigma_i(Y_n) \\
&= \frac{1}{2} \sigma'(Y_n) \Big(\sigma(Y_n) \big(W_{\frac{(n+1)T}{N}} - W_{\frac{nT}{N}} \big) \Big) \big(W_{\frac{(n+1)T}{N}} - W_{\frac{nT}{N}} \big) - \frac{T}{2N} \sum_{i=1}^{m} (\sigma_i)'(Y_n)\, \sigma_i(Y_n).
\end{aligned} \tag{5.117}$$

The proof of Proposition 5.6.6 is thus completed. □

Assumption (Commutative noise) is, for example, fulfilled in the case where $m = 1$ (cf. Example 3.5.6).

Exercise 5.6.7 (Milstein method). *In this exercise we do not distinguish between pseudo random numbers and actual random numbers. Let* $T \in (0,\infty)$, $d, N \in \mathbb{N}$, $\xi \in \mathbb{R}^d$, $\mu \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{R}^d))$, $\sigma \in C^1(\mathbb{R}^d, \mathbb{R}^d)$, $\tilde{\sigma} \in C(\mathbb{R}^d, \mathbb{R}^{d\times d})$ *satisfy for all* $x \in \mathbb{R}^d$ *that* $\tilde{\sigma}(x) = \sigma'(x)$, *let* $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$ *be a stochastic basis, let* $W \colon [0,T] \times \Omega \to \mathbb{R}$ *be a one-dimensional standard* $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$*-Brownian motion, and let* $Y \colon \{0, 1, \ldots, N\} \times \Omega \to \mathbb{R}^d$ *be a Milstein approximation for the SDE*

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t, \qquad t \in [0,T], \qquad X_0 = \xi \tag{5.118}$$

with time step size T/N *on* $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$ *(see Definition 5.6.3). Write a* Matlab *function* `Milstein(T,d,N,ξ,μ,σ,σ̃)` *with input* $T \in (0,\infty)$, $d, N \in \mathbb{N}$, $\xi \in \mathbb{R}^d$, $\mu \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{R}^d))$, $\sigma \in C^1(\mathbb{R}^d, \mathbb{R}^d)$, $\tilde{\sigma} \in C(\mathbb{R}^d, \mathbb{R}^{d\times d})$ *and output a realization of an* $Y_N(P)_{\mathcal{B}(\mathbb{R}^d)}$*-distributed random variable.*

Exercise 5.6.8 (Milstein method in two dimensions). *In this exercise we do not distinguish between pseudo random numbers and actual random numbers. Let* $T \in (0,\infty)$, $N \in \mathbb{N}$, $\xi \in \mathbb{R}^2$, *let* $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$ *be a stochastic basis, let* $W \colon [0,T] \times \Omega \to \mathbb{R}^2$ *be a two-dimensional standard* $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$*-Brownian motion, let* $A_1, A_2 \in \mathbb{R}^{2\times 2}$ *be the* 2×2*-matrices given by*

$$A_1 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \qquad \text{and} \qquad A_2 = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}, \tag{5.119}$$

let $\sigma = (\sigma_1, \sigma_2) \colon \mathbb{R}^2 \to \mathbb{R}^{2\times 2}$ *be the function which satisfies for all* $x = (x_1, x_2)$, $u = (u_1, u_2) \in \mathbb{R}^2$ *that*

$$\sigma(x)u = u_1 A_1 x + u_2 A_2 x, \tag{5.120}$$

and let $Y \colon \{0, 1, \ldots, N\} \times \Omega \to \mathbb{R}^2$ *be a Milstein approximation for the SDE*

$$dX_t = \sigma(X_t)\,dW_t, \qquad t \in [0,T], \qquad X_0 = \xi \tag{5.121}$$

with time step size T/N *on* $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$ *(see Definition 5.6.3 and Proposition 5.6.6).*

(i) Prove that for all $x \in \mathbb{R}^2$ *it holds that* $\sigma_1'(x)\,\sigma_2(x) = \sigma_2'(x)\,\sigma_1(x)$.

(ii) Write a Matlab *function* `Milstein2D(T,N,ξ)` *with input* $T \in (0,\infty)$, $N \in \mathbb{N}$, $\xi \in \mathbb{R}^2$ *and output a realization of an* $Y_N(P)_{\mathcal{B}(\mathbb{R}^2)}$*-distributed random variable.*

5.6.4 General stochastic Taylor expansions and hierarchical sets

In Subsections 5.6.2.1–5.6.2.3 three stochastic Taylor expansions have been presented. For formulating general stochastic Taylor expansions, a bit more notation is used (see, e.g., Chapter 5 in [Kloeden and Platen(1992)] and, e.g., Section 3.2.4 in [Da Prato et al.(2012)Da Prato, Jentzen, and Roeckner] for the following presentations).

5.6.4.1 Multi-indices notation and hierarchical sets

Definition 5.6.9 (Set of multi-indices)**.** *Let $S \subseteq \mathbb{N}_0$ be a non-empty subset of $\mathbb{N}_0$. Then we denote by $\mathbb{M}_S$ the set given by*

$$\mathbb{M}_S = \{\emptyset\} \cup (\cup_{n=1}^{\infty} S^n) \tag{5.122}$$

and we call $\mathbb{M}_S$ the set of S-valued multi-indices.

Note that for all $S_1, S_2 \in \mathcal{P}(\mathbb{N}_0)$ with $\emptyset \neq S_1 \subseteq S_2 \subseteq \mathbb{N}_0$ it holds that

$$\mathbb{M}_{S_1} \subseteq \mathbb{M}_{S_2} \subseteq \mathbb{M}_{\mathbb{N}_0}. \tag{5.123}$$

Definition 5.6.10 (Length of an multi-index)**.** *Let $S \subseteq \mathbb{N}_0$ be a non-empty subset of $\mathbb{N}_0$. Then we denote by $|\cdot|_{\mathbb{M}_S} \colon \mathbb{M}_S \to \mathbb{N}_0$ the function with the property that for all $n \in \mathbb{N}$, $\alpha_1, \dots, \alpha_n \in S$ it holds that*

$$|(\alpha_1, \dots, \alpha_n)|_{\mathbb{M}_S} = n \qquad \textit{and} \qquad |\emptyset|_{\mathbb{M}_S} = 0. \tag{5.124}$$

Definition 5.6.11 (Truncation of multi-indices)**.** *We denote by $(\cdot)- \colon \big(\mathbb{M}_{\mathbb{N}_0} \backslash \{\emptyset\}\big) \to \mathbb{M}_{\mathbb{N}_0}$ the function with the property that for all $n \in \mathbb{N}$, $\alpha_1, \dots, \alpha_n \in \mathbb{N}_0$ it holds that*

$$(\alpha_1, \alpha_2, \dots, \alpha_n)- = \begin{cases} (\alpha_1, \alpha_2, \dots, \alpha_{n-1}) & : n \geq 2 \\ \emptyset & : n = 1 \end{cases}. \tag{5.125}$$

Observe that Definitions 5.6.9–5.6.11, in particular, ensure that for all $m \in \mathbb{N}$ and all $\alpha = \big(\alpha_1, \dots, \alpha_{|\alpha|_{\mathbb{M}_{\{0,1,\dots,m\}}}}\big) \in \mathbb{M}_{\{0,1,\dots,m\}} \backslash \{\emptyset\}$ it holds that

$$|\alpha|_{\mathbb{M}_{\{0,1,\dots,m\}}} \geq 1 \qquad \text{and} \qquad \alpha_1, \dots, \alpha_{|\alpha|_{\mathbb{M}_{\{0,1,\dots,m\}}}} \in \{0, 1, \dots, m\}. \tag{5.126}$$

Definition 5.6.12 (Hierarchical set)**.** *A finite non-empty subset $\mathcal{A} \subseteq \mathbb{M}_{\mathbb{N}_0}$ of $\mathbb{M}_{\mathbb{N}_0}$ is called a hierarchical set if it holds for all $\alpha \in \mathcal{A} \backslash \{\emptyset\}$ that*

$$\alpha- \in \mathcal{A}. \tag{5.127}$$

Definition 5.6.13 (Remainder set)**.** *Let $S \subseteq \mathbb{N}_0$ be a non-empty subset of $\mathbb{N}_0$. Then we denote by $\mathbb{B}_S \colon \mathcal{P}(\mathbb{M}_S) \to \mathcal{P}(\mathbb{M}_S)$ the function with the property that for all $\mathcal{A} \subseteq \mathbb{M}_S$ it holds that*

$$\mathbb{B}_S(\mathcal{A}) = \{\alpha \in \mathbb{M}_S \backslash \mathcal{A} \colon \alpha- \in \mathcal{A}\}. \tag{5.128}$$

5.6.4.2 General stochastic Taylor expansions

We are now ready to present general stochastic Taylor expansions.

Theorem 5.6.14 (Stochastic Taylor expansions)**.** *Assume the setting in Section 5.1, assume that $\mu \in C^\infty(O, \mathbb{R}^d)$ and $\sigma \in C^\infty(O, \mathbb{R}^{d\times m})$, let $k \in \mathbb{N}$, $f \in C^\infty(O, \mathbb{R}^k)$. Then for all $t_0, s_0 \in [0,T]$ with $t_0 \le s_0$ and all hierarchical sets $\mathcal{A} \subseteq \mathbb{M}_{\{0,1,\dots,m\}}$ it holds P-a.s. that*

$$f(X_{s_0}) = \tag{5.129}$$
$$\underbrace{f(X_{t_0}) + \sum_{\alpha\in\mathcal{A}\setminus\{\emptyset\}} \Big(L_{\mu,\sigma}^{\alpha_{|\alpha|_{\mathbb{M}_{\mathbb{N}_0}}}} \dots L_{\mu,\sigma}^{\alpha_2} L_{\mu,\sigma}^{\alpha_1} f\Big)(X_{t_0}) \cdot \int_{t_0}^{s_0}\int_{t_0}^{s_1} \dots \int_{t_0}^{s_{|\alpha|_{\mathbb{M}_{\mathbb{N}_0}}-1}} dW_{s_{|\alpha|_{\mathbb{M}_{\mathbb{N}_0}}}}^{(\alpha_{|\alpha|_{\mathbb{M}_{\mathbb{N}_0}}})} \dots dW_{s_2}^{(\alpha_2)}\, dW_{s_1}^{(\alpha_1)}}_{\text{Taylor approximation}}$$
$$+ \underbrace{\sum_{\alpha\in\mathbb{B}_{\{0,1,\dots,m\}}(\mathcal{A})} \int_{t_0}^{s_0}\int_{t_0}^{s_1} \dots \int_{t_0}^{s_{|\alpha|_{\mathbb{M}_{\mathbb{N}_0}}-1}} \Big(L_{\mu,\sigma}^{\alpha_{|\alpha|_{\mathbb{M}_{\mathbb{N}_0}}}} \dots L_{\mu,\sigma}^{\alpha_2} L_{\mu,\sigma}^{\alpha_1} f\Big)\big(X_{s_{|\alpha|_{\mathbb{M}_{\mathbb{N}_0}}}}\big)\, dW_{s_{|\alpha|_{\mathbb{M}_{\mathbb{N}_0}}}}^{(\alpha_{|\alpha|_{\mathbb{M}_{\mathbb{N}_0}}})} \dots dW_{s_2}^{(\alpha_2)}\, dW_{s_1}^{(\alpha_1)}}_{\text{remainder of the Taylor expansion}}.$$

Let $m \in \mathbb{N}$. Then observe in the case of the hierarchical set

$$\mathcal{A} = \{\emptyset\} \tag{5.130}$$

that

$$\mathbb{B}_{\{0,1,\dots,m\}}(\mathcal{A}) = \{0, 1, \dots, m\} \tag{5.131}$$

and hence, formula (5.129) reduces to (5.95). Moreover, note in the case of the hierarchical set

$$\mathcal{A} = \{\emptyset, 0, 1, \dots, m\} \tag{5.132}$$

that

$$\mathbb{B}_{\{0,1,\dots,m\}}(\mathcal{A}) = \{0, 1, \dots, m\}^2 \tag{5.133}$$

and formula (5.129) therefore reduces to (5.99). Furthermore, in the case of the hierarchical set

$$\mathcal{A} = \{\emptyset, 0, 1, \dots, m\} \cup \{1, 2, \dots, m\}^2 \tag{5.134}$$

we obtain that

$$\begin{aligned}&\mathbb{B}_{\{0,1,\dots,m\}}(\mathcal{A})\\ &= \Big(\{0\} \times \{0,1,\dots,m\}\Big) \cup \Big(\{0,1,\dots,m\} \times \{0\}\Big) \cup \Big(\{1,2,\dots,m\}^2 \times \{0,1,\dots,m\}\Big)\end{aligned} \tag{5.135}$$

and hence, formula (5.129) reduces to (5.103). Finally, in the case of the hierarchical set

$$\mathcal{A} = \{\emptyset, 0, 1, \dots, m\} \cup \{0, 1, \dots, m\}^2 \tag{5.136}$$

we obtain that

$$\mathbb{B}_{\{0,1,\dots,m\}}(\mathcal{A}) = \{0, 1, \dots, m\}^3 \tag{5.137}$$

and formula (5.129) hence reduces to the fact that under the assumptions of Theorem 5.6.14 we have that for all $s_0, t_0 \in [0,T]$ with $s_0 \le t_0$ it holds P-a.s. that

$$\begin{aligned} f(X_{s_0}) = {} & \underbrace{f(X_{t_0}) + \sum_{\alpha_1=0}^{m} \big(L^{\alpha_1}_{\mu,\sigma} f\big)\big(X_{t_0}\big) \cdot \int_{t_0}^{s_0} dW^{(\alpha_1)}_{s_1}}_{\text{Taylor approximation}} \\ & \underbrace{+ \sum_{\alpha_1,\alpha_2=0}^{m} \big(L^{\alpha_2}_{\mu,\sigma} L^{\alpha_1}_{\mu,\sigma} f\big)\big(X_{t_0}\big) \cdot \int_{t_0}^{s_0} \int_{t_0}^{s_1} dW^{(\alpha_2)}_{s_2}\, dW^{(\alpha_1)}_{s_1}}_{\text{Taylor approximation cont'd}} \\ & + \underbrace{\sum_{\alpha_1,\alpha_2,\alpha_3=0}^{m} \int_{t_0}^{s_0} \int_{t_0}^{s_1} \int_{t_0}^{s_2} \big(L^{\alpha_3}_{\mu,\sigma} L^{\alpha_2}_{\mu,\sigma} L^{\alpha_1}_{\mu,\sigma} f\big)\big(X_{s_3}\big)\, dW^{(\alpha_3)}_{s_3}\, dW^{(\alpha_2)}_{s_2}\, dW^{(\alpha_1)}_{s_1}}_{\text{remainder of the Taylor expansion}} \end{aligned} \tag{5.138}$$

and

$$\begin{aligned} X_{s_0} = {} & \underbrace{X_{t_0} + \mu(X_{t_0})\,(s_0 - t_0) + \sigma(X_{t_0})\big(W_{s_0} - W_{t_0}\big) + (L^0_{\mu,\sigma}\mu)(X_{t_0}) \cdot \frac{(s_0 - t_0)^2}{2}}_{\text{Taylor approximation}} \\ & \underbrace{+ \sum_{i,j=1}^{m} \sigma_i'(X_{t_0})\,\sigma_j(X_{t_0}) \cdot \int_{t_0}^{s_0} \int_{t_0}^{s_1} dW^{(j)}_{s_2}\, dW^{(i)}_{s_1} + \sum_{i=1}^{m} (L^0_{\mu,\sigma}\sigma_i)(X_{t_0}) \cdot \int_{t_0}^{s_0} \int_{t_0}^{s_1} ds_2\, dW^{(i)}_{s_1}}_{\text{Taylor approximation cont'd}} \\ & \underbrace{+ \sum_{i=1}^{m} \mu'(X_{t_0})\,\sigma_i(X_{t_0}) \cdot \int_{t_0}^{s_0} \int_{t_0}^{s_1} dW^{(i)}_{s_2}\, ds_1}_{\text{Taylor approximation cont'd}} \\ & + \underbrace{\sum_{\alpha_1,\alpha_2,\alpha_3=0}^{m} \int_{t_0}^{s_0} \int_{t_0}^{s_1} \int_{t_0}^{s_2} \big(L^{\alpha_3}_{\mu,\sigma} L^{\alpha_2}_{\mu,\sigma} L^{\alpha_1}_{\mu,\sigma} \mathrm{id}_O\big)\big(X_{s_3}\big)\, dW^{(\alpha_3)}_{s_3}\, dW^{(\alpha_2)}_{s_2}\, dW^{(\alpha_1)}_{s_1}}_{\text{remainder of the Taylor expansion}}. \end{aligned} \tag{5.139}$$

In the next step the proof of Theorem 5.6.14 is presented.

Proof of Theorem 5.6.14. If $\mathcal{A} \subseteq \mathbb{M}_{\{0,1,\dots,m\}}$ is a hierarchical set, then we show (5.129) by induction on $\max_{\alpha\in\mathcal{A}} |\alpha|_{\mathbb{M}_{\mathbb{N}_0}} \in \mathbb{N}_0$. The base case $\max_{\alpha\in\mathcal{A}} |\alpha|_{\mathbb{M}_{\mathbb{N}_0}} = 0$, that is, $\mathcal{A} = \{\emptyset\}$, follows from (5.95). For the induction step we observe that if $n \in \mathbb{N}_0$ and if (5.129) holds for all hierarchical sets $\mathcal{A} \subseteq \mathbb{M}_{\{0,1,\dots,m\}}$ with $\max_{\alpha\in\mathcal{A}} |\alpha|_{\mathbb{M}_{\mathbb{N}_0}} \le n$, then combining (5.94) and (5.129) for all hierarchical sets $\mathcal{A} \subseteq \mathbb{M}_{\{0,1,\dots,m\}}$ with $\max_{\alpha\in\mathcal{A}} |\alpha|_{\mathbb{M}_{\{0,1,\dots,m\}}} = n$ results in (5.129) for all hierarchical sets $\mathcal{A} \subseteq \mathbb{M}_{\{0,1,\dots,m\}}$ with $\max_{\alpha\in\mathcal{A}} |\alpha|_{\mathbb{M}_{\{0,1,\dots,m\}}} = n+1$. The proof of Theorem 5.6.14 is thus completed. $\square$

5.6.4.3 General stochastic Taylor approximations

Theorem 5.6.14 formulates stochastic Taylor expansions for solution processes of SDEs. In particular, in the case $f = \big(O \ni x \mapsto x \in \mathbb{R}^d\big)$, Theorem 5.6.14 represents the solution process X of an SDE at some time instance $s_0 \in [0,T]$ by the sum of two terms, a stochastic *Taylor approximation* depending on the solution process X only at time $t_0 \in [0, s_0]$ and a *remainder term* of the stochastic Taylor expansion depending on the "whole" solution process X_u, $u \in [t_0, s_0]$. Omitting the remainder term does, of course, not give a representation of the solution anymore but results, under suitable assumptions, in a (in an appropriate way) good approximation of X_{s_0}. Corollary 5.6.15 below quantifies, under suitable assumptions, how "good" this approximation is. The whole identity (5.129) is referred to as *stochastic Taylor expansion.*

Corollary 5.6.15 (Estimation of the remainder terms)**.** *Assume the setting in Section 5.1, assume that $\mu \in C^\infty(O, \mathbb{R}^d)$ and $\sigma \in C^\infty(O, \mathbb{R}^{d\times m})$, let $k \in \mathbb{N}$, $f \in C^\infty(O, \mathbb{R}^k)$, $t_0 \in [0,T]$, $s_0 \in [t_0, T]$, let $\mathcal{A} \subseteq \mathbb{M}_m$ be a hierarchical set, and assume for all $\alpha \in \mathcal{A} \cup \mathbb{B}(\mathcal{A})$ that*

$$\sup_{u\in[t_0,s_0]} \big\| \big(L^{\alpha_{|\alpha|_{\mathbb{N}_0}}}_{\mu,\sigma} \dots L^{\alpha_2}_{\mu,\sigma} L^{\alpha_1}_{\mu,\sigma} f\big)\big(X_u\big)\big\|_{L^2(P;\|\cdot\|_{\mathbb{R}^k})} < \infty. \tag{5.140}$$

Then

$$\begin{aligned}&\Bigg\| f(X_{s_0}) - f(X_{t_0}) - \sum_{\alpha\in\mathcal{A}\setminus\{\emptyset\}} \big(L^{\alpha_{|\alpha|_{\mathbb{N}_0}}}_{\mu,\sigma} \dots L^{\alpha_2}_{\mu,\sigma} L^{\alpha_1}_{\mu,\sigma} f\big)\big(X_{t_0}\big) \\ &\quad \cdot \int_{t_0}^{s_0}\int_{t_0}^{s_1}\dots\int_{t_0}^{s_{|\alpha|_{\mathbb{N}_0}-1}} dW^{(\alpha_{|\alpha|_{\mathbb{N}_0}})}_{s_{|\alpha|_{\mathbb{N}_0}}} \dots dW^{(\alpha_2)}_{s_2}\, dW^{(\alpha_1)}_{s_1} \Bigg\|_{L^2(P;\|\cdot\|_{\mathbb{R}^k})} \\ &\leq \max\big\{T^{\max_{\alpha\in\mathbb{B}(\mathcal{A})}|\alpha|_{\mathbb{N}_0}}, 1\big\} \left[\sum_{\alpha\in\mathbb{B}(\mathcal{A})} \sup_{u\in[t_0,s_0]} \big\| \big(L^{\alpha_{|\alpha|}}_{\mu,\sigma} \dots L^{\alpha_2}_{\mu,\sigma} L^{\alpha_1}_{\mu,\sigma} f\big)\big(X_u\big)\big\|_{L^2(P;\|\cdot\|_{\mathbb{R}^k})}\right] \\ &\quad \cdot [s_0 - t_0]^{\left[\frac{1}{2}\min\left\{|\alpha|_{\mathbb{N}_0} + \#_{\mathbb{R}}(\{i\in\{1,2,\dots,|\alpha|_{\mathbb{N}_0}\}\colon \alpha_i = 0\})\colon \alpha\in\mathbb{B}(\mathcal{A})\right\}\right]}\end{aligned} \tag{5.141}$$

and

$$\begin{aligned}&\Bigg\| \mathbb{E}_P[f(X_{s_0})] - \mathbb{E}_P[f(X_{t_0})] - \sum_{\alpha\in\mathcal{A}\setminus\{\emptyset\}} \mathbb{E}_P\Bigg[\big(L^{\alpha_{|\alpha|_{\mathbb{N}_0}}}_{\mu,\sigma} \dots L^{\alpha_2}_{\mu,\sigma} L^{\alpha_1}_{\mu,\sigma} f\big)\big(X_{t_0}\big) \\ &\quad \cdot \int_{t_0}^{s_0}\int_{t_0}^{s_1}\dots\int_{t_0}^{s_{|\alpha|_{\mathbb{N}_0}-1}} dW^{(\alpha_{|\alpha|_{\mathbb{N}_0}})}_{s_{|\alpha|_{\mathbb{N}_0}}} \dots dW^{(\alpha_2)}_{s_2}\, dW^{(\alpha_1)}_{s_1}\Bigg] \Bigg\|_{\mathbb{R}^k} \\ &\leq \left[\sum_{\substack{\alpha\in\mathbb{B}(\mathcal{A}),\\ \alpha_1=\ldots=\alpha_{|\alpha|_{\mathbb{N}_0}}=0}} \sup_{u\in[t_0,s_0]} \big\| \mathbb{E}_P\big[\big(L^{\alpha_{|\alpha|_{\mathbb{N}_0}}}_{\mu,\sigma} \dots L^{\alpha_2}_{\mu,\sigma} L^{\alpha_1}_{\mu,\sigma} f\big)\big(X_u\big)\big]\big\|_{\mathbb{R}^k}\right] \\ &\quad \cdot \max\big\{T^{\max_{\alpha\in\mathbb{B}(\mathcal{A})}|\alpha|_{\mathbb{N}_0}}, 1\big\} \cdot [s_0 - t_0]^{\min\left\{|\alpha|_{\mathbb{N}_0}\colon \alpha\in\mathbb{B}(\mathcal{A}),\, \alpha_1=\ldots=\alpha_{|\alpha|_{\mathbb{N}_0}}=0\right\}}.\end{aligned} \tag{5.142}$$

Proof of Corollary 5.6.15. Observe that for all $\alpha \in \mathbb{M}_m \backslash \{\emptyset\}$ it holds that

$$\begin{aligned}
&\left\| \int_{t_0}^{s_0} \ldots \int_{t_0}^{s_{|\alpha|_{\mathbb{N}_0}-1}} \big(L_{\mu,\sigma}^{\alpha_{|\alpha|_{\mathbb{N}_0}}} \ldots L_{\mu,\sigma}^{\alpha_2} L_{\mu,\sigma}^{\alpha_1} f\big)\big(X_{s_{|\alpha|}}\big)\, dW_{s_{|\alpha|_{\mathbb{N}_0}}}^{(\alpha_{|\alpha|_{\mathbb{N}_0}})} \ldots dW_{s_1}^{(\alpha_1)} \right\|_{L^2(P;\|\cdot\|_{\mathbb{R}^k})}^2 \\
&\leq \left[\int_{t_0}^{s_0} \ldots \int_{t_0}^{s_{|\alpha|_{\mathbb{N}_0}-1}} \big\| \big(L_{\mu,\sigma}^{\alpha_{|\alpha|_{\mathbb{N}_0}}} \ldots L_{\mu,\sigma}^{\alpha_2} L_{\mu,\sigma}^{\alpha_1} f\big)\big(X_{s_{|\alpha|_{\mathbb{N}_0}}}\big) \big\|_{L^2(P;\|\cdot\|_{\mathbb{R}^k})}^2 \, ds_{|\alpha|_{\mathbb{N}_0}} \ldots ds_1 \right] \\
&\cdot [s_0 - t_0]^{\#_{\mathbb{R}}(\{i \in \{1,2,\ldots,|\alpha|_{\mathbb{N}_0}\} \colon \alpha_i = 0\})} \\
&\leq \left[\sup_{u \in [t_0,s_0]} \big\| \big(L_{\mu,\sigma}^{\alpha_{|\alpha|}} \ldots L_{\mu,\sigma}^{\alpha_2} L_{\mu,\sigma}^{\alpha_1} f\big)\big(X_u\big) \big\|_{L^2(P;\|\cdot\|_{\mathbb{R}^k})}^2 \right] \\
&\cdot [s_0 - t_0]^{|\alpha|_{\mathbb{N}_0} + \#_{\mathbb{R}}(\{i \in \{1,2,\ldots,|\alpha|_{\mathbb{N}_0}\} \colon \alpha_i = 0\})}.
\end{aligned} \tag{5.143}$$

Moreover, note for all $\alpha \in \mathcal{A} \cup \mathbb{B}(\mathcal{A})$ that

$$\begin{aligned}
&\left\| \mathbb{E}_P \left[\int_{t_0}^{s_0} \int_{t_0}^{s_1} \ldots \int_{t_0}^{s_{|\alpha|_{\mathbb{N}_0}-1}} \big(L_{\mu,\sigma}^{\alpha_{|\alpha|_{\mathbb{N}_0}}} \ldots L_{\mu,\sigma}^{\alpha_2} L_{\mu,\sigma}^{\alpha_1} f\big)\big(X_{s_{|\alpha|_{\mathbb{N}_0}}}\big)\, dW_{s_{|\alpha|_{\mathbb{N}_0}}}^{(\alpha_{|\alpha|})} \ldots dW_{s_2}^{(\alpha_2)}\, dW_{s_1}^{(\alpha_1)} \right] \right\|_{\mathbb{R}^k} \\
&\leq \mathbb{1}_{\{0\}}\Big(\textstyle\sum_{i=1}^{|\alpha|_{\mathbb{N}_0}} \alpha_i \Big) \\
&\cdot \int_{t_0}^{s_0} \int_{t_0}^{s_1} \ldots \int_{t_0}^{s_{|\alpha|_{\mathbb{N}_0}-1}} \Big\| \mathbb{E}_P\Big[\big(L_{\mu,\sigma}^{\alpha_{|\alpha|_{\mathbb{N}_0}}} \ldots L_{\mu,\sigma}^{\alpha_2} L_{\mu,\sigma}^{\alpha_1} f\big)\big(X_{s_{|\alpha|_{\mathbb{N}_0}}}\big) \Big] \Big\|_{\mathbb{R}^k} \, ds_{|\alpha|_{\mathbb{N}_0}} \ldots ds_2\, ds_1.
\end{aligned} \tag{5.144}$$

Combining (5.143) and (5.144) with Theorem 5.6.14 completes the proof of Corollary 5.6.15. □

In the next step we illustrate Corollary 5.6.15 in the case of the examples (5.95), (5.99), (5.103), and (5.139) above. If it holds for all $\alpha = (\alpha_1, \ldots, \alpha_{|\alpha|_{\mathbb{N}_0}}) \in \mathbb{M}_{\{0,1,\ldots,m\}} \backslash \{\emptyset\}$ that

$$\sup_{u \in [0,T]} \big\| \big(L_{\mu,\sigma}^{\alpha_{|\alpha|}} \ldots L_{\mu,\sigma}^{\alpha_2} L_{\mu,\sigma}^{\alpha_1} \mathrm{id}_O\big)\big(X_u\big) \big\|_{L^2(P;\|\cdot\|_{\mathbb{R}^d})} < \infty \tag{5.145}$$

(this is, e.g., fulfilled if $\forall\, p \in (0,\infty) \colon \sup_{u \in [0,T]} \|X_t\|_{L^p(P;\|\cdot\|_{\mathbb{R}^d})} < \infty$ and if μ and σ are infinitely often differentiable with at most polynomially growing derivatives), then Corollary 5.6.15 proves that

$$\sup_{s_0 \in (t_0,T]} \left(\frac{\|X_{s_0} - X_{t_0}\|_{L^2(P;\|\cdot\|_{\mathbb{R}^k})}}{(s_0 - t_0)^{\frac{1}{2}}} \right) < \infty \tag{5.146}$$

(Corollary 5.6.15 applied to the hierarchical set $\mathcal{A} = \{\emptyset\}$; see (5.130)), that

$$\sup_{t \in (t_0,T]} \left(\frac{\big\| X_t - \big[X_{t_0} + \mu(X_{t_0})\,(t - t_0) + \sigma(X_{t_0})\,(W_t - W_{t_0})\big] \big\|_{L^2(P;\|\cdot\|_{\mathbb{R}^k})}}{(t - t_0)} \right) < \infty \tag{5.147}$$

(Corollary 5.6.15 applied to the hierarchical set $\mathcal{A} = \{\emptyset, 0, 1, \dots, m\}$; see (5.132)), that

$$\sup_{t\in(t_0,T]}\left(\frac{\left\|X_t - \left[\begin{array}{c} X_{t_0} + \mu(X_{t_0})\,(t-t_0) + \sigma(X_{t_0})\,(W_t - W_{t_0}) \\ + \sum_{i,j=1}^{m} \sigma_i'(X_{t_0})\,\sigma_j(X_{t_0}) \int_{t_0}^{t}\int_{t_0}^{s} dW_u^{(j)} dW_s^{(i)} \end{array}\right]\right\|_{L^2(P;\|\cdot\|_{\mathbb{R}^k})}}{(t-t_0)^{\frac{3}{2}}}\right) < \infty \tag{5.148}$$

(Corollary 5.6.15 applied to the hierarchical set $\mathcal{A} = \{\emptyset, 0, 1, \dots, m\} \cup \{1, 2, \dots, m\}^2$; see (5.134)) and that

$$\sup_{t\in(t_0,T]}\left(\frac{\left\|X_t - \left[\begin{array}{c} X_{t_0} + \mu(X_{t_0})\,(t-t_0) + \sigma(X_{t_0})\,(W_t - W_{t_0}) \\ \sum_{i,j=0}^{m} \big(L^j_{\mu,\sigma}\, L^i_{\mu,\sigma} f\big)\big(X_{t_0}\big) \cdot \int_{t_0}^{t}\int_{t_0}^{s} dW_u^{(j)}\, dW_s^{(i)} \end{array}\right]\right\|_{L^2(P;\|\cdot\|_{\mathbb{R}^k})}}{(t-t_0)^{\frac{3}{2}}}\right) < \infty \tag{5.149}$$

(Corollary 5.6.15 applied to the hierarchical set $\mathcal{A} = \{\emptyset, 0, 1, \dots, m\} \cup \{0, 1, \dots, m\}^2$; see (5.136)). Theorem 5.6.14 and Corollary 5.6.15 motivate the following definition.

5.6.4.4 General higher order stochastic Taylor schemes

Definition 5.6.16 ($\mathcal{A}$-stochastic Taylor approximation)**.** *Let $T \in (0,\infty)$, $d, m, N \in \mathbb{N}$, $\mathcal{A} \in \mathcal{P}(\mathbb{M}_{\{0,1,\dots,m\}})$, $\mu \in C^\infty(\mathbb{R}^d, \mathbb{R}^d)$, $\sigma \in C^\infty(\mathbb{R}^d, \mathbb{R}^{d\times m})$, let $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$ be a stochastic basis, let $\xi \in \mathcal{L}^0(P|_{\mathbb{F}_0}; \|\cdot\|_{\mathbb{R}^d})$, let $W = (W^{(1)}, \dots, W^{(m)})\colon [0,T]\times\Omega \to \mathbb{R}^m$ be a standard $(\mathbb{F}_t)_{t\in[0,T]}$-Brownian motion, and let $Y\colon \{0,1,\dots,N\}\times\Omega \to \mathbb{R}^d$ be a stochastic process such that $Y_0 = \xi$ and such that for all $n \in \{0,1,\dots,N-1\}$ it holds P-a.s. that*

$$\begin{aligned} Y_{n+1} = Y_n + \sum_{\alpha\in\mathcal{A}\setminus\{\emptyset\}} &\Big(L^{\alpha_{|\alpha|_{\mathbb{M}_{\mathbb{N}_0}}}}_{\mu,\sigma} \dots L^{\alpha_2}_{\mu,\sigma}\, L^{\alpha_1}_{\mu,\sigma}\, id_{\mathbb{R}^d}\Big)\big(Y_n\big) \\ &\cdot \int_{\frac{nT}{N}}^{\frac{(n+1)T}{N}} \int_{\frac{nT}{N}}^{s_1} \dots \int_{\frac{nT}{N}}^{s_{|\alpha|_{\mathbb{M}_{\mathbb{N}_0}}-1}} dW^{(\alpha_{|\alpha|_{\mathbb{M}_{\mathbb{N}_0}}})}_{s_{|\alpha|_{\mathbb{M}_{\mathbb{N}_0}}}} \dots dW^{(\alpha_2)}_{s_2}\, dW^{(\alpha_1)}_{s_1}. \end{aligned} \tag{5.150}$$

*Then we call Y an $\mathcal{A}$-*stochastic Taylor approximation *for the SDE*

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t, \qquad t\in[0,T], \qquad X_0 = \xi \tag{5.151}$$

with time step size T/N.

Remark 5.6.17. *Let $T \in (0,\infty)$, $d,m,N \in \mathbb{N}$, $\mu \in C^\infty(\mathbb{R}^d,\mathbb{R}^d)$, $\sigma \in C^\infty(\mathbb{R}^d,\mathbb{R}^{d\times m})$, let $(\Omega,\mathcal{F},P,(\mathbb{F}_t)_{t\in[0,T]})$ be a stochastic basis, let $\xi \in \mathcal{L}^0(P|_{\mathbb{F}_0};\|\cdot\|_{\mathbb{R}^d})$, let $W = (W^{(1)},\dots,W^{(m)})\colon [0,T]\times\Omega \to \mathbb{R}^m$ be a standard $(\mathbb{F}_t)_{t\in[0,T]}$-Brownian motion, and let $X\colon [0,T]\times\Omega\to\mathbb{R}^d$ be a solution process of the SDE*

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t, \qquad t\in[0,T], \qquad X_0 = \xi. \tag{5.152}$$

Note

(i) that Euler-Maruyama approximations for the SDE (5.152) *with time step size T/N (see Definition 5.3.1) are $\{\emptyset,0,1,\dots,m\}$-stochastic Taylor approximations for the SDE* (5.152) *with time step size T/N,*

(ii) that Milstein approximations for the SDE (5.152) *with time step size T/N (see Definition 5.6.3) are $\big(\{\emptyset,0,1,\dots,m\}\cup\{1,2,\dots,m\}^2\big)$-stochastic Taylor approximations for the SDE* (5.152) *with time step size T/N,*

(iii) that $\big(\{\emptyset,0,1,\dots,m\}\cup\{0,1,\dots,m\}^2\big)$-stochastic Taylor approximations for the SDE (5.152) *with time step size T/N are also referred to as Milstein-Talay approximations for the SDE* (5.152) *with time step size T/N, and*

(iv) that $\big(\{\emptyset,0,1,\dots,m\}\cup\{0,1,\dots,m\}^2\cup\{1,2,\dots,m\}^3\big)$-stochastic Taylor approximations for the SDE (5.152) *with time step size T/N are also referred to as Wagner-Platen approximations for the SDE* (5.152) *with time step size T/N.*

5.6.5 Runge-Kutta schemes for SDEs

Similar as in the case of ordinary differential equations (ODEs), derivative-free Runge-Kutta type numerical approximation schemes for SDEs can be derived on the basis of stochastic Taylor schemes (such as the Milstein scheme and the Wagner-Platen scheme). For example, assume that the setting in Section 5.1, assume that $\bar\sigma$ is continuously differentiable, and let $Z^N\colon \{0,1,\dots,N\}\times\Omega\to\mathbb{R}^d$, $N\in\mathbb{N}$, be stochastic processes which satisfy for all $N\in\mathbb{N}$, $n\in\{0,1,\dots,N-1\}$ that $Z_0^N = \xi$ and P-a.s. that

$$\begin{aligned} Z_{n+1}^N &= Z_n^N + \bar\mu\big(Z_n^N\big)\tfrac{T}{N} + \bar\sigma\big(Z_n^N\big)\big(W_{\frac{(n+1)T}{N}} - W_{\frac{nT}{N}}\big) \\ &\quad + \sum_{i,j=1}^m \bar\sigma_i'(Z_n^N)\,\bar\sigma_j(Z_n^N)\cdot\int_{\frac{nT}{N}}^{\frac{(n+1)T}{N}}\int_{\frac{nT}{N}}^{s} dW_u^{(j)}\,dW_s^{(i)}. \end{aligned} \tag{5.153}$$

Then observe that the Taylor approximation

$$\frac{\Big(\bar\sigma_i\Big(Y_n^N + \bar\sigma_j(Y_n^N)\frac{\sqrt{T}}{\sqrt{N}}\Big) - \bar\sigma_i(Y_n^N)\Big)}{\sqrt{\frac{T}{N}}} \approx \bar\sigma_i'(Z_n^N)\,\bar\sigma_j(Z_n^N) \tag{5.154}$$

for $n \in \{0, 1, \ldots, N-1\}$, $N \in \mathbb{N}$ suggests the approximation

$$\begin{aligned} Z_{n+1}^N &\approx Z_n^N + \bar{\mu}\big(Z_n^N\big)\tfrac{T}{N} + \bar{\sigma}\big(Z_n^N\big)\big(W_{\frac{(n+1)T}{N}} - W_{\frac{nT}{N}}\big) \\ &+ \sqrt{\tfrac{N}{T}} \sum_{i,j=1}^m \Big(\bar{\sigma}_i\Big(Y_n^N + \bar{\sigma}_j(Y_n^N)\tfrac{\sqrt{T}}{\sqrt{N}}\Big) - \bar{\sigma}_i(Y_n^N)\Big) \int_{\frac{nT}{N}}^{\frac{(n+1)T}{N}} \int_{\frac{nT}{N}}^{s} dW_u^{(j)}\, dW_s^{(i)} \end{aligned} \tag{5.155}$$

for $n \in \{0, 1, \ldots, N-1\}$, $N \in \mathbb{N}$. This motivates the following definition (see, e.g., Section 11.1 in [Kloeden and Platen(1992)]).

Definition 5.6.18 (Milstein-Runge-Kutta approximation). *Let $T \in (0, \infty)$, $d, m, N \in \mathbb{N}$, $\mu \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{R}^d))$, $\sigma = (\sigma_j)_{j \in \{1,\ldots,m\}} \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{R}^{d \times m}))$, let $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t \in [0,T]})$ be a stochastic basis, let $\xi \in \mathcal{L}^0(P|_{\mathbb{F}_0}; \|\cdot\|_{\mathbb{R}^d})$, let $W = (W^{(1)}, \ldots, W^{(m)})\colon [0,T] \times \Omega \to \mathbb{R}^m$ be a standard $(\mathbb{F}_t)_{t \in [0,T]}$-Brownian motion, and let $Y\colon \{0, 1, \ldots, N\} \times \Omega \to \mathbb{R}^d$ be a stochastic process such that $Y_0 = \xi$ and such that for all $n \in \{0, 1, \ldots, N-1\}$ it holds P-a.s. that*

$$\begin{aligned} Y_{n+1} &= Y_n + \mu(Y_n)\tfrac{T}{N} + \sigma(Y_n)\big(W_{\frac{(n+1)T}{N}} - W_{\frac{nT}{N}}\big) \\ &+ \tfrac{\sqrt{N}}{\sqrt{T}} \sum_{i,j=1}^m \Big[\sigma_i\Big(Y_n + \sigma_j(Y_n)\tfrac{\sqrt{T}}{\sqrt{N}}\Big) - \sigma_i(Y_n)\Big] \int_{\frac{nT}{N}}^{\frac{(n+1)T}{N}} \int_{\frac{nT}{N}}^{s} dW_u^{(j)}\, dW_s^{(i)}. \end{aligned} \tag{5.156}$$

Then we call Y a Milstein-Runge-Kutta approximation *for the SDE*

$$dX_t = \mu(X_t)\, dt + \sigma(X_t)\, dW_t, \qquad t \in [0,T], \qquad X_0 = \xi \tag{5.157}$$

with time step size T/N.

Definition 5.6.19 (Linearly-interpolated Milstein-Runge-Kutta approximation). *Let $T \in (0, \infty)$, $d, m, N \in \mathbb{N}$, $\mu \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{R}^d))$, $\sigma = (\sigma_j)_{j \in \{1,\ldots,m\}} \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{R}^{d \times m}))$, let $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t \in [0,T]})$ be a stochastic basis, let $\xi \in \mathcal{L}^0(P|_{\mathbb{F}_0}; \|\cdot\|_{\mathbb{R}^d})$, let $W\colon [0,T] \times \Omega \to \mathbb{R}^m$ be a standard $(\mathbb{F}_t)_{t \in [0,T]}$-Brownian motion, and let $Y\colon [0,T] \times \Omega \to \mathbb{R}^d$ be a stochastic process with continuous sample paths such that $Y_0 = \xi$ and such that for all $n \in \{0, 1, \ldots, N-1\}$, $t \in \big[\frac{nT}{N}, \frac{(n+1)T}{N}\big]$ it holds P-a.s. that*

$$Y_t = Y_{\frac{nT}{N}} + \big(\tfrac{tN}{T} - n\big)\Big[\mu\big(Y_{\frac{nT}{N}}\big)\tfrac{T}{N} + \sigma\big(Y_{\frac{nT}{N}}\big)\big(W_{\frac{(n+1)T}{N}} - W_{\frac{nT}{N}}\big)\Big] \tag{5.158}$$
$$+ \big(\tfrac{tN}{T} - n\big)\left[\tfrac{\sqrt{N}}{\sqrt{T}} \sum_{i,j=1}^m \Big[\sigma_i\Big(Y_n + \sigma_j(Y_n)\tfrac{\sqrt{T}}{\sqrt{N}}\Big) - \sigma_i(Y_n)\Big] \int_{\frac{nT}{N}}^{\frac{(n+1)T}{N}} \int_{\frac{nT}{N}}^{s} dW_u^{(j)}\, dW_s^{(i)}\right].$$

Then we call Y a linearly-interpolated Milstein-Runge-Kutta approximation *for the SDE*

$$dX_t = \mu(X_t)\, dt + \sigma(X_t)\, dW_t, \qquad t \in [0,T], \qquad X_0 = \xi \tag{5.159}$$

with time step size T/N.

Further derivative free approximation schemes for SDEs of this type can, e.g., be found in [Kloeden and Platen(1992)] and [Rößler(2010)] and in the references mentioned therein.

6 Weak approximations for SDEs

In Chapter 5 it has, under suitable assumptions, been demonstrated that the Euler-Maruyama approximations (and other schemes) converge, under suitable assmptions, in the strong L^p-sense with order $\frac{1}{2}$ to the solution process of the SDE (5.2). This chapter, in particular, shows that the Euler-Maruyama converge, under suitable assumptions, in the numerically weak sense with order 1 to the solution process of the SDE (5.2); see Theorem 6.2.4 below. The proof of this theorem uses the *deterministic* Kolmogorov partial differential equation. This topic will be considered first; see Subsection 6.1. The content of this chapter can, e.g., be found in a bit different form in [Kloeden and Platen(1992)].

6.1 Kolmogorov backward equation

Theorem 6.1.1 (Kolmogorov backward equation)**.** *Assume the setting in the beginning of Chapter 5, let $p \in [3, \infty)$, let $f\colon \mathbb{R}^d \to \mathbb{R}$ be a three times continuously differentiable function with*

$$\sup_{x \in \mathbb{R}^d} \frac{\left\| f^{(3)}(x) \right\|_{L^{(3)}(\mathbb{R}^d, \mathbb{R})}}{(1 + \|x\|_{\mathbb{R}^d})^{(p-3)}} < \infty, \tag{6.1}$$

assume that $O = \mathbb{R}^d$ and that μ and σ are globally Lipschitz continuous and three times continuously differentiable with at most polynomially growing derivatives. Moreover, let $X^{t_0,x}\colon [t_0, T] \times \Omega \to \mathbb{R}^d$, $t_0 \in [0, T]$, $x \in \mathbb{R}^d$, be up to indistinguishability unique solution processes of the SDEs

$$dX_t = \mu(X_t)\, dt + \sigma(X_t)\, dW_t, \qquad t \in [t_0, T], \qquad X_{t_0} = x \tag{6.2}$$

for $(t_0, x) \in [0, T] \times \mathbb{R}^d$, that is, assume for every $t_0 \in [0, T]$ and every $x \in \mathbb{R}^d$ that $X^{t_0,x}$ is an $(\mathbb{F}_t)_{t \in [t_0,T]}$-adapted stochastic processes with continuous sample paths satisfying that for every $t \in [t_0, T]$ it holds that

$$X_t^{t_0,x} = x + \int_{t_0}^{t} \mu(X_s^{t_0,x})\, ds + \int_{t_0}^{t} \sigma(X_s^{t_0,x})\, dW_s \tag{6.3}$$

P-a.s. Then the function $u\colon [0, T] \times \mathbb{R}^d \to \mathbb{R}$ given by

$$u(t, x) = \mathbb{E}\big[f(X_T^{t,x})\big] \tag{6.4}$$

for all $(t, x) \in [0, T] \times \mathbb{R}^d$ is the unique function satisfying

$$\sup_{s \in [0,T]} \sup_{x \in \mathbb{R}^d} \frac{|u(s, x)|}{(1 + \|x\|_{\mathbb{R}^d})^p} < \infty, \tag{Growth condition}$$

$$u(T, x) = f(x) \tag{End value condition}$$

and

$$\begin{aligned} \tfrac{\partial}{\partial t} u(t, x) &= -\frac{1}{2} \sum_{i=1}^{m} \sum_{k,l=1}^{d} \sigma_{k,i}(x) \cdot \sigma_{l,i}(x) \cdot \tfrac{\partial^2}{\partial x_k \partial x_l} u(t, x) - \sum_{k=1}^{d} \mu_k(x) \cdot \tfrac{\partial}{\partial x_k} u(t, x) \\ &= -\frac{1}{2} \operatorname{Trace}\big(\sigma(x)\, \sigma(x)^{\top} (\operatorname{Hess}_x u)(t, x)\big) - \langle \mu(x), (\nabla_x u)(t, x) \rangle_{\mathbb{R}^d} \\ &= -\big(L^0 u(t, \cdot)\big)(x) \end{aligned}$$

(Kolmogorov partial differential equation (Kolmogorov PDE))

for all $t \in [0, T]$ and all $x \in \mathbb{R}^d$.

Sketch of the proof of Theorem 6.1.1. First of all, observe that for all $x \in \mathbb{R}^d$ it holds

that

$$u(T,x) = \mathbb{E}\Big[f(X_T^{T,x})\Big] = \mathbb{E}[f(x)] = f(x) \tag{6.5}$$

and this proves (End value condition). Next define a real number $c \in [0,\infty)$ by

$$c := \sup_{x\in\mathbb{R}^d} \frac{|f(x)|}{\big(1+\|x\|_{\mathbb{R}^d}^p\big)} < \infty \tag{6.6}$$

and observe that for all $x \in \mathbb{R}^d$ it holds that

$$|f(x)| \leq c\,\big(1+\|x\|_{\mathbb{R}^d}^p\big) \tag{6.7}$$

This implies that

$$\begin{aligned}
\sup_{t\in[0,T]}\sup_{x\in\mathbb{R}^d} \frac{|u(t,x)|}{(1+\|x\|_{\mathbb{R}^d})^p} &= \sup_{t\in[0,T]}\sup_{x\in\mathbb{R}^d} \frac{\big|\mathbb{E}\big[f(X_T^{t,x})\big]\big|}{(1+\|x\|_{\mathbb{R}^d})^p} \\
&\leq c\left(\sup_{t\in[0,T]}\sup_{x\in\mathbb{R}^d} \frac{\big(1+\mathbb{E}\big[\|X_T^{t,x}\|_{\mathbb{R}^d}^p\big]\big)}{(1+\|x\|_{\mathbb{R}^d})^p}\right) < \infty
\end{aligned} \tag{6.8}$$

and this shows (Growth condition). Next w.l.o.g. assume that for every $t_0, t \in [0,T]$ with $t_0 \leq t$ and every $\omega \in \Omega$ it holds that the function $\mathbb{R}^d \ni x \mapsto X_t^{t_0,x}(\omega) \in \mathbb{R}^d$ is continuous. Then observe that

$$X_{t_2}^{t_0,x} = X_{t_2}^{t_1,X_{t_1}^{t_0,x}} \tag{6.9}$$

P-a.s. for all $t_0,t_1,t_2 \in [0,T]$ with $t_0 \leq t_1 \leq t_2$ and all $x \in \mathbb{R}^d$. This illustrates that for all $t,h \in [0,T]$ with $t+h \leq T$ and all $x \in \mathbb{R}^d$ it holds that

$$u(t,x) = \mathbb{E}\big[f\big(X_T^{t,x}\big)\big] = \mathbb{E}\left[f\left(X_T^{t+h,X_{t+h}^{t,x}}\right)\right] = \mathbb{E}\big[u(t+h,X_{t+h}^{t,x})\big]\,. \tag{6.10}$$

Itô's formula hence proves that for all $t,h \in [0,T]$ with $t+h \leq T$ and all $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&u(t+h,x) - u(t,x) = u(t+h,x) - \mathbb{E}\big[u(t+h,X_{t+h}^{t,x})\big] \\
&= -\int_t^{t+h} \sum_{k=1}^d \mathbb{E}\Big[\Big(\tfrac{\partial}{\partial x_k}u\Big)(t+h,X_s^{t,x})\cdot\mu_k(X_s^{t,x})\Big]\,ds \\
&\quad -\int_t^{t+h} \frac{1}{2}\sum_{i=1}^m\sum_{k,l=1}^d \mathbb{E}\Big[\Big(\tfrac{\partial^2}{\partial x_k\partial x_l}u\Big)(t+h,X_s^{t,x})\cdot\sigma_{k,i}(X_s^{t,x})\cdot\sigma_{l,i}(X_s^{t,x})\Big]\,ds.
\end{aligned} \tag{6.11}$$

This implies that for all $t \in [0,T)$ and all $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&\tfrac{\partial}{\partial t}u(t,x) = \lim_{h\searrow 0}\left(\frac{u(t+h,x)-u(t,x)}{h}\right) \\
&= -\sum_{k=1}^d \Big(\tfrac{\partial}{\partial x_k}u\Big)(t,x)\cdot\mu_k(x) - \frac{1}{2}\sum_{i=1}^m\sum_{k,l=1}^d \Big(\tfrac{\partial^2}{\partial x_k\partial x_l}u\Big)(t,x)\cdot\sigma_{k,i}(x)\cdot\sigma_{l,i}(x).
\end{aligned} \tag{6.12}$$

This proves (Kolmogorov partial differential equation (Kolmogorov PDE)). The fact that u is the *unique* function satisfying (Growth condition)–(Kolmogorov partial differential equation (Kolmogorov PDE)) follows, e.g., from Corollary 4.7 in [Hairer et al.(2012)Hairer, Hutzenthaler, and Jentzen]. This completes the sketch of the proof of Theorem 6.1.1. □

In the setting of Theorem 6.1.1 it holds that the function $v\colon [0,T] \times \mathbb{R}^d \to \mathbb{R}$ given by

$$v(t,x) = u(T-t,x) = \mathbb{E}\big[f(X_t^{0,x})\big] \tag{6.13}$$

for all $(t,x) \in [0,T] \times \mathbb{R}^d$ is the unique function satisfying

$$\sup_{t\in[0,T]} \sup_{x\in\mathbb{R}^d} \frac{|v(t,x)|}{(1+\|x\|_{\mathbb{R}^d})^p} < \infty \tag{6.14}$$

and

$$v(0,x) = f(x) \tag{Initial value condition}$$

for all $x \in \mathbb{R}^d$ and

$$\begin{aligned} \tfrac{\partial}{\partial t} v(t,x) &= \frac{1}{2} \sum_{i=1}^{m} \sum_{k,l=1}^{d} \sigma_{k,i}(x) \cdot \sigma_{l,i}(x) \cdot \tfrac{\partial^2}{\partial x_k \partial x_l} v(t,x) + \sum_{k=1}^{d} \mu_k(x) \cdot \tfrac{\partial}{\partial x_k} v(t,x) \\ &= \frac{1}{2} \mathrm{Trace}\big(\sigma(x)\,\sigma(x)^\top\, (\mathrm{Hess}_x v)(t,x)\big) + \langle \mu(x), (\nabla_x v)(t,x)\rangle_{\mathbb{R}^d} \\ &= \big(L^0 v(t,\cdot)\big)(x) \end{aligned} \tag{Kolmogorov PDE revisited}$$

for all $t \in [0,T]$ and all $x \in \mathbb{R}^d$.

6.2 Numerically weak convergence of the Euler-Maruyama scheme

6.2.1 Generator and noise operator associated to the Euler scheme

Definition 6.2.1 (Generator associated to the Euler-Maruyama scheme)**.** *Let $d, m \in \mathbb{N}$, let $O \subseteq \mathbb{R}^d$ be an open set, and let $\mu \colon D \to \mathbb{R}^d$ and $\sigma \colon D \to \mathbb{R}^{d\times m}$ be continuous functions. Then we denote by*

$$\hat{L}^0_{\mu,\sigma} \colon \cup_{k\in\mathbb{N}} C^2(O, \mathbb{R}^k) \to \cup_{k\in\mathbb{N}} C(O \times O, \mathbb{R}^k) \tag{6.15}$$

the function with the property that for all $x \in O$ it holds that

$$\begin{aligned} \big(\hat{L}^0_{\mu,\sigma} f\big)(x, y) &= f'(x)\, \mu(y) + \frac{1}{2} \sum_{i=1}^{m} f''(x)\big(\sigma_i(y), \sigma_i(y)\big) \\ &= \sum_{k=1}^{d} \big(\tfrac{\partial}{\partial x_k} f\big)(x) \cdot \mu_k(y) + \frac{1}{2} \sum_{i=1}^{m} \sum_{k,l=1}^{d} \big(\tfrac{\partial^2}{\partial x_k \partial x_l} f\big)(x) \cdot \sigma_{k,i}(y) \cdot \sigma_{l,i}(y). \end{aligned} \tag{6.16}$$

Definition 6.2.2 (Noise operator associated to the Euler-Maruyama scheme)**.** *Let $d, m \in \mathbb{N}$, let $O \subseteq \mathbb{R}^d$ be an open set, and let $\mu \colon D \to \mathbb{R}^d$ and $\sigma \colon D \to \mathbb{R}^{d\times m}$ be continuous functions. Then we denote by*

$$\hat{L}^i_{\mu,\sigma} \colon \cup_{k\in\mathbb{N}} C^1(O, \mathbb{R}^k) \to \cup_{k\in\mathbb{N}} C(O \times O, \mathbb{R}^k) \tag{6.17}$$

for $i \in \{1, 2, \dots, m\}$ the functions with the property that for all $x \in O$, $i \in \{1, 2, \dots, m\}$ it holds that

$$\big(\hat{L}^i_{\mu,\sigma} f\big)(x, y) = f'(x)\, \sigma_i(y) = \sum_{k=1}^{d} \big(\tfrac{\partial}{\partial x_k} f\big)(x) \cdot \sigma_{k,i}(y). \tag{6.18}$$

6.2.2 Weak error representation for the Euler scheme

Lemma 6.2.3 (Weak error representation for the Euler-Maruyama scheme)**.** *Let* $T \in (0,\infty)$, $d, m, N \in \mathbb{N}$, *let* $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$ *be a stochastic basis, let* $W\colon [0,T]\times\Omega \to \mathbb{R}^m$ *be a standard* $(\mathbb{F}_t)_{t\in[0,T]}$*-Brownian motion, let* $\xi \in \cap_{p\in(0,\infty)}\mathcal{L}^p(P|_{\mathbb{F}_0}; \|\cdot\|_{\mathbb{R}^d})$, *let* $f\colon \mathbb{R}^d \to \mathbb{R}$, $\mu\colon \mathbb{R}^d \to \mathbb{R}^d$, *and* $\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d\times m}$ *be four times continuously differentiable functions with at most polynomially growing derivatives, assume that* μ *and* σ *are globally Lipschitz continuous, let* $X\colon [0,T]\times\Omega \to \mathbb{R}^d$ *be a solution process of the SDE*

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t, \qquad t\in[0,T], \qquad X_0 = \xi, \tag{6.19}$$

let $u \in C^2([0,T]\times\mathbb{R}^d, \mathbb{R})$ *be a function with the property that for all* $t\in[0,T]$, $x\in\mathbb{R}^d$ *it holds that* $u(T,x) = f(x)$, $\limsup_{p\nearrow\infty}\sup_{s\in[0,T]}\sup_{y\in\mathbb{R}^d}\frac{|u(s,y)|}{(1+\|y\|_{\mathbb{R}^d})^p} < \infty$, *and*

$$\begin{aligned}
\tfrac{\partial}{\partial t}u(t,x) &= -\frac{1}{2}\sum_{i=1}^m\sum_{k,l=1}^d \sigma_{k,i}(x)\cdot\sigma_{l,i}(x)\cdot\tfrac{\partial^2}{\partial x_k\partial x_l}u(t,x) - \sum_{k=1}^d \mu_k(x)\cdot\tfrac{\partial}{\partial x_k}u(t,x)\\
&= -\frac{1}{2}\operatorname{Trace}\big(\sigma(x)\,\sigma(x)^\top(\operatorname{Hess}_x u)(t,x)\big) - \langle \mu(x), (\nabla_x u)(t,x)\rangle_{\mathbb{R}^d}\\
&= -\big(L^0 u(t,\cdot)\big)(x),
\end{aligned}\tag{6.20}$$

and let $Y\colon [0,T]\times\Omega \to \mathbb{R}^d$ *be an* $(\mathbb{F}_t)_{t\in[0,T]}$*-adapted stochastic processes with continuous sample paths which satisfies that for all* $t\in[0,T]$ *it holds* P*-a.s. that*

$$Y_t = \xi + \int_0^t \mu\big(Y_{\lfloor s\rfloor_{T/N}}\big)\,ds + \int_0^t \sigma\big(Y_{\lfloor s\rfloor_{T/N}}\big)\,dW_s. \tag{6.21}$$

Then

$$\begin{aligned}
&\mathbb{E}\big[f(Y_T)\big] - \mathbb{E}\big[f(X_T)\big]\\
&= \int_0^T \int_{\lfloor t\rfloor_{T/N}}^t \mathbb{E}\Big[\big(\hat{L}^0_{\mu,\sigma}L^0_{\mu,\sigma}u_t\big)(Y_s, Y_{\lfloor t\rfloor_{T/N}}) + \big(\hat{L}^0_{\mu,\sigma}\big[\big(\hat{L}^0_{\mu,\sigma}u_t\big)(\cdot, Y_{\lfloor t\rfloor_{T/N}})\big]\big)(Y_s, Y_{\lfloor t\rfloor_{T/N}})\Big]\,ds\,dt.
\end{aligned}\tag{6.22}$$

Proof of Lemma 6.2.3. Theorem 6.1.1 and Itô's formula for time-dependent test functions (see Corollary 3.5.8) proves that

$$\begin{aligned}
&\mathbb{E}\big[f(Y_T)\big] - \mathbb{E}\big[f(X_T)\big] \\
&= \mathbb{E}[f(Y_T) - u(0, X_0)] \\
&= \mathbb{E}[u(T, Y_T) - u(0, Y_0)] \\
&= \mathbb{E}\bigg[\int_0^T \big(\tfrac{\partial}{\partial t} u\big)(t, Y_t^N) + \big(\tfrac{\partial}{\partial x} u\big)(t, Y_t^N)\, \mu(Y^N_{\lfloor t \rfloor_{T/N}}) \\
&\quad + \frac{1}{2} \sum_{i=1}^m \big(\tfrac{\partial^2}{\partial x^2} u\big)(t, Y_t)\big(\sigma_i(Y_{\lfloor t \rfloor_{T/N}}), \sigma_i(Y_{\lfloor t \rfloor_{T/N}})\big)\, dt\bigg] \\
&= \int_0^T \mathbb{E}\Big[\big(\tfrac{\partial}{\partial t} u\big)(t, Y_t^N) + \big(\hat{L}^0_{\mu,\sigma} u_t\big)(Y_t, Y_{\lfloor t \rfloor_{T/N}})\Big]\, dt.
\end{aligned} \tag{6.23}$$

This and (6.20) show that

$$\mathbb{E}\big[f(X_T)\big] - \mathbb{E}\big[f(Y_T)\big] = \int_0^T \mathbb{E}\Big[\big(L^0_{\mu,\sigma} u_t\big)(Y_t) - \big(\hat{L}^0_{\mu,\sigma} u_t\big)(Y_t, Y_{\lfloor t \rfloor_{T/N}})\Big]\, dt. \tag{6.24}$$

Therefore, we obtain that

$$\begin{aligned}
&\mathbb{E}\big[f(X_T)\big] - \mathbb{E}\big[f(Y_T)\big] \\
&= \int_0^T \mathbb{E}\Big[\big(L^0_{\mu,\sigma} u_t\big)(Y_t) - \big(L^0_{\mu,\sigma} u_t\big)(Y_{\lfloor t \rfloor_{T/N}})\Big]\, dt \\
&\quad + \int_0^T \mathbb{E}\Big[\big(\hat{L}^0_{\mu,\sigma} u_t\big)(Y_{\lfloor t \rfloor_{T/N}}, Y_{\lfloor t \rfloor_{T/N}}) - \big(\hat{L}^0_{\mu,\sigma} u_t\big)(Y_t, Y_{\lfloor t \rfloor_{T/N}})\Big]\, dt.
\end{aligned} \tag{6.25}$$

Again Itô's formula for time-independent test functions (see Corollary 3.5.8) hence shows that

$$\begin{aligned}
&\mathbb{E}\big[f(X_T)\big] - \mathbb{E}\big[f(Y_T)\big] \\
&= \int_0^T \int_{\lfloor t \rfloor_{T/N}}^t \mathbb{E}\Big[\big(\hat{L}^0_{\mu,\sigma} L^0_{\mu,\sigma} u_t\big)(Y_s, Y_{\lfloor t \rfloor_{T/N}})\Big]\, ds\, dt \\
&\quad + \int_0^T \int_{\lfloor t \rfloor_{T/N}}^t \mathbb{E}\Big[\Big(\hat{L}^0_{\mu,\sigma}\big[\big(\hat{L}^0_{\mu,\sigma} u_t\big)(\cdot, Y_{\lfloor t \rfloor_{T/N}})\big]\Big)(Y_s, Y_{\lfloor t \rfloor_{T/N}})\Big]\, ds\, dt.
\end{aligned} \tag{6.26}$$

The proof of Lemma 6.2.3 is thus completed. □

6.2.3 Weak convergence analysis

Theorem 6.2.4 (Numerically weak convergence with order 1 of the Euler-Maruyama scheme). *Let* $T \in (0,\infty)$, $d, m \in \mathbb{N}$, *let* $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$ *be a stochastic basis, let* $W\colon [0,T]\times\Omega \to \mathbb{R}^m$ *be a standard* $(\mathbb{F}_t)_{t\in[0,T]}$*-Brownian motion, let* $\xi \in \cap_{p\in(0,\infty)} \mathcal{L}^p(P|_{\mathbb{F}_0}; \|\cdot\|_{\mathbb{R}^d})$, *let* $f\colon \mathbb{R}^d \to \mathbb{R}$, $\mu\colon \mathbb{R}^d \to \mathbb{R}^d$, *and* $\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d\times m}$ *be four times continuously differentiable functions with at most polynomially growing derivatives, assume that* μ *and* σ *are globally Lipschitz continuous, let* $X\colon [0,T]\times\Omega \to \mathbb{R}^d$ *be a solution process of the SDE*

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t, \qquad t\in[0,T], \qquad X_0 = \xi, \tag{6.27}$$

and for every $N \in \mathbb{N}$ *let* $Y^N\colon [0,T]\times\Omega \to \mathbb{R}^d$ *be a linearly interpolated Euler-Maruyama approximation for the SDE* (6.27) *with time step size* T/N*. Then there exists a real number* $C \in \mathbb{R}$ *such that for all* $N \in \mathbb{N}$ *it holds that*

$$\big|\mathbb{E}\big[f(X_T)\big] - \mathbb{E}\big[f(Y_T^N)\big]\big|_{\mathbb{R}} \le C \cdot N^{-1}. \tag{6.28}$$

Proof of Theorem 6.2.4. Throughouht this proof let $\tilde{Y}^N\colon [0,T]\times\Omega \to \mathbb{R}^d$, $N \in \mathbb{N}$, be $(\mathbb{F}_t)_{t\in[0,T]}$-adapted stochastic processes with continuous sample paths satisfying

$$\tilde{Y}_t^N = \xi + \int_0^t \mu\big(\bar{Y}^N_{\lfloor s\rfloor_N}\big)\,ds + \int_0^t \sigma\big(\bar{Y}^N_{\lfloor s\rfloor_N}\big)\,dW_s \tag{6.29}$$

P-a.s. for all $t \in [0,T]$ and all $N \in \mathbb{N}$. Observe that $\tilde{Y}^N$, $N \in \mathbb{N}$, are *Itô processes* and that

$$\tilde{Y}_t^N = \bar{Y}_t^N \tag{6.30}$$

P-a.s. for all $t \in \{0, \frac{T}{N}, \ldots, T\}$ and all $N \in \mathbb{N}$ and that

$$\tilde{Y}_t^N = \xi + \int_0^t \mu\big(\tilde{Y}^N_{\lfloor s\rfloor_N}\big)\,ds + \int_0^t \sigma\big(\tilde{Y}^N_{\lfloor s\rfloor_N}\big)\,dW_s \tag{6.31}$$

P-a.s. for all $t \in [0,T]$ and all $N \in \mathbb{N}$. Lemma 6.2.3 implies that for all $N \in \mathbb{N}$ it holds that

$$\begin{aligned}
&\big|\mathbb{E}\big[f(X_T)\big] - \mathbb{E}\big[f(\bar{Y}_T^N)\big]\big| \\
&\le \left|\int_0^T \int_{\lfloor t\rfloor_N}^t \mathbb{E}\Big[\big(\hat{L}^0_{\mu,\sigma} L^0_{\mu,\sigma} u_t\big)(\tilde{Y}_s^N, \tilde{Y}^N_{\lfloor t\rfloor_N}) + \Big(\hat{L}^0_{\mu,\sigma}\big[\big(\hat{L}^0_{\mu,\sigma} u_t\big)(\cdot, \tilde{Y}^N_{\lfloor t\rfloor_N})\big]\Big)(\tilde{Y}_s^N, \tilde{Y}^N_{\lfloor t\rfloor_N})\Big]\,ds\,dt\right| \\
&\le \left(\sup_{s,t\in[0,T]} \Big|\mathbb{E}\Big[\big(\hat{L}^0_{\mu,\sigma} L^0_{\mu,\sigma} u_t\big)(\tilde{Y}_s^N, \tilde{Y}^N_{\lfloor t\rfloor_N}) + \Big(\hat{L}^0_{\mu,\sigma}\big[\big(\hat{L}^0_{\mu,\sigma} u_t\big)(\cdot, \tilde{Y}^N_{\lfloor t\rfloor_N})\big]\Big)(\tilde{Y}_s^N, \tilde{Y}^N_{\lfloor t\rfloor_N})\Big]\Big|\right)\frac{T^2}{N} \\
&< \infty.
\end{aligned} \tag{6.32}$$

The proof of Theorem 6.2.4 is thus completed. □

6.3 Higher order weak numerical methods

Theorem 6.3.1 (Numerically weak convergence with order 2 of the Milstein-Talay scheme). *Assume that the setting in Section 5.1, let* $\bar{Y}^N \colon [0,T] \times \Omega \to \mathbb{R}^d$*,* $N \in \mathbb{N}$*, be linearly interpolated* $(\{\emptyset, 0, 1, \ldots, m\} \cup \{0, 1, \ldots, m\}^2)$*-stochastic Taylor approximations (see Definition 5.6.16), assume that* $O = \mathbb{R}^d$*, assume that* $\forall\, p \in [1, \infty) \colon \mathbb{E}\big[\|\xi\|_{\mathbb{R}^d}^p\big] < \infty$*, assume that* $f \colon \mathbb{R}^d \to \mathbb{R}$*,* $\mu \colon \mathbb{R}^d \to \mathbb{R}^d$*, and* $\sigma \colon \mathbb{R}^d \to \mathbb{R}^{d \times m}$ *are six times continuously differentiable with at most polynomially growing derivatives, and assume that* μ *and* σ *are globally Lipschitz continuous. Then there exists a real number* $C \in [0, \infty)$ *such that for all* $N \in \mathbb{N}$ *it holds that*

$$\big|\mathbb{E}\big[f(X_T)\big] - \mathbb{E}\big[f(\bar{Y}_T^N)\big]\big|_{\mathbb{R}} \leq C \cdot N^{-2}. \tag{6.33}$$

The proof of Theorem 6.3.1 is omitted. Under a bit different assumptions, a proof of (6.33) can be found in Section 14.5 in [Kloeden and Platen(1992)] (see Theorem 14.5.2 in Chapter 14 in [Kloeden and Platen(1992)]). There also exists a simplified version of the $\{\emptyset, 0, 1, \ldots, m\} \cup \{0, 1, \ldots, m\}^2$-stochastic Taylor scheme in which the iterated stochastic integrals appearing in it are replaced by random variables that can be simulated easily. More details on this issue can be found in Section 14.2 in [Kloeden and Platen(1992)].

7 Monte Carlo integration methods for SDEs

We refer, e.g., to [Milstein(1995)] and [Glasserman(2004)] as basic references for the content of this chapter.

7.1 Monte Carlo Euler method

Combining the Monte Carlo method from Chapter 2 and the Euler-Maruyama method from Chapter 5 results in the so-called *Monte Carlo Euler method*. This is the subject of the next definition.

Definition 7.1.1 (Monte Carlo Euler method)**.** *Let $T \in (0,\infty)$, $d,m,N,K \in \mathbb{N}$, $\mu \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d),\mathcal{B}(\mathbb{R}^d))$, $\sigma \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d),\mathcal{B}(\mathbb{R}^{d\times m}))$, let $(\Omega,\mathcal{F},P,(\mathbb{F}_t)_{t\in[0,T]})$ be a stochastic basis, let $\xi \in \mathcal{M}(\mathbb{F}_0,\mathcal{B}(\mathbb{R}^d))$, let $W\colon [0,T]\times\Omega \to \mathbb{R}^m$ be an m-dimensional standard $(\Omega,\mathcal{F},P,(\mathbb{F}_t)_{t\in[0,T]})$-Brownian motion, let $X\colon [0,T]\times\Omega\to\mathbb{R}^d$ be a solution process of the SDE*

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t, \qquad t\in[0,T], \qquad X_0=\xi \tag{7.1}$$

on $(\Omega,\mathcal{F},P,(\mathbb{F}_t)_{t\in[0,T]})$, let $\xi^{[k]} \in \mathcal{M}(\mathbb{F}_0,\mathcal{B}(\mathbb{R}^d))$, $k\in\mathbb{N}$, be i.i.d. random variables with $\xi^{[1]}(P)_{\mathcal{B}(\mathbb{R}^d)} = \xi(P)_{\mathcal{B}(\mathbb{R}^d)}$, let $W^{[k]}\colon [0,T]\times\Omega\to\mathbb{R}^m$, $k\in\{1,\dots,K\}$, be P-independent m -dimensional standard $(\Omega,\mathcal{F},P,(\mathbb{F}_t)_{t\in[0,T]})$-Brownian motions, for every $k\in\{1,\dots,K\}$ let $Y^k\colon [0,T]\times\Omega\to\mathbb{R}^d$ be a linearly-interpolated Euler-Maruyama approximation for the SDE

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t^{[k]}, \qquad t\in[0,T], \qquad X_0=\xi^{[k]} \tag{7.2}$$

with time step size T/N, and let $f\colon C([0,T],\mathbb{R}^d)\to\mathbb{R}$ be a $\mathcal{B}(C([0,T],\mathbb{R}^d))/\mathcal{B}(\mathbb{R}^d)$-measurable function with $\mathbb{E}_P\big[|f(X)|_{\mathbb{R}}\big]<\infty$. Then we call the random variable

$$\frac{1}{K}\sum_{k=1}^{K} f\big(Y^k\big) \tag{7.3}$$

a Monte Carlo Euler approximation (of $\mathbb{E}_P\big[f(X)\big]$ based on K samples (Monte Carlo runs) and time step size T/N).

Exercise 7.1.2 (Monte Carlo Euler method)**.** *In this exercise we do not distinguish between pseudo random numbers and actual random numbers. Let $T\in(0,\infty)$,*

$d, m, N, K \in \mathbb{N}$, $\xi \in \mathbb{R}^d$, $\mu \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{R}^d))$, $\sigma \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{R}^{d\times m}))$, let $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$ be a stochastic basis, let $W\colon [0,T]\times\Omega \to \mathbb{R}^m$ be an m-dimensional standard $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$-Brownian motion, let $X\colon [0,T]\times\Omega\to\mathbb{R}^d$ be a solution process of the SDE

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t, \qquad t\in[0,T], \qquad X_0 = \xi \tag{7.4}$$

on $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t\in[0,T]})$ and let $f \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{R}))$ satisfy $\mathbb{E}_P\big[|f(X_T)|_{\mathbb{R}}\big] < \infty$. Write a Matlab *function* **MonteCarloEuler(T,d,m,N,K,ξ,μ,σ,f)** *with input $T\in(0,\infty)$, $d,m,N,K\in\mathbb{N}$, $\xi\in\mathbb{R}^d$, $\mu \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{R}^d))$, $\sigma \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{R}^{d\times m}))$, $f \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{R}))$ and output a realization of a Monte Carlo Euler approximation of $\mathbb{E}_P\big[f(X_T)\big]$ based on K samples and time step size T/N (see Definition 7.1.1).*

The root mean square approximation error of the Monte Carlo Euler method is estimated in the following theorem.

Theorem 7.1.3 (Monte Carlo Euler method). *Assume the setting in Section 5.1, assume that $O = \mathbb{R}^d$, let $\xi^{[k]}\colon\Omega\to\mathbb{R}^d$, $k\in\mathbb{N}$, be i.i.d. $\mathbb{F}_0/\mathcal{B}(\mathbb{R}^d)$-measurable mappings with $\xi^{[1]}(P)_{\mathcal{B}(\mathbb{R}^d)} = \xi(P)_{\mathcal{B}(\mathbb{R}^d)}$ and $\forall p\in(0,\infty)\colon \mathbb{E}_P\big[\|\xi\|^p_{\mathbb{R}^d}\big] < \infty$, let $W^{[k]}\colon[0,T]\times\Omega\to\mathbb{R}^m$, $k\in\mathbb{N}$, be independent standard $(\mathbb{F}_t)_{t\in[0,T]}$-Brownian motions, let $Y^{N,k}\colon\{0,1,\ldots,N\}\times\Omega\to\mathbb{R}^d$, $N,k\in\mathbb{N}$, be stochastic processes which satisfy for all $N,k\in\mathbb{N}$, $n\in\{0,1,\ldots,N-1\}$ that $Y_0^{N,k} = \xi^{[k]}$ and*

$$Y_{n+1}^{N,k} = Y_n^{N,k} + \mu\big(Y_n^{N,k}\big)\tfrac{T}{N} + \sigma\big(Y_n^{N,k}\big)\Big(W^{[k]}_{\frac{(n+1)T}{N}} - W^{[k]}_{\frac{nT}{N}}\Big) \tag{7.5}$$

(i.i.d. Euler-Maruyama approximations), assume that $f\colon\mathbb{R}^d\to\mathbb{R}$, $\mu\colon\mathbb{R}^d\to\mathbb{R}^d$, and $\sigma\colon\mathbb{R}^d\to\mathbb{R}^{d\times m}$ are four times continuously differentiable with at most polynomially growing derivatives, and assume that μ and σ are globally Lipschitz continuous. Then there exists a real number $C\in[0,\infty)$ such that for all $N,K\in\mathbb{N}$ it holds that

$$\left\|\mathbb{E}_P\Big[f(X_T)\Big] - \frac{1}{K}\sum_{k=1}^{K} f\big(Y_N^{N,k}\big)\right\|_{L^2(P;|\cdot|_{\mathbb{R}})} \le C\left(\frac{1}{\sqrt{K}} + \frac{1}{N}\right). \tag{7.6}$$

Proof of Theorem 7.1.3. First of all, note that Theorem 6.2.4 implies that there exists a real number $C\in[0,\infty)$ such that for all $N\in\mathbb{N}$ it holds that

$$\Big|\mathbb{E}_P\Big[f(X_T)\Big] - \mathbb{E}_P\Big[f(Y_N^{N,1})\Big]\Big|_{\mathbb{R}} \le \frac{C}{N}. \tag{7.7}$$

This together with Theorem 2.4.8 proves that for all $N, K \in \mathbb{N}$ it holds that

$$\begin{aligned}
&\mathbb{E}_P\left[\left|\mathbb{E}_P\Big[f(X_T)\Big] - \frac{1}{K}\sum_{k=1}^{K} f\big(Y_N^{N,k}\big)\right|_{\mathbb{R}}^2\right] \\
&= \underbrace{\left|\mathbb{E}_P\Big[f(X_T)\Big] - \mathbb{E}_P\Big[f\big(Y_N^{N,1}\big)\Big]\right|_{\mathbb{R}}^2}_{\text{squared bias}} + \mathbb{E}_P\left[\left|\mathbb{E}_P\Big[f\big(Y_N^{N,1}\big)\Big] - \frac{1}{K}\sum_{k=1}^{K} f\big(Y_N^{N,k}\big)\right|_{\mathbb{R}}^2\right] \\
&= \underbrace{\left|\mathbb{E}_P\Big[f(X_T)\Big] - \mathbb{E}_P\Big[f\big(Y_N^{N,1}\big)\Big]\right|_{\mathbb{R}}^2}_{\text{squared bias}} + \frac{\mathsf{Var}(f(Y_N^{N,1}))}{K} \\
&\leq \frac{C^2}{N^2} + \frac{\mathsf{Var}(f(Y_N^{N,1}))}{K} \leq \max\Big\{C^2, \sup_{M\in\mathbb{N}} \mathsf{Var}(f(Y_M^{M,1}))\Big\}\left[\frac{1}{N^2} + \frac{1}{K}\right].
\end{aligned} \tag{7.8}$$

Next note that by assumption there exists a real number $c \in [1, \infty)$ such that for all $x \in \mathbb{R}^d$ it holds that

$$|f(x)|_{\mathbb{R}} \leq c\,(1 + \|x\|_{\mathbb{R}^d}^c)\,. \tag{7.9}$$

This implies that for all $M \in \mathbb{N}$ it holds that

$$\begin{aligned}
&\sqrt{\mathsf{Var}(f(Y_M^{M,1}))} \leq \left\|f(Y_M^{M,1})\right\|_{L^2(P;|\cdot|)} \leq c\left(1 + \left\|Y_M^{M,1}\right\|_{L^{2c}(P;\|\cdot\|_{\mathbb{R}^d})}^c\right) \\
&\leq c\left[1 + \left\|Y_M^{M,1}\right\|_{L^{2c}(P;\|\cdot\|_{\mathbb{R}^d})}\right]^c \\
&\leq c\left[1 + \|X_T\|_{L^{2c}(P;\|\cdot\|_{\mathbb{R}^d})} + \left\|X_T - Y_M^{M,1}\right\|_{L^{2c}(P;\|\cdot\|_{\mathbb{R}^d})}\right]^c \\
&\leq c\left[1 + \|X_T\|_{L^{2c}(P;\|\cdot\|_{\mathbb{R}^d})} + \sup_{N\in\mathbb{N}}\left\|X_T - Y_N^{N,1}\right\|_{L^{2c}(P;\|\cdot\|_{\mathbb{R}^d})}\right]^c < \infty
\end{aligned} \tag{7.10}$$

where the last inequality follows from Theorem 5.3.10. Inequality (7.10) proves that

$$\sup_{M\in\mathbb{N}} \mathsf{Var}\Big(f(Y_M^{M,1})\Big) < \infty. \tag{7.11}$$

Combining this with (7.8) shows that for all $N, K \in \mathbb{N}$ it holds that

$$\begin{aligned}
&\left\|\mathbb{E}_P\Big[f(X_T)\Big] - \frac{1}{K}\sum_{k=1}^{K} f\big(Y_N^{N,k}\big)\right\|_{L^2(P;|\cdot|_{\mathbb{R}})} \\
&\leq \underbrace{\max\Big\{C, \sup_{M\in\mathbb{N}} \sqrt{\mathsf{Var}(f(Y_M^{M,1}))}\Big\}}_{<\infty}\left[\frac{1}{\sqrt{K}} + \frac{1}{N}\right].
\end{aligned} \tag{7.12}$$

The proof of Theorem 7.1.3 is thus completed. $\square$

The term $\frac{1}{\sqrt{K}}$ on the right hand side of (7.6) appears due to approximating the expectation in (7.6) with Monte Carlo approximations (cf. Theorem 2.4.8). The term $\frac{1}{N}$ on the right hand side of (7.6) appears due to approximating the solution process X of the SDE (5.2) with Euler-Maruyama approximations (cf. Theorem 6.2.4). The right hand side of (7.6) converges to zero as both K and N on the right hand side of (7.6) tend to ∞. In order to balance the error on the right hand side of (7.6), it turns out to be asymptotically optimal to choose $K \approx N^2$ (cf., e.g., the references mentioned in the introductory section in [Hutzenthaler and Jentzen(2011)]). The choice $K = N^2$ in (7.6) in Theorem 7.1.3 proves in the setting of Theorem 7.1.3 that there exists a real number $C \in [0, \infty)$ such that for all $N \in \mathbb{N}$ it holds that

$$\left\| \mathbb{E}_P\Big[f(X_T)\Big] - \frac{1}{N^2} \sum_{k=1}^{N^2} f\big(Y_N^{N,k}\big) \right\|_{L^2(P;|\cdot|_{\mathbb{R}})} \leq \frac{C}{N}. \tag{7.13}$$

Observe that if the initial random variable ξ is deterministic, that is, $\xi(\omega) = \xi(\tilde{\omega})$ for all $\omega, \tilde{\omega} \in \Omega$, then

$$N \cdot K \cdot m = m \cdot N^3 \tag{7.14}$$

realizations of independent standard normal random variables (cf. "*randn*" calls in Matlab) are needed to compute a realization of

$$\frac{1}{N^2} \sum_{k=1}^{N^2} f\big(Y_N^{N,k}\big) \tag{7.15}$$

for $N \in \mathbb{N}$ and in that case, the Monte Carlo Euler method converges under the assumptions of Theorem 7.1.3 with the order $\frac{1}{3}$ with respect to the number of used independent standard normal random variables. In other words, $O(\varepsilon^{-3})$ independent standard normal random variables are used in (7.15) to compute an approximation of

$$\mathbb{E}_P\big[f(X)\big] \tag{7.16}$$

with a root mean square approximation error of size $\varepsilon > 0$ (cf. (7.6) in Theorem 7.1.3). In the next step two simple illustrative Matlab codes for the Monte Carlo Euler method are presented (cf. Exercises 3.3.9 and 3.3.11).

```
function mc = MonteCarloEuler(mu, sigma, BM_dim, T, x0, f, N, M)
  mc = 0;
  h = T/N;
  sqrth = sqrt(h);
  for m = 1:M
    Y = x0;
    for n = 1:N
      Y = Y + mu(Y)*h + sigma(Y)*sqrth*randn(BM_dim, 1);
    end
```



```
      mc = mc + f(Y);
    end
    mc = mc/M;
end
```

Matlab code 7.1: A Matlab function for the Monte Carlo Euler method. Matlab/MonteCarloEuler.m

```
mu = @(x) log(1.06)*x;
sigma = @(x) x/10;
noise_dim = 1;
T = 1;
x0 = 92;
f = @(x) max(x-100,0);
N = 100;
M = N^2;
tic
MonteCarloEuler(mu,sigma,noise_dim,T,x0,f,N,M)
MonteCarloEuler(mu,sigma,noise_dim,T,x0,f,N,M)
MonteCarloEuler(mu,sigma,noise_dim,T,x0,f,N,M)
toc
```

Matlab code 7.2: A Matlab code for the Euler-Maruyama method. Matlab/RunMonteCarloEuler.m

The Matlab code 7.2 prints three realizations of a Monte Carlo Euler approximation of

$$\mathbb{E}_P\left[\max\{X_1 - 100, 0\}\right] \tag{7.17}$$

based on 10000 samples and time step size $1/100$ where $(X_t)_{t\in[0,1]}$ is a solution process of the SDE (4.55) with $\alpha = \ln(1.06)$, $\beta = 1/10$, and $T = 1$. We would like to point out that the Matlab codes 7.1 and 7.2 respectively can be significantly improved by vectorization to obtain more efficient computations for the Monte Carlo Euler method. This is illustrated in the following two Matlab codes.

```
function mc = MonteCarloEuler2(mu,sigma,T,x0,f,N,M)
    h = T/N;
    sqrth = sqrt(h);
    Y = ones(M,1)*x0;
    for n = 1:N
      Y = Y + mu(Y)*h + sigma(Y).*randn(M,1)*sqrth;
```

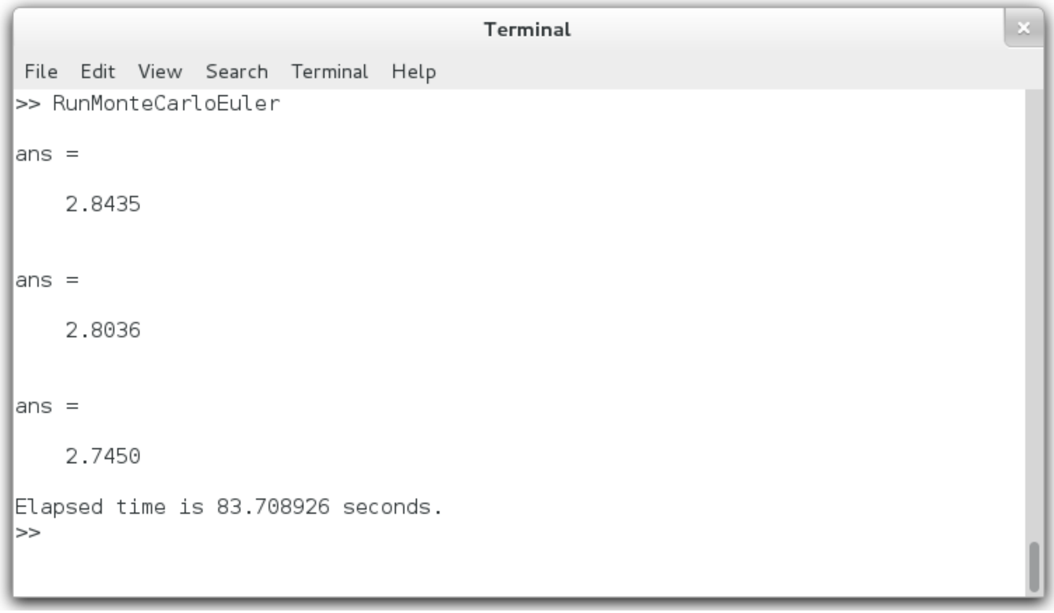


Figure 7.1: Result of a call of the Matlab code 7.2.

```
    end
    mc = sum(f(Y))/M;
end
```

Matlab code 7.3: A Matlab function for the Monte Carlo Euler method. Matlab/MonteCarloEuler2.m

```
mu = @(x) log(1.06)*x;
sigma = @(x) x/10;
T = 1;
x0 = 92;
f = @(x) max(x-100,0);
N = 100;
M = N^2;
tic
MonteCarloEuler2(mu,sigma,T,x0,f,N,M)
MonteCarloEuler2(mu,sigma,T,x0,f,N,M)
MonteCarloEuler2(mu,sigma,T,x0,f,N,M)
toc
```

Matlab code 7.4: A Matlab code for the Euler-Maruyama method. Matlab/RunMonteCarloEuler2.m

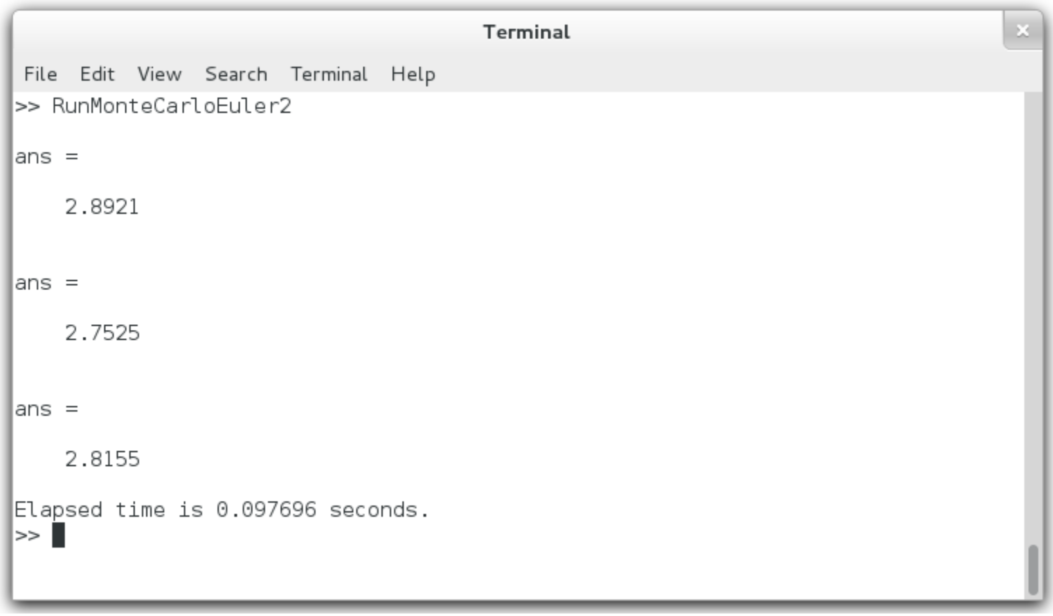


Figure 7.2: Result of a call of the Matlab code 7.4.

7.2 Further one-level Monte Carlo methods for SDEs

The next theorem, Theorem 7.2.1, is proved similar as Theorem 7.1.3 above (cf. Theorem 6.3.1 above).

Theorem 7.2.1 (Monte Carlo $\{\emptyset, 0, 1, \ldots, m\} \cup \{0, 1, \ldots, m\}^2$-stochastic Taylor method)**.** *Assume the setting in the beginning of Chapter 5, assume that* $O = \mathbb{R}^d$, *that* $f\colon \mathbb{R}^d \to \mathbb{R}$, $\mu\colon \mathbb{R}^d \to \mathbb{R}^d$ *and* $\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d\times m}$ *are six times continuously differentiable with at most polynomially growing derivatives and that* μ *and* σ *are globally Lipschitz continuous. Moreover, let* $\xi^{[k]}\colon \Omega \to \mathbb{R}^d$, $k \in \mathbb{N}$, *be i.i.d.* $\mathbb{F}_0/\mathcal{B}(\mathbb{R}^d)$*-measurable mappings with* $P_{\xi^{[1]}} = P_\xi$ *and* $\mathbb{E}_P\big[\|\xi\|^p_{\mathbb{R}^d}\big] < \infty$ *for all* $p \in (0,\infty)$, *let* $W^{[k]} = (W^{[k],(1)}, \ldots, W^{[k],(m)})\colon [0,T] \times \Omega \to \mathbb{R}^m$, $k \in \mathbb{N}$, *be independent standard* $(\mathbb{F}_t)_{t\in[0,T]}$*-Brownian motions, let* $Y^{N,k}\colon \{0,1,\ldots,N\} \times \Omega \to \mathbb{R}^d$, $N, k \in \mathbb{N}$, *be stochastic processes satisfying* $Y_0^{N,k} = \xi^{[k]}$ *and*

$$\begin{aligned}
&Y_{n+1}^{N,k}\\
&= Y_n^{N,k} + \mu(Y_n^{N,k})\,\tfrac{T}{N} + \sigma(Y_n^N)\,\big(W^{[k]}_{\frac{(n+1)T}{N}} - W^{[k]}_{\frac{nT}{N}}\big) + \big(L^0_{\mu,\sigma}\,L^0_{\mu,\sigma} id_{\mathbb{R}^d}\big)\big(Y_n^{N,k}\big)\cdot \tfrac{T^2}{2N^2}\\
&\quad + \sum_{i=1}^m \big(L^0_{\mu,\sigma}\,L^i_{\mu,\sigma} id_{\mathbb{R}^d}\big)\big(Y_n^{N,k}\big)\cdot \int_{\frac{nT}{N}}^{\frac{(n+1)T}{N}} \int_{\frac{nT}{N}}^{s_1} ds_2\, dW_{s_1}^{[k],(i)}\\
&\quad + \sum_{j=1}^m \big(L^j_{\mu,\sigma}\,L^0_{\mu,\sigma} id_{\mathbb{R}^d}\big)\big(Y_n^{N,k}\big)\cdot \int_{\frac{nT}{N}}^{\frac{(n+1)T}{N}} \int_{\frac{nT}{N}}^{s_1} dW_{s_2}^{[k],(j)}\, ds_1\\
&\quad + \sum_{i,j=1}^m \big(L^j_{\mu,\sigma}\,L^i_{\mu,\sigma} id_{\mathbb{R}^d}\big)\big(Y_n^{N,k}\big)\cdot \int_{\frac{nT}{N}}^{\frac{(n+1)T}{N}} \int_{\frac{nT}{N}}^{s_1} dW_{s_2}^{[k],(j)}\, dW_{s_1}^{[k],(i)}
\end{aligned} \tag{7.18}$$

P*-a.s. for all* $n \in \{0, 1, \ldots, N-1\}$, $k, N \in \mathbb{N}$ *(i.i.d.* $\{\emptyset, 0, 1, \ldots, m\} \cup \{0, 1, \ldots, m\}^2$*-stochastic Taylor approximations). Then there exists a real number* $C \in [0, \infty)$ *such that for all* $N, K \in \mathbb{N}$ *it holds that*

$$\left\| \mathbb{E}_P\Big[f(X_T)\Big] - \frac{1}{K}\sum_{k=1}^K f\big(Y_N^{N,k}\big) \right\|_{L^2(P;|\cdot|)} \le C\left(\frac{1}{\sqrt{K}} + \frac{1}{N^2}\right). \tag{7.19}$$

Proof of Theorem 7.2.1. First of all, note that Theorem 6.3.1 implies that there exists a real number $C \in [0,\infty)$ such that for all $N \in \mathbb{N}$ it holds that

$$\Big|\mathbb{E}_P\Big[f(X_T)\Big] - \mathbb{E}_P\Big[f(Y_N^{N,1})\Big]\Big| \le \frac{C}{N^2}. \tag{7.20}$$

This together with Theorem 2.4.8 proves that for all $N, K \in \mathbb{N}$ it holds that

$$\begin{aligned}
&\mathbb{E}_P\left[\left|\mathbb{E}_P\Big[f(X_T)\Big] - \frac{1}{K}\sum_{k=1}^{K} f\big(Y_N^{N,k}\big)\right|^2\right] \\
&= \underbrace{\left|\mathbb{E}_P\Big[f(X_T)\Big] - \mathbb{E}_P\Big[f\big(Y_N^{N,1}\big)\Big]\right|^2}_{\text{squared bias}} + \mathbb{E}_P\left[\left|\mathbb{E}_P\Big[f\big(Y_N^{N,1}\big)\Big] - \frac{1}{K}\sum_{k=1}^{K} f\big(Y_N^{N,k}\big)\right|^2\right] \\
&= \underbrace{\left|\mathbb{E}_P\Big[f(X_T)\Big] - \mathbb{E}_P\Big[f\big(Y_N^{N,1}\big)\Big]\right|^2}_{\text{squared bias}} + \frac{\mathsf{Var}(f(Y_N^{N,1}))}{K} \qquad (7.21)\\
&\leq \frac{C^2}{N^4} + \frac{\mathsf{Var}(f(Y_N^{N,1}))}{K} \leq \max\Big\{C^2, \sup_{M\in\mathbb{N}} \mathsf{Var}(f(Y_M^{M,1}))\Big\}\left[\frac{1}{N^4} + \frac{1}{K}\right] \\
&\leq \max\Big\{C^2, \sup_{M\in\mathbb{N}} \mathbb{E}_P\Big[\big(f(Y_M^{M,1})\big)^2\Big]\Big\}\left[\frac{1}{N^4} + \frac{1}{K}\right].
\end{aligned}$$

Next observe that again Theorem 6.3.1 implies that there exists a real number $\hat{C} \in [0, \infty)$ such that for all $N \in \mathbb{N}$ it holds that

$$\left|\mathbb{E}_P\Big[\big(f(X_T)\big)^2\Big] - \mathbb{E}_P\Big[\big(f(Y_N^{N,1})\big)^2\Big]\right| \leq \frac{\hat{C}}{N^2}. \qquad (7.22)$$

This ensures that

$$\sup_{N\in\mathbb{N}} \left|\mathbb{E}_P\Big[\big(f(X_T)\big)^2\Big] - \mathbb{E}_P\Big[\big(f(Y_N^{N,1})\big)^2\Big]\right| < \infty \qquad (7.23)$$

and therefore, we obtain that

$$\sup_{N\in\mathbb{N}} \mathbb{E}_P\Big[\big(f(Y_N^{N,1})\big)^2\Big] < \infty. \qquad (7.24)$$

Putting this into (7.21) completes the proof of Theorem 7.2.1. □

If we choose $K = N^4$ in the setting of Theorem 7.2.1, then (7.19) proves that there exists a real number $C \in [0, \infty)$ such that for all $N \in \mathbb{N}$ it holds that

$$\left\|\mathbb{E}_P\Big[f(X_T)\Big] - \frac{1}{N^4}\sum_{k=1}^{N^4} f\big(Y_N^{N,k}\big)\right\|_{L^2(P;|\cdot|)} \leq \frac{C}{N^2}. \qquad (7.25)$$

Note that if the initial random variable ξ is deterministic, that is, $\xi(\omega) = \xi(\tilde{\omega})$ for all $\omega, \tilde{\omega} \in \Omega$, then

$$N \cdot K \cdot m = m \cdot N^5 \qquad (7.26)$$

realizations of independent standard normal random variables (cf. "*randn*" calls in Matlab) are needed to compute a realization of

$$\frac{1}{N^4}\sum_{k=1}^{N^4} f\big(Y_N^{N,k}\big) \tag{7.27}$$

for $N \in \mathbb{N}$ and in that case, the Monte Carlo method in (7.19) converges under the assumptions of Theorem 7.1.3 with the order $\frac{2}{5} = 0.4\,\big(> 0.333\ldots = \frac{1}{3}\big)$ with respect to the number of used independent standard normal random variables. In other words, $O(\varepsilon^{-\frac{5}{2}}) = O(\varepsilon^{-2.5})$ independent standard normal random variables are used in (7.27) to compute an approximation of

$$\mathbb{E}_P\big[f(X_T)\big] \tag{7.28}$$

with a root mean square approximation error of size $\varepsilon > 0$ (cf. (7.19) in Theorem 7.2.1).

7.3 Multilevel Monte Carlo Euler method

This subsection presents and briefly investigates multilevel Monte Carlo Euler approximations; see [Giles(2008)] and [Heinrich(2001)].

Definition 7.3.1 (Multilevel Monte Carlo Euler method). *Assume the setting in the beginning of Chapter 5, let $f\colon C([0,T],\mathbb{R}^d) \to \mathbb{R}$ be a Borel measurable function with $\mathbb{E}_P\big[|f(X)|\big] < \infty$, let $\xi^{k,l}\colon \Omega \to \mathbb{R}^d$, $k \in \mathbb{N}$, $l \in \mathbb{N}_0$, be i.i.d. $\mathbb{F}_0/\mathcal{B}(\mathbb{R}^d)$-measurable mappings with $P_{\xi^{1,0}} = P_\xi$, let $W^{k,l}\colon [0,T]\times\Omega \to \mathbb{R}^m$, $k \in \mathbb{N}$, $l \in \mathbb{N}_0$, be independent standard $(\mathbb{F}_t)_{t\in[0,T]}$-Brownian motions, let $Y^{N,k,l}\colon \{0,1,\ldots,N\}\times\Omega \to \mathbb{R}^d$, $N,k \in \mathbb{N}$, $l \in \mathbb{N}_0$, be stochastic processes satisfying $Y_0^{N,k,l} = \xi^{k,l}$ and*

$$Y_{n+1}^{N,k,l} = Y_n^{N,k,l} + \bar{\mu}\big(Y_n^{N,k,l}\big)\tfrac{T}{N} + \bar{\sigma}\big(Y_n^{N,k,l}\big)\Big(W^{k,l}_{\frac{(n+1)T}{N}} - W^{k,l}_{\frac{nT}{N}}\Big) \tag{7.29}$$

*for all $n \in \{0,1,\ldots,N-1\}$, $N,k \in \mathbb{N}$, $l \in \mathbb{N}_0$ (*i.i.d. Euler-Maruyama approximations*) and let $\bar{Y}^{N,k,l}\colon [0,T]\times\Omega \to \mathbb{R}^d$, $N,k \in \mathbb{N}$, $l \in \mathbb{N}_0$, be stochastic processes satisfying*

$$\bar{Y}_t^{N,k,l} = \big(\tfrac{tN}{T} - n\big)\,Y_{n+1}^{N,k,l} + \big(n+1-\tfrac{tN}{T}\big)\,Y_n^{N,k,l} \tag{7.30}$$

*for all $t \in \big[\frac{nT}{N}, \frac{(n+1)T}{N}\big]$, $n \in \{0,1,\ldots,N-1\}$, $N,k \in \mathbb{N}$, $l \in \mathbb{N}_0$ (*i.i.d. linearly interpolated Euler-Maruyama approximations*). Then for every $L \in \mathbb{N}_0$, $N_0,N_1,\ldots,N_L,K_0,K_1,\ldots,K_L \in \mathbb{N}$ the random variable*

$$\frac{1}{K_0}\sum_{k=1}^{K_0} f\big(\bar{Y}^{N_0,k,0}\big) + \sum_{l=1}^{L}\frac{1}{K_l}\left[\sum_{k=1}^{K_l} f\big(\bar{Y}^{N_l,k,l}\big) - f\big(\bar{Y}^{N_{l-1},k,l}\big)\right] \tag{7.31}$$

is called a Multilevel Monte Carlo Euler approximation *(of $\mathbb{E}_P\big[f(X)\big]$ based on L Levels, K_0 samples on level zero, K_1 samples on level one, …, K_L samples on level L and time step sizes $\frac{T}{N_0}$, $\frac{T}{N_1}$, …, $\frac{T}{N_L}$).*

The root mean square approximation error of the multilevel Monte Carlo Euler method is estimated in Theorem 7.3.2 below. Theorem 7.3.2 is, e.g., similar to Proposition 6.2 in [Hutzenthaler et al.(2011b)Hutzenthaler, Jentzen, and Kloeden].

Theorem 7.3.2 (Multilevel Monte Carlo Euler method)**.** *Assume the setting in the beginning of Chapter 5, assume that $O = \mathbb{R}^d$, let $\xi^{k,l}\colon \Omega \to \mathbb{R}^d$, $k \in \mathbb{N}$, $l \in \mathbb{N}_0$, be i.i.d. $\mathbb{F}_0/\mathcal{B}(\mathbb{R}^d)$-measurable mappings with $P_{\xi^{1,0}} = P_\xi$ and $\mathbb{E}_P\big[\|\xi\|^p_{\mathbb{R}^d}\big] < \infty$ for all $p \in (0,\infty)$, let $W^{k,l}\colon [0,T] \times \Omega \to \mathbb{R}^m$, $k \in \mathbb{N}$, $l \in \mathbb{N}_0$, be independent standard $(\mathbb{F}_t)_{t\in[0,T]}$-Brownian motions, let $Y^{N,k,l}\colon \{0,1,\ldots,N\} \times \Omega \to \mathbb{R}^d$, $N,k \in \mathbb{N}$, $l \in \mathbb{N}_0$, be stochastic processes satisfying $Y_0^{N,k,l} = \xi^{k,l}$ and*

$$Y_{n+1}^{N,k,l} = Y_n^{N,k,l} + \mu\big(Y_n^{N,k,l}\big)\tfrac{T}{N} + \sigma\big(Y_n^{N,k,l}\big)\Big(W^{k,l}_{\frac{(n+1)T}{N}} - W^{k,l}_{\frac{nT}{N}}\Big) \tag{7.32}$$

for all $n \in \{0,1,\ldots,N-1\}$, $N,k \in \mathbb{N}$, $l \in \mathbb{N}_0$ (i.i.d. Euler-Maruyama approximations) and let $\bar{Y}^{N,k,l}\colon [0,T] \times \Omega \to \mathbb{R}^d$, $N,k \in \mathbb{N}$, $l \in \mathbb{N}_0$, be stochastic processes satisfying

$$\bar{Y}_t^{N,k,l} = \big(\tfrac{tN}{T} - n\big)\, Y_{n+1}^{N,k,l} + \big(n + 1 - \tfrac{tN}{T}\big)\, Y_n^{N,k,l} \tag{7.33}$$

*for all $t \in \big[\frac{nT}{N}, \frac{(n+1)T}{N}\big]$, $n \in \{0,1,\ldots,N-1\}$, $N,k \in \mathbb{N}$, $l \in \mathbb{N}_0$ (*i.i.d. linearly interpolated Euler-Maruyama approximations*). Moreover, let μ and σ be globally Lipschitz continuous, let $c \in [0,\infty)$ be a real number and let $f\colon C([0,T],\mathbb{R}^d) \to \mathbb{R}$ be a function satisfying*

$$\begin{aligned} |f(v) - f(w)| &\leq c\left(1 + \|v\|^c_{C([0,T],\mathbb{R}^d)} + \|w\|^c_{C([0,T],\mathbb{R}^d)}\right)\|v - w\|_{C([0,T],\mathbb{R}^d)} \\ &= c\left(1 + \sup_{t\in[0,T]} \|v(t)\|^c_{\mathbb{R}^d} + \sup_{t\in[0,T]} \|w(t)\|^c_{\mathbb{R}^d}\right)\left[\sup_{t\in[0,T]} \|v(t) - w(t)\|_{\mathbb{R}^d}\right] \end{aligned} \tag{7.34}$$

for all $v, w \in C([0,T],\mathbb{R}^d)$. Then there exists a real number $C \in [0,\infty)$ such that for all $L \in \mathbb{N}_0$, $N_0, N_1, \ldots, N_L, K_0, K_1, \ldots, K_L \in \mathbb{N}$ with $N_0 \leq N_1 \leq \cdots \leq N_L$ it holds that

$$\begin{aligned} &\left\| \mathbb{E}_P\big[f(X)\big] - \frac{1}{K_0}\sum_{k=1}^{K_0} f\big(\bar{Y}^{N_0,k,0}\big) - \sum_{l=1}^{L}\frac{1}{K_l}\left[\sum_{k=1}^{K_l} f\big(\bar{Y}^{N_l,k,l}\big) - f\big(\bar{Y}^{N_{l-1},k,l}\big)\right]\right\|_{L^2(P;|\cdot|)} \\ &\leq C\sqrt{(1 + \ln(N_L))}\sqrt{\left(\frac{1}{K_0} + \left[\sum_{l=1}^{L}\frac{1}{K_l N_{(l-1)}}\right] + \frac{1}{N_L}\right)}. \end{aligned} \tag{7.35}$$

Proof of Theorem 7.3.2. First of all, note that Theorem 5.3.11 implies that there exists a real number $C \in [0, \infty)$ such that for all $N \in \mathbb{N}$ it holds that

$$\begin{aligned}
&\left(\mathbb{E}_P\left[\sup_{t\in[0,T]}\left\|X_t - \bar{Y}_t^{N,1,0}\right\|_{\mathbb{R}^d}^4\right]\right)^{\frac{1}{4}} = \left\|\sup_{t\in[0,T]}\left\|X_t - \bar{Y}_t^{N,1,0}\right\|_{\mathbb{R}^d}\right\|_{L^4(P;|\cdot|)} \\
&= \left\|\left\|X - \bar{Y}^{N,1,0}\right\|_{C([0,T],\mathbb{R}^d)}\right\|_{L^4(P;|\cdot|)} = \left\|X - \bar{Y}^{N,1,0}\right\|_{L^4(P;C([0,T],\mathbb{R}^d))} \\
&\leq \frac{C\sqrt{1+\ln(N)}}{\sqrt{N}}.
\end{aligned} \tag{7.36}$$

Assumption (7.34) and Hölder's inequality hence show that for all $N \in \mathbb{N}$ it holds that

$$\begin{aligned}
&\left\|f(X) - f(\bar{Y}^{N,1,0})\right\|_{L^2(P;|\cdot|)} \\
&\leq \left\|\left(1 + \|X\|_{C([0,T],\mathbb{R}^d)}^c + \left\|\bar{Y}^{N,1,0}\right\|_{C([0,T],\mathbb{R}^d)}^c\right)\left\|X - \bar{Y}^{N,1,0}\right\|_{C([0,T],\mathbb{R}^d)}\right\|_{L^2(P;|\cdot|)} \\
&\leq \left\|\left(1 + \|X\|_{C([0,T],\mathbb{R}^d)}^c + \left\|\bar{Y}^{N,1,0}\right\|_{C([0,T],\mathbb{R}^d)}^c\right)\right\|_{L^4(P;|\cdot|)}\left\|\left\|X - \bar{Y}^{N,1,0}\right\|_{C([0,T],\mathbb{R}^d)}\right\|_{L^4(P;|\cdot|)} \\
&\leq \left(1 + \|X\|_{L^{4c}(P;C([0,T],\mathbb{R}^d))}^c + \left\|\bar{Y}^{N,1,0}\right\|_{L^{4c}(P;C([0,T],\mathbb{R}^d))}^c\right)\left\|X - \bar{Y}^{N,1,0}\right\|_{L^4(P;C([0,T],\mathbb{R}^d))} \\
&\leq \left(1 + \|X\|_{L^{4c}(P;C([0,T],\mathbb{R}^d))}^c + \sup_{M\in\mathbb{N}}\left\|\bar{Y}^{M,1,0}\right\|_{L^{4c}(P;C([0,T],\mathbb{R}^d))}^c\right)\frac{C\sqrt{1+\ln(N)}}{\sqrt{N}}.
\end{aligned} \tag{7.37}$$

This implies that there exists a real number $\hat{C} \in [0, \infty)$ such that for all $N \in \mathbb{N}$ it holds that

$$\left\|f(X) - f(\bar{Y}^{N,1,0})\right\|_{L^2(P;|\cdot|)}^2 = \mathbb{E}_P\left[\left|f(X) - f(\bar{Y}^{N,1,0})\right|^2\right] \leq \frac{\hat{C}\,(1+\ln(N))}{N}. \tag{7.38}$$

Next observe that for all $L \in \mathbb{N}_0$, $K_0, K_1, \ldots, K_L, N_0, N_1, \ldots, N_L \in \mathbb{N}$ it holds that

$$\begin{aligned}
&\mathbb{E}_P\left[\frac{1}{K_0}\sum_{k=1}^{K_0} f\big(\bar{Y}^{N_0,k,0}\big) + \sum_{l=1}^{L}\frac{1}{K_l}\left[\sum_{k=1}^{K_l} f\big(\bar{Y}^{N_l,k,l}\big) - f\big(\bar{Y}^{N_{l-1},k,l}\big)\right]\right] \\
&= \mathbb{E}_P\big[f\big(\bar{Y}^{N_0,1,0}\big)\big] + \sum_{l=1}^{L}\mathbb{E}_P\big[f\big(\bar{Y}^{N_l,1,0}\big) - f\big(\bar{Y}^{N_{l-1},1,0}\big)\big] = \mathbb{E}_P\big[f\big(\bar{Y}^{N_L,1,0}\big)\big].
\end{aligned} \tag{7.39}$$

Therefore, we obtain that for all $L \in \mathbb{N}_0$, $K_0, K_1, \ldots, K_L, N_0, N_1, \ldots, N_L \in \mathbb{N}$ it holds

that

$$
\begin{aligned}
&\mathbb{E}_P\left[\left|\mathbb{E}_P\big[f(X)\big]-\frac{1}{K_0}\sum_{k=1}^{K_0} f\big(\bar{Y}^{N_0,k,0}\big)-\sum_{l=1}^{L}\frac{1}{K_l}\left[\sum_{k=1}^{K_l} f\big(\bar{Y}^{N_l,k,l}\big)-f\big(\bar{Y}^{N_{l-1},k,l}\big)\right]\right|^2\right]\\
&=\underbrace{\Big|\mathbb{E}_P\big[f(X)\big]-\mathbb{E}_P\big[f\big(\bar{Y}^{N_L,1,0}\big)\big]\Big|^2}_{\text{squared bias}}\\
&\quad+\mathsf{Var}\left(\frac{1}{K_0}\sum_{k=1}^{K_0} f\big(\bar{Y}^{N_0,k,0}\big)+\sum_{l=1}^{L}\frac{1}{K_l}\left[\sum_{k=1}^{K_l} f\big(\bar{Y}^{N_l,k,l}\big)-f\big(\bar{Y}^{N_{l-1},k,l}\big)\right]\right)\\
&=\underbrace{\Big|\mathbb{E}_P\big[f(X)\big]-\mathbb{E}_P\big[f\big(\bar{Y}^{N_L,1,0}\big)\big]\Big|^2}_{\text{squared bias}}+\mathsf{Var}\left(\frac{1}{K_0}\sum_{k=1}^{K_0} f\big(\bar{Y}^{N_0,k,0}\big)\right)\\
&\quad+\sum_{l=1}^{L}\mathsf{Var}\left(\frac{1}{K_l}\left[\sum_{k=1}^{K_l} f\big(\bar{Y}^{N_l,k,l}\big)-f\big(\bar{Y}^{N_{l-1},k,l}\big)\right]\right)
\end{aligned}
\tag{7.40}
$$

and hence, we get that for all $L\in\mathbb{N}_0$, $K_0,K_1,\dots,K_L,N_0,N_1,\dots,N_L\in\mathbb{N}$ it holds that

$$
\begin{aligned}
&\mathbb{E}_P\left[\left|\mathbb{E}_P\big[f(X)\big]-\frac{1}{K_0}\sum_{k=1}^{K_0} f\big(\bar{Y}^{N_0,k,0}\big)-\sum_{l=1}^{L}\frac{1}{K_l}\left[\sum_{k=1}^{K_l} f\big(\bar{Y}^{N_l,k,l}\big)-f\big(\bar{Y}^{N_{l-1},k,l}\big)\right]\right|^2\right]\\
&=\underbrace{\Big|\mathbb{E}_P\big[f(X)\big]-\mathbb{E}_P\big[f\big(\bar{Y}^{N_L,1,0}\big)\big]\Big|^2}_{\text{squared bias}}+\frac{\mathsf{Var}\big(f\big(\bar{Y}^{N_0,1,0}\big)\big)}{K_0}\\
&\quad+\sum_{l=1}^{L}\frac{\mathsf{Var}\big(f\big(\bar{Y}^{N_l,1,0}\big)-f\big(\bar{Y}^{N_{l-1},1,0}\big)\big)}{K_l}\\
&\le\mathbb{E}_P\Big[\big|f(X)-f\big(\bar{Y}^{N_L,1,0}\big)\big|^2\Big]+\frac{\sup_{M\in\mathbb{N}}\big\|f\big(\bar{Y}^{M,1,0}\big)\big\|^2_{L^2(P;|\cdot|)}}{K_0}\\
&\quad+\sum_{l=1}^{L}\frac{\big\|f\big(\bar{Y}^{N_l,1,0}\big)-f\big(\bar{Y}^{N_{l-1},1,0}\big)\big\|^2_{L^2(P;|\cdot|)}}{K_l}.
\end{aligned}
\tag{7.41}
$$

The triangle inequality and the estimate $(a+b)^2\le 2a^2+2b^2$ for all $a,b\in\mathbb{R}$ hence imply that for all $L \in \mathbb{N}_0$, $K_0, K_1, \dots, K_L, N_0, N_1, \dots, N_L \in \mathbb{N}$ it holds that

$$\begin{aligned}
&\mathbb{E}_P\left[\left|\mathbb{E}_P\big[f(X)\big] - \frac{1}{K_0}\sum_{k=1}^{K_0} f\big(\bar{Y}^{N_0,k,0}\big) - \sum_{l=1}^{L}\frac{1}{K_l}\left[\sum_{k=1}^{K_l} f\big(\bar{Y}^{N_l,k,l}\big) - f\big(\bar{Y}^{N_{l-1},k,l}\big)\right]\right|^2\right] \\
&\leq \big\|f(X) - f\big(\bar{Y}^{N_L,1,0}\big)\big\|^2_{L^2(P;|\cdot|)} + \frac{\sup_{M\in\mathbb{N}}\big\|f\big(\bar{Y}^{M,1,0}\big)\big\|^2_{L^2(P;|\cdot|)}}{K_0} \\
&\quad + \sum_{l=1}^{L}\frac{2\,\big\|f\big(\bar{Y}^{N_l,1,0}\big) - f\big(X\big)\big\|^2_{L^2(P;|\cdot|)} + 2\,\big\|f\big(X\big) - f\big(\bar{Y}^{N_{l-1},1,0}\big)\big\|^2_{L^2(P;|\cdot|)}}{K_l}.
\end{aligned} \tag{7.42}$$

Inequality (7.38) therefore proves that for all $L \in \mathbb{N}_0$, $K_0, K_1, \dots, K_L, N_0, N_1, \dots, N_L \in \mathbb{N}$ it holds that

$$\begin{aligned}
&\mathbb{E}_P\left[\left|\mathbb{E}_P\big[f(X)\big] - \frac{1}{K_0}\sum_{k=1}^{K_0} f\big(\bar{Y}^{N_0,k,0}\big) - \sum_{l=1}^{L}\frac{1}{K_l}\left[\sum_{k=1}^{K_l} f\big(\bar{Y}^{N_l,k,l}\big) - f\big(\bar{Y}^{N_{l-1},k,l}\big)\right]\right|^2\right] \\
&\leq \frac{\hat{C}\,(1+\ln(N_L))}{N_L} + \frac{\sup_{M\in\mathbb{N}}\big\|f\big(\bar{Y}^{M,1,0}\big)\big\|^2_{L^2(P;|\cdot|)}}{K_0} \\
&\quad + \sum_{l=1}^{L}\frac{1}{K_l}\left(\frac{2\hat{C}\,(1+\ln(N_l))}{N_l} + \frac{2\hat{C}\,(1+\ln(N_{l-1}))}{N_{l-1}}\right).
\end{aligned} \tag{7.43}$$

The fact that the function

$$[1,\infty) \ni x \mapsto \frac{1+\ln(x)}{x} \in (0,\infty) \tag{7.44}$$

is strictly decreasing hence proves that for all $L \in \mathbb{N}_0$, $K_0, K_1, \dots, K_L, N_0, N_1, \dots, N_L \in \mathbb{N}$ with $N_0 \leq N_1 \leq \dots \leq N_L$ it holds that

$$\begin{aligned}
&\mathbb{E}_P\left[\left|\mathbb{E}_P\big[f(X)\big] - \frac{1}{K_0}\sum_{k=1}^{K_0} f\big(\bar{Y}^{N_0,k,0}\big) - \sum_{l=1}^{L}\frac{1}{K_l}\left[\sum_{k=1}^{K_l} f\big(\bar{Y}^{N_l,k,l}\big) - f\big(\bar{Y}^{N_{l-1},k,l}\big)\right]\right|^2\right] \\
&\leq \frac{\hat{C}\,(1+\ln(N_L))}{N_L} + \frac{\sup_{M\in\mathbb{N}}\big\|f\big(\bar{Y}^{M,1,0}\big)\big\|^2_{L^2(P;|\cdot|)}}{K_0} + \sum_{l=1}^{L}\frac{4\hat{C}\,(1+\ln(N_{l-1}))}{K_l N_{l-1}} \\
&\leq \left(1 + 4\hat{C} + \sup_{M\in\mathbb{N}}\big\|f\big(\bar{Y}^{M,1,0}\big)\big\|^2_{L^2(P;|\cdot|)}\right)(1+\ln(N_L))\left(\frac{1}{N_L} + \frac{1}{K_0} + \sum_{l=1}^{L}\frac{1}{K_l N_{l-1}}\right).
\end{aligned} \tag{7.45}$$

The proof of Theorem 7.1.3 is thus completed. □

If we choose

$$L = \mathrm{ld}(N), \quad N_0 = 2^0 = 1, \quad N_1 = 2^1 = 2, \quad N_2 = 2^2 = 4, \quad \ldots, \quad N_L = 2^L = N \tag{7.46}$$

and

$$K_0 = \frac{N}{2^0} = N, \quad K_1 = \frac{N}{2^1} = \frac{N}{2}, \quad K_2 = \frac{N}{2^2} = \frac{N}{4}, \quad \ldots, \quad K_L = \frac{N}{2^L} = 1 \tag{7.47}$$

for $N \in \{2^0, 2^1, 2^2, 2^3, \ldots\}$ in the setting of Theorem 7.3.2, then (7.35) proves that there exists a real number $C \in [0, \infty)$ such that for all $N \in \{2^0, 2^1, 2^2, 2^3, \ldots\}$ it holds that

$$\begin{aligned}
&\left\| \mathbb{E}_P\Big[f(X_T)\Big] - \frac{1}{N} \sum_{k=1}^{N} f\big(\bar{Y}^{1,k,0}\big) - \sum_{l=1}^{\mathrm{ld}(N)} \frac{2^l}{N} \left[\sum_{k=1}^{\frac{N}{2^l}} f\big(\bar{Y}^{2^l,k,l}\big) - f\big(\bar{Y}^{2^{l-1},k,l}\big) \right] \right\|_{L^2(P;|\cdot|)} \\
&\leq C\sqrt{(1+\ln(N))} \left(\frac{1}{\sqrt{N}} + \sqrt{\sum_{l=1}^{\mathrm{ld}(N)} \frac{1}{\left(\frac{N}{2}\right)}} \right) = C\sqrt{(1+\ln(N))} \left(\frac{1}{\sqrt{N}} + \frac{\sqrt{2\,\mathrm{ld}(N)}}{\sqrt{N}} \right) \\
&= \frac{C\sqrt{(1+\ln(N))}\left(1+\sqrt{2\,\mathrm{ld}(N)}\right)}{\sqrt{N}} \leq \frac{\left(\sqrt{2}C\right)\sqrt{(1+\ln(N))}\left(1+\sqrt{1+\mathrm{ld}(N)}\right)}{\sqrt{N}} \\
&\leq \frac{\left(\sqrt{8}C\right)\sqrt{(1+\ln(N))}\sqrt{(1+\mathrm{ld}(N))}}{\sqrt{N}} \leq \frac{\left(\sqrt{8}C\right)(1+\mathrm{ld}(N))}{\sqrt{N}}.
\end{aligned} \tag{7.48}$$

Observe that if the initial random variable ξ is deterministic, that is, $\xi(\omega) = \xi(\tilde{\omega})$ for all $\omega, \tilde{\omega} \in \Omega$, then

$$\begin{aligned}
&\left(N + \sum_{l=1}^{\mathrm{ld}(N)} \frac{N}{2^l} \left(2^l + 2^{l-1}\right) \right) m = \left(N + \sum_{l=1}^{\mathrm{ld}(N)} \frac{3N}{2} \right) m \\
&= Nm \left(1 + \frac{3\mathrm{ld}(N)}{2} \right) \leq \left(\frac{3m}{2} \right) N \left(1 + \mathrm{ld}(N)\right)
\end{aligned} \tag{7.49}$$

realizations of independent standard normal random variables (cf. *randn* calls in Matlab) are needed to compute a realization of

$$\frac{1}{N} \sum_{k=1}^{N} f\big(\bar{Y}^{1,k,0}\big) + \sum_{l=1}^{\mathrm{ld}(N)} \frac{2^l}{N} \left[\sum_{k=1}^{\frac{N}{2^l}} f\big(\bar{Y}^{2^l,k,l}\big) - f\big(\bar{Y}^{2^{l-1},k,l}\big) \right] \tag{7.50}$$

for $N \in \{2^0, 2^1, 2^2, 2^3, \ldots\}$ and in that case, for every arbitrarily small $\delta \in (0, \infty)$, the multilevel Monte Carlo Euler method converges with the order $\frac{1}{2} - \delta$ with respect to the number of used independent standard normal random variables. In other words, for every $\delta \in (0, \infty)$ it holds that $O(\varepsilon^{-(2+\delta)})$ independent standard normal random variables are used in (7.50) to compute an approximation of

$$\mathbb{E}_P\big[f(X)\big] \tag{7.51}$$

with a root mean square approximation error of size $\varepsilon > 0$ (cf. (7.35) in Theorem 7.3.2).

8 Solutions to the exercises

In this chapter we do not distinguish between pseudo random numbers and actual random numbers.

8.1 Chapter 0

8.1.1 Solution to Exercise 0.2.8

Lemma 8.1.1. *Let* $(\Omega, \mathcal{A})$ *be a measurable space and let* $A \subseteq \Omega$ *be a subset of* Ω*. Then* $(A, A \Cap \mathcal{A})$ *is a measurable space.*

Proof of Lemma 8.1.1. First of all, observe that the fact that $\mathcal{A}$ is a sigma-algebra on Ω ensures that $\emptyset \in \mathcal{A}$ and this implies that

$$\emptyset = A \cap \underbrace{\emptyset}_{\in \mathcal{A}} \in A \Cap \mathcal{A}. \tag{8.1}$$

Next we observe that for all $B \in \mathcal{A}$ it holds that $\Omega \backslash B \in \mathcal{A}$. This shows that for all $B \in \mathcal{A}$ it holds that

$$A \backslash \underbrace{(A \cap B)}_{\in A \Cap \mathcal{A}} = A \backslash B = A \cap \underbrace{(\Omega \backslash B)}_{\in \mathcal{A}} \in A \Cap \mathcal{A}. \tag{8.2}$$

It thus remains to verify that $A \Cap \mathcal{A}$ is closed under countable unions; see Definition 0.2.1. To see this observe that for all functions $B \colon \mathbb{N} \to \mathcal{A}$ it holds that

$$\cup_{n \in \mathbb{N}} (A \cap B(n)) = A \cap (\cup_{n \in \mathbb{N}} B(n)). \tag{8.3}$$

Combining (8.1), (8.2), and (8.3) proves that $A \Cap \mathcal{A}$ is a sigma-algebra on A. The proof of Lemma 8.1.1 is thus completed. □

8.1.2 Solution to Exercise 0.2.43

Lemma 8.1.2. *Let* $(E, \mathcal{E})$ *be a topological space and let* $\mu \colon \mathcal{B}(\mathcal{E}) \to [0, \infty]$ *be a measure on* $(E, \mathcal{B}(\mathcal{E}))$*. Then it holds that* $E \backslash \operatorname{supp}(\mu) \in \mathcal{E}$*.*

Proof of Lemma 8.1.2. First, note that

$$\begin{aligned} \operatorname{supp}(\mu) &= \Big\{x \in E \colon \Big(\forall U \in \mathcal{E} \colon (x \in U \Rightarrow \mu(U) > 0)\Big)\Big\} \\ &= \Big\{x \in E \colon \Big(\forall U \in \mathcal{E} \colon (x \notin U) \vee (\mu(U) > 0)\Big)\Big\}. \end{aligned} \tag{8.4}$$

This implies that

$$\begin{aligned} E \backslash \operatorname{supp}(\mu) &= \Big\{x \in E \colon \neg\Big(\forall U \in \mathcal{E} \colon (x \notin U) \vee (\mu(U) > 0)\Big)\Big\} \\ &= \Big\{x \in E \colon \Big(\exists U \in \mathcal{E} \colon \neg\left[(x \notin U) \vee (\mu(U) > 0)\right]\Big)\Big\} \\ &= \Big\{x \in E \colon \Big(\exists U \in \mathcal{E} \colon (x \in U) \wedge (\mu(U) = 0)\Big)\Big\} \\ &= \cup_{U \in \mathcal{E}, \mu(U) = 0} U \in \mathcal{E}. \end{aligned} \tag{8.5}$$

The proof of Lemma 8.1.2 is thus completed. □

8.1.3 Solution to Exercise 0.4.7

Lemma 8.1.3 (Approximations of the exponential function)**.** *Let $a_l \in \mathbb{R}$, $l \in \mathbb{N}$, be a convergent sequence and let $n_l \in \mathbb{N}$, $l \in \mathbb{N}$, satisfy $\lim_{l\to\infty} n_l = \infty$. Then*

$$\lim_{l\to\infty} \left[\left[1 + \tfrac{a_l}{n_l}\right]^{n_l} \right] = \exp\Big(\lim_{l\to\infty} a_l\Big). \tag{8.6}$$

Proof of Lemma 8.1.3. Let $f_l \colon \mathbb{N}_0 \to \mathbb{R}$, $l \in \mathbb{N}$, be the functions which satisfy for all $l \in \mathbb{N}$, $k \in \mathbb{N}_0$ that

$$f_l(k) = \begin{cases} \frac{n_l (n_l - 1)\cdot \ldots \cdot (n_l - k + 1)}{(n_l)^k} \cdot \frac{(a_l)^k}{k!} & : k \le n_l \\ 0 & : k > n_l \end{cases}. \tag{8.7}$$

Next observe that the binomial theorem proves that

$$\left[1 + \tfrac{a_l}{n_l}\right]^{n_l} = \sum_{k=0}^{n_l} \binom{n_l}{k} \left[\tfrac{a_l}{n_l}\right]^k = \sum_{k=0}^{\infty} f_l(k) = \int_{\mathbb{N}_0} f_l(k) \, \#_{\mathbb{N}_0}(dk). \tag{8.8}$$

Moreover, note that for all $k \in \mathbb{N}_0$ it holds that

$$\lim_{l\to\infty} f_l(k) = \frac{\left[\lim_{l\to\infty} a_l\right]^k}{k!} \qquad \text{and} \qquad \sup_{l\in\mathbb{N}} |f_l(k)| \le \frac{\left[\sup_{l\in\mathbb{N}} |a_l|\right]^k}{k!}. \tag{8.9}$$

Lebesgue's theorem of dominated convergence hence proves that

$$\begin{aligned} &\lim_{l\to\infty} \int_{\mathbb{N}_0} f_l(k) \, \#_{\mathbb{N}_0}(dk) = \int_{\mathbb{N}_0} \lim_{l\to\infty} f_l(k) \, \#_{\mathbb{N}_0}(dk) = \int_{\mathbb{N}_0} \frac{\left[\lim_{l\to\infty} a_l\right]^k}{k!} \, \#_{\mathbb{N}_0}(dk) \\ &= \sum_{k=0}^{\infty} \frac{\left[\lim_{l\to\infty} a_l\right]^k}{k!} = \exp\Big(\lim_{l\to\infty} a_l\Big). \end{aligned} \tag{8.10}$$

This and (8.8) ensure that

$$\lim_{l\to\infty}\left[\left[1+\tfrac{a_l}{n_l}\right]^{n_l}\right]=\exp\!\Big(\lim_{l\to\infty}a_l\Big). \tag{8.11}$$

The proof of Lemma 8.1.3 is thus completed. □

8.1.4 Solution to Exercise 0.4.14

Lemma 8.1.4. *Let $d\in\mathbb{N}$. Then it holds that*

$$\mathcal{N}_{0,I_{\mathbb{R}^d}}(\mathbb{R}^d)=1. \tag{8.12}$$

Proof of Lemma 8.1.4. First of all, observe that a polar coordinate transform ensures that

$$\begin{aligned}\mathcal{N}_{0,I_{\mathbb{R}^2}}(\mathbb{R}^2)&=\int_{\mathbb{R}^2}1\,\mathcal{N}_{0,I_{\mathbb{R}^2}}(dx)=\frac{1}{2\pi}\int_{\mathbb{R}^2}\exp\!\Big(\tfrac{-\|x\|_{\mathbb{R}^2}^2}{2}\Big)\,\lambda_{\mathbb{R}^2}(dx)\\&=\frac{1}{2\pi}\int_0^\infty\int_0^{2\pi}r\,\exp\!\Big(\tfrac{-r^2}{2}\Big)\,d\alpha\,dr=\int_0^\infty r\,\exp\!\Big(\tfrac{-r^2}{2}\Big)\,dr\\&=\left[-\exp\!\Big(-\frac{r^2}{2}\Big)\right]_{r=0}^{r=\infty}=1.\end{aligned} \tag{8.13}$$

This implies that

$$\begin{aligned}\mathcal{N}_{0,I_{\mathbb{R}^d}}(\mathbb{R}^d)&=\int_{\mathbb{R}^d}1\,\mathcal{N}_{0,I_{\mathbb{R}^d}}(dx)=\frac{1}{(2\pi)^{d/2}}\int_{\mathbb{R}^d}e^{\frac{-1}{2}\|x\|_{\mathbb{R}^d}^2}\,dx\\&=\left[\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}e^{\frac{-x^2}{2}}\,dx\right]^d=\left[\frac{1}{2\pi}\int_{\mathbb{R}^2}e^{\frac{-1}{2}\|x\|_{\mathbb{R}^2}^2}\,dx\right]^{d/2}\\&=\big[\mathcal{N}_{0,I_{\mathbb{R}^2}}(\mathbb{R}^2)\big]^{d/2}=1.\end{aligned} \tag{8.14}$$

The proof of Lemma 8.1.4 is thus completed. □

Lemma 8.1.5 (First and second moments)**.** *Let $d\in\mathbb{N}$. Then it holds for all $i,j\in\{1,2,\dots,d\}$ that*

$$\int_{\mathbb{R}^d}x_i\,\mathcal{N}_{0,I_{\mathbb{R}^d}}(dx_1,\dots,dx_d)=0 \tag{8.15}$$

and

$$\int_{\mathbb{R}^d}x_i\cdot x_j\,\mathcal{N}_{0,I_{\mathbb{R}^d}}(dx_1,\dots,dx_d)=\begin{cases}1 & : i=j\\ 0 & : \textit{else}\end{cases}. \tag{8.16}$$

Proof of Lemma 8.1.5. First, observe that for all $x \in \mathbb{R}$ it holds that

$$\tfrac{d}{dx}\left[e^{\frac{-1}{2}x^2}\right] = (-x)\, e^{\frac{-1}{2}x^2}. \tag{8.17}$$

This ensures that

$$\int_{\mathbb{R}} x\, e^{\frac{-1}{2}x^2} dx = -\int_{\mathbb{R}} (-x)\, e^{\frac{-1}{2}x^2} dx = \left[e^{\frac{-1}{2}x^2}\right]_{x=-\infty}^{x=\infty} = 0. \tag{8.18}$$

Moreover, note that (8.17), Lemma 8.1.4, and integration by parts imply that

$$\begin{aligned} &\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x^2\, e^{-\frac{1}{2}x^2} dx = -\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x \cdot \left[(-x)\, e^{-\frac{1}{2}x^2}\right] dx \\ &= -\frac{1}{\sqrt{2\pi}} \left[x \cdot e^{-\frac{1}{2}x^2}\right]_{x=-\infty}^{x=\infty} + \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{1}{2}x^2}\, dx = 1. \end{aligned} \tag{8.19}$$

Combining Lemma 8.1.4, (8.18), and (8.19) proves that for all $k \in \{0, 1, 2\}$ it holds that

$$\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x^k\, e^{-\frac{1}{2}x^2} dx = \begin{cases} 0 & : k = 1 \\ 1 & : k \in \{0, 2\} \end{cases}. \tag{8.20}$$

This shows that for all $i \in \{1, 2, \ldots, d\}$, $k \in \{1, 2\}$ it holds that

$$\begin{aligned} &\int_{\mathbb{R}^d} [x_i]^k\, \mathcal{N}_{0, I_{\mathbb{R}^d}}(dx_1, \ldots, dx_d) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} [x_i]^k\, e^{-\frac{1}{2}\|x\|_{\mathbb{R}^d}^2}\, dx \\ &= \left[\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{1}{2}x^2} dx\right]^{(d-1)} \left[\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x^k\, e^{-\frac{1}{2}x^2} dx\right] \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x^k\, e^{-\frac{1}{2}x^2} dx = \begin{cases} 0 & : k = 1 \\ 1 & : k = 2 \end{cases}. \end{aligned} \tag{8.21}$$

Next note that (8.20) ensures that for all $i, j \in \{1, 2, \ldots, d\}$ with $i \neq j$ it holds that

$$\begin{aligned} &\int_{\mathbb{R}^d} x_i \cdot x_j\, \mathcal{N}_{0, I_{\mathbb{R}^d}}(dx_1, \ldots, dx_d) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} x_i\, x_j\, e^{-\frac{1}{2}\|x\|_{\mathbb{R}^d}^2}\, dx \\ &= \left[\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{1}{2}\|x\|_{\mathbb{R}^d}^2}\, dx\right]^{(d-2)} \left[\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x\, e^{-\frac{1}{2}x^2}\, dx\right]^2 \\ &= \left[\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x\, e^{-\frac{1}{2}x^2}\, dx\right]^2 = 0. \end{aligned} \tag{8.22}$$

This and (8.21) complete the proof of Lemma 8.1.5. □

8.2 Chapter 1

8.2.1 Solution to Exercise 1.2.16

Lemma 8.2.1 (∩-stability of the class of "southwest rectangles")**.** *Let $d \in \mathbb{N}$ and let $\mathcal{A} \subseteq \mathcal{P}(\mathbb{R}^d)$ be the set given by $\mathcal{A} = \cup_{x_1, \ldots, x_d \in \mathbb{R}} \{(-\infty, x_1) \times \cdots \times (-\infty, x_d)\}$. Then it holds that $\mathcal{A}$ is ∩-stable.*

Proof of Lemma 8.2.1. Let $(a_1, \dots, a_d), (b_1, \dots, b_d) \in \mathbb{R}^d$ and let $A, B \in \mathcal{A}$ be the sets given by

$$A = \times_{j=1}^{d} (-\infty, a_j) \qquad \text{and} \qquad B = \times_{j=1}^{d} (-\infty, b_j). \tag{8.23}$$

Then it holds for all $x = (x_1, \dots, x_d) \in A \cap B$, $j \in \{1, \dots, d\}$ that

$$x_j < a_j \qquad \text{and} \qquad x_j < b_j. \tag{8.24}$$

Moreover, it holds for all $x \in \times_{j=1}^{d} (-\infty, \min\{a_j, b_j\})$ that $x \in A \cap B$. This and (8.24) imply that

$$A \cap B = \times_{j=1}^{d} (-\infty, \min\{a_j, b_j\}) \in \mathcal{A}. \tag{8.25}$$

The proof of Lemma 8.2.1 is thus completed. □

Lemma 8.2.2 (Generation of $\mathcal{B}(\mathbb{R}^d)$ by rectangles)**.** *Let $d \in \mathbb{N}$ and let $\mathcal{R} \subseteq \mathcal{P}(\mathbb{R}^d)$ be the set given by*

$$\mathcal{R} = \bigcup_{\substack{a_1, \dots, a_d, \\ b_1, \dots, b_d \in \mathbb{Q}}} \big\{ [a_1, b_1) \times \dots \times [a_d, b_d) \big\}. \tag{8.26}$$

Then it holds that $\mathcal{B}(\mathbb{R}^d) = \sigma_{\mathbb{R}^d}(\mathcal{R})$.

Proof of Lemma 8.2.2. Throughout this proof for every open set $G \subseteq \mathbb{R}^d$ and every $y \in G$ let $A_y^G \in \mathcal{R}$ be a set with the property that $y \in A_y^G$ and $A_y^G \subseteq G$. Hence, we obtain for all open sets $G \subseteq \mathbb{R}^d$ that

$$G = \cup_{y \in G} A_y^G. \tag{8.27}$$

Combining (8.27) with the fact that $\mathcal{R}$ is countable implies that for every open set $G \subseteq \mathbb{R}^d$ there exists a countable set $J \subseteq G$ such that

$$G = \cup_{y \in G} A_y^G = \cup_{y \in J} A_y^G. \tag{8.28}$$

This and the fact that $\mathcal{B}(\mathbb{R}^d)$ is generated by the open sets in the normed $\mathbb{R}$-vector space $(\mathbb{R}^d, \|\cdot\|_{\mathbb{R}^d})$ shows that $\mathcal{B}(\mathbb{R}^d) \subseteq \sigma_{\mathbb{R}^d}(\mathcal{R})$. Next observe that for all $a_1, \dots, a_d, b_1, \dots, b_d \in \mathbb{Q}$ it holds that

$$\times_{j=1}^{d} \, [a_j, b_j) = \cap_{n \in \mathbb{N}} \times_{j=1}^{d} \, (a_j - \tfrac{1}{n}, b_j). \tag{8.29}$$

This, the fact that for all $a_1, \dots, a_d, b_1, \dots, b_d \in \mathbb{Q}$ it holds that the set $\times_{j=1}^{d} (a_j, b_j)$ is open in $\mathbb{R}^d$, and the fact that $\mathcal{B}(\mathbb{R}^d)$ is generated by the open sets in the normed $\mathbb{R}$-vector space $(\mathbb{R}^d, \|\cdot\|_{\mathbb{R}^d})$ show that

$$\sigma_{\mathbb{R}^d}(\mathcal{R}) \subseteq \mathcal{B}(\mathbb{R}^d). \tag{8.30}$$

This completes the proof of Lemma 8.2.2. □

Proposition 8.2.3 (An $\cap$-stable generating system for the Borel sigma-algebra)**.** *Let $d \in \mathbb{N}$ and let $\mathcal{A} \subseteq \mathcal{P}(\mathbb{R}^d)$ be the set given by $\mathcal{A} = \cup_{x_1, \dots, x_d \in \mathbb{R}} \{ (-\infty, x_1) \times \dots \times (-\infty, x_d) \}$. Then it holds that $\mathcal{B}(\mathbb{R}^d) = \sigma_{\mathbb{R}^d}(\mathcal{A})$.*

Proof of Proposition 8.2.3. Throughout this proof let $S_{(x_1,\dots,x_d)} \subseteq \mathbb{R}^d$, $(x_1,\dots,x_d) \in \mathbb{R}^d$, be the sets with the property that for all $(x_1,\dots,x_d) \in \mathbb{R}^d$ it holds that

$$S_{(x_1,\dots,x_d)} = \times_{j=1}^d (-\infty, x_j) \tag{8.31}$$

and let $\mathcal{R}, \mathcal{Q} \subseteq \mathcal{P}(\mathbb{R}^d)$ be the sets given by

$$\mathcal{R} = \bigcup_{\substack{a_1,\dots,a_d,\\ b_1,\dots,b_d \in \mathbb{Q}}} \{[a_1,b_1) \times \cdots \times [a_d,b_d)\} \qquad \text{and} \qquad \mathcal{Q} = \bigcup_{x_1,\dots,x_d \in \mathbb{Q}} \left\{S_{(x_1,\dots,x_d)}\right\}. \tag{8.32}$$

We first prove that $\mathcal{B}(\mathbb{R}^d) = \sigma_{\mathbb{R}^d}(\mathcal{Q})$. Observe that every element of $\mathcal{Q}$ is open in $\mathbb{R}^d$. This implies that $\sigma_{\mathbb{R}^d}(\mathcal{Q}) \subseteq \mathcal{B}(\mathbb{R}^d)$. Moreover, note that for all $a_1,\dots,a_d,b_1,\dots,b_d \in \mathbb{Q}$ it holds that

$$\times_{j=1}^d [a_j,b_j) = S_{(b_1,\dots,b_d)} \backslash \left[S_{(a_1,b_2,\dots,b_d)} \cup S_{(b_1,a_2,\dots,b_d)} \cup \cdots \cup S_{(b_1,b_2,\dots,a_d)}\right]. \tag{8.33}$$

This and Lemma 8.2.2 imply that

$$\mathcal{B}(\mathbb{R}^d) = \sigma_{\mathbb{R}^d}(\mathcal{R}) \subseteq \sigma_{\mathbb{R}^d}(\mathcal{Q}). \tag{8.34}$$

Therefore, we obtain that $\mathcal{B}(\mathbb{R}^d) = \sigma_{\mathbb{R}^d}(\mathcal{Q})$. In addition, observe that

$$\mathcal{B}(\mathbb{R}^d) = \sigma_{\mathbb{R}^d}(\mathcal{Q}) \subseteq \sigma_{\mathbb{R}^d}(\mathcal{A}). \tag{8.35}$$

Next note that every element of $\mathcal{A}$ is open in $\mathbb{R}^d$. This implies that $\mathcal{B}(\mathbb{R}^d) \supseteq \sigma_{\mathbb{R}^d}(\mathcal{A})$. Hence, we obtain that

$$\mathcal{B}(\mathbb{R}^d) = \sigma_{\mathbb{R}^d}(\mathcal{A}). \tag{8.36}$$

The proof of Proposition 8.2.3 is thus completed. □

8.2.2 Solution to Exercise 1.2.24

Lemma 8.2.4. *Let $a, b \in \mathbb{R}$ be real numbers with $a < b$ and let $F\colon \mathbb{R} \to [0,1]$ be a distribution function which satisfies for all $y \in (0,1)$ that*

$$I_F(y) = yb + (1-y)a. \tag{8.37}$$

Then it holds for all $x \in \mathbb{R}$ that

$$F(x) = \max\left\{0, \min\left\{1, \frac{(x-a)}{(b-a)}\right\}\right\} = \begin{cases} 0 & : x \leq a \\ \frac{(x-a)}{(b-a)} & : a \leq x \leq b \\ 1 & : b \leq x \end{cases}. \tag{8.38}$$

Proof of Lemma 8.2.4. Throughout this proof let Ω and $\mathcal{F}$ be the sets given by $\Omega = (0,1)$ and $\mathcal{F} = \mathcal{B}((0,1))$, let $P\colon \mathcal{F} \to [0,\infty]$ be a measure given by

$$P = B_{(0,1)}, \tag{8.39}$$

and let $U\colon \Omega \to \mathbb{R}$ be the function which satisfies for all $\omega \in \Omega$ that

$$U(\omega) = \omega. \tag{8.40}$$

Then it holds that $(\Omega, \mathcal{F}, P)$ is a probability space and it holds that $U\colon \Omega \to \mathbb{R}$ is an $\mathcal{U}_{(0,1)}$-distributed random variable which satisfies that

$$U(\Omega) \subseteq (0,1). \tag{8.41}$$

Proposition 1.2.7 hence proves that for all $x \in \mathbb{R}$ it holds that

$$\begin{aligned} F(x) &= P\big(I_F(U) \leq x\big) = P\big(Ub + (1-U)\,a \leq x\big) = P\big(U\,(b-a) + a \leq x\big) \\ &= P\big(U\,(b-a) \leq x - a\big) = P\left(U \leq \frac{(x-a)}{(b-a)}\right) \\ &= \max\left\{0, \min\left\{1, \frac{(x-a)}{(b-a)}\right\}\right\}. \end{aligned} \tag{8.42}$$

The proof of Lemma 8.2.4 is thus completed. □

8.2.3 Solution to Exercise 1.2.25

```
function X=Cauchy(N,mu,lambda)
  U=rand(1,N);
  X=lambda*tan(pi*(U - 0.5)) + mu;
end
```

Matlab code 8.1: A Matlab function `Cauchy(`N`,`μ`,`λ`)` with input $N \in \mathbb{N}$, $\mu \in \mathbb{R}$, $\lambda \in (0,\infty)$ and output a realization of an $(\mathrm{Cau}_{\mu,\lambda})^{\otimes N}$-distributed random variable generated with the inversion method. The Matlab function `Cauchy(`N`,`μ`,`λ`)` uses exactly N realizations of an $\mathcal{U}_{(0,1)}$-distributed random variable.

8.2.4 Solution to Exercise 1.2.26

Lemma 8.2.5 (Distribution function of the Laplace distribution)**.** *Let* $\lambda \in (0,\infty)$ *and let* $F\colon \mathbb{R} \to [0,1]$ *be the distribution function of* $\operatorname{Laplace}_\lambda$*. Then*

(i) it holds for all $x \in \mathbb{R}$ *that*

$$F(x) = \operatorname{Laplace}_\lambda((-\infty,x]) = \begin{cases} \frac{1}{2}e^{\lambda x} & : x < 0 \\ 1 - \frac{1}{2}e^{-\lambda x} & : x \geq 0 \end{cases} \tag{8.43}$$

and

(ii) it holds for all $y \in (0,1)$ *that*

$$I_F(y) = \begin{cases} \frac{1}{\lambda}\ln(2y) & : 0 < y < \frac{1}{2} \\ -\frac{1}{\lambda}\ln(2-2y) & : \frac{1}{2} \leq y < 1 \end{cases}. \tag{8.44}$$

Proof of Lemma 8.2.5. First of all, observe that for all $x \in \mathbb{R}$ it holds that

$$\begin{aligned} F(x) = \operatorname{Laplace}_\lambda((-\infty,x]) &= \frac{\lambda}{2}\int_{-\infty}^{x} e^{-\lambda|u|}\,du \\ &= \frac{\lambda}{2}\left[\int_{-\infty}^{\min\{x,0\}} e^{\lambda u}\,du + \int_{\min\{x,0\}}^{x} e^{-\lambda u}\,du\right] \\ &= \frac{\lambda}{2}\left[\frac{1}{\lambda}\,e^{\lambda u}\right]_{u=-\infty}^{u=\min\{x,0\}} + \frac{\lambda}{2}\left[-\frac{1}{\lambda}\,e^{-\lambda u}\right]_{u=\min\{x,0\}}^{u=x} \\ &= \frac{1}{2}\left[e^{\lambda u}\right]_{u=-\infty}^{u=\min\{x,0\}} - \frac{1}{2}\left[e^{-\lambda u}\right]_{u=\min\{x,0\}}^{u=x} \\ &= \frac{1}{2}\left[e^{\lambda\min\{x,0\}} + e^{-\lambda\min\{x,0\}} - e^{-\lambda x}\right] = \begin{cases} \frac{1}{2}e^{\lambda x} & : x < 0 \\ 1 - \frac{1}{2}e^{-\lambda x} & : x \geq 0 \end{cases}. \end{aligned} \tag{8.45}$$

This, in particular, ensures that F is strictly increasing. Lemma 1.2.4 hence proves that for all $y \in (0,1) = F(\mathbb{R})$ it holds that

$$y = F(F^{-1}(y)) = F(I_F(y)). \tag{8.46}$$

The fact that $F([0,\infty)) = [\frac{1}{2},1)$ and (8.45) therefore ensure that for all $y \in (0,\frac{1}{2})$ it holds that

$$\frac{1}{2} > y = F(I_F(y)) = \tfrac{1}{2}e^{\lambda I_F(y)}. \tag{8.47}$$

This shows that for all $y \in (0,\frac{1}{2})$ it holds that

$$\tfrac{1}{\lambda}\ln(2y) = I_F(y). \tag{8.48}$$

In addition, we observe that (8.45), (8.46), and the fact that $F((-\infty,0)) \subseteq (0,\frac{1}{2})$ imply that for all $y \in [\frac{1}{2},1)$ it holds that

$$\frac{1}{2} \leq y = F(I_F(y)) = 1 - \frac{1}{2}e^{-\lambda I_F(y)}. \tag{8.49}$$

Hence, we obtain that for all $y \in [\frac{1}{2},1)$ it holds that

$$\ln(2-2y) = -\lambda I_F(y). \tag{8.50}$$

This proves that for all $y \in [\frac{1}{2},1)$ it holds that

$$-\tfrac{1}{\lambda}\ln(2-2y) = I_F(y). \tag{8.51}$$

Combining this with (8.48) completes the proof of Lemma 8.2.5. □

8.2.5 Solution to Exercise 1.2.27

```
function X=Laplace(N,lambda)
  U=rand(1,N);
  X=InvDFLaplace(U,lambda);
end
```

Matlab code 8.2: A Matlab function `Laplace(`N`,`λ`)` with input $N \in \mathbb{N}$, $\lambda \in (0,\infty)$ and output a realization of an $(\text{Laplace}_\lambda)^{\otimes N}$-distributed random variable generated with the inversion method. The Matlab function `Laplace(`N`, `λ`)` uses exactly N realizations of an $\mathcal{U}_{(0,1)}$-distributed random variable.

```
function x=InvDFLaplace(y,lambda)
  x=zeros(size(y));
  I=(y<0.5);
  x(I)=log(2*y(I))/lambda;
  I=logical(1-I);
  x(I)=-log(2*(1-y(I)))/lambda;
end
```

```
function LaplacePlot()
  rng('default');
  N=10^5;
  lambda=0.1;
  X=Laplace(N,lambda);
  hist(X,10^3)
end
```

Matlab code 8.3: A Matlab function `LaplacePlot()` which plots 10^5 realizations of an $\text{Laplace}_{0.1}$-distributed random variable generated with the Matlab function `Laplace(`N`,`λ`)` in a histogram with 1000 bins.

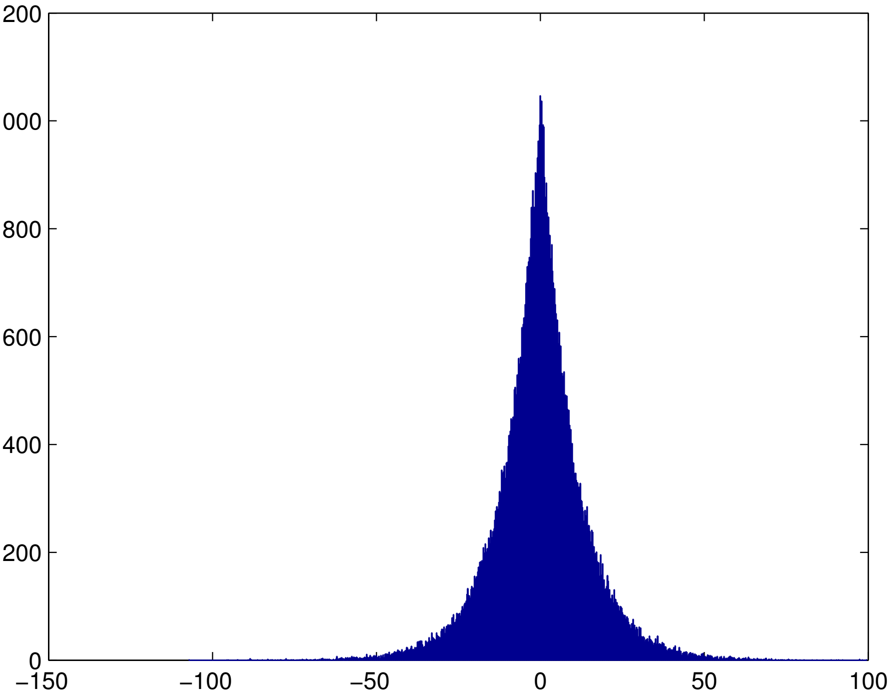


Figure 8.1: Result of a call of the Matlab function 8.3.

8.2.6 Solution to Exercise 1.2.30

```
function X = AcceptanceRejection(N)
  X = [];
  while ( length(X) < N )
    U = rand(2, N-length(X));
    U = [8 0; 0 sqrt(8)]*U;
    U(1,:) = U(1,:) - 4;
    U(2,:) = U(2,:) - sqrt(2);
    I = ( U(1,:).^2/8 + U(2,:).^2 <= 2 );
```

```
      X = [ X, U(:,I) ];
   end
end
```

Matlab code 8.4: A Matlab function `AcceptanceRejection(`N`)` with input $N \in \mathbb{N}$ and output a realization of an $(\mathcal{U}_A)^{\otimes N}$-distributed random variable generated with the acceptance-rejection method.

```
function AcceptanceRejectionPlot()
   rng('default');
   N=10^5;
   X=AcceptanceRejection(N);
   plot(X(1,:),X(2,:),'*')
   axis([-4.5 4.5 -1.7 1.7]);
   daspect([1 1 1]);
end
```

Matlab code 8.5: A Matlab function `AcceptanceRejectionPlot()` which uses the Matlab function `AcceptanceRejection(`N`)` and the built-in Matlab function `plot(...)` to plot 10^5 realizations of an $\mathcal{U}_A$-distributed random variable in a coordinate plane.

8.2.7 Solution to Exercise 1.2.38

Lemma 8.2.6. *Let $f, \tilde{f} \colon \mathbb{R} \to [0,\infty)$ be the functions which satisfy for all $x \in \mathbb{R}$ that*

$$f(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2} \qquad \text{and} \qquad \tilde{f}(x) = \frac{1}{\pi\,(1+x^2)} \tag{8.52}$$

and let $\tilde{A} \subseteq \mathbb{R}$ be the set given by $\tilde{A} = \big\{C \in \mathbb{R} \colon \big(\forall\, x \in \mathbb{R} \colon f(x) \leq C\tilde{f}(x)\big)\big\}$. Then it holds that $C \in \tilde{A}$ if and only if $\big(\forall\, y \in [0,\infty) \colon 1 + 2y \leq \frac{\sqrt{2}\,C\,e^y}{\sqrt{\pi}}\big)$.

Proof of Lemma 8.2.6. The definition of $\tilde{A}$ and (8.52) ensure that $C \in \tilde{A}$ if and only if for all $x \in \mathbb{R}$ it holds that

$$\frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2} \leq C\,\frac{1}{\pi(1+x^2)} = C\,\frac{1}{\pi\left(1 + 2\left[\frac{1}{2}x^2\right]\right)}. \tag{8.53}$$

The fact that $\left\{\frac{1}{2}x^2 \in \mathbb{R} \colon x \in \mathbb{R}\right\} = [0,\infty)$ hence proves that $C \in \tilde{A}$ if and only if for all $y \in [0,\infty)$ it holds that

$$\frac{1}{\sqrt{2\pi}} e^{-y} \leq C \frac{1}{\pi(1+2y)}. \tag{8.54}$$

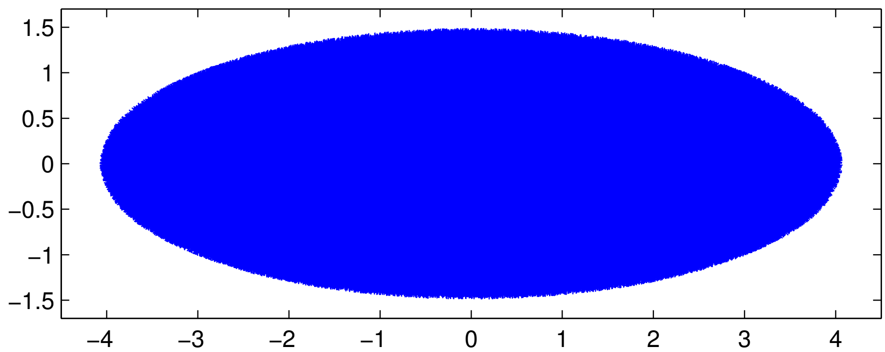


Figure 8.2: Result of a call of the Matlab function 8.5.

Reorganizing the terms in (8.54) shows that $C \in \tilde{A}$ if and only if for all $y \in [0, \infty)$ it holds that

$$1 + 2y \leq \frac{\sqrt{2}\, C\, e^{y}}{\sqrt{\pi}}. \tag{8.55}$$

The proof of Lemma 8.2.6 is thus completed. □

Lemma 8.2.7. *Let $f, \tilde{f} \colon \mathbb{R} \to [0, \infty)$ be the functions which satisfy for all $x \in \mathbb{R}$ that*

$$f(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2} \qquad \textit{and} \qquad \tilde{f}(x) = \frac{1}{\pi\,(1 + x^2)} \tag{8.56}$$

and let $\tilde{A} \subseteq \mathbb{R}$ be the set given by $\tilde{A} = \big\{C \in \mathbb{R} \colon \big(\forall\, x \in \mathbb{R} \colon f(x) \leq C \tilde{f}(x)\big)\big\}$. Then it holds that

$$\tilde{A} = \left[\sup_{x \in \mathbb{R}} \left(\frac{f(x)}{\tilde{f}(x)}\right), \infty\right). \tag{8.57}$$

Proof of Lemma 8.2.7. Throughout this proof let $\kappa \in \mathbb{R}$ be the real number given by

$$\kappa = \sup_{y \in \mathbb{R}} \frac{f(y)}{\tilde{f}(y)}. \tag{8.58}$$

The definition of the set $\tilde{A}$ assures that for all $C \in \tilde{A}$, $x \in \mathbb{R}$ it holds that

$$\frac{f(x)}{\tilde{f}(x)} \leq C. \tag{8.59}$$

Hence, we obtain for all $C \in \tilde{A}$ that

$$\kappa \leq C. \tag{8.60}$$

This implies that

$$\tilde{A} \subseteq [\kappa, \infty). \tag{8.61}$$

Next observe that the definition of κ ensures that for all $x \in \mathbb{R}$, $C \in [\kappa, \infty)$ it holds that

$$\frac{f(x)}{\tilde{f}(x)} \leq \sup_{y \in \mathbb{R}} \left(\frac{f(y)}{\tilde{f}(y)}\right) = \kappa \leq C. \tag{8.62}$$

Therefore, we obtain that for all $x \in \mathbb{R}$, $C \in [\kappa, \infty)$ it holds that

$$f(x) \leq C \tilde{f}(x). \tag{8.63}$$

This proves that

$$[\kappa, \infty) \subseteq \tilde{A}. \tag{8.64}$$

Hence, we obtain that $\tilde{A} = [\kappa, \infty)$. The proof of Lemma 8.2.7 is thus completed. □

Lemma 8.2.8. *Let $f, \tilde{f} \colon \mathbb{R} \to [0, \infty)$ be the functions which satisfy for all $x \in \mathbb{R}$ that*

$$f(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2} \qquad \text{and} \qquad \tilde{f}(x) = \frac{1}{\pi\,(1+x^2)}. \tag{8.65}$$

Then it holds that

$$\sup_{x\in\mathbb{R}} \frac{f(x)}{\tilde{f}(x)} = \sqrt{\frac{2\pi}{e}} \qquad \text{and} \qquad \frac{\int_{\mathbb{R}} \left[\sup_{y\in\mathbb{R}} \frac{f(y)}{\tilde{f}(y)}\right] \tilde{f}(x)\, dx}{\int_{\mathbb{R}} f(x)\, dx} = \sqrt{\frac{2\pi}{e}}. \tag{8.66}$$

Proof of Lemma 8.2.8. Throughout this proof let $\kappa \in \mathbb{R}$ be the real number given by

$$\kappa = \sup_{y\in\mathbb{R}} \frac{f(y)}{\tilde{f}(y)}, \tag{8.67}$$

let $\tilde{A} \subseteq \mathbb{R}$ be the set given by $\tilde{A} = \big\{C \in \mathbb{R} \colon \big(\forall\, x \in \mathbb{R} \colon f(x) \leq C\tilde{f}(x)\big)\big\}$, and let $g : [0, \infty) \to \mathbb{R}$ be the function with the property that for all $y \in [0, \infty)$ it holds that

$$g(y) = e^{-y}(1 + 2y). \tag{8.68}$$

Lemma 8.2.7 implies that $\min(\tilde{A}) = \kappa$. Hence, Lemma 8.2.6 ensures for all $\varepsilon \in (0, \infty)$ that there exists a real number $y_0 \in [0, \infty)$ such that

$$1 + 2y_0 > \frac{\sqrt{2}(\kappa - \varepsilon)e^{y_0}}{\sqrt{\pi}}. \tag{8.69}$$

This implies that for every $\varepsilon \in (0, \infty)$ there exists a real number $y_0 \in [0, \infty)$ such that

$$\sqrt{\frac{\pi}{2}} e^{-y_0}(1 + 2y_0) > \kappa - \varepsilon. \tag{8.70}$$

Similarly, Lemma 8.2.6 proves that for all $y \in [\kappa, \infty)$ it holds that

$$\sqrt{\frac{\pi}{2}} e^{-y}(1 + 2y) \leq \kappa. \tag{8.71}$$

Combining (8.70) and (8.71) proves that

$$\sqrt{\frac{\pi}{2}} \sup_{y\in[0,\infty)} e^{-y}(1 + 2y) = \sqrt{\frac{\pi}{2}} \sup_{y\in[0,\infty)} g(y) = \kappa. \tag{8.72}$$

The fact that g is smooth enables us to find its extrema by analyzing its derivatives. Note that for all $y \in (0, \infty)$ it holds that

$$g'(y) = e^{-y}(1 - 2y) \tag{8.73}$$

and

$$g''(y) = e^{-y}(-3 + 2y). \tag{8.74}$$

Next note that $g'(\frac{1}{2}) = 0$ and $g''(\frac{1}{2}) < 0$. This implies that g has a local maximum point at $\frac{1}{2}$. Observe that g' has no roots in $(0,\infty)\backslash\{\frac{1}{2}\}$. This shows that $\frac{1}{2}$ is the only local maximum point of the function $g|_{(0,\infty)}$. Combining the fact that $g(0) = 1 \leq g(\frac{1}{2}) = \frac{2}{\sqrt{e}}$ and

$$\lim_{y\to\infty} e^{-y}(1+2y) = \lim_{y\to\infty} \frac{1+2y}{e^y} = 0 \tag{8.75}$$

with (8.72) proves that

$$\kappa = \sqrt{\frac{\pi}{2}} \sup_{y\in[0,\infty)} e^{-y}(1+2y) = \sqrt{\frac{\pi}{2}} \frac{2}{\sqrt{e}} = \sqrt{\frac{2\pi}{e}}. \tag{8.76}$$

Furthermore, Lemma 8.2.7 and the fact that f and $\tilde{f}$ are probability density functions imply that

$$\frac{\int_{\mathbb{R}} \left[\sup_{y\in\mathbb{R}} \frac{f(y)}{\tilde{f}(y)}\right] \tilde{f}(x)\,dx}{\int_{\mathbb{R}} f(x)\,dx} = \sup_{y\in\mathbb{R}} \frac{f(y)}{\tilde{f}(y)} = \kappa = \sqrt{\frac{2\pi}{e}}. \tag{8.77}$$

The proof of the Lemma 8.2.8 is thus completed. □

Please be aware that the following Matlab function makes use of the Matlab function `Cauchy(`$N, mu, lambda$`)`, which is part of the solution to Exercise 1.2.25.

```
function  X = AcceptanceRejectionGaussianCauchy(N)
  X = [];
  while( length(X) < N )
    Y = Cauchy(N - length(X) ,0,1);
    U = rand(1, N - length(X) );
    kappa = sqrt( 2*pi/exp(1) );
    I = ( U / pi./(1 + Y.^2) * kappa...
           <= 1/sqrt(2*pi)*exp(-1/2*Y.^2) );
    X = [ X, Y(I) ];
  end
end
```

Matlab code 8.6: A Matlab function `AcceptanceRejectionGaussianCauchy(`N`)` with input $N \in \mathbb{N}$ and output a realization of an $(\mathcal{N}_{0,I_{\mathbb{R}}})^{\otimes N}$-distributed random variable generated with the acceptance-rejection method with f as the density of the target distribution and $\kappa\tilde{f}$ as the unnormalized density of the proposal distribution $\mathrm{Cau}_{0,1}$.

```
function AcceptanceRejectionGaussianCauchyPlot()
  rng('default');
  N=10^5;
  X=AcceptanceRejectionGaussianCauchy(N);
  hist(X,1000);
```

```
end
```

Matlab code 8.7: A Matlab function `AcceptanceRejectionGaussianCauchyPlot()` which plots 10^5 realizations of an $\mathcal{N}_{0,I_{\mathbb{R}}}$-distributed random variable generated with the Matlab function `AcceptanceRejectionGaussianCauchy(`N`)` in a histogram with 1000 bins.

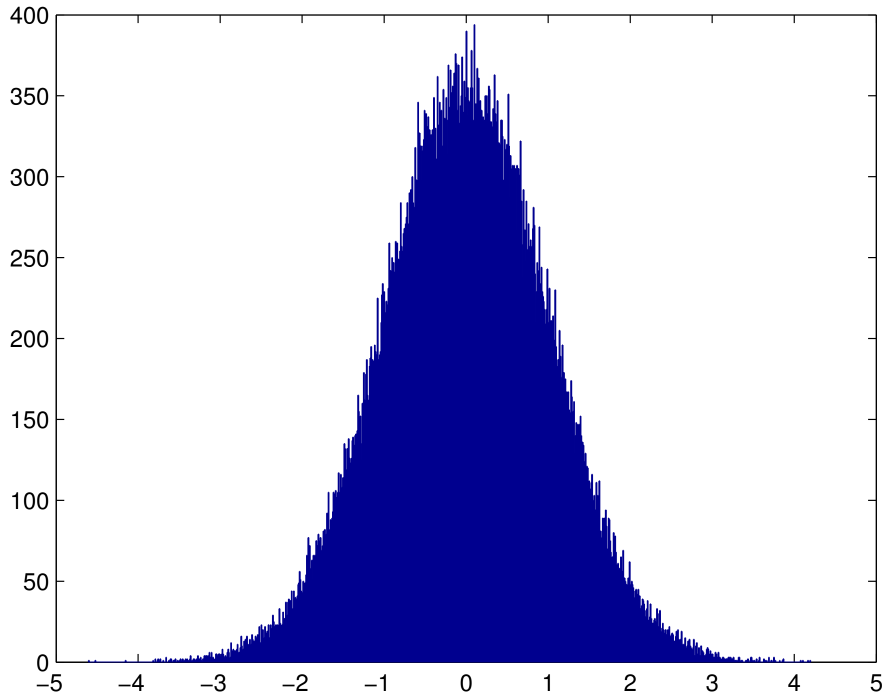


Figure 8.3: Result of a call of the Matlab function 8.7.

8.2.8 Solution to Exercise 1.2.39

Lemma 8.2.9. *Let $f, \hat{f} \colon \mathbb{R} \to [0, \infty)$ be the functions which satisfy for all $x \in \mathbb{R}$ that*

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \qquad \text{and} \qquad \hat{f}(x) = \frac{e^{-|x|}}{2} \tag{8.78}$$

and let $\hat{A} \subseteq \mathbb{R}$ be the set given by $\hat{A} = \big\{C \in \mathbb{R} \colon \big(\forall\, x \in \mathbb{R} \colon f(x) \leq C\hat{f}(x)\big)\big\}$. Then

$$\hat{A} = \left[\sup_{x \in \mathbb{R}} \left(\frac{f(x)}{\hat{f}(x)}\right), \infty\right). \tag{8.79}$$

Proof of Lemma 8.2.9. Throughout this proof let $\hat{\kappa} \in \mathbb{R}$ be the real number given by $\hat{\kappa} = \sup_{y \in \mathbb{R}} \frac{f(y)}{\hat{f}(y)}$. The definition of the set $\hat{A}$ implies that for all $C \in \hat{A}$, $x \in \mathbb{R}$ it holds that

$$\frac{f(x)}{\hat{f}(x)} \leq C. \tag{8.80}$$

Hence, we obtain that for all $C \in \hat{A}$ it holds that

$$\hat{\kappa} \leq C. \tag{8.81}$$

This implies that $\hat{A} \subseteq [\hat{\kappa}, \infty)$. Moreover, the definition of $\hat{\kappa}$ implies that for all $C \in [\hat{\kappa}, \infty)$, $x \in \mathbb{R}$ it holds that

$$\frac{f(x)}{\hat{f}(x)} \leq \sup_{y \in \mathbb{R}} \frac{f(y)}{\hat{f}(y)} = \hat{\kappa} \leq C. \tag{8.82}$$

This shows that for all $x \in \mathbb{R}$, $C \in [\hat{\kappa}, \infty)$ it holds that

$$f(x) \leq C\hat{f}(x). \tag{8.83}$$

This proves that $[\hat{\kappa}, \infty) \subseteq \hat{A}$. Hence, we obtain that $\hat{A} = [\hat{\kappa}, \infty)$. The proof of Lemma 8.2.9 is thus completed. □

Lemma 8.2.10. *Let $f, \hat{f} \colon \mathbb{R} \to [0, \infty)$ be the functions which satisfy for all $x \in \mathbb{R}$ that*

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \qquad \text{and} \qquad \hat{f}(x) = \frac{e^{-|x|}}{2}. \tag{8.84}$$

Then it holds that

$$\sup\nolimits_{x \in \mathbb{R}} \frac{f(x)}{\hat{f}(x)} = \sqrt{\frac{2e}{\pi}} \qquad \text{and} \qquad \frac{\int_{\mathbb{R}} \Big[\sup_{y \in \mathbb{R}} \frac{f(y)}{\hat{f}(y)}\Big] \hat{f}(x)\, dx}{\int_{\mathbb{R}} f(x)\, dx} = \sqrt{\frac{2e}{\pi}}. \tag{8.85}$$

Proof of Lemma 8.2.10. Throughout this proof let $h\colon \mathbb{R} \to \mathbb{R}$ be the function which satisfies for all $x \in \mathbb{R}$ that

$$h(x) = \frac{f(x)}{\hat{f}(x)} = \sqrt{\frac{2}{\pi}}\, e^{-\frac{1}{2}x^2} e^{|x|}. \tag{8.86}$$

We are interested in the local maxima of the function h. The fact that h is an even function, i.e., $\forall\, x \in \mathbb{R}\colon h(x) = h(-x)$, implies that it is enough to analyze h over $(0,\infty)$. First of all, note that $h|_{(0,\infty)}\colon (0,\infty) \to \mathbb{R}$ is smooth. Next observe that for all $x \in (0,\infty)$ it holds that

$$h'(x) = \sqrt{\frac{2}{\pi}}\,(1-x) e^{-\frac{1}{2}x^2} e^{x} \tag{8.87}$$

and

$$\begin{aligned} h''(x) &= \frac{\mathrm{d}}{\mathrm{d}x}\left[\sqrt{\frac{2}{\pi}}\,(1-x)e^{-\frac{1}{2}x^2}e^{x}\right] \\ &= \sqrt{\frac{2}{\pi}}\left[\frac{\mathrm{d}}{\mathrm{d}x}(1-x)\right]e^{-\frac{1}{2}x^2}e^{x} + \sqrt{\frac{2}{\pi}}\,(1-x)\left[\frac{\mathrm{d}}{\mathrm{d}x}e^{-\frac{1}{2}x^2}e^{x}\right] \\ &= \sqrt{\frac{2}{\pi}}\left[-1 + (1-x)^2\right]e^{-\frac{1}{2}x^2}e^{x}. \end{aligned} \tag{8.88}$$

Additionally, note that $h'(1) = 0$ and $h''(1) < 0$. Combining (8.87) with (8.88) hence proves that h has a local maximum point at 1 with value

$$h(1) = \sqrt{\frac{2e}{\pi}}. \tag{8.89}$$

The function $h'|_{(0,\infty)\backslash\{1\}}\colon (0,\infty)\backslash\{1\} \to \mathbb{R}$ has no roots. This implies that 1 is the only local maximum point of $h|_{(0,\infty)}$. Combining this with the fact that

$$h(0) = \sqrt{\frac{2}{\pi}} < h(1) \tag{8.90}$$

and

$$\lim_{|x|\to\infty} h(x) = 0 \tag{8.91}$$

implies that

$$\sup_{x\in\mathbb{R}} \frac{f(x)}{\hat{f}(x)} = h(1) = \sqrt{\frac{2e}{\pi}}. \tag{8.92}$$

The fact that f and $\hat{f}$ are normalized density functions hence proves that

$$\frac{\int_{\mathbb{R}} \left[\sup_{y\in\mathbb{R}} \frac{f(y)}{\hat{f}(y)}\right] \hat{f}(x)\, dx}{\int_{\mathbb{R}} f(x)\, dx} = \frac{\int_{\mathbb{R}} \left[\sup_{y\in\mathbb{R}} \frac{f(y)}{\hat{f}(y)}\right] \hat{f}(x)\, dx}{\int_{\mathbb{R}} f(x)\, dx} = \sqrt{\frac{2e}{\pi}}. \tag{8.93}$$

The proof of Lemma 8.2.10 is thus completed. □

Please be aware that the following Matlab function makes use of the Matlab function `Laplace(`N`,`$lambda$`)`, which itself uses the Matlab function `InvDFLaplace(`$y, lambda$`)`. These are both part of the solution to Exercise 1.2.27.

```
function X = AcceptanceRejectionGaussianLaplace(N)
  X = [];
  while ( length(X) < N )
    Y = Laplace(N - length(X) , 1 );
    U = rand(1,N - length(X) );
    kappa = sqrt(2*exp(1)/pi);
    I = ( U .* exp(-abs(Y))/2 * kappa ...
          <= 1/sqrt(2*pi)*exp(-1/2*Y.^2) );
    X = [ X, Y(I) ];
  end
end
```

Matlab code 8.8: A Matlab function `AcceptanceRejectionGaussianLaplace(`N`)` with input $N \in \mathbb{N}$ and output a realization of an $(\mathcal{N}_{0,I_{\mathbb{R}}})^{\otimes N}$-distributed random variable generated with the acceptance-rejection method with f as the density of the target distribution and $\hat{\kappa}\hat{f}$ as the unnormalized density of the proposal distribution $\mathrm{Cau}_{0,1}$.

```
function AcceptanceRejectionGaussianLaplacePlot()
  rng('default');
  N=10^5;
  X=AcceptanceRejectionGaussianLaplace(N);
  hist(X,1000);
end
```

Matlab code 8.9: A Matlab function `AcceptanceRejectionGaussianLaplacePlot()` which plots 10^5 realizations of an $\mathcal{N}_{0,I_{\mathbb{R}}}$-distributed random variable generated with the Matlab function `AcceptanceRejectionGaussianLaplace(`N`)` in a histogram with 1000 bins.

8.2.9 Solution to Exercise 1.3.7

```
function X = BoxMuller( N )
% allocating memory for the output
% and simultaneously generating uniformly distributed
% random variables
```

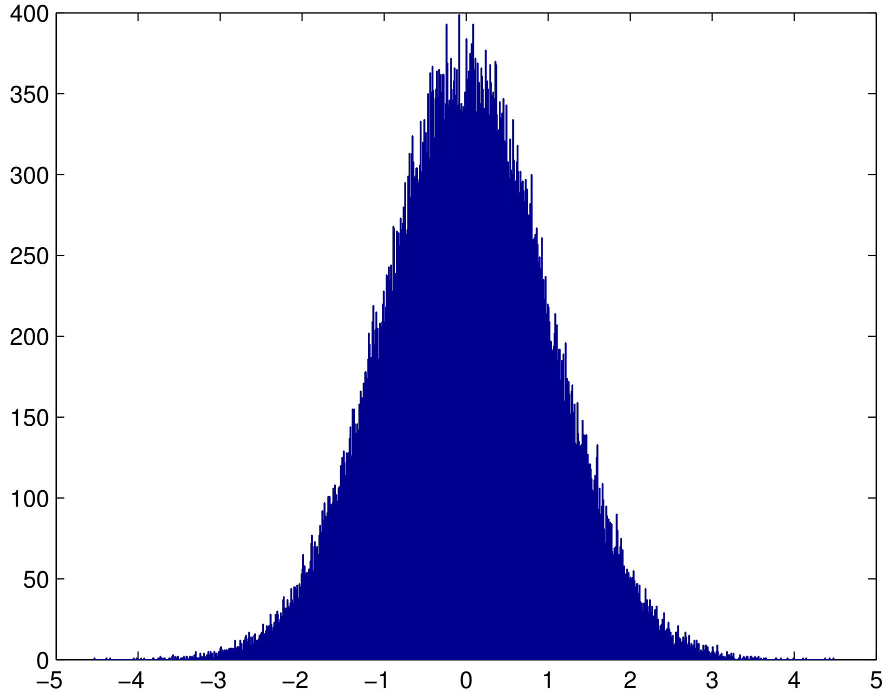


Figure 8.4: Result of a call of the Matlab function 8.9.

```
% (one in addition to N if N is odd)
  X = rand( 2, ceil( N / 2 ) );

  X = [ sqrt( -2*log( X(1,:) ) ) .* ...
        cos( 2*pi * X(2,:) ), ...
        sqrt( -2*log( X( 1, 1 : floor( N / 2 ) ) ) ) .* ...
        sin( 2*pi * X( 2, 1: floor( N / 2 ) ) ) ];
end
```

Matlab code 8.10: A Matlab function `BoxMuller(`N`)` with input $N \in \mathbb{N}$ and output a realization of an $\mathcal{N}_{0,I_{\mathbb{R}^N}}$-distributed random variable generated with the Box-Muller method.

```
function BoxMullerPlot()
  rng('default');
  N = 10^5;
% Probability density function of the normal distribution
  f = @(x) exp(-0.5*x.^2)/sqrt(2*pi);
% Generating the normalized histogram for the Box-Muller
% sampling method
  X_BM = BoxMuller(N);
  figure(1)
  clf
  [N_BM,x]=hist(X_BM,1000);
% Normalizing the bins such that they have area 1
  bar(x,N_BM/N/(x(2)-x(1)))
  hold on
  plot(x,f(x),'r')
  legend('normalized histogram','normal distribution')
  hold off
end
```

Matlab code 8.11: A Matlab function `BoxMullerPlot()` which plots 10^5 realizations of an $\mathcal{N}_{0,I_{\mathbb{R}}}$-distributed random variable generated with the Matlab function `BoxMuller(`N`)` in a *normalized* histogram with 1000 bins and which also plots the density of $\mathcal{N}_{0,I_{\mathbb{R}}}$ in this histogram.

8.2.10 Solution to Exercise 1.3.9

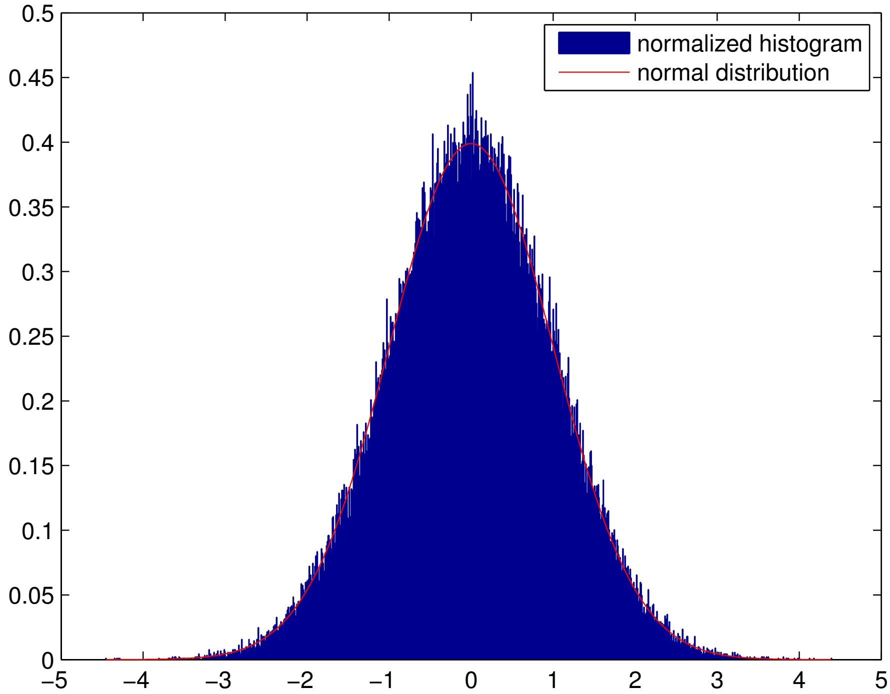


Figure 8.5: Result of a call of the Matlab function 8.11.

```
function X = MarsagliaPolar( N )
% Sampling from uniformly distributed random vectors
% over the unit disk
% in the 2-dimensional real space
  X = AcceptanceRejectionUnitDisk( ceil( N / 2 ) );
  q = sum(X.^2,1);
  q = sqrt( -2*log(q)./q );
  X = [ q.*X(1,:), q( 1: floor(N/2) ).* ...
        X( 2, 1: floor(N/2) ) ];
end
```

Matlab code 8.12: A Matlab function `MarsagliaPolar(`N`)` with input $N \in \mathbb{N}$ and output a realization of an $\mathcal{N}_{0,I_{\mathbb{R}^N}}$-distributed random variable generated with the Marsaglia polar method.

```
function X = AcceptanceRejectionUnitDisk(N)
  X = [];
  while ( length(X) < N )
    RV = rand( 2, N-length(X) );
    RV = 2*RV - 1;
    I = ( RV(1,:).^2 + RV(2,:).^2 < 1 );
    X = [ X, RV(:,I) ];
  end
end
```

Matlab code 8.13: A Matlab function `AcceptanceRejectionUnitDisk(`N`)` with input $N \in \mathbb{N}$ and output N realizations of uniformly-distributed random variables over the unit disk generated with the acceptance-rejection method.

```
function MarsagliaPolarPlot()
  rng('default');
  N = 10^5;
% Probability density function of the normal distribution
  f = @(x) exp(-0.5*x.^2)/sqrt(2*pi);
% Generating the normalized histogram for the Marsaglia
% polar sampling method
  X_MP = MarsagliaPolar(N);
  figure(1)
  clf
  [N_MP,x]=hist(X_MP,1000);
% Normalizing the bins such that they have area 1
  bar(x,N_MP/N/(x(2)-x(1)))
```

```
    hold on
    plot(x,f(x),'r')
    legend('normalized histogram','normal distribution')
    hold off
end
```

Matlab code 8.14: A Matlab function `MarsagliaPolarPlot()` which plots 10^5 realizations of an $\mathcal{N}_{0,I_{\mathbb{R}}}$-distributed random variable generated with the Matlab function `MarsagliaPolar(`N`)` in a *normalized* histogram with 1000 bins and which also plots the density of $\mathcal{N}_{0,I_{\mathbb{R}}}$ in this histogram.

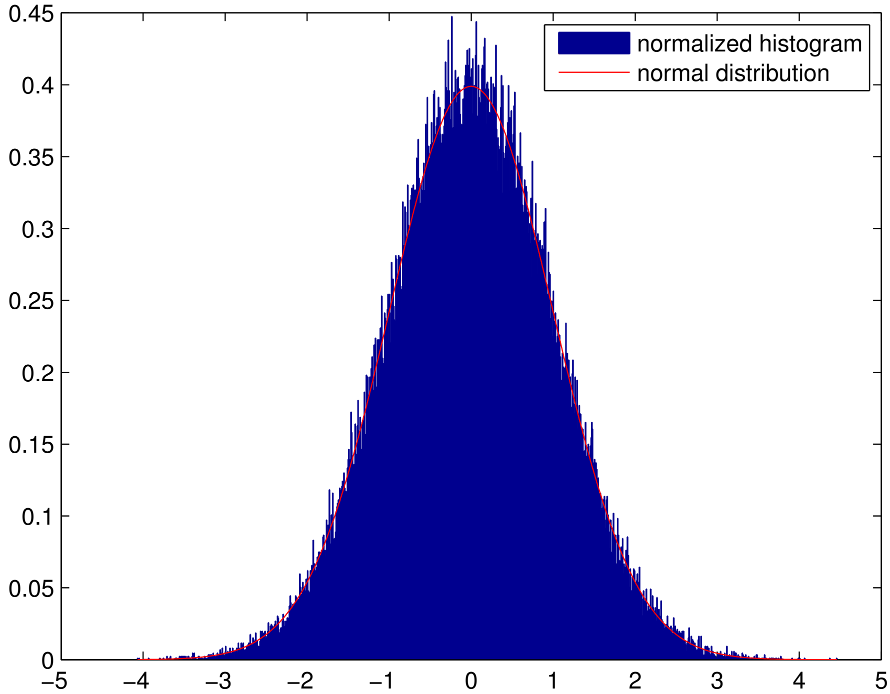


Figure 8.6: Result of a call of the Matlab function 8.14.

8.2.11 Solution to Exercise 1.3.11

```
function X = StandardBrownianMotion(t)
%getting rid of 0-valued t's
```

```
    I = (t>0);     Q = t(I);
%constructing the desired covariance matrix
    Q = repmat( Q, size(Q') );
    Q = min(Q,Q');
%assigning the Brownian Motion at times t to X
    Q = chol(Q)'*randn(sum(I),1);
    X = zeros(length(I),1);
    X(I) = Q;
end
```

Matlab code 8.15: A Matlab function `StandardBrownianMotion(t)` with input $\mathbf{t} \in A$ and output a realization of an $\mathcal{N}_{0,Q(\mathbf{t})}$-distributed random variable.

```
rng('default');
N=10^3;
preimage = (0:1/N:1);
X=StandardBrownianMotion(preimage);
plot(preimage,X);
hold on
X=StandardBrownianMotion(preimage);
plot(preimage,X,'r');
X=StandardBrownianMotion(preimage);
plot(preimage,X,'g');
```

Matlab code 8.16: A Matlab function `StandardBrownianMotionPlot()` which plots linearly interpolated 3 realizations of an $\mathcal{N}_{0,Q(\mathbf{t})}$-distributed random variable generated with the Matlab function `StandardBrownianMotion(t)`.

8.3 Chapter 2

8.3.1 Solution to Exercise 2.1.12

Lemma 8.3.1. *Let (E, d_E) and (F, d_F) be metric spaces and let $f: E \to F$ be a function. Then it holds that f is uniformly continuous if and only if*

$$\lim_{h \searrow 0} w_f(h) = w_f(0). \tag{8.94}$$

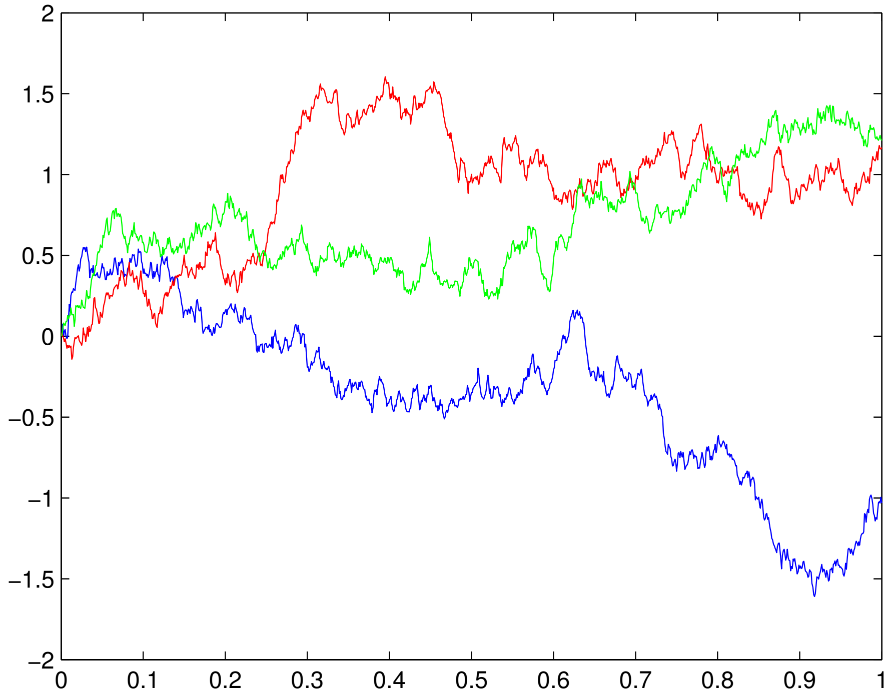


Figure 8.7: Result of a call of the Matlab function 8.16.

Proof of Lemma 8.3.1. Without loss of generality we assume that $E \neq \emptyset$. Next observe that the fact that $w_f(0) = 0$ ensures that it holds that $\lim_{h \searrow 0} w_f(h) = \lim_{h \searrow 0} \sup_{x,y \in E, d_E(x,y) \leq h} d_F\big(f(x), f(y)\big) = w_f(0)$ if and only if it holds that

$$\forall\, \varepsilon \in (0,\infty) \colon \exists\, h \in (0,\infty) \colon \sup_{x,y \in E, d_E(x,y) \leq h} d_F\big(f(x), f(y)\big) \leq \varepsilon. \tag{8.95}$$

This implies that $\lim_{h \searrow 0} w_f(h) = w_f(0)$ if and only if

$$\forall\, \varepsilon \in (0,\infty) \colon \exists\, \delta \in (0,\infty) \colon \forall\, x, y \in E \text{ with } d_E(x,y) \leq \delta \colon d_F\big(f(x), f(y)\big) \leq \varepsilon. \tag{8.96}$$

This completes the proof of Lemma 8.3.1. □

8.3.2 Solution to Exercise 2.1.13

Lemma 8.3.2. *Let (E, d_E) and (F, d_F) be metric spaces, let $\alpha \in (0,1]$ be a real number, and let $f \colon E \to F$ be a function. Then it holds that*

$$|f|_{C^\alpha(E,F)} = \sup_{h \in (0,\infty)} \left[\frac{w_f(h)}{h^\alpha} \right]. \tag{8.97}$$

Proof of Lemma 8.3.2. Without loss of generality we assume that $\#_E > 1$. Note that for all $x, y \in E$, $h \in (0,\infty)$ with $x \neq y$ it holds that $d_E(x,y) \leq h$ implies that $\frac{1}{|d_E(x,y)|^\alpha} \geq \frac{1}{h^\alpha}$. This together with the definition of $|\cdot|_{C^\alpha(E,F)}$ and w_f imply for all $h \in (0,\infty)$ that

$$|f|_{C^\alpha(E,F)} = \sup_{\substack{x,y \in E,\\ x \neq y}} \left[\frac{d_F(f(x), f(y))}{|d_E(x,y)|^\alpha} \right] \geq \frac{1}{h^\alpha} \left[\sup_{\substack{x,y \in E,\\ d_E(x,y) \leq h}} d_F(f(x), f(y)) \right] = \frac{w_f(h)}{h^\alpha}. \tag{8.98}$$

This implies that

$$|f|_{C^\alpha(E,F)} \geq \sup_{h \in (0,\infty)} \left[\frac{w_f(h)}{h^\alpha} \right]. \tag{8.99}$$

Again the definition of $|\cdot|_{C^\alpha(E,F)}$ and w_f show that for all $x', y' \in E$ with $x' \neq y'$ it holds that

$$\begin{aligned} \sup_{h \in (0,\infty)} \left[\frac{w_f(h)}{h^\alpha} \right] &\geq \frac{w_f(d_E(x',y'))}{(d_E(x',y'))^\alpha} \\ &= \frac{1}{(d_E(x',y'))^\alpha} \sup_{\substack{x,y \in E,\\ d_E(x,y) \leq d_E(x',y')}} d_F(f(x), f(y)) \\ &\geq \frac{d_F(f(x'), f(y'))}{|d_E(x',y')|^\alpha}. \end{aligned} \tag{8.100}$$

This implies that

$$\sup_{h \in (0,\infty)} \left[\frac{w_f(h)}{h^\alpha} \right] \geq |f|_{C^\alpha(E,F)}. \tag{8.101}$$

The proof of Lemma 8.3.2 is thus completed. □

8.3.3 Solution to Exercise 2.2.8

Lemma 8.3.3. *Let $n \in \mathbb{N}$, $a, b \in \mathbb{R}$ with $a < b$ and let $f \in \mathcal{L}^1(B_{[a,b]}; |\cdot|_{\mathbb{R}})$. Then*

$$T^n_{[a,b]}[f] - \int_a^b f(x)\,dx = \sum_{i=0}^{n-1} \Bigg[\int_{a+\frac{i\,(b-a)}{n}}^{a+\frac{(i+1/2)(b-a)}{n}} f\big(a + \tfrac{i}{n}(b-a)\big) - f(x)\,dx + \int_{a+\frac{(i+1/2)(b-a)}{n}}^{a+\frac{(i+1)(b-a)}{n}} f\Big(a + \tfrac{(i+1)}{n}(b-a)\Big) - f(x)\,dx \Bigg]. \tag{8.102}$$

Proof of Lemma 8.3.3. Note that (2.44) implies that

$$\begin{aligned} &T^n_{[a,b]}[f] - \int_a^b f(x)\,dx \\ &= \frac{(b-a)}{n} \left(\sum_{i=0}^{n-1} \frac{f\big(a + \frac{i}{n}(b-a)\big) + f\Big(a + \frac{(i+1)}{n}(b-a)\Big)}{2} \right) \\ &- \sum_{i=0}^{n-1} \left(\int_{a+\frac{i\,(b-a)}{n}}^{a+\frac{(i+1/2)(b-a)}{n}} f(x)\,dx + \int_{a+\frac{(i+1/2)(b-a)}{n}}^{a+\frac{(i+1)(b-a)}{n}} f(x)\,dx \right) \\ &= \sum_{i=0}^{n-1} \left(\int_{a+\frac{i\,(b-a)}{n}}^{a+\frac{(i+1/2)(b-a)}{n}} f\left(a + \frac{i}{n}(b-a)\right) - f(x)\,dx \right. \\ &\left. + \int_{a+\frac{(i+1/2)(b-a)}{n}}^{a+\frac{(i+1)(b-a)}{n}} f\left(a + \frac{(i+1)}{n}(b-a)\right) - f(x)\,dx \right). \end{aligned} \tag{8.103}$$

The proof of Lemma 8.3.3 is thus completed. ☐

Proposition 8.3.4. *Let $\alpha \in (0,1]$, $n \in \mathbb{N}$, $a, b \in \mathbb{R}$ with $a < b$ and let $f \in \mathcal{L}^1(B_{[a,b]}; |\cdot|_{\mathbb{R}})$. Then*

$$\left| T^n_{[a,b]}[f] - \int_a^b f(x)\,dx \right| \le (b-a) \cdot w_f\big(\tfrac{(b-a)}{2n}\big) \le \frac{(b-a)^{(1+\alpha)}\,|f|_{\mathcal{C}^\alpha([a,b],\mathbb{R})}}{(2n)^\alpha}. \tag{8.104}$$

Proof of Proposition 8.3.4. Note that for all $i \in \{0, 1, \ldots, n-1\}$, $x \in (a + \frac{i\,(b-a)}{n}, a + \frac{(i+1/2)(b-a)}{n})$, $y \in (a + \frac{(i+1/2)(b-a)}{n}, a + \frac{(i+1)(b-a)}{n})$ it holds that

$$\left| x - \left(a + \frac{i\,(b-a)}{n}\right) \right| \le \frac{(b-a)}{2n}, \qquad \left| y - \left(a + \frac{(i+1)\,(b-a)}{n}\right) \right| \le \frac{(b-a)}{2n}. \tag{8.105}$$

Combining this with Lemma 8.3.3 and the fact that w_f is non-decreasing implies that

$$
\begin{aligned}
&\left|T^n_{[a,b]}[f] - \int_a^b f(x)\,dx\right| \\
&\leq \sum_{i=0}^{n-1}\left[\int_{a+\frac{i\,(b-a)}{n}}^{a+\frac{(i+1/2)(b-a)}{n}} \left|f\Big(a+\frac{i}{n}(b-a)\Big) - f(x)\right| dx\right. \\
&\left.+ \int_{a+\frac{(i+1/2)(b-a)}{n}}^{a+\frac{(i+1)(b-a)}{n}} \left|f\Big(a+\frac{(i+1)}{n}(b-a)\Big) - f(x)\right| dx\right] \\
&\leq \sum_{i=0}^{n-1}\left[\int_{a+\frac{i\,(b-a)}{n}}^{a+\frac{(i+1/2)(b-a)}{n}} w_f\Big(\frac{(b-a)}{2n}\Big)\,dx + \int_{a+\frac{(i+1/2)(b-a)}{n}}^{a+\frac{(i+1)(b-a)}{n}} w_f\Big(\frac{(b-a)}{2n}\Big)\,dx\right] \\
&= w_f\Big(\frac{(b-a)}{2n}\Big)\sum_{i=0}^{n-1}\left[\int_{a+\frac{i\,(b-a)}{n}}^{a+\frac{(i+1/2)(b-a)}{n}} dx + \int_{a+\frac{(i+1/2)(b-a)}{n}}^{a+\frac{(i+1)(b-a)}{n}} dx\right] \\
&= (b-a)\cdot w_f\Big(\frac{(b-a)}{2n}\Big).
\end{aligned}
\tag{8.106}
$$

Moreover, observe that Lemma 8.3.2 proves that

$$
(b-a)\cdot w_f(\tfrac{(b-a)}{2n}) \leq \frac{(b-a)^{(1+\alpha)}\,|f|_{\mathcal{C}^\alpha([a,b],\mathbb{R})}}{(2n)^\alpha}. \tag{8.107}
$$

This and (8.106) complete the proof of Proposition 8.3.4. □

8.3.4 Solution to Exercise 2.2.10

Proposition 8.3.5. *Let $f\colon [0,1] \to \mathbb{R}$ be a function with the property that for all $x \in [0,1]$ it holds that $f(x) = x^2$. Then*

(i) it holds that f is infinitely often differentiable and

(ii) it holds for all $n \in \mathbb{N}$ that

$$
n^2\left[T^n_{[0,1]}[f] - \int_0^1 f(x)\,dx\right] = \frac{1}{6}. \tag{8.108}
$$

Proof of Proposition 8.3.5. The fact that f is a polynomial implies that it is infinitely often differentiable. Next note that Lemma 8.3.3 and the integral transform theorem show that for all $n \in \mathbb{N}$ it holds that

$$\begin{aligned}
&T^n_{[0,1]}[f] - \int_0^1 f(x)\,dx \\
&= \sum_{i=0}^{n-1} \left(\int_{\frac{i}{n}}^{\frac{(i+1/2)}{n}} \left[\left(\tfrac{i}{n}\right)^2 - x^2 \right] dx + \int_{\frac{(i+1/2)}{n}}^{\frac{(i+1)}{n}} \left[\left(\tfrac{(i+1)}{n}\right)^2 - x^2 \right] dx \right) \\
&= \sum_{i=0}^{n-1} \left(\int_{\frac{i}{n}}^{\frac{(i+1/2)}{n}} \left[\tfrac{i}{n} - x\right] \left[\tfrac{i}{n} + x\right] dx + \int_{\frac{(i+1/2)}{n}}^{\frac{(i+1)}{n}} \left[\tfrac{(i+1)}{n} - x\right] \left[\tfrac{(i+1)}{n} + x\right] dx \right) \\
&= \sum_{i=0}^{n-1} \left(\int_{-\frac{1}{2n}}^{0} u \left[\tfrac{2i}{n} - u\right] du + \int_0^{\frac{1}{2n}} u \left[\tfrac{2(i+1)}{n} - u\right] du \right) \\
&= \sum_{i=0}^{n-1} \left(\frac{2i}{n} \left[\frac{u^2}{2}\right]_{u=-\frac{1}{2n}}^{u=0} - \left[\frac{u^3}{3}\right]_{u=-\frac{1}{2n}}^{u=0} + \frac{2(i+1)}{n} \left[\frac{u^2}{2}\right]_{u=0}^{u=\frac{1}{2n}} - \left[\frac{u^3}{3}\right]_{u=0}^{u=\frac{1}{2n}} \right) \\
&= \sum_{i=0}^{n-1} \left(-\frac{i}{4n^3} - \frac{1}{24n^3} + \frac{(i+1)}{4n^3} - \frac{1}{24n^3} \right) = n\left(\frac{1}{6n^3}\right) = \frac{1}{6n^2}.
\end{aligned} \tag{8.109}$$

This shows that for all $n \in \mathbb{N}$ it holds that

$$n^2 \left[T^n_{[0,1]}[f] - \int_0^1 f(x)\,dx \right] = \frac{1}{6}. \tag{8.110}$$

The proof of Proposition 8.3.5 is thus completed. □

8.3.5 Solution to Exercise 2.2.13

```
function R_n=RecRule(f,a,b,d,n)
  grid1d = linspace(a,b,n+1);
  grid1d = grid1d(1:end-1);
  h = (b-a)/n;
  % empty array for the recursion
  x = [];
  R_n = RecRuleRecursion(f,x,grid1d,h,d);
end

function R_n = RecRuleRecursion(f,x,grid1d,h,d)
  % Subfunction, which contains the recursive
  % evaluation of f over the
  % d-dimensional grid.
  R_n = 0;
  if d > 1
    y = [x 0];
```

```
    for i= 1:length(grid1d)
        y(end) = grid1d(i);
        R_n = R_n + h*RecRuleRecursion(f,y,grid1d,h,d-1);
    end
  else
    a=repmat(x,length(grid1d),1);
    R_n = R_n + h * sum( f([a,grid1d']) );
  end
end
```

Matlab code 8.17: A Matlab function `RecRule`(a, b, d, n, f) with input $a \in \mathbb{R}, b \in (a, \infty)$, $d, n \in \mathbb{N}$, and a function $f\colon [a,b]^d \to \mathbb{R} \in \mathcal{L}^1(B_{[a,b]^d}; |\cdot|_{\mathbb{R}})$ and output $R^n_{[a,b]}[f]$.

```
function RecRuleTest()
  % Warning: with the given parameter values
  % the programs runs for about 30 mins.
  % parameter
  a = 0;
  b = 2;
  f = @(x) x(:,1);
  d_list =1:8;
  n_list =5:10;
  % Array for the error
  err_list = zeros(length(d_list),length(n_list));
  time_list = zeros(length(d_list),length(n_list));

  for i = 1:length(d_list)
    % exact value of the integral
    int = 0.5*(b-a)^d_list(i)*(b+a);
    for j = 1:length(n_list)
        % computing the error and measure run time
        tic
        err_list(i,j) = abs( RecRule(f,a,b,d_list(i),...
            n_list(j)) - int);
        time_list(i,j) = toc;
    end
  end

  format short;
  f1 = figure;
  cnames = n_list;
  rnames = d_list;
  t_err = uitable('Parent',f1,'Data',err_list,...
```

```
              'ColumnName',cnames,'RowName',rnames,...
              'Position',[0 210 525 200]);
set(t_err,'columnwidth',{80});
t_time = uitable('Parent',f1,'Data',time_list,...
              'ColumnName',cnames,...
              'RowName',rnames,'Position',[0 0 525 200]);
set(t_time,'columnwidth',{80});


d_list = 9;
n_list = 3:8;
% Array for the error
err_list = zeros(length(d_list),length(n_list));
time_list = zeros(length(d_list),length(n_list));
% exact value of the integral
int = 0.5*(b-a)^d_list*(b+a);
  for j = 1:length(n_list)
      % computing the error and measure run time
      tic
      err_list(j) = abs( RecRule(f,a,b,d_list,...
                  n_list(j)) - int);
      time_list(j) = toc;
  end

f2 = figure;
cnames = n_list;
rnames = d_list;
t_err = uitable('Parent',f2,'Data',err_list,...
              'ColumnName',cnames,'RowName',rnames,...
              'Position',[0 60 525 50]);
set(t_err,'columnwidth',{80});
t_time = uitable('Parent',f2,'Data',time_list,...
              'ColumnName',cnames,'RowName',rnames,...
              'Position',[0 0 525 50]);
set(t_time,'columnwidth',{80});

d_list =10;
n_list =1:6;
% Array for the error
err_list = zeros(length(d_list),length(n_list));
time_list = zeros(length(d_list),length(n_list));
% exact value of the integral
int = 0.5*(b-a)^d_list*(b+a);
for j = 1:length(n_list)
```

```
        % computing the error and measure run time
        tic
        err_list(j) = abs( RecRule(f,a,b,d_list ,...
                        n_list(j)) - int);
        time_list(j) = toc;
    end

    f3 = figure;
    cnames = n_list;
    rnames = d_list;
    t_err = uitable('Parent',f3,'Data',err_list ,...
                'ColumnName',cnames,'RowName',rnames,...
                'Position',[0 60 525 50]);
    set(t_err,'columnwidth',{80});
    t_time = uitable('Parent',f3,'Data',time_list ,...
                'ColumnName',cnames,'RowName',rnames,...
                'Position',[0 0 525 50]);
    set(t_time,'columnwidth',{80});
end
```

Matlab code 8.18: A Matlab function `RecRuleTest()` which outputs error and measured run time of the Matlab function 8.17 in the case of $f = [0,2]^d \ni (x_1,\dots,x_n) \mapsto x_1 \in \mathbb{R}$, $a = 0, b = 2$, $(d,n) \in \{1,\dots,8\} \times \{5,\dots,10\} \cup \{9\} \times \{3,\dots,8\} \cup \{10\} \times \{1,\dots,6\}$.

	5	6	7	8	9	10
1	0.4000	0.3333	0.2857	0.2500	0.2222	0.2000
2	0.8000	0.6667	0.5714	0.5000	0.4444	0.4000
3	1.6000	1.3333	1.1429	1	0.8889	0.8000
4	3.2000	2.6667	2.2857	2	1.7778	1.6000
5	6.4000	5.3333	4.5714	4	3.5556	3.2000
6	12.8000	10.6667	9.1429	8	7.1111	6.4000
7	25.6000	21.3333	18.2857	16	14.2222	12.8000
8	51.2000	42.6667	36.5714	32	28.4444	25.6000

	5	6	7	8	9	10
1	0.0048	3.1600e-04	9.3000e-05	6.6000e-05	5.9000e-05	6.1000e-05
2	0.0011	3.1000e-04	3.0900e-04	3.0900e-04	4.0600e-04	3.7000e-04
3	0.0010	0.0014	0.0023	0.0026	0.0030	0.0035
4	0.0049	0.0083	0.0129	0.0189	0.0269	0.0367
5	0.0239	0.0477	0.0872	0.1488	0.2334	0.3600
6	0.1184	0.2884	0.6132	1.1990	2.2677	3.7622
7	0.6036	1.8265	4.4076	9.5715	19.3589	36.9045
8	3.0910	10.5999	30.2362	77.8631	175.6351	362.0270

Figure 8.8: Partial result of a call of the Matlab function 8.18 for $(d, n) \in \{1, \ldots, 8\} \times \{5, \ldots, 10\}$. The upper table presents the approximation errors and the lower table presents the measured run times in seconds, where $d \in \{1, \ldots, 8\}$ is the row index and where $n \in \{5, \ldots, 10\}$ is the column index.

	3	4	5	6	7	8
9	170.6667	128	102.4000	85.3333	73.1429	64

	3	4	5	6	7	8
9	0.2784	2.6465	15.0062	63.2750	215.3077	621.4268

Figure 8.9: Partial result of a call of the Matlab function 8.18 for $(d, n) \in \{9\} \times \{3, \ldots, 8\}$. The upper table presents the approximation errors and the lower table presents the measured run times in seconds, where $n \in \{3, \ldots, 8\}$ is the column index.

	1	2	3	4	5	6
10	1024	512	341.3333	256	204.8000	170.6667

	1	2	3	4	5	6
10	4.0100e-04	0.0259	0.8239	10.1803	74.1359	380.1321

Figure 8.10: Partial result of a call of the Matlab function 8.18 for $(d, n) \in \{10\} \times \{1, \ldots, 6\}$. The upper table presents the approximation errors and the lower table presents the measured run times in seconds, where $n \in \{1, \ldots, 6\}$ is the column index.

8.3.6 Solution to Exercise 2.3.9

Lemma 8.3.6. *Let $A, B \subseteq \mathbb{R}^2$ be the sets given by*

$$A = \left\{(x,y) \in \mathbb{R}^2 \colon\, |x-2|_{\mathbb{R}}^2 + y^2 \leq 4\right\}, \tag{8.111}$$

$$B = \left\{(x,y) \in \mathbb{R}^2 \colon\, x^2 + |y-2|_{\mathbb{R}}^2 \leq 4\right\}, \tag{8.112}$$

let $f \colon \mathbb{R}^2 \to \mathbb{R}$ be the function with the property that for all $x, y \in \mathbb{R}$ it holds that $f(x,y) = \mathbb{1}_{(A \cap B)}(x,y) \cdot |x|_{\mathbb{R}}^{2/3}$, let $(\Omega, \mathcal{F}, P)$ be a probability space, let $Y_n, Z_n \colon \Omega \to \mathbb{R}$, $n \in \mathbb{N}$, be independent $\mathcal{U}_{(0,1)}$-distributed random variables, and let $I_N \colon \Omega \to \mathbb{R}$, $N \in \mathbb{N}$, be functions with the property that for all $N \in \mathbb{N}$ it holds that

$$I_N = \frac{4}{N}\left[\sum_{n=1}^{N} f(2Y_n, 2Z_n)\right]. \tag{8.113}$$

Then it holds for all $N \in \mathbb{N}$ that I_N is P-unbiased with respect to $\int_0^2 \int_0^2 f(x,y)\,dx\,dy$.

Proof of Lemma 8.3.6. Observe that for all $N \in \mathbb{N}$ it holds that $(2Y_n, 2Z_n)$, $n \in \{1, \ldots, N\}$, are independent $\mathcal{U}_{(0,2)^2}$-distributed random variables. This implies that for all $N \in \mathbb{N}$ it holds that

$$\mathbb{E}_P[I_N] = \frac{4}{N} \sum_{n=1}^{N} \int_{(0,2)^2} f(x,y)\, \frac{1}{4}\, B_{\mathbb{R}^2}(dx, dy) = \int_0^2 \int_0^2 f(x,y)\,dx\,dy. \tag{8.114}$$

Hence, we obtain for all $N \in \mathbb{N}$ that the random variable I_N is P-unbiased with respect to $\int_0^2 \int_0^2 f(x,y)\,dx\,dy$. The proof of Lemma 8.3.6 is thus completed. □

```
function Int=MonteCarlo(N)
  RV = 2 * rand(2,N);
  I = ( RV(1,:).^2 + (RV(2,:) - 2 ).^2 <=4)...
      & ( (RV(1,:) - 2).^2 + RV(2,:).^2 <=4);
```

```
  Int = 4/N * sum(RV(1,I).^(2/3));
end
```

Matlab code 8.19: A Matlab function `MonteCarlo(N)` with input $N \in \mathbb{N}$ and output a realization of I_N.

```
function MonteCarloPlot()
  rng( 'default' );
  Int = zeros( 25, 1);
  k = zeros( 25, 1);
  for j = 2:6
      for i = 1:5
        Int( (j-2)*5 + i ) = MonteCarlo( 10^j );
        k( (j-2)*5 + i ) = j;
      end
   end
   plot(k,Int,'*')
end
```

Matlab code 8.20: A Matlab function `MonteCarloPlot()` which plots for every $k \in \{2, 3, 4, 5, 6\}$ five realizations of I_{10^k}, each marked by a blue star, in a coordinate plane.

8.3.7 Solution to Exercise 2.3.10

```
function I = intMC(a,b,d,f,N)
 RV = (b-a) * rand(N,d) + a;
 I = (b - a)^d/N * sum( f(RV) );
end
```

Matlab code 8.21: A Matlab function `intMC(`a, b, d, f, N`)` for a Monte Carlo estimator of the integral $\int_{[a,b]^d} f(x)\, dx$ with N i.i.d. samples of $\mathcal{U}_{[a,b]^d}$-distributed random vectors.

```
function intMCTest()
  a = 0;
  b = 2;
  f = @(x) x(:,1);
  d_list = 3:2:7;
  N_list = [5.^d_list, 10.^d_list];
```

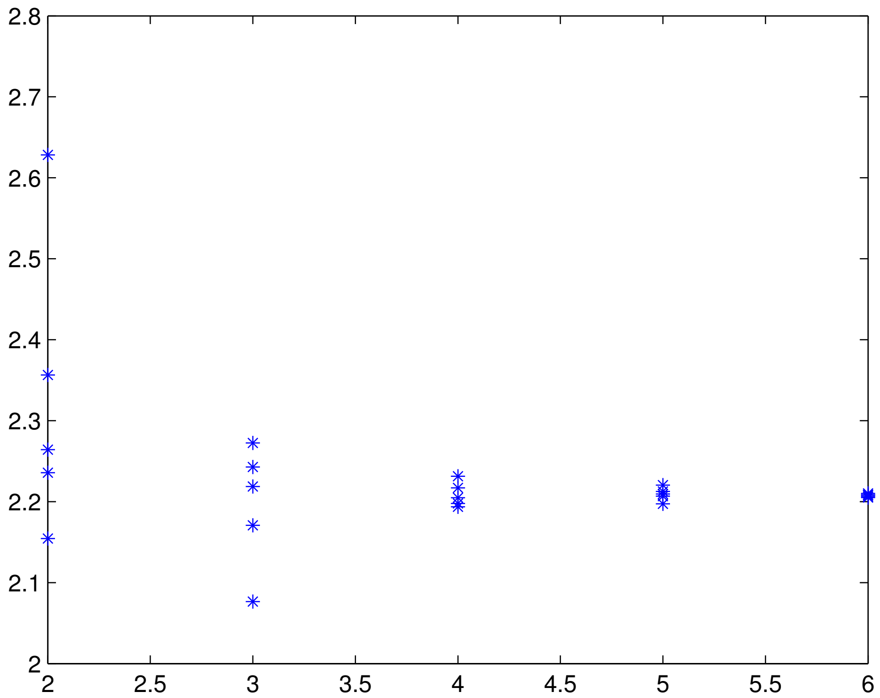


Figure 8.11: Result of a call of the Matlab function 8.20.

```
  % Array for the error
  err_list = zeros(length(d_list),length(N_list));
  time_list = zeros(length(d_list),length(N_list));

  for i = 1:length(d_list)
    % exact value of the integral
    int = 0.5*(b-a)^d_list(i)*(b+a);
    for j = 1:length(N_list)
        % computing the error and measure run time
        tic
        err_list(i,j) = abs( intMC(a,b,d_list(i),f,...
            N_list(j)) - int);
        time_list(i,j) = toc;
    end
  end

  format short;
  f1 = figure;
  cnames = N_list;
  rnames = d_list;
  t_err = uitable('Parent',f1,'Data',err_list,...
            'ColumnName',cnames,'RowName',rnames,...
            'Position',[0 100 525 90]);
  set(t_err,'columnwidth',{80});
  t_time = uitable('Parent',f1,'Data',time_list,...
            'ColumnName',cnames,...
            'RowName',rnames,'Position',[0 0 525 90]);
  set(t_time,'columnwidth',{80});


end
```

Matlab code 8.22: A Matlab function `intMCTest()` which outputs error and measured run time of the Matlab function 8.21 with $a = 0$, $b = 2$, $f = [0,2]^d \ni (x_1, \dots, x_d) \mapsto x_1 \in \mathbb{R}$, $(d, N) \in \{3, 5, 7\} \times \{5^3, 5^5, 5^7, 10^3, 10^5, 10^7\}$.

	125	3125	78125	1000	100000	1e+07
3	0.1125	0.1566	0.0052	0.2178	0.0170	5.4451e-04
5	0.6627	0.4222	0.1144	0.8768	0.1409	0.0076
7	0.7740	1.3507	0.2025	1.5960	0.0563	0.0278

	125	3125	78125	1000	100000	1e+07
3	1.5800e-04	2.6000e-04	0.0041	1.1400e-04	0.0055	0.4554
5	1.1300e-04	1.9000e-04	0.0055	1.1700e-04	0.0070	0.6901
7	1.1400e-04	2.8600e-04	0.0076	1.3900e-04	0.0097	0.9409

Figure 8.12: Result of a call of the Matlab function 8.22. The upper table presents the approximation errors and the lower table presents the measured run times in seconds, where the dimension $d \in \{3, 5, 7\}$ is the row index and where the number of used samples $N \in \{5^3, 5^5, 5^7, 10^3, 10^5, 10^7\}$ is the column index.

8.3.8 Solution to Exercise 2.4.10

Lemma 8.3.7. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $f \in \mathcal{M}(\mathcal{B}(\mathbb{R}), \mathcal{B}(\mathbb{R}))$ be a globally bounded function, and let $U_n \in \mathcal{M}(\mathcal{F}, \mathcal{B}(\mathbb{R}))$, $n \in \mathbb{N}$, be independent $\mathcal{U}_{(-1,1)}$-distributed random variables. Then*

$$\left(\mathbb{E}_P\left[\left|\frac{f(U_1) + \ldots + f(U_{5000})}{2500} - \int_{-1}^{1} f(x)\,dx\right|_{\mathbb{R}}^2\right]\right)^{1/2} \leq \frac{\sup_{x\in\mathbb{R}} |f(x)|_{\mathbb{R}}}{30}. \tag{8.115}$$

Proof of Lemma 8.3.7. The assumption that f is globally bounded ensures that for all $n \in \mathbb{N}$ it holds that $f \circ U_n \in \mathcal{L}^2(P; |\cdot|_{\mathbb{R}})$. Theorem 2.4.8 hence implies that for all $n \in \mathbb{N}$ it holds that

$$\left(\mathbb{E}_P\left[\left|\mathbb{E}_P[f(U_1)] - \frac{f(U_1) + \ldots + f(U_n)}{n}\right|_{\mathbb{R}}^2\right]\right)^{1/2} = \frac{\sqrt{\mathrm{Var}_P(f(U_1))}}{\sqrt{n}}. \tag{8.116}$$

This proves that for all $n \in \mathbb{N}$ it holds that

$$\begin{aligned}\left(\mathbb{E}_P\left[\left|\frac{f(U_1) + \ldots + f(U_n)}{n} - \frac{\int_{-1}^{1} f(x)\,dx}{2}\right|_{\mathbb{R}}^2\right]\right)^{1/2} &= \frac{\sqrt{\mathrm{Var}_P(f(U_1))}}{\sqrt{n}} \\ = \frac{\sqrt{\mathbb{E}_P[|f(U_1)|_{\mathbb{R}}^2] - |\mathbb{E}_P[f(U_1)]|_{\mathbb{R}}^2}}{\sqrt{n}} \leq \frac{\sqrt{\mathbb{E}_P[|f(U_1)|_{\mathbb{R}}^2]}}{\sqrt{n}} &\leq \frac{\sup_{x\in\mathbb{R}} |f(x)|_{\mathbb{R}}}{\sqrt{n}}.\end{aligned} \tag{8.117}$$

Hence, we obtain that for all $n \in \mathbb{N}$ it holds that

$$\left(\mathbb{E}_P\left[\left|\frac{f(U_1) + \ldots + f(U_n)}{[n/2]} - \int_{-1}^{1} f(x)\,dx\right|_{\mathbb{R}}^2\right]\right)^{1/2} \leq \frac{2\,[\sup_{x\in\mathbb{R}} |f(x)|_{\mathbb{R}}]}{\sqrt{n}}. \tag{8.118}$$

This and the fact that $\sqrt{5000} \geq \sqrt{3600} = 60 = 2 \cdot 30$ show that

$$\left(\mathbb{E}_P\left[\left|\frac{f(U_1)+\ldots+f(U_{5000})}{2500} - \int_{-1}^{1} f(x)\,dx\right|_{\mathbb{R}}^2\right]\right)^{1/2} \leq \frac{\sup_{x\in\mathbb{R}}|f(x)|_{\mathbb{R}}}{\left[\sqrt{5000}/2\right]} \leq \frac{\sup_{x\in\mathbb{R}}|f(x)|_{\mathbb{R}}}{30}. \tag{8.119}$$

The proof of Lemma 8.3.7 is thus completed. □

8.4 Chapter 3

8.4.1 Solution to Exercise 3.1.9

Lemma 8.4.1. *Let Ω, $\mathcal{F}$ be the sets given by $\Omega = \{1,2\}$ and $\mathcal{F} = \{\emptyset, \{1,2\}\}$. Then it holds that $(\Omega, \mathcal{F})$ is a measurable space and it holds that*

$$\{(\omega,\omega)\in\Omega^2\colon \omega\in\Omega\} = \{(1,1),(2,2)\} \notin \{\emptyset, \Omega\times\Omega\} = \mathcal{F}\otimes\mathcal{F}. \tag{8.120}$$

Proof of Lemma 8.4.1. Clearly, it holds that $(\Omega, \mathcal{F})$ is a measurable space. Moreover, observe that

$$\begin{split}\mathcal{F}\otimes\mathcal{F} &= \sigma_{\Omega\times\Omega}(\{A\times B\in\mathcal{P}(\Omega\times\Omega)\colon A,B\in\mathcal{F}\}) = \sigma_{\Omega\times\Omega}(\{\Omega\times\Omega\}) \\ &= \{\emptyset,\Omega\times\Omega\} = \Big\{\emptyset,\big\{(1,1),(1,2),(2,1),(2,2)\big\}\Big\} \not\ni \big\{(1,1),(2,2)\big\}.\end{split} \tag{8.121}$$

The proof of Lemma 8.4.1 is thus completed. □

8.4.2 Solution to Exercise 3.1.10

Lemma 8.4.2. *Let $\Omega, \mathcal{F}, S, \mathcal{S}$ be the sets given by $\Omega = S = \{1,2\}$ and $\mathcal{F} = \mathcal{S} = \big\{\emptyset, \{1,2\}\big\}$ and let $X, Y : \Omega \to S$ be the functions with the property that for all $\omega \in \Omega$ it holds that*

$$X(\omega) = \omega \qquad \text{and} \qquad Y(\omega) = 1. \tag{8.122}$$

Then

(i) it holds that $(\Omega, \mathcal{F})$ and $(S, \mathcal{S})$ are measurable spaces,

(ii) it holds that X and Y are $\mathcal{F}/\mathcal{S}$-measurable functions, and

(iii) it holds that

$$\{\omega\in\Omega\colon X(\omega) = Y(\omega)\} \notin \mathcal{F}. \tag{8.123}$$

Proof of Lemma 8.4.2. Clearly, it holds that $(\Omega, \mathcal{F})$ and $(S, \mathcal{S})$ are measurable spaces. Moreover, note that

$$X^{-1}(\emptyset) = Y^{-1}(\emptyset) = \emptyset \in \mathcal{F} \qquad \text{and} \qquad X^{-1}(S) = Y^{-1}(S) = \Omega \in \mathcal{F}. \tag{8.124}$$

It thus holds that X and Y are $\mathcal{F}/\mathcal{S}$-measurable functions. Furthemore, observe that

$$\{\omega \in \Omega \colon X(\omega) = Y(\omega)\} = \{1\} \notin \Big\{\emptyset, \{1, 2\}\Big\} = \mathcal{F}. \tag{8.125}$$

The proof of Lemma 8.4.2 is thus completed. □

8.4.3 Solution to Exercise 3.2.23

Lemma 8.4.3 (Product measurable random fields)**.** *Let $(I, \mathcal{I})$, $(\Omega, \mathcal{F})$, and $(S, \mathcal{S})$ be measurable spaces, let $X \colon I \times \Omega \to S$ be an $(\mathcal{I} \otimes \mathcal{F})/\mathcal{S}$-measurable function, and let $\omega \in \Omega$. Then it holds that $I \ni i \mapsto X(i, \omega) \in S$ is $\mathcal{I}/\mathcal{S}$-measurable.*

Proof of Lemma 8.4.3. First of all, let $R \colon I \to (I \times \Omega)$ be the function with the property that for all $i \in I$ it holds that

$$R(i) = (i, \omega). \tag{8.126}$$

Next note that for all $A \in \mathcal{I}$, $B \in \mathcal{F}$ it holds that

$$R^{-1}(A \times B) = \{i \in I \colon R(i) \in A \times B\} = \{i \in I \colon (i, \omega) \in A \times B\} = \begin{cases} A & : \omega \in B \\ \emptyset & : \omega \notin B \end{cases}. \tag{8.127}$$

This and the fact that

$$\mathcal{I} \otimes \mathcal{F} = \sigma_{I \times \Omega}(\{A \times B \colon A \in \mathcal{I}, B \in \mathcal{F}\}) \tag{8.128}$$

prove that R is $\mathcal{I}/(\mathcal{I} \otimes \mathcal{F})$-measurable. The assumption that X is $(\mathcal{I} \otimes \mathcal{F})/\mathcal{S}$-measurable hence ensures that

$$X \circ R = \big[I \ni i \mapsto (X \circ R)(i) = X(R(i)) = X(i, \omega) \in S\big] \tag{8.129}$$

is $\mathcal{I}/\mathcal{S}$-measurable. The proof of Lemma 8.4.3 is thus completed. □

8.4.4 Solution to Exercise 3.3.9

```
function E_N = MonteCarloGBM(T, alpha, beta, x0, f, N)
  W_T = sqrt(T) * randn(1,N);
  X_T = exp( alpha*T + beta*W_T ) * x0;
  E_N = 1/N * sum( f(X_T) );
end
```

Matlab code 8.23: A Matlab function `MonteCarloGBM(`T`,`α`,`β`,`x_0`,`f`,`N`)` with input $T \in (0,\infty)$, $\alpha,\beta \in \mathbb{R}$, $x_0 \in (0,\infty)$, $f \in \mathcal{L}^1(X_T(P)_{\mathcal{B}(\mathbb{R})}; |\cdot|_{\mathbb{R}})$, $N \in \mathbb{N}$ and output a Monte Carlo approximation of $\mathbb{E}\big[f(X_T)\big]$ based on $N \in \mathbb{N}$ samples; see Exercise 3.3.9.

```
T = 1;
beta = 1/10;
alpha = log(1.06) - beta^2/2;
f = @(x) subplus( x - 100 );
x0 = 92;
N = 10^4;
MonteCarloGBM(T, alpha, beta, x0, f, N)

ans =

    2.7708
```

Matlab code 8.24: A script in the Matlab console to test the Matlab function 8.23 in the case $T = 1$, $\beta = \frac{1}{10}$, $\alpha = \ln(1.06) - \frac{\beta^2}{2}$, $x_0 = 92$, $f = \mathbb{R} \ni x \mapsto [x - 100]^+ \in \mathbb{R}$, $N = 10^4$.

8.4.5 Solution to Exercise 3.3.10

```
function BM = BrownianMotion(T,m,N)
  BM = cumsum( [ zeros(m,1), randn(m,N) ]*sqrt(T/N), 2 );
end
```

Matlab code 8.25: A Matlab function `BrownianMotion(`T`,`m`,`N`)` with input $T \in (0,\infty)$, $m, N \in \mathbb{N}$ and output a realization of an $\big(W_0, W_{\frac{T}{N}}, W_{\frac{2T}{N}}, \dots, W_{\frac{(N-1)T}{N}}, W_T\big)(P)_{\mathcal{B}(\mathbb{R}^{m\times(N+1)})}$-distributed random variable; see Exercise 3.3.10.

```
function BrownianMotion2DPlot()
  rng('default');
  T = 1; N = 1000;
  BM = BrownianMotion(T,2,N);
  InterpolatedBMx = @(t) (floor(t*N/T) + 1 - t*N/T).*...
      BM(1,floor(t*N/T)+1) + (t*N/T - floor(t*N/T)).*...
      BM(1,ceil(t*N/T)+1);
  InterpolatedBMy = @(t) (floor(t*N/T) + 1 - t*N/T).*...
```

```
        BM(2,floor(t*N/T)+1) + (t*N/T - floor(t*N/T)).*...
        BM(2,ceil(t*N/T)+1);
    tgrid = [0:T/(1e3*N):T];
    clf
    hold on
    plot3(InterpolatedBMx(0),0,InterpolatedBMy(0),'r*');
    plot3(InterpolatedBMx(T),T,InterpolatedBMy(T),'r*');
    plot3(InterpolatedBMx(tgrid),tgrid,...
        InterpolatedBMy(tgrid));
    xlabel('x'); ylabel('t'); zlabel('y');
    grid on; daspect([1 1 1]); view(-30,15);
    hold off
end
```

Matlab code 8.26: A Matlab function `BrownianMotion2DPlot()` which uses the Matlab function 8.25 to plot in the case $T = 1$, $m = 2$ one realization of an $(\tilde{W}^{1000})(P)_{\otimes_{t\in[0,1]}\mathcal{B}(\mathbb{R}^2)}$-distributed random variable in a three-dimensional coordinate system; see Exercise 3.3.10.

8.4.6 Solution to Exercise 3.3.11

Lemma 8.4.4 (Geometric Brownian motion revisited). *Let $T, x_0, \beta \in (0,\infty)$, $\alpha \in \mathbb{R}$, let $(\Omega, \mathcal{F}, P)$ be a probability space, let $W\colon [0,T]\times\Omega \to \mathbb{R}$ be a standard Brownian motion, let $X\colon [0,T]\times\Omega\to\mathbb{R}$ be the function which satisfies for all $t\in[0,T]$ that $X_t = e^{(\alpha t+\beta W_t)}\,x_0$, and let $\Phi\colon \mathbb{R}\to\mathbb{R}$ be the function which satisfies for all $y\in\mathbb{R}$ that $\Phi(y) = \int_{-\infty}^{y}\frac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}x^2}\,dx$. Then for all $K\in\mathbb{R}$ it holds that*

$$\begin{aligned}&\mathbb{E}\big[\max\{X_T-K,0\}\big]\\&=\begin{cases} e^{(\alpha+\frac{1}{2}\beta^2)T}\,x_0 - K & : K\le 0\\ e^{(\alpha+\frac{1}{2}\beta^2)T}\,x_0\,\Phi\Big(\frac{\alpha T+\ln(\frac{x_0}{K})}{\beta\sqrt{T}}+\beta\sqrt{T}\Big) - K\,\Phi\Big(\frac{\alpha T+\ln(\frac{x_0}{K})}{\beta\sqrt{T}}\Big) & : K>0\end{cases}.\end{aligned}\tag{8.130}$$

Proof of Lemma 8.4.4. Observe that $\ln(x_0)+\alpha T+\beta W_T$ is an $\mathcal{N}_{\ln(x_0)+\alpha T,\beta^2 T}$-distributed random variable and note that

$$\max\{X_T-K,0\} = \max\{e^{\ln(x_0)+\alpha T+\beta W_T}-K,0\}.\tag{8.131}$$

This and Lemma 4.7.2 complete the proof of Lemma 8.4.4. □

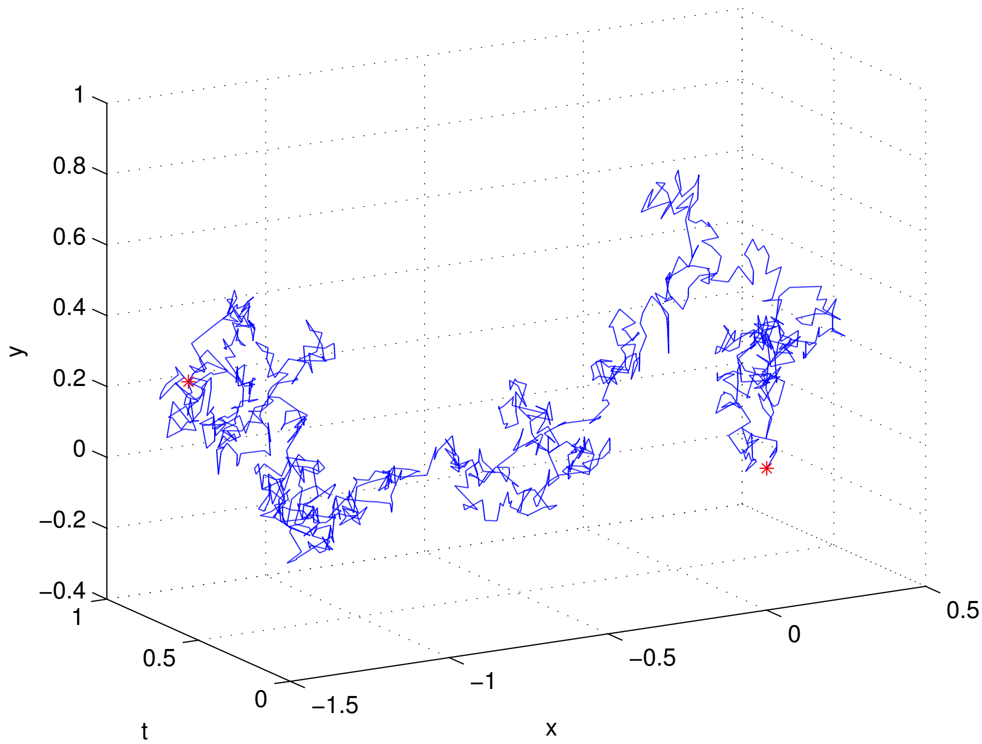


Figure 8.13: Result of a call of the Matlab function 8.26. The two red stars mark the starting point and the end point of the plotted sample path.

```
  T = 1;
  beta = 1/10;
  alpha = log(1.06) - beta^2/2;
 K = 100;
  x0 = 92;

  Z1 = ( alpha*T + log(x0/K) )/beta/sqrt(T);
  Z2 = Z1 + beta*sqrt(T);

  price = exp( (alpha + beta^2/2)*T ) * x0 * ...
      1/2 * ( erf( Z2/sqrt(2) ) + 1 ) - K * ...
      1/2 * ( erf( Z1/sqrt(2) ) + 1 )

price =

    2.8217
```

Matlab code 8.27: A Matlab script in the Matlab console to price the European call option according to the Black–Scholes model with the parameters specified in Exercise 3.3.11.

8.4.7 Solution to Exercise 3.3.15

Lemma 8.4.5 (Quadratic variation of Brownian motion)**.** *Let $T \in (0,\infty)$, $N \in \mathbb{N}$, $0 = t_0 < t_1 < \cdots < t_N = T$, let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $W\colon [0,T] \times \Omega \to \mathbb{R}$ be a standard Brownian motion. Then*

$$\left\| T - \sum_{n=0}^{N-1} \big(W_{t_{n+1}} - W_{t_n}\big)^2 \right\|_{\mathcal{L}^2(P;|\cdot|_{\mathbb{R}})} \leq \sqrt{2}\,T \left[\max_{n\in\{0,1,\dots,N-1\}} |t_{n+1} - t_n| \right]^{1/2}. \tag{8.132}$$

Proof of Lemma 8.4.5. Observe that Corollary 2.4.7 and the fact that $T =$

$\sum_{n=0}^{N-1}(t_{n+1}-t_n)$ prove that

$$\begin{aligned}
&\left\| T - \sum_{n=0}^{N-1} \big(W_{t_{n+1}} - W_{t_n}\big)^2 \right\|_{\mathcal{L}^2(P;|\cdot|_{\mathbb{R}})}^2 \\
&= \mathbb{E}_P\!\left[\left| \sum_{n=0}^{N-1} \Big[\big(W_{t_{n+1}} - W_{t_n}\big)^2 - (t_{n+1}-t_n) \Big] \right|_{\mathbb{R}}^2 \right] \\
&= \mathsf{Var}_P\!\left(\sum_{n=0}^{N-1} \Big[\big(W_{t_{n+1}} - W_{t_n}\big)^2 - (t_{n+1}-t_n) \Big] \right) \\
&= \sum_{n=0}^{N-1} \mathsf{Var}_P\Big(\big(W_{t_{n+1}} - W_{t_n}\big)^2 - (t_{n+1}-t_n) \Big).
\end{aligned} \tag{8.133}$$

This and Lemma 3.3.12 imply that

$$\begin{aligned}
&\left\| T - \sum_{n=0}^{N-1} \big(W_{t_{n+1}} - W_{t_n}\big)^2 \right\|_{\mathcal{L}^2(P;|\cdot|_{\mathbb{R}})}^2 = \sum_{n=0}^{N-1} \mathbb{E}_P\!\left[\Big| \big(W_{t_{n+1}} - W_{t_n}\big)^2 - (t_{n+1}-t_n) \Big|_{\mathbb{R}}^2 \right] \\
&= \sum_{n=0}^{N-1} \Big(\mathbb{E}_P\Big[\big(W_{t_{n+1}} - W_{t_n}\big)^4 \Big] - 2\,\mathbb{E}_P\Big[\big(W_{t_{n+1}} - W_{t_n}\big)^2 (t_{n+1}-t_n) \Big] + (t_{n+1}-t_n)^2 \Big) \\
&= \sum_{n=0}^{N-1} \left(3 \left| \mathbb{E}_P\Big[\big(W_{t_{n+1}} - W_{t_n}\big)^2 \Big] \right|_{\mathbb{R}}^2 - (t_{n+1}-t_n)^2 \right) = 2 \sum_{n=0}^{N-1} (t_{n+1}-t_n)^2 \\
&\leq 2\,T \left[\max_{n\in\{0,1,\dots,N-1\}} |t_{n+1}-t_n|_{\mathbb{R}} \right].
\end{aligned} \tag{8.134}$$

The proof of Lemma 8.4.5 is thus completed. □

8.4.8 Solution to Exercise 3.4.7

Lemma 8.4.6. *Let (E,ρ) be a metric space, let $c \in (0,\infty)$, and let $\varrho\colon E\times E \to [0,\infty)$ be the function with the property that for all $x,y\in E$ it holds that*

$$\varrho(x,y) = \min\{c, \rho(x,y)\}. \tag{8.135}$$

Then ϱ is a globally bounded metric on E.

Proof of Lemma 8.4.6. First, we observe that for all $x,y\in E$ it holds that

$$\varrho(x,y) \leq c. \tag{8.136}$$

Therefore, it holds that ϱ is a globally bounded function. It thus remains to prove that ϱ is a metric on E. Note that for all $x, y \in E$ it holds that

$$\varrho(x,y) = \min\{c, \rho(x,y)\} = \min\{c, \rho(y,x)\} = \varrho(y,x). \tag{8.137}$$

This shows that ϱ is symmetric. Moreover, we note that for all $x, y \in E$ it holds that

$$\varrho(x,y) = 0 \Leftrightarrow \rho(x,y) = 0. \tag{8.138}$$

The positive definiteness of the metric ρ thus shows that ϱ is positive definite. In addition, we observe that the fact that the metric ρ satisfies the triangle inequality shows that for all $x, y, z \in E$ with $\big((\rho(x,y) \leq c) \text{ and } (\rho(y,z) \leq c)\big)$ it holds that

$$\begin{aligned} \varrho(x,z) &= \min\{c, \rho(x,z)\} \leq \min\{c, \rho(x,y) + \rho(y,z)\} \leq \rho(x,y) + \rho(y,z) \\ &= \min\{c, \rho(x,y)\} + \min\{c, \rho(y,z)\}. \end{aligned} \tag{8.139}$$

Similarly, we note that for all $x, y, z \in E$ with $\big((\rho(x,y) > c) \text{ or } (\rho(y,z) > c)\big)$ it holds that

$$\varrho(x,z) = \min\{c, \rho(x,z)\} \leq c \leq \min\{c, \rho(x,y)\} + \min\{c, \rho(y,z)\}. \tag{8.140}$$

Combining (8.139) and (8.140) proves that for all $x, y, z \in E$ it holds that

$$\varrho(x,z) \leq \varrho(x,y) + \varrho(y,z). \tag{8.141}$$

This proves that ϱ satisfies the triangle inequality. The proof of Lemma 8.4.6 is thus completed. □

Lemma 8.4.7. *Let $d \in \mathbb{N}$, $p \in [1,\infty)$, let $(\Omega, \mathcal{F}, P)$ be a probability space, let $\rho\colon \mathbb{R}^d \times \mathbb{R}^d \to [0,\infty)$ be a globally bounded $\mathcal{B}(\mathbb{R}^d \times \mathbb{R}^d)/\mathcal{B}([0,\infty))$-measurable metric, and let $\varrho\colon L^0(P; \|\cdot\|_{\mathbb{R}^d}) \times L^0(P; \|\cdot\|_{\mathbb{R}^d}) \to [0,\infty)$ be the function with the property that for all $X, Y \in L^0(P; \|\cdot\|_{\mathbb{R}^d})$ it holds that*

$$\varrho(X,Y) = \|\rho(X,Y)\|_{L^p(P;|\cdot|_{\mathbb{R}})} = \left(\mathbb{E}[|\rho(X,Y)|^p]\right)^{1/p}. \tag{8.142}$$

Then ϱ is a metric on $L^0(P; \|\cdot\|_{\mathbb{R}^d})$.

Proof of Lemma 8.4.7. First, we note that the symmetry of ϱ follows immediately from the symmetry of ρ. Next we observe that for all $X, Y \in L^0(P; \|\cdot\|_{\mathbb{R}^d})$ with $\varrho(X,Y) = \|\rho(X,Y)\|_{L^p(P;|\cdot|_{\mathbb{R}})} = 0$ it holds P-a.s. that $\rho(X,Y) = 0$. Combining this with the positive definiteness of the metric ρ proves the positive definiteness of ϱ. Finally, we observe that for all $X, Y, Z \in L^0(P; \|\cdot\|_{\mathbb{R}^d})$ it holds that

$$\begin{aligned} \varrho(X,Z) &= \|\rho(X,Z)\|_{L^p(P;|\cdot|_{\mathbb{R}})} \leq \|\rho(X,Y) + \rho(Y,Z)\|_{L^p(P;|\cdot|_{\mathbb{R}})} \\ &\leq \|\rho(X,Y)\|_{L^p(P;|\cdot|_{\mathbb{R}})} + \|\rho(Y,Z)\|_{L^p(P;|\cdot|_{\mathbb{R}})} = \varrho(X,Y) + \varrho(Y,Z). \end{aligned} \tag{8.143}$$

This shows that ϱ satisfies the triangle inequality and therefore is a metric on $L^0(P; \|\cdot\|_{\mathbb{R}^d})$. The proof of Lemma 8.4.7 is thus completed. □

Proposition 8.4.8 (Metrization of convergence in probability). *Let $d \in \mathbb{N}$, $p \in [1,\infty)$, $c \in (0,\infty)$, let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $\rho\colon L^0(P; \|\cdot\|_{\mathbb{R}^d}) \times L^0(P; \|\cdot\|_{\mathbb{R}^d}) \to [0,\infty)$ be the function with the property that for all $X, Y \in L^0(P; \|\cdot\|_{\mathbb{R}^d})$ it holds that*

$$\rho(X,Y) = \|\min\{c, \|X-Y\|_{\mathbb{R}^d}\}\|_{L^p(P;|\cdot|_{\mathbb{R}})} = \big(\mathbb{E}\big[\min\{c^p, \|X-Y\|_{\mathbb{R}^d}^p\}\big]\big)^{1/p}. \tag{8.144}$$

Then $\big(L^0(P; \|\cdot\|_{\mathbb{R}^d}), \rho\big)$ is a metric space.

Proof of Proposition 8.4.8. Combining Lemma 8.4.6 and Lemma 8.4.7 completes the proof of Proposition 8.4.8. □

Proposition 8.4.9 (Metrization of convergence in probability). *Let $d \in \mathbb{N}$, $p \in [1,\infty)$, $c \in (0,\infty)$, let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $\rho\colon L^0(P; \|\cdot\|_{\mathbb{R}^d}) \times L^0(P; \|\cdot\|_{\mathbb{R}^d}) \to [0,\infty)$ be the function with the property that for all $X, Y \in L^0(P; \|\cdot\|_{\mathbb{R}^d})$ it holds that*

$$\rho(X,Y) = \|\min\{c, \|X-Y\|_{\mathbb{R}^d}\}\|_{L^p(P;|\cdot|_{\mathbb{R}})} = \big(\mathbb{E}\big[\min\{c^p, \|X-Y\|_{\mathbb{R}^d}^p\}\big]\big)^{1/p}. \tag{8.145}$$

Then for all $X_n \in L^0(P; \|\cdot\|_{\mathbb{R}^d})$, $n \in \mathbb{N}_0$, it holds that $\limsup_{n\to\infty} \rho(X_n, X_0) = 0$ if and only if $\forall \epsilon \in (0,\infty)\colon \limsup_{n\to\infty} P\big(\|X_n - X_0\|_{\mathbb{R}^d} \geq \epsilon\big) = 0$.

Proof of Proposition 8.4.9. Observe that the Markov inequality proves the for all $\epsilon \in (0,\infty)$, $X, Y \in L^0(P; \|\cdot\|_{\mathbb{R}^d})$ it holds that

$$\begin{aligned} P\big(\|X-Y\|_{\mathbb{R}^d} \geq \epsilon\big) &\leq P\big(\min\{c^p, \|X-Y\|_{\mathbb{R}^d}^p\} \geq \min\{c^p, \epsilon^p\}\big) \\ &\leq \frac{\mathbb{E}\big[\min\{c^p, \|X-Y\|_{\mathbb{R}^d}^p\}\big]}{\min\{c^p, \epsilon^p\}} = \left|\frac{\rho(X,Y)}{\min\{c,\epsilon\}}\right|^p. \end{aligned} \tag{8.146}$$

This proves that for all $X_n \in L^0(P; \|\cdot\|_{\mathbb{R}^d})$, $n \in \mathbb{N}_0$, with $\limsup_{n\to\infty} \rho(X_n, X_0) = 0$ it holds that $\forall \epsilon \in (0,\infty)\colon \limsup_{n\to\infty} P\big(\|X_n - X_0\|_{\mathbb{R}^d} \geq \epsilon\big) = 0$. Next we note that for all $\epsilon, \delta \in (0,\infty)$ and for all $X_n \in L^0(P; \|\cdot\|_{\mathbb{R}^d})$, $n \in \mathbb{N}_0$, with $\forall \tilde{\epsilon} \in (0,\infty)\colon \limsup_{n\to\infty} P\big(\|X_n - X_0\|_{\mathbb{R}^d} \geq \tilde{\epsilon}\big) = 0$ there exists an $N_0 \in \mathbb{N}$ such that for all $n \in \{N_0, N_0+1, \ldots\}$ it holds that

$$P\big(\|X_n - X_0\|_{\mathbb{R}^d} \geq \epsilon\big) < \delta. \tag{8.147}$$

This ensures that for all $\epsilon \in (0,\infty)$ and all $X_n \in L^0(P; \|\cdot\|_{\mathbb{R}^d})$, $n \in \mathbb{N}_0$, with $\forall \tilde{\epsilon} \in (0,\infty)\colon \limsup_{n\to\infty} P\big(\|X_n - X_0\|_{\mathbb{R}^d} \geq \tilde{\epsilon}\big) = 0$ there exists an $N_0 \in \mathbb{N}$ such that for all

$n \in \{N_0, N_0+1, \ldots\}$ it holds that $P\big(\|X_n - X_0\|_{\mathbb{R}^d} \geq \frac{\epsilon}{2^{1/p}}\big) < \frac{\epsilon^p}{2\,c^p}$ and

$$\begin{aligned}\rho(X_n, X_0) &= \big|\mathbb{E}\big[\min\{c^p, \|X_n - X_0\|_{\mathbb{R}^d}^p\}\big]\big|^{1/p} \\ &= \bigg|\,\mathbb{E}\Big[\min\{c^p, \|X_n - X_0\|_{\mathbb{R}^d}^p\}\mathbb{1}_{\{\|X_n - X_0\|_{\mathbb{R}^d} \geq \frac{\epsilon}{2^{1/p}}\}}\Big] \\ &\quad + \mathbb{E}\Big[\min\{c^p, \|X_n - X_0\|_{\mathbb{R}^d}^p\}\mathbb{1}_{\{\|X_n - X_0\|_{\mathbb{R}^d} < \frac{\epsilon}{2^{1/p}}\}}\Big]\,\bigg|^{1/p} \\ &\leq \Big|c^p \cdot P\big(\|X_n - X_0\|_{\mathbb{R}^d} \geq \tfrac{\epsilon}{2^{1/p}}\big) + \min\{c^p, \tfrac{\epsilon^p}{2}\} \cdot P\big(\|X_n - X_0\|_{\mathbb{R}^d} < \tfrac{\epsilon}{2^{1/p}}\big)\Big|^{1/p} \\ &< \Big|\tfrac{\epsilon^p}{2} + \tfrac{\epsilon^p}{2}\Big|^{1/p} = \epsilon.\end{aligned} \tag{8.148}$$

The proof of Proposition 8.4.9 is now completed. □

8.4.9 Solution to Exercise 3.4.8

Lemma 8.4.10. *Let $d, m \in \mathbb{N}$, $T \in (0,\infty)$, $p \in [1,\infty)$, let $(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{t\in[0,T]})$ be a stochastic basis, let $\rho\colon L^2(\lambda_{[0,T]}; \mathbb{R}^{d\times m}) \times L^2(\lambda_{[0,T]}; \mathbb{R}^{d\times m}) \to [0,\infty)$ be a globally bounded* $\operatorname{Pred}((\mathcal{F}_t)_{t\in[0,T]}) \otimes \operatorname{Pred}((\mathcal{F}_t)_{t\in[0,T]})/\mathcal{B}([0,\infty))$*-measurable metric, and let $\varrho\colon L^0(\mathcal{P}_{(\mathcal{F}_t)_{t\in[0,T]}}; \|\cdot\|_{\mathbb{R}^{d\times m}}) \times L^0(\mathcal{P}_{(\mathcal{F}_t)_{t\in[0,T]}}; \|\cdot\|_{\mathbb{R}^{d\times m}}) \to [0,\infty)$ be the function with the property that for all $X, Y \in L^0(\mathcal{P}_{(\mathcal{F}_t)_{t\in[0,T]}}; \|\cdot\|_{\mathbb{R}^{d\times m}})$ it holds that*

$$\varrho(X,Y) = \|\rho(X,Y)\|_{L^p(\mathcal{P}_{(\mathcal{F}_t)_{t\in[0,T]}}; |\cdot|_{\mathbb{R}})} = \left(\mathbb{E}[|\rho(X,Y)|^p]\right)^{1/p}. \tag{8.149}$$

Then ϱ is a metric on $L^0(\mathcal{P}_{(\mathcal{F}_t)_{t\in[0,T]}}; \|\cdot\|_{\mathbb{R}^{d\times m}})$.

Proof of Lemma 8.4.10. The proof of Lemma 8.4.7 can be copied line by line with $L^0(P; \|\cdot\|_{\mathbb{R}^d})$ replaced by $L^0(\mathcal{P}_{(\mathcal{F}_t)_{t\in[0,T]}}; \|\cdot\|_{\mathbb{R}^{d\times m}})$. The proof of Lemma 8.4.10 is thus completed. □

Proposition 8.4.11 (Metrization of convergence in probability)**.** *Let $d, m \in \mathbb{N}$, $T \in (0,\infty)$, $p \in [1,\infty)$, $c \in (0,\infty)$, let $(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{t\in[0,T]})$ be a stochastic basis, and let*

$$d_{p,c}\colon \left\{X \in L^0(\mathcal{P}_{(\mathcal{F}_t)_{t\in[0,T]}}; \|\cdot\|_{\mathbb{R}^{d\times m}})\colon P\Big(\int_0^T \|X_s\|_{\mathbb{R}^{d\times m}}^2\, ds < \infty\Big) = 1\right\}^2 \to [0,\infty), \tag{8.150}$$

be the function with the property that for all $X, Y \in \operatorname{dom}(d_{p,c})$ it holds that

$$d_{p,c}(X,Y) = \left\|\sqrt{\min\left\{c, \int_0^T \|X_s - Y_s\|_{\mathbb{R}^{d\times m}}^2\, ds\right\}}\right\|_{L^p(P; |\cdot|_{\mathbb{R}})}. \tag{8.151}$$

Then $\big(L^0(\mathcal{P}_{(\mathcal{F}_t)_{t\in[0,T]}}; \|\cdot\|_{\mathbb{R}^{d\times m}}), d_{p,c}\big)$ is a metric space.

Proof of Proposition 8.4.11. Note that for all $a \in [0,\infty)$ it holds that $\sqrt{\min\{c,a\}} = \min\{\sqrt{c},\sqrt{a}\}$. Combining the fact that the function $\rho\colon \big(L^2(\lambda_{[0,T]};\mathbb{R}^{d\times m})\big)^2 \to [0,\infty)$ with the property that for all $f,g \in L^2(\lambda_{[0,T]};\mathbb{R}^{d\times m})$ it holds that

$$\rho(f,g) = \sqrt{\int_0^T \|f(s)-g(s)\|^2_{\mathbb{R}^{d\times m}}\,ds} \tag{8.152}$$

is a metric, Lemma 8.4.7, and Lemma 8.4.10 completes the proof of Proposition 8.4.11. □

Proposition 8.4.12 (Metrization of convergence in probability)**.** *Let $d,m \in \mathbb{N}$, $T \in (0,\infty)$, $p \in [1,\infty)$, $c \in (0,\infty)$, let $(\Omega,\mathcal{F},P,(\mathcal{F}_t)_{t\in[0,T]})$ be a stochastic basis, and let*

$$d_{p,c}\colon \left\{X \in L^0(\mathcal{P}_{(\mathcal{F}_t)_{t\in[0,T]}};\|\cdot\|_{\mathbb{R}^{d\times m}})\colon P\big(\int_0^T \|X_s\|^2_{\mathbb{R}^{d\times m}}\,ds < \infty\big) = 1\right\}^2 \to [0,\infty), \tag{8.153}$$

be the function with the property that for all $X,Y \in \operatorname{dom}(d_{p,c})$ it holds that

$$d_{p,c}(X,Y) = \left\|\sqrt{\min\left\{c, \int_0^T \|X_s - Y_s\|^2_{\mathbb{R}^{d\times m}}\,ds\right\}}\right\|_{L^p(P;|\cdot|_{\mathbb{R}})}. \tag{8.154}$$

Then for all $X_n \in L^0(\mathcal{P}_{(\mathcal{F}_t)_{t\in[0,T]}};\|\cdot\|_{\mathbb{R}^{d\times m}})$, $n \in \mathbb{N}_0$, it holds that $\limsup_{n\to\infty} d_{p,c}(X_n,X_0) = 0$ if and only if $\forall \epsilon \in (0,\infty)\colon \limsup_{n\to\infty} P\big(\int_0^T \|X_n - X_0\|^2_{\mathbb{R}^{d\times m}} \geq \epsilon\big) = 0$.

Proof of Proposition 8.4.12. The proof of Proposition 8.4.9 can be copied line by line with the the usual metric on $\mathbb{R}^d$ replaced by the metric defined in (8.152). Thus the proof of Proposition 8.4.12 is completed. □

8.4.10 Solution to Exercise 3.4.20

See, e.g., Lemma 4.4.25 in [Jentzen(2014)] for the statement and the proof of the following lemma.

Lemma 8.4.13. *Let $T \in [0,\infty)$, $t \in [0,T]$, let $(\Omega,\mathcal{F},P,(\mathcal{F}_s)_{s\in[0,T]})$ be a stochastic basis, let $(S,\mathcal{S})$ be a measurable space, let $X\colon \Omega \to S$ be an $\mathcal{F}/\mathcal{S}$-measurable function, let $Y\colon \Omega \to S$ be an $\mathcal{F}_t/\mathcal{S}$-measurable function, and let $A \in \mathcal{F}$ satisfy $P(A) = 1$ and*

$$A \subseteq \{\omega \in \Omega\colon X(\omega) = Y(\omega)\}. \tag{8.155}$$

Then it holds that X is an $\mathcal{F}_t/\mathcal{S}$-measurable function.

Proof of Lemma 8.4.13. First, note that for all $\omega \in A$ it holds that $X(\omega) = Y(\omega)$. Next observe that for all $B \in \mathcal{S}$ it holds that

$$\begin{aligned} X^{-1}(B) &= \left[X^{-1}(B) \cap A\right] \cup \left[X^{-1}(B)\backslash A\right] \\ &= \{\omega \in A \colon X(\omega) \in B\} \cup \left[X^{-1}(B)\backslash A\right] \\ &= \{\omega \in A \colon Y(\omega) \in B\} \cup \left[X^{-1}(B)\backslash A\right] \\ &= \left[Y^{-1}(B) \cap A\right] \cup \left[X^{-1}(B)\backslash A\right]. \end{aligned} \tag{8.156}$$

Moreover, observe that the assumption that $(\mathcal{F}_s)_{s\in[0,T]}$ is a normal filtration together with the fact that $P(A) = 1$ implies that

$$A, A^c \in \mathcal{F}_0 \subseteq \mathcal{F}_t \subseteq \mathcal{F}. \tag{8.157}$$

This and the assumption that Y is $\mathcal{F}_t/\mathcal{S}$-measurable prove that for all $B \in \mathcal{S}$ it holds that

$$Y^{-1}(B) \cap A \in \mathcal{F}_t. \tag{8.158}$$

Furthemore, note that the monotonicity of the probability measure P ensures that for all $B \in \mathcal{S}$ it holds that $P(X^{-1}(B)\backslash A) = 0$. The assumption that $(\mathcal{F}_s)_{s\in[0,T]}$ is normal hence shows that for all $B \in \mathcal{S}$ it holds that

$$X^{-1}(B)\backslash A \in \mathcal{F}_t. \tag{8.159}$$

Combining (8.156) with (8.158) and (8.159) proves that for all $B \in \mathcal{S}$ it holds that $X^{-1}(B) \in \mathcal{F}_t$. The proof of Lemma 8.4.13 is thus completed. □

8.4.11 Solution to Exercise 3.4.22

Lemma 8.4.14. *Let $T \in (0,\infty)$, $d, m \in \mathbb{N}$, let $(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{t\in[0,T]})$ be a stochastic basis, let $W \colon [0,T] \times \Omega \to \mathbb{R}^m$ be an m-dimensional standard $(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{t\in[0,T]})$-Brownian motion, let $a, b, c, d \in [0,T]$ with $a \le b \le c \le d$, and let $X, Y \colon [0,T] \times \Omega \to \mathbb{R}^{d\times m}$ be $(\mathcal{F}_t)_{t\in[0,T]}/\mathcal{B}(\mathbb{R}^{d\times m})$-predictable functions with $X, Y \in C([0,T], L^2(P; \|\cdot\|_{\mathbb{R}^{d\times m}}))$. Then $\int_a^b \mathbb{E}_P\big[\|X_s\|^2_{\mathrm{HS}(\mathbb{R}^m,\mathbb{R}^d)} + \|Y_s\|^2_{\mathrm{HS}(\mathbb{R}^m,\mathbb{R}^d)}\big]\, ds < \infty$ and*

$$\mathbb{E}_P\left[\left\langle \int_a^b X_s\, dW_s, \int_c^d Y_s\, dW_s \right\rangle_{\mathbb{R}^d}\right] = 0. \tag{8.160}$$

Proof of Lemma 8.4.14. Throughout this proof let $Z, R \colon [0,T] \times \Omega \to \mathbb{R}^{d\times m}$ be stochastic processes which satisfy for all $t \in [0,T]$ that

$$Z_t = \mathbb{1}_{(a,b]}(t) \cdot X_t + \mathbb{1}_{(c,d]}(t) \cdot Y_t \tag{8.161}$$

and

$$R_t = \mathbb{1}_{(a,b]}(t) \cdot X_t - \mathbb{1}_{(c,d]}(t) \cdot Y_t. \tag{8.162}$$

Then note that Item (iv) in Theorem 3.4.21, the fact that $\|\cdot\|_{\mathbb{R}^{d\times m}}$ and $\|\cdot\|_{\mathrm{HS}(\mathbb{R}^m,\mathbb{R}^d)}$ are equivalent norms on $\mathbb{R}^{d\times m}$, and the assumption that $X, Y \in C([0,T], L^2(P; \|\cdot\|_{\mathbb{R}^{d\times m}}))$ imply that $\int_a^b \mathbb{E}_P\big[\|X_s\|^2_{\mathrm{HS}(\mathbb{R}^m,\mathbb{R}^d)} + \|Y_s\|^2_{\mathrm{HS}(\mathbb{R}^m,\mathbb{R}^d)}\big]\, ds < \infty$, $\int_a^b \mathbb{E}_P\big[\|Z_s\|^2_{\mathrm{HS}(\mathbb{R}^m,\mathbb{R}^d)} + \|R_s\|^2_{\mathrm{HS}(\mathbb{R}^m,\mathbb{R}^d)}\big]\, ds < \infty$, and

$$
\begin{aligned}
&\mathbb{E}_P\bigg[\bigg\langle \int_a^b X_s\, dW_s, \int_c^d Y_s\, dW_s \bigg\rangle_{\mathbb{R}^d}\bigg] \\
&= \tfrac{1}{4}\Bigg(\mathbb{E}_P\Bigg[\bigg\| \int_a^b X_s\, dW_s + \int_c^d Y_s\, dW_s \bigg\|^2_{\mathbb{R}^d}\Bigg] - \mathbb{E}_P\Bigg[\bigg\| \int_a^b X_s\, dW_s - \int_c^d Y_s\, dW_s \bigg\|^2_{\mathbb{R}^d}\Bigg]\Bigg) \\
&= \tfrac{1}{4}\Bigg(\mathbb{E}_P\Bigg[\bigg\| \int_a^d Z_s\, dW_s \bigg\|^2_{\mathbb{R}^d}\Bigg] - \mathbb{E}_P\Bigg[\bigg\| \int_a^d R_s\, dW_s \bigg\|^2_{\mathbb{R}^d}\Bigg]\Bigg) \\
&= \tfrac{1}{4}\bigg(\int_a^d \mathbb{E}_P\Big[\|Z_s\|^2_{\mathrm{HS}(\mathbb{R}^m,\mathbb{R}^d)}\Big]\, ds - \int_a^d \mathbb{E}_P\Big[\|R_s\|^2_{\mathrm{HS}(\mathbb{R}^m,\mathbb{R}^d)}\Big]\, ds\bigg) \\
&= \tfrac{1}{4}\bigg(\int_a^b \mathbb{E}_P\Big[\|X_s\|^2_{\mathrm{HS}(\mathbb{R}^m,\mathbb{R}^d)}\Big]\, ds + \int_c^d \mathbb{E}_P\Big[\|Y_s\|^2_{\mathrm{HS}(\mathbb{R}^m,\mathbb{R}^d)}\Big]\, ds \\
&\quad - \int_a^b \mathbb{E}_P\Big[\|X_s\|^2_{\mathrm{HS}(\mathbb{R}^m,\mathbb{R}^d)}\Big]\, ds - \int_c^d \mathbb{E}_P\Big[\|-Y_s\|^2_{\mathrm{HS}(\mathbb{R}^m,\mathbb{R}^d)}\Big]\, ds\bigg) = 0.
\end{aligned}
\tag{8.163}
$$

The proof of Lemma 8.4.14 is thus completed. □

Lemma 8.4.15. *Let $T \in (0,\infty)$, $d, m \in \mathbb{N}$, let $(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{t\in[0,T]})$ be a stochastic basis, let $W\colon [0,T]\times\Omega \to \mathbb{R}^m$ be an m-dimensional standard $(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{t\in[0,T]})$-Brownian motion, let $a \in [0,T]$, $b \in [a,T]$, and let $X\colon [0,T]\times\Omega \to \mathbb{R}^{d\times m}$ be an $(\mathcal{F}_t)_{t\in[0,T]}/\mathcal{B}(\mathbb{R}^{d\times m})$-predictable function with $X \in C([0,T], L^2(P; \|\cdot\|_{\mathbb{R}^{d\times m}}))$. Then $\int_a^b \mathbb{E}_P\big[\|X_s\|^2_{\mathrm{HS}(\mathbb{R}^m,\mathbb{R}^d)}\big] < \infty$ and*

$$
\limsup_{n\to\infty} \left\| \int_a^b X_s\, dW_s - \left[\sum_{k=0}^{n-1} X_{(a+\frac{k(b-a)}{n})}\Big(W_{a+\frac{(k+1)(b-a)}{n}} - W_{a+\frac{k(b-a)}{n}}\Big)\right]\right\|_{L^2(P;\|\cdot\|_{\mathbb{R}^d})} = 0.
\tag{8.164}
$$

Proof of Lemma 8.4.15. Lemma 8.4.14 proves for all $n \in \mathbb{N}$ that

$\int_a^b \mathbb{E}_P\big[\|X_s\|^2_{\mathrm{HS}(\mathbb{R}^m,\mathbb{R}^d)}\big] < \infty$ and

$$\begin{aligned}
&\left\|\int_a^b X_s\,dW_s - \left[\sum_{k=0}^{n-1} X_{(a+\frac{k(b-a)}{n})}\Big(W_{a+\frac{(k+1)(b-a)}{n}} - W_{a+\frac{k(b-a)}{n}}\Big)\right]\right\|^2_{L^2(P;\|\cdot\|_{\mathbb{R}^d})}\\
&= \left\|\sum_{k=0}^{n-1}\int_{a+\frac{k(b-a)}{n}}^{a+\frac{(k+1)(b-a)}{n}} \Big(X_s - X_{(a+\frac{k(b-a)}{n})}\Big)\,dW_s\right\|^2_{L^2(P;\|\cdot\|_{\mathbb{R}^d})}\\
&= \mathbb{E}_P\left[\sum_{k=0}^{n-1}\left\|\int_{a+\frac{k(b-a)}{n}}^{a+\frac{(k+1)(b-a)}{n}} \Big(X_s - X_{(a+\frac{k(b-a)}{n})}\Big)\,dW_s\right\|^2_{\mathbb{R}^d}\right]\\
&= \sum_{k=0}^{n-1}\left\|\int_{a+\frac{k(b-a)}{n}}^{a+\frac{(k+1)(b-a)}{n}} \Big(X_s - X_{(a+\frac{k(b-a)}{n})}\Big)\,dW_s\right\|^2_{L^2(P;\|\cdot\|_{\mathbb{R}^d})}.
\end{aligned} \tag{8.165}$$

Item (iv) in Theorem 3.4.21 and the fact that $\|\cdot\|_{\mathbb{R}^{d\times m}}$ and $\|\cdot\|_{\mathrm{HS}(\mathbb{R}^m,\mathbb{R}^d)}$ are equivalent norms on $\mathbb{R}^{d\times m}$ hence prove for all $n \in \mathbb{N}$ that

$$\begin{aligned}
&\left\|\int_a^b X_s\,dW_s - \left[\sum_{k=0}^{n-1} X_{(a+\frac{k(b-a)}{n})}\Big(W_{a+\frac{(k+1)(b-a)}{n}} - W_{a+\frac{k(b-a)}{n}}\Big)\right]\right\|^2_{L^2(P;\|\cdot\|_{\mathbb{R}^d})}\\
&= \sum_{k=0}^{n-1}\int_{a+\frac{k(b-a)}{n}}^{a+\frac{(k+1)(b-a)}{n}} \big\|X_s - X_{(a+\frac{k(b-a)}{n})}\big\|^2_{L^2(P;\|\cdot\|_{\mathrm{HS}(\mathbb{R}^m,\mathbb{R}^d)})}ds\\
&\le (b-a)\left[w_X\big(\tfrac{b-a}{n}\big)\right]^2 \sup_{x\in\mathbb{R}^{d\times m}\setminus\{0\}}\left[\frac{\|x\|_{\mathrm{HS}(\mathbb{R}^m,\mathbb{R}^d)}}{\|x\|_{\mathbb{R}^{d\times m}}}\right]^2 < \infty.
\end{aligned} \tag{8.166}$$

The fact that $X \in C([0,T], L^2(P;\|\cdot\|_{\mathbb{R}^{d\times m}}))$ and the compactness of the interval $[0,T]$ imply that the function $[0,T] \ni t \mapsto X_t \in L^2(P;\|\cdot\|_{\mathbb{R}^{d\times m}})$ is uniformly continuous. Combining this with (8.166) and Lemma 8.3.1 completes the proof of Lemma 8.4.15. □

8.5 Chapter 5

8.5.1 Solution to Exercise 5.2.4

Lemma 8.5.1. *Let (E,d_E) and (F,d_F) be metric spaces, let $v \in E$, $w \in F$, $c \in [0,\infty)$, and let $f\colon E \to F$ be a function which satisfies for all $x \in E$ that*

$$d_F(w, f(x)) \le c\big(1 + d_E(v,x)\big)^c. \tag{8.167}$$

Then f grows at most polynomially from (E,d_E) to (F,d_F).

Proof of Lemma 8.5.1. Observe that (8.167) and the triangle inequality ensure that for all $\tilde{v}, x \in E$, $\tilde{w} \in F$ it holds that

$$\begin{aligned} d_F(\tilde{w}, f(x)) &\leq d_F(\tilde{w}, w) + d_F(w, f(x)) \\ &\leq d_F(\tilde{w}, w) + c\left(1 + d_E(v, x)\right)^c \\ &\leq d_F(\tilde{w}, w) + c\left(1 + d_E(v, \tilde{v}) + d_E(\tilde{v}, x)\right)^c \\ &\leq \left(d_F(\tilde{w}, w) + c\right)\left[1 + \left(1 + d_E(v, \tilde{v}) + d_E(\tilde{v}, x)\right)^c\right] \\ &\leq \left(d_F(\tilde{w}, w) + c\right)\left(1 + d_E(v, \tilde{v})\right)^c\left[1 + \left(1 + d_E(\tilde{v}, x)\right)^c\right] \\ &\leq 2\left(d_F(\tilde{w}, w) + c\right)\left(1 + d_E(v, \tilde{v})\right)^c\left(1 + d_E(\tilde{v}, x)\right)^c \\ &\leq \left[2\left(d_F(\tilde{w}, w) + c\right)\left(1 + d_E(v, \tilde{v})\right)^c\right]\left(1 + d_E(\tilde{v}, x)\right)^{\left[2\,(d_F(\tilde{w}, w)+c)\,(1+d_E(v,\tilde{v}))^c\right]}. \end{aligned} \tag{8.168}$$

The proof of Lemma 8.5.1 is thus completed. □

Proposition 8.5.2. *Let (E, d_E) and (F, d_F) be metric spaces with $E \neq \emptyset$. Then for all functions $f\colon E \to F$ it holds that f grows at most polynomially from (E, d_E) to (F, d_F) if and only if there exist $v \in E$, $w \in F$ such that*

$$\limsup_{c \to \infty} \sup_{x \in E} \left[\frac{d_F(w, f(x))}{[1 + d_E(v, x)]^c}\right] < \infty. \tag{8.169}$$

Proof of Proposition 8.5.2. Throughout this proof let $f\colon E \to F$ be an at most polynomially growing function and let $g \in \mathbb{M}(E, F)$, $\tilde{v} \in E$, $\tilde{w} \in F$ satisfy that

$$\limsup_{c \to \infty} \sup_{x \in E} \left[\frac{d_F(\tilde{w}, g(x))}{[1 + d_E(\tilde{v}, x)]^c}\right] < \infty. \tag{8.170}$$

Observe that (8.170) implies that there exist real numbers $K, \tilde{c} \in [0, \infty)$ such that for all $x \in E$ it holds that

$$d_F(\tilde{w}, g(x)) \leq K\left(1 + d_E(\tilde{v}, x)\right)^{\tilde{c}} \leq \max\{K, \tilde{c}\}\left(1 + d_E(\tilde{v}, x)\right)^{\max\{K, \tilde{c}\}}. \tag{8.171}$$

Lemma 8.5.1 hence shows that g grows at most polynomially from (E, d_E) to (F, d_F). It thus remains to prove that there exist $v \in E$, $w \in F$ such that

$$\limsup_{c \to \infty} \sup_{x \in E} \left[\frac{d_F(w, f(x))}{[1 + d_E(v, x)]^c}\right] < \infty. \tag{8.172}$$

The assumption that f is polynomially growing ensures that for all $v \in E$, $w \in F$ there exists a real number $c \in [0, \infty)$ such that

$$\sup_{x \in E} \left[\frac{d_F(w, f(x))}{[1 + d_E(v, x)]^c}\right] < \infty. \tag{8.173}$$

Combining this with the fact that for all $v, x \in E$, $w \in F$ it holds that the function

$$(0, \infty) \ni r \mapsto \left[\frac{d_F(w, f(x))}{[1 + d_E(v, x)]^r}\right] \in [0, \infty) \tag{8.174}$$

is non-increasing implies (8.172). The proof of Proposition 8.5.2 is thus completed. □

8.5.2 Solution to Exercise 5.2.7

Lemma 8.5.3. *Let $k, l \in \mathbb{N}$ and let $f\colon \mathbb{R}^k \to \mathbb{R}^l$ be a continuously differentiable function with at most polynomially growing derivative. Then f grows at most polynomially.*

Proof of Lemma 8.5.3. Observe that the fundamental theorem of calculus ensures that for all $x, y \in \mathbb{R}^k$ it holds that

$$f(y) - f(x) = \int_0^1 f'(x + r\,(y-x))\,(y-x)\,dr. \tag{8.175}$$

The assumption that f' grows at most polynomially hence implies that there exists a real number $c \in [0,\infty)$ such that for all $x \in \mathbb{R}^k$ it holds that

$$\begin{split}
&\|f(x)\|_{\mathbb{R}^l} \leq \|f(x) - f(0)\|_{\mathbb{R}^l} + \|f(0)\|_{\mathbb{R}^l} \\
&\leq \int_0^1 \|f'(rx)\,x\|_{\mathbb{R}^l}\,dr + \|f(0)\|_{\mathbb{R}^l} \\
&\leq \int_0^1 \|f'(rx)\|_{L(\mathbb{R}^k,\mathbb{R}^l)}\,\|x\|_{\mathbb{R}^k}\,dr + \|f(0)\|_{\mathbb{R}^l} \\
&\leq \int_0^1 c\left(1 + r\,\|x\|_{\mathbb{R}^k}\right)^c \|x\|_{\mathbb{R}^k}\,dr + \|f(0)\|_{\mathbb{R}^l} \\
&\leq c\left(1 + \|x\|_{\mathbb{R}^k}\right)^c \|x\|_{\mathbb{R}_k} + \|f(0)\|_{\mathbb{R}^l} \\
&\leq (c + \|f(0)\|_{\mathbb{R}^l})\left(1 + \|x\|_{\mathbb{R}^k}\right)^{(c+1)} \\
&\leq (c + \|f(0)\|_{\mathbb{R}^l} + 1)\left(1 + \|x\|_{\mathbb{R}^k}\right)^{(c+\|f(0)\|_{\mathbb{R}^l}+1)}.
\end{split} \tag{8.176}$$

This and Lemma 8.5.1 imply that f grows at most polynomially. The proof of Lemma 8.5.3 is thus completed. ☐

Proposition 8.5.4. *Let $k, l, v \in \mathbb{N}$ and let $f\colon \mathbb{R}^k \to \mathbb{R}^l$ be a v–times continuously differentiable function with at most polynomially growing derivatives. Then it holds for all $w \in \{0, 1, \ldots, v\}$ that $f^{(w)}$ grows at most polynomially.*

Proof of Proposition 8.5.4. The proof of Proposition 8.5.4 in the case $v = 1$ follows directly from Lemma 8.5.3. Therefore, assume without loss of generality that $v \in \{2, 3, \ldots\}$. Next we show by induction that for all $j \in \{1, 2, \ldots, v\}$ it holds that $f^{(j)}$ grows at most polynomially. It is clear that $f^{(v)}$ grows at most polynomially. Let $w \in \{1, \ldots, v-1\}$ satisfy that $f^{(w+1)}$ grows at most polynomially. Note that the function

$$\mathbb{R}^k \ni x \mapsto f^{(w)}(x) \in L^{(w)}(\mathbb{R}^k, \mathbb{R}^l) \tag{8.177}$$

is continuously differentiable with at most polynomially growing derivative. This, the fact that $L^{(w)}(\mathbb{R}^k, \mathbb{R}^l) \cong \mathbb{R}^{k^w \cdot l}$, and Lemma 8.5.3 show that $f^{(w)}$ grows at most polynomially. Induction implies that for all $j \in \{1, \ldots, v\}$ it holds that $f^{(j)}$ grows at most polynomially. In addition, Lemma 8.5.3 proves that f grows at most polynomially. The proof of Proposition 8.5.4 is thus completed. ☐

8.5.3 Solution to Exercise 5.3.3

```
function  Y = EulerMaruyama(T,d,m,N,xi,mu,sigma)
  h = T/N;
  sqrth = sqrt(h);
  Y = xi;
  for i=1:N
      Y = Y + mu(Y)*h + sigma(Y)*sqrth*randn(m,1);
  end
end
```

Matlab code 8.28: A Matlab function EulerMaruyama(T,d,m,N,ξ,μ,σ) with input $T \in (0,\infty)$, $d,m,N \in \mathbb{N}$, $\xi \in \mathbb{R}^d$, $\mu \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d),\mathcal{B}(\mathbb{R}^d))$, $\sigma \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d),\mathcal{B}(\mathbb{R}^{d\times m}))$ and output a realization of an $Y_N(P)_{\mathcal{B}(\mathbb{R}^d)}$-distributed random variable.

8.5.4 Solution to Exercise 5.3.4

```
function E = MonteCarloEulerGBM(T,alpha,beta,xi,K,N,M)
  Y = xi;
  h = T/N;
  sqrth = sqrt(h);
  for i=1:N
      Y = Y + alpha*Y*h + beta*Y.*randn(M,1)*sqrth;
  end
  E = sum(max( Y - K,0 ))/M;
end
```

Matlab code 8.29: A Matlab function MonteCarloEulerGBM(T,α,β,ξ,K,N,M) with input $T,\alpha,\beta,\xi,K \in (0,\infty)$, $N,M \in \mathbb{N}$ and output a realization of an $\big(\frac{1}{M}\sum_{k=1}^{M}\max\{Y_N^k - K,0\}\big)(P)_{\mathcal{B}(\mathbb{R})}$-distributed random variable.

```
MonteCarloEulerGBM(1,log(1.06)-1/200,1/10,92,100,100,10000)

ans =

    2.6618
```

Matlab code 8.30: Result of the Matlab function 8.29 with specified values in the Matlab console.

8.5.5 Solution to Exercise 5.5.6

Lemma 8.5.5. *Let $L \in \mathbb{R}$. Then for all $\mu \in C^1(\mathbb{R}, \mathbb{R})$ it holds that $\sup_{x\in\mathbb{R}} \mu'(x) \leq L$ if and only if $\forall\, x, y \in \mathbb{R}$: $(x - y) \cdot (\mu(x) - \mu(y)) \leq L\,(x - y)^2$.*

Proof of Lemma 5.5.6. Throughout this proof let $\mu, \eta \in C^1(\mathbb{R}, \mathbb{R})$ satisfy for all $x, y \in \mathbb{R}$ that $\sup_{t\in\mathbb{R}} \mu'(t) \leq L$ and $(x - y) \cdot (\eta(x) - \eta(y)) \leq L\,(x - y)^2$. Observe that for all $x \in \mathbb{R}, y \in [x, \infty)$ it holds that

$$\mu(x) - \mu(y) = -\int_x^y \mu'(t)\,dt \geq -\int_x^y \sup\nolimits_{s\in\mathbb{R}} \mu'(s)\,dt = (x - y) \sup\nolimits_{t\in\mathbb{R}} \mu'(t). \tag{8.178}$$

Next note that for all $x \in \mathbb{R}, y \in (-\infty, x]$ it holds that

$$\mu(x) - \mu(y) = \int_y^x \mu'(t)\,dt \leq \int_y^x \sup\nolimits_{s\in\mathbb{R}} \mu'(s)\,dt = (x - y) \sup\nolimits_{t\in\mathbb{R}} \mu'(t). \tag{8.179}$$

Combining (8.178) and (8.179) implies that for all $x, y \in \mathbb{R}$ it holds that

$$(x - y)(\mu(x) - \mu(y)) \leq (x - y)^2 \sup\nolimits_{t\in\mathbb{R}} \mu'(t). \tag{8.180}$$

This shows for all $x, y \in \mathbb{R}$ that $(x - y) \cdot (\mu(x) - \mu(y)) \leq L(x - y)^2$. Furthermore, for all $x, y \in \mathbb{R}$ with $x \neq y$ it holds that $\frac{\eta(x)-\eta(y)}{x-y} \leq L$. Therefore, for all $x \in \mathbb{R}$ it holds that $\eta'(x) = \lim_{y\to x} \frac{\eta(x)-\eta(y)}{x-y} \leq L$. Hence, it holds that $\sup_{x\in\mathbb{R}} \eta'(x) \leq L$. The proof of Lemma 5.5.6 is thus completed. □

8.5.6 Solution to Exercise 5.5.9

```
function Y = IncrementTamed(T,d,m,N,xi,mu,sigma)
  h = T/N;
  sqrth = sqrt(h);
  Y = xi;
  for i=1:N
      Z = mu(Y)*h + sigma(Y)*sqrth*randn(m,1);
      Y = Y + Z/max([1,h*norm(Z)]);
  end
end
```

Matlab code 8.31: A Matlab function `IncrementTamed(`T`,`d`,`m`,`N`,`ξ`,`μ`,`σ`)` with input $T \in (0, \infty)$, $d, m, N \in \mathbb{N}$, $\xi \in \mathbb{R}^d$, $\mu \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{R}^d))$, $\sigma \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{R}^{d\times m}))$ and output a realization of an $Y_N(P)_{\mathcal{B}(\mathbb{R}^d)}$-distributed random variable.

8.5.7 Solution to Exercise 5.6.7

```
function Y = Milstein(T,d,N,xi,mu,sigma,sigma_t)
  h = T/N;
  sqrth = sqrt(h);
  Y = xi;
  for i=1:N
      Delta_W = randn(1,1)*sqrth;
      Y = Y + mu(Y)*h + sigma(Y)*Delta_W ...
          + 1/2*sigma_t(Y)*sigma(Y)*Delta_W^2 ...
          - h/2*sigma_t(Y)*sigma(Y);
  end
end
```

Matlab code 8.32: A Matlab function `Milstein(`T`,`d`,`N`,`ξ`,`μ`,`σ`,`$\tilde{\sigma}$`)` with input $T \in (0, \infty)$, $d, N \in \mathbb{N}$, $\xi \in \mathbb{R}^d$, $\mu \in \mathcal{M}(\mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{R}^d))$, $\sigma \in C^1(\mathbb{R}^d, \mathbb{R}^d)$, $\tilde{\sigma} \in C^1(\mathbb{R}^d, \mathbb{R}^{d\times d})$ and output a realization of an $Y_N(P)_{\mathcal{B}(\mathbb{R}^d)}$-distributed random variable.

8.5.8 Solution to Exercise 5.6.8

Lemma 8.5.6. *Let $A_1, A_2 \in \mathbb{R}^{2\times 2}$ be the 2×2-matrices given by*

$$A_1 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \qquad \textit{and} \qquad A_2 = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} \tag{8.181}$$

and let $\sigma = (\sigma_1, \sigma_2) \colon \mathbb{R}^2 \to \mathbb{R}^{2\times 2}$ be the function which satisfies for all $x = (x_1, x_2)$, $u = (u_1, u_2) \in \mathbb{R}^2$ that

$$\sigma(x)u = u_1 A_1 x + u_2 A_2 x. \tag{8.182}$$

Then for all $x \in \mathbb{R}^2$ it holds that $\sigma_1'(x)\,\sigma_2(x) = \sigma_2'(x)\,\sigma_1(x)$.

Proof of Lemma 8.5.6. Observe that for all $x \in \mathbb{R}^2$ it holds that

$$\sigma_1(x) = A_1 x\,, \qquad \sigma_2(x) = A_2 x\,, \qquad \sigma_1'(x) = A_1 \qquad \text{and} \qquad \sigma_2'(x) = A_2\,. \tag{8.183}$$

Observe that

$$A_1 A_2 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix}. \tag{8.184}$$

and

$$A_2 A_1 = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix}. \tag{8.185}$$

The proof of Lemma 8.5.6 is thus completed. □

```
function Y = Milstein2D(T,N,xi)
  A1 = [1, 1; 1, 0];
  A2 = [1, -1; -1, 2];
  sigma1 = @(x) A1*x;
  sigma2 = @(x) A2*x;
  sigma = @(x) [sigma1(x),sigma2(x)];
  h = T/N;
  sqrth = sqrt(h);
  Y = xi;
  for i=1:N
      Delta_W = randn(2,1)*sqrth;
      Y = Y + sigma(Y)*Delta_W ...
            - h/2*( A1*sigma1(Y) + A2*sigma2(Y) ) ...
          + 1/2*Delta_W(1)*A1*(sigma1(Y)*Delta_W(1) ...
              + sigma2(Y)*Delta_W(2)) ...
          + 1/2*Delta_W(2)*A2*(sigma1(Y)*Delta_W(1) ...
              + sigma2(Y)*Delta_W(2));
  end
end
```

Matlab code 8.33: A Matlab function `Milstein2D(`T`,`N`,`ξ`)` with input $T \in (0,\infty)$, $N \in \mathbb{N}$, $\xi \in \mathbb{R}^2$ and output a realization of an $Y_N(P)_{\mathcal{B}(\mathbb{R}^2)}$-distributed random variable.

Bibliography

[Ahrens and Dieter(1974)] J. H. Ahrens and U. Dieter. Computer methods for sampling from gamma, beta, poisson and binomial distributions. *Computing (Arch. Elektron. Rechnen)*, 12:223–246, 1974.

[Ait-Sahalia(1996)] Y. Ait-Sahalia. Testing continuous-time models of the spot interest rate. *Review of Financial Studies*, 9(2):385–426, 1996.

[Ash(2013)] R. B. Ash. *Lectures on Statistics.* Lecture notes, University of Illinois. http://www.math.uiuc.edu/ r-ash/Stat.html, 2013.

[Asmussen and Glynn(2007)] S. Asmussen and P. W. Glynn. *Stochastic Simulation: Algorithms and Analysis*, volume 57 of *Stochastic Modelling and Applied Probability.* New York, NY: Springer, 2007.

[Atkinson(1989)] K. E. Atkinson. *An introduction to numerical analysis.* John Wiley & Sons Inc., New York, second edition, 1989. ISBN 0-471-62489-6.

[Bauer(1991)] H. Bauer. *Probability Theory. (Wahrscheinlichkeitstheorie.) 4., völlig überarb. u. neugestaltete Aufl. des Werkes: Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie.* De Gruyter Lehrbuch. Berlin etc.: Walter de Gruyter, 1991.

[Black and Scholes(1973)] F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654, 1973.

[Box and Muller(1958)] G. E. P. Box and M. E. Muller. A note on the generation of random normal deviates. *The Annals of Mathematical Statistics*, 21:610–611, 1958.

[Cox et al.(1985)Cox, Ingersoll, and Ross] J. C. Cox, J. E. Ingersoll, Jr., and S. A. Ross. A theory of the term structure of interest rates. *Econometrica*, 53(2):385–407, 1985. ISSN 0012-9682. doi: 10.2307/1911242. URL `http://dx.doi.org/10.2307/1911242`.

[Da Prato and Zabczyk(1992)] G. Da Prato and J. Zabczyk. *Stochastic equations in infinite dimensions*, volume 44 of *Encyclopedia of Mathematics and its Applications.* Cambridge University Press, Cambridge, 1992. ISBN 0-521-38529-6. doi: 10.1017/CBO9780511666223. URL `http://dx.doi.org/10.1017/CBO9780511666223`.

[Da Prato et al.(2012)Da Prato, Jentzen, and Roeckner] G. Da Prato, A. Jentzen, and M. Roeckner. A mild ito formula for spdes. *arXiv:1009.3526*, pages 1–39, 2012.

[Fishman(1996)] G. S. Fishman. *Monte Carlo. Concepts, Algorithms, and Applications.* New York, NY: Springer, 1996.

[FrankfurterAllgemeineZeitung(2006)] FrankfurterAllgemeineZeitung. Fusion der grössten amerikanischen terminbörsen. *http://www.faz.net/aktuell/finanzen/boersen-fusion-der-groessten-amerikanischen-terminboersen-1385544.html*, 2006.

[Garcke(2008)] J. Garcke. Sparse grid tutorial. pages 1–24, 2008.

[Giles(2008)] M. B. Giles. Multilevel Monte Carlo path simulation. *Oper. Res.*, 56(3): 607–617, 2008. doi: 10.1287/opre.1070.0496.

[Glasserman(2004)] P. Glasserman. *Monte Carlo methods in financial engineering.* Applications of Mathematics (New York). Springer-Verlag, 2004.

[Gyöngy(1998)] I. Gyöngy. A note on Euler's approximations. *Potential Anal.*, 8(3): 205–216, 1998. ISSN 0926-2601. doi: 10.1023/A:1008605221617. URL `http://dx.doi.org/10.1023/A:1008605221617`.

[Gyöngy and Krylov(1996)] I. Gyöngy and N. Krylov. Existence of strong solutions for Itô's stochastic equations via approximations. *Probab. Theory Related Fields*, 105(2):143–158, 1996. ISSN 0178-8051. doi: 10.1007/BF01203833. URL `http://dx.doi.org/10.1007/BF01203833`.

[Hairer et al.(2012)Hairer, Hutzenthaler, and Jentzen] M. Hairer, M. Hutzenthaler, and A. Jentzen. Loss of regularity for Kolmogorov equations. *arXiv:1209.6035*, pages 1–27, 2012.

[Halmos(1950)] P. R. Halmos. *Measure Theory.* D. Van Nostrand Company, Inc., New York, N. Y., 1950.

[Heinrich(2001)] S. Heinrich. Multilevel Monte Carlo methods. In *Large-Scale Scientific Computing*, volume 2179 of *Lect. Notes Comput. Sci.*, pages 58–67. Springer, Berlin, 2001.

[Henry-Labordère(2007)] P. Henry-Labordère. Solvable local and stochastic volatility models: supersymmetric methods in option pricing. *Quant. Finance*, 7(5):525–535, 2007. ISSN 1469-7688. doi: 10.1080/14697680601103045. URL `http://dx.doi.org/10.1080/14697680601103045`.

[Heston(1993)] S. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Commun. Comput. Phys.*, 6(2): 327–343, 1993.

[Higham(2004)] D. J. Higham. *An introduction to financial option valuation.* Cambridge University Press, 2004.

[Higham(2011)] D. J. Higham. Stochastic ordinary differential equations in applied and computational mathematics. *IMA J. Appl. Math.*, 76(3):449–474, 2011. ISSN 0272-4960. doi: 10.1093/imamat/hxr016. URL `http://dx.doi.org/10.1093/imamat/hxr016`.

[Higham et al.(2002)Higham, Mao, and Stuart] D. J. Higham, X. Mao, and A. M. Stuart. Strong convergence of Euler-type methods for nonlinear stochastic differential equations. *SIAM J. Numer. Anal.*, 40(3):1041–1063, 2002. doi: 10.1137/S0036142901389530.

[Hu(1996)] Y. Hu. Semi-implicit Euler-Maruyama scheme for stiff stochastic equations. In *Stochastic analysis and related topics, V (Silivri, 1994)*, volume 38 of *Progr. Probab.*, pages 183–202. Birkhäuser Boston, Boston, MA, 1996.

[Hutzenthaler and Jentzen(2011)] M. Hutzenthaler and A. Jentzen. Convergence of the stochastic Euler scheme for locally Lipschitz coefficients. *Found. Comput. Math.*, 11 (6):657–706, 2011. ISSN 1615-3375. doi: 10.1007/s10208-011-9101-9. URL `http://dx.doi.org/10.1007/s10208-011-9101-9`.

[Hutzenthaler and Jentzen(2012)] M. Hutzenthaler and A. Jentzen. Numerical approximations of stochastic differential equations with non-globally Lipschitz continuous coefficients. *arXiv:1203.5809*, pages 1–59, 2012.

[Hutzenthaler et al.(2011a)Hutzenthaler, Jentzen, and Kloeden] M. Hutzenthaler, A. Jentzen, and P. E. Kloeden. Strong and weak divergence in finite time of Euler's method for stochastic differential equations with non-globally Lipschitz continuous coefficients. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 467:1563–1576, 2011a. doi: 10.1098/rspa.2010.0348. URL `http://rspa.royalsocietypublishing.org/content/early/2010/12/08/rspa.2010.0348.abstract`.

[Hutzenthaler et al.(2011b)Hutzenthaler, Jentzen, and Kloeden] M. Hutzenthaler, A. Jentzen, and P. E. Kloeden. Divergence of the multilevel Monte Carlo Euler method for nonlinear stochastic differential equations. *arXiv:1105.0226v2*, page 31 pages, 2011b. To appear in Ann. Appl. Probab..

[Hutzenthaler et al.(2012)Hutzenthaler, Jentzen, and Kloeden] M. Hutzenthaler, A. Jentzen, and P. E. Kloeden. Strong convergence of an explicit numerical method for sdes with non-globally lipschitz continuous coefficients. *Ann. Appl. Probab.*, 22 (4):1611–1641, 2012.

[Ikeda and Watanabe(1989)] N. Ikeda and S. Watanabe. *Stochastic Differential Equations and Diffusion Processes. 2nd ed.*, volume 24 of *North-Holland Mathematical Library.* Amsterdam etc.: North-Holland; Tokyo: Kodansha Ltd., 1989.

[Jacod and Shiryaev(2003)] J. Jacod and A. N. Shiryaev. *Limit Theorems for Stochastic Processes. 2nd ed.*, volume 288 of *Grundlehren der Mathematischen Wissenschaften.* Berlin: Springer, 2003.

[Jentzen(2014)] A. Jentzen. *Numerical Analysis of Stochastic Partial Differential Equations.* Lecture notes, ETH Zurich. https://www.math.ethz.ch/education/bachelor/lectures/fs2014/math/numsol, 2014.

[Jentzen and Kloeden(2011)] A. Jentzen and P. E. Kloeden. *Taylor approximations for stochastic partial differential equations*, volume 83 of *CBMS-NSF Regional Conference Series in Applied Mathematics.* Society for Industrial and Applied

Mathematics (SIAM), Philadelphia, PA, 2011. ISBN 978-1-611972-00-9. doi: 10.1137/1.9781611972016. URL `http://dx.doi.org/10.1137/1.9781611972016`.

[Kallenberg(2002)] O. Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002. ISBN 0-387-95313-2.

[Kallsen(2009)] J. Kallsen. Option pricing. In *T. Andersen, R. Davis, J. Kreiss, and T. Mikosch, Handbook of Financial Time Series*, London Math. Soc. Lecture Note Ser., pages 599–613. Springer, Berlin, 2009.

[Karatzas and Shreve(1988)] I. Karatzas and S. Shreve. *Brownian Motion and Stochastic Calculus*. Berlin: Springer, 1988.

[Klenke(2008)] A. Klenke. *Probability theory*. Universitext. Springer-Verlag London Ltd., 2008.

[Kloeden and Neuenkirch(2007)] P. E. Kloeden and A. Neuenkirch. The pathwise convergence of approximation schemes for stochastic differential equations. *LMS J. Comput. Math.*, 10:235–253, 2007. ISSN 1461-1570.

[Kloeden and Platen(1992)] P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*, volume 23 of *Applications of Mathematics*. Berlin: Springer, 1992.

[Knuth(1998)] D. E. Knuth. *The Art of Computer Programming. Vol. 2: Seminumerical Algorithms. 3rd ed.* Bonn: Addison-Wesley, 1998.

[Kühn(2004)] C. Kühn. *Stochastische Analysis mit Finanzmathematik*. Lecture notes. http://www.math.uni-frankfurt.de, 2004.

[Kuo(2006)] H.-H. Kuo. *Introduction to Stochastic Integration*. Universitext. Springer, New York, 2006. ISBN 978-0387-28720-1; 0-387-28720-5.

[Lewis(2000)] A. L. Lewis. *Option valuation under stochastic volatility*. Finance Press, Newport Beach, CA, 2000. ISBN 0-9676372-0-1. With Mathematica code.

[Lotka(1920)] A. J. Lotka. Undamped oscillations derived from the law of mass action. *Journal of the American Chemical Society*, 42(8):1595–1599, 1920.

[Maruyama(1953)] G. Maruyama. Markov processes and stochastic equations. *Nat. Sci. Rep. Ochanomizu Univ.*, 4:40–43, 1953. ISSN 0029-8190.

[Maruyama(1955)] G. Maruyama. Continuous Markov processes and stochastic equations. *Rend. Circ. Mat. Palermo (2)*, 4:48–90, 1955. ISSN 0009-725X.

[Mattingly et al.(2002)Mattingly, Stuart, and Higham] J. C. Mattingly, A. M. Stuart, and D. J. Higham. Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stochastic Process. Appl.*, 101(2):185–232, 2002. ISSN 0304-4149. doi: 10.1016/S0304-4149(02)00150-3. URL `http://dx.doi.org/10.1016/S0304-4149(02)00150-3`.

[Merton(1973)] R. C. Merton. Theory of rational option pricing. *The Bell Journal of Economics and Management Science*, 4:141–183, 1973.

[Métivier(1982)] M. Métivier. *Semimartingales: A Course on Stochastic Processes*, volume 2 of *De Gruyter Studies in Mathematics*. Berlin - New York: Walter de Gruyter, 1982.

[Milstein(1974)] G. N. Milstein. Approximate integration of stochastic differential equations. *Teor. Verojatnost. i Primenen.*, 19:583–588, 1974. ISSN 0040-361x.

[Milstein(1995)] G. N. Milstein. *Numerical integration of stochastic differential equations*, volume 313 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1995. ISBN 0-7923-3213-X. Translated and revised from the 1988 Russian original.

[Milstein and Tretyakov(2004)] G. N. Milstein and M. V. Tretyakov. *Stochastic numerics for mathematical physics*. Scientific Computation. Springer-Verlag, Berlin, 2004. ISBN 3-540-21110-1.

[Milstein and Tretyakov(2005)] G. N. Milstein and M. V. Tretyakov. Numerical integration of stochastic differential equations with nonglobally Lipschitz coefficients. *SIAM J. Numer. Anal.*, 43(3):1139–1154 (electronic), 2005. ISSN 0036-1429. doi: 10.1137/040612026. URL `http://dx.doi.org/10.1137/040612026`.

[Milstein et al.(1998)Milstein, Platen, and Schurz] G. N. Milstein, E. Platen, and H. Schurz. Balanced implicit methods for stiff stochastic systems. *SIAM J. Numer. Anal.*, 35(3):1010–1019 (electronic), 1998. ISSN 0036-1429. doi: 10.1137/S0036142994273525. URL `http://dx.doi.org/10.1137/S0036142994273525`.

[Mörters and Peres(2010)] P. Mörters and Y. Peres. *Brownian motion*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2010. ISBN 978-0-521-76018-8. With an appendix by Oded Schramm and Wendelin Werner.

[Müller-Gronbach(2002)] T. Müller-Gronbach. The optimal uniform approximation of systems of stochastic differential equations. *Ann. Appl. Probab.*, 12(2):664–690, 2002. ISSN 1050-5164. doi: 10.1214/aoap/1026915620. URL `http://dx.doi.org/10.1214/aoap/1026915620`.

[Müller-Gronbach et al.(2012)Müller-Gronbach, Novak, and Ritter] T. Müller-Gronbach, E. Novak, and K. Ritter. *Monte Carlo-Algorithmen*. Springer-Lehrbuch. [Springer Textbook]. Heidelberg: Springer, 2012.

[Øksendal(2003)] B. Øksendal. *Stochastic Differential Equations. An Introduction with Applications. 6th ed.* Universitext. Berlin: Springer, 2003.

[Prévôt and Röckner(2007)] C. Prévôt and M. Röckner. *A concise course on stochastic partial differential equations*, volume 1905 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007. ISBN 978-3-540-70780-6; 3-540-70780-8.

[Protter(2004)] P. E. Protter. *Stochastic Integration and Differential Equations. 2nd ed.*, volume 21 of *Applications of Mathematics*. Berlin: Springer, 2004.

[Revuz and Yor(1999)] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion. 3rd ed.*, volume 293 of *Grundlehren der Mathematischen Wissenschaften.* Berlin: Springer-Verlag, 1999.

[Roberts and Tweedie(1996)] G. O. Roberts and R. L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4): 341–363, 1996. ISSN 1350-7265. doi: 10.2307/3318418. URL http://dx.doi.org/10.2307/3318418.

[Ross(2006)] S. M. Ross. *Simulation. 4th revised edition.* Amsterdam: Elsevier/Academic Press, 2006.

[Rößler(2010)] A. Rößler. Strong and weak approximation methods for stochastic differential equations—some recent developments. In *Recent developments in applied probability and statistics*, pages 127–153. Physica, Heidelberg, 2010. doi: 10.1007/978-3-7908-2598-5_6. URL http://dx.doi.org/10.1007/978-3-7908-2598-5_6.

[Szpruch et al.(2011)Szpruch, Mao, Higham, and Pan] L. Szpruch, X. Mao, D. J. Higham, and J. Pan. Numerical simulation of a strongly nonlinear Ait-Sahalia-type interest rate model. *BIT*, 51(2):405–425, 2011. ISSN 0006-3835. doi: 10.1007/s10543-010-0288-y. URL http://dx.doi.org/10.1007/s10543-010-0288-y.

[Volterra(1926)] V. Volterra. Fluctuations in the abundance of a species considered mathematically. *Nature*, 118(2972):558–560, 1926.

[Wakolbinger(2004)] A. Wakolbinger. *Stochastic Processes.* Course notes. http://ismi.math.uni-frankfurt.de/wakolbinger/teaching/, 2004.

[Wichura(2001)] M. J. Wichura. Statistics 304: Autumn 2001 Distribution Theory Lecture Notes. *Lecture Notes*, 2001. http://galton.uchicago.edu/ wichura/Stat304/handouts.html.