universität wien

DIPLOMARBEIT

Titel der Diplomarbeit

Numerical solution of stochastic differential equations

angestrebter akademischer Grad

Magister der Naturwissenschaften (Mag. rer. nat.)

Verfasser:	Johannes Temme
Matrikel-Nummer:	0306409
Studienrichtung:	A 405 Mathematik
Betreuer:	O. Univ.-Prof. Dr. Arnold Neumaier

Wien, im September 2009

Abstract

This work presents the mathematical aspects of numerical solution of stochastic differential equations by means of time discrete approximations, i.e., one-step methods, and Gaussian approximation, also known as mean field approximation. It is intended for both a theoretical as well as an applied and practical introduction to the wide field of numerical solution of stochastic differential equations.

Chapter 1 gives an introduction to the theory of stochastic differential equations from the very beginning. In Section 1.1 we start with a brief summary of central results of probability and measure theory that we shall need – among others – in Section 1.2 and Section 1.3 to discuss Wiener processes and Diffusion processes. Section 1.4 introduces the stochastic integral by means of the Itô integral and treats the Stratonovich integral at the end of the section. Furthermore, we state a multi-dimensional version of the famous Itô–Doeblin formula. The theory of Itô differential equations and their main properties are handled in Section 1.5. The connections between Itô differential equations and diffusion processes as well as Stratonovich differential equations are also covered.

Chapter 2 starts with a short overview of various approaches of numerical solution to stochastic differential equations. The numerical schemes that we treat are based on Itô SDEs, but using the theory of Section 1.5 of Chapter 1 they can be easily extended to Stratonovich SDEs. Furthermore, we only discuss weak approximations as opposed to strong approximations of Itô SDEs. In Section 2.2 we introduce time discrete approximations based on the truncation of the Itô–Taylor expansion. We treat the calculus of multiple Itô integrals as well as a general convergence theorem for Itô–Taylor schemes. The numerical schemes that we use are the Euler–Maruyama scheme, the simplified Euler–Maruyama scheme, the order 2.0 weak Taylor scheme and the order 3.0 weak Taylor scheme. The Gaussian approximation, which is another approach to weak approximations of SDEs, is treated in Section 2.3. We discuss the Gaussian approximation in terms of expolynomial SDEs and formulate the Theorem of Wick and Isserlis, that we shall apply frequently in Chapter 3.

In **Chapter 3** we first discuss the adequate MATLAB implementation of the time discrete approximations of Chapter 2 in terms of random number generators. In Section

3.2 we compare for four SDEs the results of the time discrete approximations to the Gaussian approximation. Each SDE is different in structure and contains either one- or two-dimensional noise, multiplicative or additive noise and assumes values in $\mathbb{R}$ or $\mathbb{R}^2$.
The **Appendix** contains the MATLAB program headers of each program that is used in Chapter 3.
This thesis is submitted with all MATLAB programs that are mentioned in the Appendix.

Deutsche Zusammenfassung

Diese Arbeit ist sowohl als theoretische als auch praktische Einführung in das Gebiet der Numerik von stochastischen Differentialgleichungen gedacht, wobei die numerische Methoden dieser Arbeit Einschrittverfahren und die Gaußsche Näherung umfassen.
Im **1. Kapitel** werden notwendige Grundlagen als auch die Theorie der stochastischen Differentialgleichungen behandelt. In Abschnitt 1.1 besprechen wir kurz die wichtigsten Resultate und Sätze aus Wahrscheinlichkeits- und Maßtheorie, die wir im Weiteren unter anderem in Abschnitt 1.2 und 1.3 bei der Behandlung von Wiener- und Diffusionsprozessen benötigen. Der Konstruktion des stochastischen Integrals an Hand des Itô Integrals ist Abschnitt 1.4 gewidmet. Weiters behandeln wir in diesem Abschnitt die Itô–Doeblin Formel in einer multi-dimensionalen Form und besprechen kurz die Definition des Stratonovich Integrals. Abschnitt 1.5 umfasst eine Einführung in die Theorie von stochastischen Itô Differentialgleichungen. Auf die Zusammenhänge zwischen Itô Differentialgleichungen und Diffusionsprozessen als auch Stratonovich Differentialgleichungen wird ebenfalls eingegangen.
Das **2. Kapitel** beginnt mit einem Überblick über verschiedene numerische Methoden zum Lösen von stochastischen Differentialgleichungen. Die numerischen Methoden, die wir in dieser Arbeit behandeln, sind auf Itô Differentialgleichungen zugeschnitten, mit dem Wissen aus Abschnitt 1.5 aus Kapitel 1 können sie jedoch auch problemlos auf Stratonovich Differentialgleichungen umgewandelt werden. Weiters beschränken wir uns bei der Entwicklung von numerischen Methoden auf schwache Näherungen im Gegensatz zu starken Näherungen. In Abschnitt 2.2 behandeln wir Einschrittverfahren die auf dem Abbruch der Itô–Taylor Entwicklung basieren. Wir führen mehrfache Itô Integrale ein und formulieren einen allgemeinen Hauptsatz über die Konvergenz von Itô–Taylor Näherungen. In dieser Arbeit verwenden wir das Euler–Maruyama Verfahren, das vereinfachte Euler–Maruyama Verfahren, das order 2.0 weak Taylor Verfahren und das order 3.0 weak Taylor Verfahren. Ein anderer Zugang zum numerischen Lösen von stochastischen Differentialgleichungen ist die Gaußsche Näherung, die wir in Abschnitt 2.3 diskutieren werden. Wir behandeln die Gaußsche Näherung für stochastische Differentialgleichungen

mit expolynomialen Koeffizienten und formulieren den Satz von Wick und Isserlis, welchen wir in den Beispielen von Kapitel 3 oft verwenden werden.
Kapitel 3 ist der MATLAB Implementierung und dem Vergleich zwischen Einschrittverfahren aus dem 2. Kapitel und der Gaußschen Näherung gewidmet. Zuerst besprechen wir MATLAB Zufallszahlengeneratoren, die für die Einschrittverfahren notwendig sind. Im Weiteren vergleichen wir die Ergebnisse der Einschrittverfahren mit jenen der Gaußschen Näherung anhand von 4 stochastischen Differentialgleichungen, die sich alle in ihrer Form unterscheiden: Sie enthalten entweder ein- oder zweidimensionales Rauschen, multiplikatives oder additives Rauschen und nehmen Werte entweder in $\mathbb{R}$ oder $\mathbb{R}^2$ an.
Der **Appendix** umfasst die MATLAB Programmköpfe all jener Programme, die wir in Kapitel 3 verwenden.
Diese Arbeit wird mit einer Multimediabeilage eingereicht, die alle im Appendix genannten MATLAB Programme beinhaltet.

Danksagung

An dieser Stelle möchte ich mich herzlich bei Prof. Arnold Neumaier für seine bereitwillige Unterstützung, seine Vorschläge, seine Ideen und Literaturhinweise bedanken. Prof. Hermann Schichl danke ich für seine Ratschläge und Hinweise zur Diplomarbeit.

Ich danke Prof. Peter Kloeden und Eckhard Platen für ihre Hilfe zur Portierung des order 3.0 weak Taylor Verfahrens (2.2.40) vom skalaren in den n-dimensionalen Fall. Mathias Beiglböck danke ich für seine hilfreichen Erklärungen zur Definition der bedingten Wahrscheinlichkeit (1.1.9). Ich danke Prof. Walter Schachermayer für seine Hinweise zur Umwandlung des Euler–Maruyama Verfahrens (2.2.2) in das vereinfachte Euler–Maruyama Verfahren (2.2.7).

Ein besonderer Dank gilt meinen Eltern Uschi und Peter sowie meinen Geschwistern Anne, Inger, Liza, Moritz, Sören und meiner Familie für ihre jahrelange Unterstützung und dafür, dass sie immer an mich geglaubt haben. Meinen Freunden danke ich für die notwendige Zerstreuung und Ablenkung während des Schreibens.
Meinen Katzen danke ich ausdrücklich dafür, dass sie nicht immer auf der Tastatur schliefen als der Computer rechnete.

Meiner Freundin Ada möchte ich für ihre Unterstützung, ihre Ideen, ihre Geduld, ihre Aufmunterungen und dafür danken, dass mich jeder Tag im Leben mit ihr so glücklich macht.

Contents

List of frequently used notation and symbols

$\mathbb{N}$ ..... the set of natural numbers, $\mathbb{N} = \{0, 1, 2, \dots\}$

$\mathbb{P}$ ..... a probability measure, see p. 2

$\mathbb{P}(t, B; s, x)$ ..... the transition probabilities, see p. 19

$\mathbb{R}^n$ ..... the n-dimensional Euclidean space

$\mathcal{B}(\mathbb{R}^n)$ ..... the Borel σ-algebra of $\mathbb{R}^n$, see p. 3

$\mathcal{P}(\mathbb{R}^n, \mathbb{R})$ ..... the set of all polynomial functions $g : \mathbb{R}^n \to \mathbb{R}$

$\mathcal{P}_{ex}(\mathbb{R}^n, \mathbb{R})$ ..... the set of all expolynomials $g : \mathbb{R}^n \to \mathbb{R}$, see p. 70

$\mathbb{P}(A|Y = s)$ ..... the probability of a set A conditioned on the event $\{Y = s\}$ for a random variable Y, see p. 12

$\Delta(t, \omega) \circ dW_t(\omega)$ ..... the Stratonovich integral of Δ, see p. 22

$\Delta(t, \omega) dW_t(\omega)$ ..... the Itô integral of Δ, see p. 22

ΔW_ℓ^j ..... the ℓth increment of the jth component of a Wiener process $(W_t)_t$, see p. 50

Δ_ℓ ..... the ℓth stepsize of a time discretization, see p. 49

$\det(A)$ ..... the determinant of a matrix A

$\emptyset$ ..... the empty set

$(\mathcal{F}_t)_t$ ..... the filtration associated to $(W_t)_t$, see p. 18

Γ_β ..... the set of all multi-indices of length less or equal to β, see p. 57

$\mathbb{P}_{X;t_1,\dots,t_k}$ ..... the finite dimensional distributions of a stochastic process $(X)_t$, see p. 10

$\mathbb{P}_X$ ..... the probability measure induced by a random variable X, see p. 3

$[a, b]$ ..... the closed interval $a \leq x \leq b$ in $\mathbb{R}$

(a, b) ..... the open interval $a < x < b$ in $\mathbb{R}$

$\mathcal{I}_\alpha [f(\cdot)]_{t_0,t}$ ..... the multiple Itô integral based on a multi-index α of a stochastic process f, see p. 58

$\mathcal{I}[\Delta]$ ..... the Itô integral of Δ, see p. 23

$L^p(\Omega)$ ..... the space of L^p-integrable functions $f : \Omega \to \mathbb{R}^n$, see p. 6

$\mathbf{1}$ ..... the function $\mathbf{1} : \Omega \to \mathbb{R}^n$, $\mathbf{1}(\omega) = (1, \ldots, 1)$

$\mathcal{C}(\mathbb{R}^m, \mathbb{R}^n)$ ..... the space of continuous functions $f : \mathbb{R}^m \to \mathbb{R}^n$

$\mathcal{C}^l(\mathbb{R}^m, \mathbb{R}^n)$ ..... the space of l times continuously differentiable functions $f : \mathbb{R}^m \to \mathbb{R}^n$

$\mathcal{C}^l_P(\mathbb{R}^m, \mathbb{R})$ ..... the space of l times continuously differentiable functions $f : \mathbb{R}^m \to \mathbb{R}$ with polynomial growth, see p. 51

$\mathcal{M}$ ..... the set of all multi-indices, see p. 56

$\mathcal{R}(\mathcal{N})$ ..... the remainder set of a hierarchical set $\mathcal{N}$, see p. 57

$\mathbb{1}_A$ ..... the indicator function of a set A, see p. 11

∇_x ..... the gradient with respect to the vector $x \in \mathbb{R}^n$: $\nabla_x f = (\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n})$

$\|x\|_2$ ..... the Euclidean norm of a vector $x \in \mathbb{R}^n$, see p. 3

∂^α ..... the partial derivative w.r.t a multi-index $\alpha = (\alpha_1, \ldots, \alpha_k)$, i.e., $\partial^\alpha = \frac{\partial^k}{\partial x_{\alpha_1} \cdots \partial x_{\alpha_k}}$

$[X, Y](T, \omega)$ ..... the cross variation of the stochastic processes $(X_t)_t$ and $(Y_t)_t$, see p. 28

$[X, X](T, \omega)$ ..... the quadratic variation of the stochastic process $(X_t)_t$, see p. 28

$(s_n)_n \downarrow x; (s_n)_n \searrow x$ .. the series $(s_n)_n$ converges to x with $x \leq s_n$

$(s_n)_n \to x$ ..... the series $(s_n)_n$ converges to x

$(s_n)_n \uparrow x; (s_n)_n \nearrow x$ .. the series $(s_n)_n$ converges to x with $s_n \leq x$

$(W_t)_t$ ..... a Brownian motion\Wiener process, see p. 15

$A = (a^{i,j})_{i,j}$ ..... a matrix A with ijth component $a^{i,j}$

$A^\top$ ..... the transpose of a matrix A

f_α ..... the Itô coefficient function of a function f, see p. 59

i ..... the imaginary unit $i = \sqrt{-1}$

I_n ..... the n-dimensional identity matrix

$l(\alpha)$ ..... the length of a multi-index α, see p. 56

L^i ..... differential operator associated to an Itô SDE for $i = 0, 1, \ldots$, see p. 55

$n(\alpha)$ ..... the number of components of a multi-index α which are equal to 0, see p. 56

n_t ..... the maximum index n of a time discretization $(\tau)_\delta$ such that $\tau_n \leq t$, see p. 49

$p(\mathbf{t}'; \mathbf{z}|\mathbf{t}; \mathbf{x})$ ..... the transition density functions, see p. 12

$p_{X;t_1,\ldots,t_k}$ ..... densities of $F_{X;t_1,\ldots,t_k}$, see p. 10

v ..... the multi-index of length 0, see p. 56

$x = (x^1, \ldots, x^n)$ ..... a vector $x \in \mathbb{R}^n$ with ith component x^i for $i = 1, \ldots, n$

$x^\top$ ..... the transpose of a vector $x \in \mathbb{R}^n$
$\mathbb{C}\mathrm{ov}[X;Y]$ ..... the covariance matrix of two random variables X and Y, see p. 4
$\mathbb{C}\mathrm{ov}[X]$ ..... the covariance matrix of a random variable X, see p. 4
$F_{X;t_1,\dots,t_k}$ ..... the joint distribution functions of a stochastic process $(X_t)_t$, see p. 10
F_X ..... the distribution function of a real-valued random variable X, see p. 3
$\mathbb{E}[X|\mathcal{G}]$ ..... the expectation of a random variable X conditioned on $\mathcal{G}$, see p. 7
$\mathbb{E}_C[\cdot]$ ..... see p. 70
$\mathbb{E}[X]$ ..... the expectation of a random variable X, see p. 3
$\mathcal{O}(\Delta^p)$ ..... expression divided by Δ^p remains bounded as $\Delta \to 0$
$\|X\|_{L^p}$ ..... the L^p-norm of a random variable X, see p. 6
$(\Omega, \mathcal{F}, \mathbb{P})$ ..... a probability space, see p. 2
$(\tau)_\delta$ ..... a time discretization with maximum step size δ, see p. 49
$\mathbb{V}\mathrm{ar}[X]$ ..... the variance of a real-valued random variable X, see p. 4
a.s. ..... almost surely, see p. 2
iff ..... if and only if
w.p.1 ..... with probability one, see p. 2

Note that vectors and matrices are indexed with superscripts and parantheses will be used when taking powers of their components. For example with $(x^i)^2$ we denote the square of x^i, the ith component of a vector $x \in \mathbb{R}^n$.

Chapter 1

Stochastic calculus

In this chapter we state the theory of stochastic differential equations from the very beginning. We start by summarizing main results of probability and measure theory that we shall need to define Wiener processes, diffusion processes and stochastic integrals. We construct the stochastic integral by means of the Itô integral and briefly summarize the construction of the Stratonovich integral. We also formulate the Itô–Doeblin formula in its multi-dimensional version, since we shall frequently use it in the following chapters.
In the last section we summarize the main results of the theory of stochastic differential equations (SDE): We formulate an existence and uniqueness theorem, discuss connections between solutions and diffusion processes and the relation between Itô and Stratonovich SDEs.
Throughout this introduction, we mainly follow Chapter 2, 3 and 5 of Øksendal (2006) and Chapter 2, 3, 4 of Kloeden & Platen (2000).

1.1 Probability spaces, stochastic processes

In this section we state the main results and definitions of probability and measure theory that we shall need throughout this exposition.

Definition 1.1.1. Let Ω be a set. We call a subset $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ of the power set $\mathcal{P}(\Omega)$ of Ω a **σ-algebra** of Ω if

$$\begin{aligned} \Omega &\in \mathcal{F}, \\ \Omega \backslash A =: A^c &\in \mathcal{F} \quad \text{if } A \in \mathcal{F}, \\ \bigcup_{n=1}^{\infty} A_n &\in \mathcal{F} \quad \text{if } A_1, A_2, \ldots, A_n, \ldots \in \mathcal{F}. \end{aligned}$$

The pair $(\Omega, \mathcal{F})$ is called a **measurable space**.

Let $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ be two measurable spaces, then we define the **product σ-algebra** $\mathcal{F} \otimes \mathcal{F}'$ on $\Omega \times \Omega'$ to be the σ-algebra generated by the sets $A \times A'$ for all $A \in \mathcal{F}$ and $A' \in \mathcal{F}'$.
We call a function $\mu : \mathcal{F} \to \mathbb{R}$ a **signed measure** if

$$\mu(\emptyset) = 0, \tag{1.1.1}$$

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n) \tag{1.1.2}$$

for any sequences $A_1, A_2, \ldots, A_n, \ldots \in \mathcal{F}$ with $A_i \cap A_j = \emptyset$ for $i \neq j$ where $i, j \in \{1, 2, 3, \ldots\}$. Le $\mathbb{P}$ be a signed measure $\mathbb{P} : \mathcal{F} \to [0, 1]$ and additionally $\mathbb{P}(\Omega) = 1$. Then we call this measure a **probability measure**.
If $\mathbb{P}$ is a probability measure on the measurable space $(\Omega, \mathcal{F})$, we call the triple $(\Omega, \mathcal{F}, \mathbb{P})$ a **probability space**.

Remark 1.1.2. The definition of a σ-algebra $\mathcal{F}$ implies that $\emptyset \in \mathcal{F}$ and $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$ whenever $A_1, A_2, \ldots, A_n, \ldots \in \mathcal{F}$.

We call the elements of a σ-algebra **measurable sets** and **possbile events**. We call a set $A \in \mathcal{F}$ a set of **measure zero** if $\mathbb{P}(A) = 0$. If for $A \in \mathcal{F}$ we have $\mathbb{P}(A) = 1$ we say that the event A happens **almost surely** (a.s.) or **with probability one** (w.p.1). More generally we say that a statement s, i.e., a function $s : \Omega \to \{\text{true}, \text{false}\}$ is "true *almost surely*" or *with probability one*, if $s^{-1}(\{\text{false}\})$ is a set of measure zero. For example, for two given functions $X, Y : \Omega \to \Omega'$ we define s as

$$s(\omega) := \begin{cases} \text{false}; & \text{if } X(\omega) \neq Y(\omega), \\ \text{true}; & \text{if } X(\omega) = Y(\omega). \end{cases}$$

Then we say that $s = \text{true}$, i.e., $X = Y$ *almost surely* or *with probability one* if $s^{-1}(\{\text{false}\}) \in \mathcal{F}$ and if $s^{-1}(\{\text{false}\})$ is a set of measure zero.

Definition 1.1.3 (cf. ØKSENDAL (2006, Definition 2.1.3))**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a measurable space. We call two sets $A, B \in \mathcal{F}$ **independent** if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B).$$

Let $\mathcal{A} = \{\mathcal{H}_i \,|\, i \in I\}$ be a collection of families $\mathcal{H}_i$ of measurable sets. We call $\mathcal{A}$ **independent** if

$$\mathbb{P}(H_{i_1} \cap H_{i_2} \cap \ldots \cap H_{i_k}) = \mathbb{P}(H_{i_1}) \cdot \mathbb{P}(H_{i_2}) \cdots \mathbb{P}(H_{i_k})$$

for all choices $H_{i_1} \in \mathcal{H}_{i_1}, H_{i_2} \in \mathcal{H}_{i_2}, \ldots, H_{i_k} \in \mathcal{H}_{i_k}$ for all indices $i_1, \ldots, i_k \in I$ and $k \in \mathbb{N}$.

1.1.1 Random variable, expectation

In this subsection we introduce random variables, the expectation of a random variable and the covariance matrix of two random variables. We also formulate main results of probability and measure theory: The Theorem of Radon–Nikodým, the Theorem of Fubini, the Lemma of Doob–Dynkin and the strong law of large numbers.

Given any family $\mathcal{U} \in \mathcal{P}(\Omega)$, we denote by $\mathcal{H}_U$ the smallest σ-algebra containing $\mathcal{U}$, i.e.,

$$\mathcal{H}_U := \bigcap \left\{ \mathcal{H} \,\middle|\, \mathcal{H} \text{ } \sigma\text{-algebra of } \Omega,\ \mathcal{U} \subseteq \mathcal{H} \right\}.$$

If $\mathcal{U}$ is the collection of all open subsets of a topological space Ω, we write $\mathcal{B}(\Omega) := \mathcal{H}_U$ and call $\mathcal{B}(\Omega)$ the **Borel σ-algebra** on Ω.
If $(\Omega, \mathcal{F}, \mathbb{P})$ is a given probability space, then a function $Y : \Omega \to \mathbb{R}^n$ is called $\mathcal{F} : \mathcal{B}(\mathbb{R}^n)$-**measurable** or short **$\mathcal{F}$-measurable** if

$$Y^{-1}(U) := \left\{ \omega \in \Omega \,\middle|\, Y(\omega) \in U \right\} \in \mathcal{F}; \quad \forall U \in \mathcal{B}(\mathbb{R}^n).$$

For a vector $x = (x^1, \ldots, x^n) \in \mathbb{R}^n$ we denote by $\|x\|_2 := \sqrt{(x^1)^2 + \ldots + (x^n)^2}$ the **euclidean norm** of x.

Definition 1.1.4 (cf. Øksendal (2006, p. 9))**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a given probability space. A **random variable** X is a $\mathcal{F}$-measurable function $X : \Omega \to \mathbb{R}^n$. We define by

$$\mathbb{P}_X : \mathcal{B}(\mathbb{R}^n) \to \mathbb{R}^n, \quad \mathbb{P}_X(B) := \mathbb{P}(X^{-1}(B))$$

the **probability measure $\mathbb{P}_X$ induced by** X and also call it the **distribution** of X. If X takes values in $\mathbb{R}$, the function

$$F_X : \mathbb{R} \to [0,1], \quad F_X(t) := \mathbb{P}_X\left((-\infty, t]\right)$$

is called the **distribution function** of X. If $\int_\Omega \|X(\omega)\|_2 \, d\mathbb{P}(\omega) < \infty$ then the real vector

$$\mathbb{E}[X] := \int_\Omega X(\omega) d\mathbb{P}(\omega) = \int_{\mathbb{R}^n} x d\mathbb{P}_X(x)$$

is called the **expectation** of X with respect to $\mathbb{P}$.
If $f : \mathbb{R}^n \to \mathbb{R}$ is a Borel-measurable function, i.e., f is $\mathcal{B}(\mathbb{R}^n) : \mathcal{B}(\mathbb{R})$-measurable and if

$\int_\Omega |f(X(\omega))| d\mathbb{P}(\omega) < \infty$, we define

$$\mathbb{E}[f(X)] := \int_\Omega f(X(\omega)) d\mathbb{P}(\omega) = \int_{\mathbb{R}^n} f(x) d\mathbb{P}_X(x).$$

Remark 1.1.5. Due to the definition of a Lebesgue integral, the expectation is a linear operator, i.e., for two random variables X, Y and real numbers α and β we have

$$\mathbb{E}[\alpha X + \beta Y] = \alpha \mathbb{E}[X] + \beta \mathbb{E}[Y].$$

Similarly, we find

$$\mathbb{E}[X] \leq \mathbb{E}[Y],$$

given that $X \leq Y$ w.p.1.

Definition 1.1.6. Let $X, Y : \Omega \to \mathbb{R}^n$ be two random variables, then we define the **covariance matrix** $\mathbb{C}\mathrm{ov}[X;Y] = (c^{i,j})_{i,j}$ **of** X **and** Y by

$$c^{i,j} := \mathbb{E}\left[X^i Y^j\right] - \mathbb{E}\left[X^i\right] \mathbb{E}\left[Y^j\right]$$

for $i, j = 1, \ldots, n$, where $X = (X^1, \ldots, X^n), Y = (Y^1, \ldots, Y^n)$.
We define the **covariance matrix** $\mathbb{C}\mathrm{ov}[X]$ **of** X by $\mathbb{C}\mathrm{ov}[X] := \mathbb{C}\mathrm{ov}[X;X]$. If X assumes values in $\mathbb{R}$ we define by $\mathbb{V}\mathrm{ar}[X] := \mathbb{C}\mathrm{ov}[X]$ the **variance of** X.

Theorem 1.1.7 (**Radon–Nikodým**). *Let $(\Omega, \mathcal{F})$ be a measurable space. Let $\mathbb{P}$ and $\tilde{\mathbb{P}}$ be two probability measures defined on $\mathcal{F}$, and let $\mathbb{P}$ be absolutely continuous with respect to $\tilde{\mathbb{P}}$, i.e.,*

$$\forall A \in \mathcal{F}: \quad \tilde{\mathbb{P}}(A) = 0 \Longrightarrow \mathbb{P}(A) = 0.$$

Then there exists an $\tilde{\mathbb{P}}$-a.s. positive and $\tilde{\mathbb{P}}$-a.s. unique $\mathcal{F}$-measurable random variable Z called the **Radon–Nikodým derivative** *such that*

$$\mathbb{P}(A) = \int_A Z(\omega) d\tilde{\mathbb{P}}(\omega) \quad \textit{for every } A \in \mathcal{F}.$$

Proof. See Kallenberg (2002, Theorem 2.10). □

The following theorem gives us the possibility to evaluate integrals over a product space $\Omega \times [a, b]$, $a, b \in \mathbb{R}$:

Theorem 1.1.8 (**Fubini**). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $[a, b]$ be a compact interval in $\mathbb{R}$ and let λ denote the Lebesgue measure on $\mathcal{B}([a, b])$. Then there exists a unique*

measure $\mathbb{P} \otimes \lambda$ *on* $(\Omega \times [a,b], \mathcal{F} \otimes \mathcal{B}([a,b]))$ *satisfying*

$$(\mathbb{P} \otimes \lambda)(A \times B) = \mathbb{P}(A) \cdot \lambda(B)$$

for all $A \in \mathcal{F}$ *and* $B \in \mathcal{B}([a,b])$*. Furthermore, for any measurable function* $f : \Omega \times [a,b] \to [0,\infty)$*,*

$$\begin{aligned}\int_{\Omega\times[a,b]} f(\omega,x)d(\mathbb{P}\otimes\lambda)(\omega,x) &= \int_\Omega \left(\int_{[a,b]} f(\cdot,x)d\lambda(x)\right)(\omega)d\mathbb{P}(\omega)\\ &= \int_{[a,b]} \left(\int_\Omega f(\omega,\cdot)d\mathbb{P}(\omega))\right)(x)d\lambda(x).\end{aligned}$$

The last relation remains valid for functions $f : \Omega \times [a,b] \to \mathbb{R}$ *satisfying*

$$\int_{\Omega\times[a,b]} |f(\omega,x)|d(\mathbb{P}\otimes\lambda)(\omega,x) < \infty. \tag{1.1.3}$$

Proof. See Kallenberg (2002, Theorem 1.27). □

Given an arbitrary function $X : \Omega \to \mathbb{R}^n$ we denote by $\mathcal{H}_X$ the smallest σ-algebra containing all the sets $\left\{X^{-1}(U) \,\middle|\, U \in \mathcal{B}(\mathbb{R}^n)\right\}$. It is easy to prove that X is $\mathcal{H}_X$-measurable and $\mathcal{H}_X$ is the smallest σ-algebra such that X is measurable.

Definition 1.1.9. Let X and Y be two random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G}$ a sub-σ-algebra of $\mathcal{F}$. We call X and Y **independent** if the collection $\mathcal{A} = \left\{\mathcal{H}_X, \mathcal{H}_Y\right\}$ is independent. We call X and $\mathcal{G}$ **independent** if the collection $\mathcal{A}' = \left\{\mathcal{H}_X, \mathcal{G}\right\}$ is independent.

Remark 1.1.10. Let X and Y be independent random variables with $\mathbb{E}\left[\|X\|_2\right] < \infty$ and $\mathbb{E}\left[\|Y\|_2\right] < \infty$. Then

$$\mathbb{E}\left[X^\top Y\right] = \mathbb{E}\left[X\right]^\top \mathbb{E}\left[Y\right].$$

The following Lemma shall be necessary later to define conditional probabilities and expectations:

Lemma 1.1.11. (**Doob–Dynkin**, cf. Øksendal (2006, Lemma 2.1.2)) If $X, Y : \Omega \to \mathbb{R}^n$ are two given functions, then Y is $\mathcal{H}_X$-measurable iff there exists a Borel measurable function $h : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$Y = h(X).$$

Proof. See Rao (1977, Proposition 3). □

We finally state an important result that we shall use in Chapter 2 to approximate the expectation operate $\mathbb{E}[\cdot]$:

Theorem 1.1.12. (**strong law of large numbers**, cf. KALLENBERG (2002, Theorem 4.23) Let $X_1, X_2, \ldots$ be independent and identically distributed random variables, i.e., they have the same distribution $\mathbb{P}_{X_1}$ and thus the same moments $\mathbb{E}[f(X_1)]$. Further, let $\mathbb{E}[\|X_1\|_2] < \infty$. Then

$$\frac{1}{n}\sum_{i=1}^{n} X_i$$

converges a.s. to $\mathbb{E}[X_1]$ as $n \to \infty$.

Proof. See KALLENBERG (2002, pp. 73–74). □

1.1.2 L^p-spaces, conditional expectation

In this subsection we introduce L^p-spaces and the L^p-norm that we shall use frequently. The notion of the conditional expectation of a random variable gives us a stochastic tool to approximate a given random variable.

Definition 1.1.13 (cf. ØKSENDAL (2006, p. 9))**.** Let $(\Omega, \mathcal{F})$ be a measurable space and $\mu : \mathcal{F} \to [0, \infty)$ be a signed measure on $\mathcal{F}$. Further let $X : \Omega \to \mathbb{R}^n$ be a $\mathcal{F}$-measurable function and let $p \in [1, \infty)$ be a constant. We define the L^p**-norm** of X, $\|X\|_{L^p}$, by

$$\|X\|_{L^p} := \|X\|_{L^p(\Omega)} := \left(\int_\Omega \|X(\omega)\|_2^p \, d\mu(\omega)\right)^{\frac{1}{p}}.$$

If $p = \infty$, we set

$$\|X\|_\infty := \|X\|_{L^\infty(\Omega)} := \inf\left\{M \in \mathbb{R} \,\middle|\, \|X(\omega)\|_2 \leq M \text{ a.s.}\right\}.$$

We define the corresponding L^p-spaces for $p \in [1, \infty]$ on Ω by

$$L^p(\Omega) := \left\{X : \Omega \to \mathbb{R}^n, X \text{ is } \mathcal{F}\text{-measurable} \,\middle|\, \|X\|_{L^p} < \infty\right\}.$$

Remark 1.1.14. For fixed p the pair $(L^p, \|.\|_{L^p})$ is a Banach space, i.e., a complete normed linear space. $(L^2, \|.\|_{L^2})$ is even a Hilbert space with inner product

$$\langle X; Y\rangle_{L^2(\Omega)} := \mathbb{E}\left[X^\top Y\right].$$

We call the elements X of $L^1(\Omega)$ **integrable** random variables.

Given a $\mathcal{F}$-measurable random variable $X : \Omega \to \mathbb{R}^n$ we can consider $\mathcal{F}$ or even $\mathcal{H}_X \subseteq \mathcal{F}$ to contain all the information needed to evaluate X. If we are given a sub-

σ-algebra $\mathcal{G} \subseteq \mathcal{F}$ we then want to find a random variable Y that is an appropriate approximation to X:

Definition 1.1.15 (Kloeden & Platen (2000, p. 60))**.** Let $X : \Omega \to \mathbb{R}^n$ be a random variable on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $X \in L^1(\Omega)$. Let $\mathcal{G} \subseteq \mathcal{F}$ be a sub-σ-algebra, thus representing a coarser profile of information than is available in $\mathcal{F}$. We call the **conditional expectation** of X with respect to $\mathcal{G}$ or the **expectation conditioned on** $\mathcal{G}$, which we denote by $\mathbb{E}[X|\mathcal{G}]$, as any $\mathcal{G}$-measurable random variable Y satisfying

$$\int_G Y d\mathbb{P} = \int_G X d\mathbb{P}; \quad \text{w.p.1, for all } G \in \mathcal{G}. \tag{1.1.4}$$

Remark 1.1.16.

(i) We again note that $\mathbb{E}[X|\mathcal{G}]$ is a $\mathcal{G}$-measurable random variable.

(ii) The existence and uniqueness (w.p.1) of $Y = \mathbb{E}[X|\mathcal{G}]$ are guaranteed by the Radon–Nikodým Theorem 1.1.7.

We now consider special cases for $\mathcal{G}$:
Let $\mathcal{G} = \{\Omega, \emptyset\}$. $\mathbb{E}[X|\mathcal{G}]$ has to be $\mathcal{G}$-measurable, i.e., it has to be constant: $\mathbb{E}[X|\mathcal{G}] = c \in \mathbb{R}^n$. $\mathbb{E}[X|\mathcal{G}]$ also has to fulfill (w.p.1) the equation

$$c = \int_\Omega \mathbb{E}[X|\mathcal{G}]\, d\mathbb{P} = \int_\Omega X d\mathbb{P}.$$

By Definition 1.1.4 we conclude that $c = \mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$.
Let now $\mathcal{G} = \mathcal{H}_X$. The random variable X is of course $\mathcal{H}_X$-measurable. By setting $Y = X$, equation (1.1.4) is satisfied. We conclude that $\mathbb{E}[X|\mathcal{H}_X] = X$. More generally, for any σ-algebra $\mathcal{G}$ with $\mathcal{H}_X \subseteq \mathcal{G}$ we conclude that $\mathbb{E}[X|\mathcal{G}] = X$. We summarize several important properties of conditional expectations in the following theorem:

Theorem 1.1.17 (cf. Shreve (2004, Theorem 2.3.2))**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathcal{G} \subseteq \mathcal{F}$ be a sub-σ-algebra of $\mathcal{F}$. Further, let X and Y be integrable random variables assuming values in $\mathbb{R}^n$.*

(i) **(Linearity)** *For $a, b \in \mathbb{R}$*

$$\mathbb{E}[aX + bY|\mathcal{G}] = a\mathbb{E}[X|\mathcal{G}] + b\mathbb{E}[Y|\mathcal{G}].$$

(ii) **(Taking out what is known)** *If $X^\top Y$ is integrable and X is $\mathcal{G}$-measurable, then*

$$\mathbb{E}\left[X^\top Y|\mathcal{G}\right] = X^\top \mathbb{E}[Y|\mathcal{G}].$$

(iii) **(Iterated conditioning)** *Let $\mathcal{H} \subseteq \mathcal{G}$ be a sub-σ-algebra of $\mathcal{G}$, then*

$$\mathbb{E}\left[\mathbb{E}\left[X|\mathcal{G}\right]|\mathcal{H}\right] = \mathbb{E}\left[X|\mathcal{H}\right].$$

(iv) **(Independence)** *Let X be independent of $\mathcal{G}$, then*

$$\mathbb{E}\left[X|\mathcal{G}\right] = \mathbb{E}\left[X\right].$$

Proof. See SHREVE (2004, pp. 70–72) for X and Y assuming values in $\mathbb{R}$. The proof of (i), (iii)-(iv) for the more general case – i.e., assuming values in $\mathbb{R}^n$ – is the same. The proof of (ii) can be reduced to the real-valued case by $X^\top Y = \sum_i^n X^i Y^i$ and the linearity (i). □

1.1.3 Stochastic process, Markov property

In this subsection we introduce stochastic processes and define sample path continuity of a stochastic process, a property that every stochastic process investigated in this exposition shall satisfy. Kolmogorov's continuity theorem tells us conditions under which a stochastic process is sample path continuous. We apply Kolmogorov's extension theorem later to construct a stochastic process from given probability measures. After introducing conditional probabilities we define Markov processes.

Definition 1.1.18 (cf. ØKSENDAL (2006, p. 10); cf. ØKSENDAL (2006, p. 14))**.** Let T denote an interval of $\mathbb{R}$. A **stochastic process** is a parametrized collection of random variables

$$(X_t)_{t\in T},$$

defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and assuming values in $\mathbb{R}^n$.
Let $(X_t)_{t\in T}$ and $(Y_t)_{t\in T}$ be two stochastic processes. We say that $(X_t)_{t\in T}$ is a **version** of $(Y_t)_{t\in T}$ if for all $t \in T$ the two random variables X_t and Y_t only differ on a set of measure zero.

The parameter interval T in this exposition shall be in most of the cases the halfline $[0;\infty)$ or a compact interval $[a,b]; a,b \in \mathbb{R}$. By definition we have for each fixed $t \in T$ a random variable

$$\omega \mapsto X_t(\omega).$$

If we fix $\omega \in \Omega$ we can consider the function

$$t \mapsto X_t(\omega),$$

which we call a **realization** or **path** of $(X_t)_t$. In applications the parameter $t \in T$ is considered as "time" and each $\omega \in \Omega$ is thought of as a "random experiment" whose outcome is the realization $X_t(\omega)$.
We can also consider a stochastic process $(X_t)_{t\in T}$ as a function $X : T \times \Omega \to \mathbb{R}^n$; $(t, \omega) \mapsto X_t(\omega) =: X(t, \omega)$. This viewpoint shall be necessary later when we talk about jointly measurable stochastic processes.

Definition 1.1.19. Let $(X_t)_{t\in T}$ and $(Y_t)_{t\in T}$ be two stochastic processes on $(\Omega, \mathcal{F}, \mathbb{P})$. We call $(X_{t\in T})_t$ and $(Y_{t\in T})_t$ **independent** if for any arbitrary finite sequence $0 \leq t_1 < t_2 < \ldots < t_n$ the two random variables $X := (X_{t_1}, \ldots, X_{t_n})$ and $Y := (Y_{t_1}, \ldots, Y_{t_n})$ are independent.

An important question in discussing stochastic processes is whether their realizations are continuous functions of t. We first clarify the definition of continuity of a stochastic process:

Definition 1.1.20 (cf. Kloeden & Platen (2000, p. 39))**.** Let $(X_t)_{t\in T}$ be a stochastic process. For each $t \in T$ we define

$$A_t = \left\{\omega \in \Omega \,\middle|\, \lim_{s\to t} \|X(s, \omega) - X(t, \omega)\|_2 \neq 0\right\}.$$

We call $(X_t)_{t\in T}$ **sample path continuous** if $A := \bigcup_{t\in T} A_t$ satisfies $A \in \mathcal{F}$ and

$$\mathbb{P}(A) = 0.$$

Remark 1.1.21. We have included the condition $A \in \mathcal{F}$, as $A := \bigcup_{t\in T} A_t$ is an uncountable union of measurable sets and thus does not necessarily have to be measurable.

We now state a very important criterion for a stochastic process to be sample path continuous.

Theorem 1.1.22. (**Kolmogorov's continuity theorem**; cf. Øksendal (2006, Theorem 2.2.3)) Let $X = (X_t)_{t\in T}$ be a stochastic process. Suppose that for all $S > 0$ there exist *positive* constants α, β, D such that

$$\mathbb{E}\left[\|X_t - X_s\|_2^\alpha\right] \leq D|t - s|^{1+\beta}; \quad 0 \leq s, t \leq S.$$

Then there exists a continuous version of X.

Proof. See Stroock & Varadhan (1979, p. 51) □

Definition 1.1.23 (cf. Øksendal (2006, p. 11) and Kloeden & Platen (2000, p. 18))**.** Let $X = (X_t)_{t\in T}$ be a stochastic process taking values in $\mathbb{R}^n$ and let $t_1, \ldots, t_k \in T$, $k \in \mathbb{N}$. The **finite dimensional distributions** of X are the measures defined on $\mathbb{R}^{nk}$ by

$$\mathbb{P}_{X;t_1,\ldots,t_k}(B_1 \times \ldots \times B_k) := \mathbb{P}(X_{t_1} \in B_1, \ldots, X_{t_k} \in B_k); \quad B_i \in \mathcal{B}(\mathbb{R}^n).$$

For a real-valued stochastic process (i.e., $n = 1$) we can consider the **finite dimensional distribution function**, also called **joint distribution function**, $F_{X;t_1,\ldots,t_k} : \mathbb{R}^k \mapsto [0,1]$, defined by

$$\begin{aligned} F_{X;t_1,\ldots,t_k}(x^1, \ldots, x^k) &:= \mathbb{P}(X_{t_1} \leq x^1, \ldots, X_{t_k} \leq x^k) \\ &= \mathbb{P}_{X;t_1,\ldots,t_k}\left((-\infty, x^1] \times \ldots \times (-\infty, x^k]\right). \end{aligned}$$

In certain cases the joint distribution functions $F_{X;t_1,\ldots,t_k}$ are absolutely continuous with respect to the Lebesgue-measure λ on $\mathbb{R}^n$. In this case the Theorem of Radon–Nikodým ensures the existence and λ-w.p.1 uniqueness of a Radon–Nikodým derivative $p_{X;t_1,\ldots,t_k}$ such that

$$F_{X;t_1,\ldots,t_k}(B_1 \times \ldots \times B_k) = \int_{B_1\times\ldots\times B_k} p_{X;t_1,\ldots,t_k} d\lambda(x^1, \ldots, x^k); \quad \forall B_i \in \mathcal{B}(\mathbb{R}^n).$$

We call the functions $p_{X;t_1,\ldots,t_k}$ the **densities** or **density functions** of $F_{X;t_1,\ldots,t_k}$.

For each stochastic process $X = (X_t)_t$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ we are able to construct a family $\left\{\mathbb{P}_{X;t_1,\ldots t_k} \,\middle|\, k \in \mathbb{N}; t_i \in T\right\}$ of finite dimensional distributions. The converse is also true under certain assumptions and known as *Kolmogorov's extension theorem*:

Theorem 1.1.24. (**Kolmogorov's extension theorem**, cf. Øksendal (2006, p. 11)) Let T be an interval of $\mathbb{R}$. For all $t_1, \ldots, t_k \in T$, $k \in \mathbb{N}$ let $\mu_{t_1,\ldots,t_k}$ be probability measures on $\mathcal{B}(\mathbb{R}^{nk})$ such that

$$\mu_{t_{\sigma(1)},\ldots,t_{\sigma(k)}}(B_1\times, \ldots, \times B_k) = \mu_{t_1,\ldots,t_k}(B_{\sigma^{-1}(1)} \times \ldots \times B_{\sigma^{-1}(k)}) \tag{1.1.5}$$

for all permutations σ on $\left\{1, 2, \ldots, k\right\}$ and

$$\mu_{t_1,\ldots,t_k}(B_1 \times \ldots \times B_k) = \mu_{t_1,\ldots,t_k,t_{k+1},\ldots,t_{k+m}}(B_1 \times \ldots \times B_k \times \underbrace{\mathbb{R}^n \times \ldots \times \mathbb{R}^n}_{\text{m times}}), \tag{1.1.6}$$

for all $m \in \mathbb{N}$.
Then there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a stochastic process $X = (X_t)_{t\in T}$ on

Ω, $X_t : \Omega \mapsto \mathbb{R}^n$, such that

$$\mu_{t_1,\ldots,t_k}(B_1 \times \ldots \times B_k) = \mathbb{P}(X_{t_1} \in B_1, \ldots, X_{t_k} \in B_k) = \mathbb{P}_{X;t_1,\ldots,t_k}(B_1 \times \ldots \times B_k)$$

for all $t_i \in T, k \in \mathbb{N}$ and all $B_i \in \mathcal{B}(\mathbb{R}^n)$

Remark 1.1.25. Equation (1.1.5) is known as the **permutation property** and equation (1.1.6) is known as the **restriction property**.

In the following we want to give sense to so called conditional probabilities $\mathbb{P}(A|Y = s)$, where $A \in \mathcal{F}$, $s \in \mathbb{R}^n$ and Y is a random variable assuming values in $\mathbb{R}^n$. From basic probability theory we know the definition

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{B} \tag{1.1.7}$$

for $A, B \in \mathcal{F}$ given that $\mathbb{P}(B) > 0$. But in our case we usually have $\mathbb{P}\left(\{Y = s\}\right) = 0$. To extend the definition to sets $B = \{Y = s\}$ of measure 0, we first give sense to conditional expectations

$$\mathbb{E}[X|Y = s]$$

with X, Y random variables assuming values in $\mathbb{R}^n$. By Definition 1.1.15 we know that $\mathbb{E}[X|Y] := \mathbb{E}[X|\mathcal{H}_Y] : \Omega \to \mathbb{R}^n$ is a $\mathcal{H}_Y$-measurable function. By Lemma 1.1.11 there exists a function $h : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$h(Y) = \mathbb{E}[X|Y].$$

We note that the function h has to be unique on $Y(\Omega) \subseteq \mathbb{R}^n$ and define for all $s \in \mathbb{R}^n$

$$\mathbb{E}[X|Y = s] := h(s). \tag{1.1.8}$$

Let $A \in \mathcal{F}$. We define by $\mathbb{1}_A : \Omega \to \{0, 1\}$ the **indicator function** of A, i.e., $\mathbb{1}_A(\omega) = 0$ if $\omega \notin A$ and $\mathbb{1}_A(\omega) = 1$ if $\omega \in A$. Now we set

$$X := (\underbrace{\mathbb{1}_A, \ldots, \mathbb{1}_A}_{n-\text{times}})$$

and note $X(\omega) \in \mathbb{R}^n$. By definition we see that $\mathbb{E}[X|Y] = \left(\mathbb{E}[X|Y]^1, \mathbb{E}[X|Y]^2, \ldots, \mathbb{E}[X|Y]^n\right)$ assumes values in $\mathbb{R}^n$ where $\mathbb{E}[X|Y]^i$ denotes the ith component of $\mathbb{E}[X|Y]$ and also that $\mathbb{E}[X|Y]^i = \mathbb{E}[X|Y]^j$ for $i, j = 1, \ldots, n$. This means we are able to apply the above conclusions to X and an arbitrary random variable Y assuming values in $\mathbb{R}^n$, i.e., there exists a function $h : \mathbb{R}^n \to \mathbb{R}^n$ such that $h(Y) = \mathbb{E}[X|Y]$. If we denote by $h^i(s)$

the ith component of $h(s) \in \mathbb{R}^n$ we again see that $h^i(s) = h^j(s)$ for all $s \in Y(\Omega)$ and $i, j = 1, \ldots, n$. For $s \in Y(\Omega)$ we now define the conditional probability

$$\mathbb{P}(A|Y = s) := h^1(s) \tag{1.1.9}$$

and for $s \notin Y(\Omega)$ we set $\mathbb{P}(A|Y = s) = 0$. In order for this definition to be reasonable we have to check whether it agrees with the definition of $\mathbb{P}(A|B)$ in (1.1.7) in the case of $B := \{Y = s\}$ and $\mathbb{P}(\{Y = s\}) > 0$: By $h^1(Y(\omega)) = \mathbb{E}[\mathbb{1}_A|Y](\omega)$ we deduce that $h^1(s) = \mathbb{E}[\mathbb{1}_A|Y](Y^{-1}(\{s\}))$. Further, by definition of conditional expectations we find

$$\int_{Y^{-1}(\{s\})} \mathbb{E}[\mathbb{1}_A|Y]\, d\mathbb{P} = \int_{Y^{-1}(\{s\})} \mathbb{1}_A d\mathbb{P} = \mathbb{P}(A \cap \{Y = s\}).$$

As $\mathbb{E}[\mathbb{1}_A|Y]$ is $\mathcal{H}_Y$-measurable it has to be constant to a real number C on $Y^{-1}(\{s\})$. Thus we further deduce from the last equation that $C \cdot \mathbb{P}(Y = s) = \mathbb{P}(A \cap \{Y = s\})$ and thus because of $\mathbb{P}(\{Y = s\}) > 0$

$$h^1(s) = h^1\left(Y(Y^{-1}(\{s\}))\right) = \mathbb{E}[\mathbb{1}_A|Y]\left(Y^{-1}(\{s\})\right) = C = \frac{\mathbb{P}(A \cap \{Y = s\})}{\mathbb{P}(\{Y = s\})}.$$

Thus the definition of $\mathbb{P}(A|Y = s)$ is reasonable.

Let $X = (X_s)_{s \in T}$ be a stochastic process assuming values in $\mathbb{R}^n$. Further let $0 < t_1 < t_2 < \ldots < t_k < t_{k+1} < \ldots < t_{k+m}$ be given time instants in T, $x^1, \ldots, x^k \in \mathbb{R}^n$ and $B_1, \ldots, B_m \in \mathcal{B}(\mathbb{R}^n)$. We set $\mathbf{t} := (t_1, \ldots, t_k)$, $\mathbf{t}' := (t_{k+1}, \ldots, t_{k+m})$, $\mathbf{B}' := B_1 \times \ldots \times B_m$, $\mathbf{x} = (x^1, \ldots, x^k)$ and $\mathbf{X}_{\mathbf{t}'} = (X_{t_{k+1}}, \ldots, X_{t_{k+m}})$, $\mathbf{X}_{\mathbf{t}} = (X_{t_1}, \ldots, X_{t_k})$. We then define the conditional probability

$$\mathbb{P}(\mathbf{X}_{\mathbf{t}'} \in \mathbf{B}'|\mathbf{X}_{\mathbf{t}} = \mathbf{x}) = \mathbb{P}(A|Y = \mathbf{x})$$

by equation (1.1.9) with $A := \{\omega \in \Omega \,|\, \mathbf{X}_{\mathbf{t}'} \in \mathbf{B}'\}$ and $Y = \mathbf{X}_{\mathbf{t}}$.

Remark 1.1.26. If X has a family of density functions $\{p_{X;t_1,\ldots,t_k} \,|\, t_1 < t_2 < \ldots < t_k; t_i \in T\}$, it is also possible to define conditional probabilities via density functions, cf. Gardiner (1990, p. 42): To alleviate notation we set $p(\mathbf{t}, \mathbf{y}) := p_{X;t_1,\ldots,t_k}(\mathbf{y})$ for $\mathbf{t} = (t_1, \ldots, t_k)$ and $\mathbf{y} = (y^1, \ldots, y^k)$. We then define **transition density functions** by

$$p(\mathbf{t}'; \mathbf{z}|\mathbf{t}; \mathbf{x}) := p(\mathbf{t}', \mathbf{z}|\mathbf{t}, \mathbf{x}) := \frac{p((\mathbf{t}, \mathbf{t}'); (\mathbf{x}, \mathbf{z}))}{p(\mathbf{t}; \mathbf{x})}; \quad \forall \mathbf{z} \in \mathbb{R}^{NM},$$

and the conditional probabilities by

$$\mathbb{P}\left(X_{\mathbf{t}'} \in \mathbf{B}' \middle| X_{\mathbf{t}} = \mathbf{x}\right) := \int_{\mathbf{B}'} p(\mathbf{t}', \mathbf{z}|\mathbf{t}, \mathbf{x}) d\lambda(\mathbf{z}), \tag{1.1.10}$$

given that the denominator within $p(\mathbf{t}'; \mathbf{z}|\mathbf{t}; \mathbf{x})$ is not zero. Otherwise we set the conditional probability to be zero.

Definition 1.1.27. Let $X := (X_t)_{t\in T}$ be a stochastic process. We call X a **Markov process** if for any time instants $0 < t_1 < t_2 < \ldots < t_k < t_{k+1} < \ldots < t_{k+m}$, $x^1, \ldots, x^k \in \mathbb{R}^n$ and $B_1, \ldots, B_m \in \mathcal{B}(\mathbb{R}^n)$ we have

$$\mathbb{P}\left(X_{\mathbf{t}'} \in \mathbf{B}' | X_{\mathbf{t}} = \mathbf{x}\right) = \mathbb{P}\left(X_{\mathbf{t}'} \in \mathbf{B}' | X_{t_k} = x^k\right), \tag{1.1.11}$$

where $\mathbf{x}, \mathbf{t}, \mathbf{t}', \mathbf{B}'$ are defined as above. This property is called the **Markov property**.

Remark 1.1.28. A Markov process can be seen as a process that does not "have memory": Equation (1.1.11) says that if we condition the probability that the random variable $X_{\mathbf{t}'}$ takes values within $\mathbf{B}'$ on our information of the "past" that $X_{\mathbf{t}} = \mathbf{x}$, the only important information is given by the last past time instant t_k, i.e., $X_{t_k} = x^k$.

1.1.4 Martingale property

In this subsection we define the martingale property which is related to the mathematical definition of a fair game. We shall see that the important stochastic processes of this exposition, like Wiener processes or Itô-integrals, satisfy the martingale property.

By definition of a stochastic process $(X_t)_{t\in T}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we see that the random variable X_t is $\mathcal{F} : \mathcal{B}(\mathbb{R}^n)$-measurable for each $t \in T$. The σ-algebra $\mathcal{F}$ usually contains much more information necessary to evaluate the random variable X_t, i.e., in general $\mathcal{H}_{X_t} \subsetneq \mathcal{F}$. We denote similar to Kloeden & Platen (2000, p. 65) by $\mathcal{A}_t$ the sub-σ-algebra of $\mathcal{F}$ generated by the subsets

$$A := \left\{\omega \in \Omega \,\middle|\, X_{t_i}(\omega) \in B_i \quad \text{for } i = 1, \ldots, k\right\}$$

of Ω for any $t_1 \leq t_2 \leq \ldots \leq t_k \leq t$ with $t_1, t_2, \ldots, t_k \in T$ and $B_1, \ldots, B_k \in \mathcal{B}(\mathbb{R}^n)$ where $k \in \mathbb{N}$. By definition we conclude that $\mathcal{A}_s \subseteq \mathcal{A}_t$ for $s \leq t$, which is usually referred to as that *information accumulates*. Also again by definition, we see that for each $s \in T$ and $s \leq t$ the random variable X_s is $\mathcal{A}_t : \mathcal{B}(\mathbb{R}^n)$-measurable.
Conversely, we can start with an increasing family of sub-σ-algebras and relate it to a stochastic process $(X_t)_{t\in T}$:

Definition 1.1.29 (cf. ØKSENDAL (2006, Definition 3.2.2))**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a measurable space, let T be an interval of $\mathbb{R}$ and let $(X_t)_{t\in T}$ be a stochastic process with values in $\mathbb{R}^n$. A **filtration** on $(\Omega, \mathcal{F})$ is a family $(\mathcal{A}_t)_{t\in T}$ of σ-algebras $\mathcal{A}_t \subseteq \mathcal{F}$ such that

$$\mathcal{A}_s \subseteq \mathcal{A}_t \quad \text{for } 0 \leq s < t.$$

We call the stochastic process $(X_t)_{t\in t}$ **adapted** to $(\mathcal{A}_t)_{t\in T}$ if for all $t \in T$ the random variable X_t is $\mathcal{A}_t : \mathcal{B}(\mathbb{R}^n)$-measurable.

Definition 1.1.30 (cf. ØKSENDAL (2006, Definition 3.2.2))**.** Let $(X_t)_{t\in T}$ be a stochastic process defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and adapted to a given filtration $(\mathcal{A}_t)_{t\in T}$. $(X_t)_{t\in T}$ is called a **martingale** with respect to $(\mathcal{A}_t)_{t\in T}$ and with respect to $\mathbb{P}$ if

(i) $\mathbb{E}\left[\|X_t\|_2\right] < \infty$ for all $t \in T$ and

(ii) $\mathbb{E}\left[X_t|\mathcal{A}_s\right] = X_s$ w.p.1 for all $s \leq t$; $s, t \in T$.

Remark 1.1.31. The condition $\mathbb{E}\left[\|X_t\|_2\right] < \infty$ is necessary to be able to define conditional expectations.

We now give an explanation why the martingale property is related to the mathematical definition of a fair game. Let $(X_t)_{t\in T}$ be a martingale. As before, we consider the interval T as a time-interval and fix two time instants $s \leq t$, with $s, t \in T$. The σ-algebra $\mathcal{A}_s$ is sufficiently large to evaluate the random variable X_s, but it could possibly contain more information. Nonetheless this possibly higher amount of information $\mathcal{A}_s \backslash \mathcal{H}_{X_s}$ does not help us to find a better approximation for X_t: Because of $\mathbb{E}\left[X_t|\mathcal{A}_s\right] = X_s$ the best assumption on the evolution of the process at time t based on our information $\mathcal{A}_s$ at time s is the process X_s itself.

1.2 Wiener process – Brownian motion

In this section we introduce Wiener processes, also known as Brownian motion, and summarize their main properties. Wiener processes play the keyrole in the construction of stochastic integrals.

In the literature, there are two ways of introducing a Wiener process, resp. Brownian motion. Either by defining it through its density as a product of transition densities, such as in ØKSENDAL (2006) or GARDINER (1990), or through its properties of expectation, variance and independence of its increments, such as in KLOEDEN & PLATEN (2000) or SHREVE (2004). We shall pursue the first way and give the following definition:

Definition 1.2.1 (cf. ØKSENDAL (2006, Definition 2.2.1))**.** Fix $x_0 \in \mathbb{R}^n$ and define

$$p(t,x,y) := (2\pi t)^{-\frac{n}{2}} \cdot exp\left(-\frac{\|x-y\|_2^2}{2t}\right); \quad \forall x,y \in \mathbb{R}^n, t>0, \tag{1.2.1}$$

with the convention that

$$p(0,x,y) = \delta_x(y),$$

where δ_x denotes the **unit point mass at** x or the **Dirac delta function at** x. Let $T = [0,\infty)$ and $t_1 \leq t_2 \leq \ldots \leq t_k$ with $t_i \in T$, $k \in \mathbb{N}$. We define a measure $\mu_{t_1,\ldots,t_k}$ on $\mathcal{B}(\mathbb{R}^n)$ by

$$\begin{aligned}&\mu_{t_1,\ldots,t_k}(B_1 \times \ldots \times B_k) := \\ &:= \int_{B_1\times\ldots\times B_k} p(t_1,x^0,x^1)p(t_2-t_1,x^1,x^2)\cdots p(t_k-t_{k-1},x^{k-1},x^k)d\lambda(x^1,\ldots,x^k).\end{aligned}$$

We extend this definition to finite sequences of t_i using equation (1.1.5). If there is a probability space $(\Omega,\mathcal{F},\mathbb{P})$ and a stochastic process $(W_t)_{t\in T}$ such that

$$\begin{aligned}&\mathbb{P}(W_{t_1} \in B_1, \cdots, W_{t_k} \in B_k) = \\ &= \int_{B_1\times\ldots\times B_k} p(t_1,x^0,x^1)p(t_2-t_1,x^1,x^2)\cdots p(t_k-t_{k-1},x^{k-1},x^k)d\lambda(x^1,\ldots,x^k),\end{aligned} \tag{1.2.2}$$

we call this process a version of **Brownian motion starting at** x or a **Wiener process starting at** x.

Remark 1.2.2.

(i) The Dirac delta function δ_x is the distribution defined by

$$\int_B f(y)\delta_x(y)d\lambda(y) = f(x)$$

for any continuous function f and $B \in \mathcal{B}(\mathbb{R}^n)$ such that $x \in B$.

(ii) Assuming that there is a probability space $(\Omega,\mathcal{F},\mathbb{P})$ and a stochastic process $(W_t)_{t\in T}$ starting at x_0 such that equation (1.2.2) holds, we conclude by the previous remark (choosing $f \equiv 1$) that

$$\mathbb{P}(W_0 = x_0) = 1.$$

(iii) Brownian motion is not *unique*, i.e., there are more quadruples $(W_t,\Omega,\mathcal{F},\mathbb{P})$ such that equation (1.2.2) holds (cf. ØKSENDAL (2006, p. 12)).

We now show that there really exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a stochastic process $(W_t)_t$ such that equation (1.2.2) holds. We prove the existence by checking whether the assumptions of Kolmogorov's extension theorem 1.1.24 are fulfilled. Clearly, the family of functions $\left\{\mu_{t_1,\dots,t_k} \,\middle|\, t_i \in T, k \in \mathbb{N}\right\}$ satisfies equation (1.1.5) by definition. By introducing polar coordinates we can show that

$$\int_{\mathbb{R}^n} p(t,x,y)d\lambda(y) = 1; \quad t \geq 0. \tag{1.2.3}$$

We now check whether $\left\{\mu_{t_1,\dots,t_k} \,\middle|\, t_i \in T, k \in \mathbb{N}\right\}$ are indeed measures: Equation (1.1.1) is of course fulfilled. Equation (1.1.2) holds because of the definition of integrals. Equation (1.2.3) then implies that each $\mu_{t_1,\dots,t_k}$ takes values inside of $[0,1]$ and $\mu_{t_1,\dots,t_k}(\mathbb{R}^{nk}) = 1$. Thus $\left\{\mu_{t_1,\dots,t_k} \,\middle|\, t_i \in T, k \in \mathbb{N}\right\}$ is a family of measures. Similar to the proof of $\mu_{t_1,\dots,t_k}(\mathbb{R}^{nk}) = 1$ we can conclude that equation (1.1.5) also holds, thus all assumptions of Kolmogorov's extension theorem are fulfilled.

Definition 1.2.3. Let $X = (X_t)_{t\in T}$ be a stochastic process. We call X a **Gaussian process** if for all $0 \leq t_1 \leq \dots \leq t_k$, $t_i \in T$, the random variable $Z = (X_{t_1}, \dots, X_{t_k}) \in \mathbb{R}^{nk}$ is **normally distributed**, i.e., there exists a vector $\mu \in \mathbb{R}^{nk}$ and a positive semidefinite matrix $C = (c^{j,m})_{jm} \in \mathbb{R}^{nk\times nk}$ such that

$$\mathbb{E}\left[\exp\left(iu^\top Z\right)\right] = \exp\left(-\frac{1}{2}u^\top (Cu) + iu^\top \mu\right) \quad \forall u \in \mathbb{R}^{nk},$$

where i denotes the imaginary unit.

Remark 1.2.4. Let $Z = (Z^1, \dots, Z^n)$ be a random variable assuming values in $\mathbb{R}^n$. The function $\phi : u \mapsto \mathbb{E}\left[\exp\left(iu^\top Z\right)\right]$ is called the **characteristic function** of Z and can be seen as the inverse Fourier transformation of the probability measure $\mathbb{P}_Z$. Higher moments of Z can be determined via

$$\mathbb{E}\left[\left(Z^1\right)^{m_1} \cdots \left(Z^n\right)^{m_n}\right] = \prod_{j=1}^{n} (-i)^{m_j} \frac{\partial^{m_j}\phi}{\partial u_j^{m_j}}(0)$$

for $m_i \in \left\{0, 1, 2, \dots\right\}$, cf. Gardiner (1990, Section 2.6).

The following theorem states several important properties of a Wiener process:

Theorem 1.2.5 (cf. Øksendal (2006, p. 13))**.** *Let $(W_t)_{t\in T}$ be a Wiener process starting at x on $(\Omega, \mathcal{F}, \mathbb{P})$, then $(W_t)_t$ is a Gaussian process and with the notation of Definition*

1.2.3 we have

$$\mu = (x, \dots, x) \in \mathbb{R}^{nk}; \quad C = \begin{pmatrix} t_1 I_n & t_1 I_n & \cdots & t_1 I_n \\ t_1 I_n & t_2 I_n & \cdots & t_2 I_n \\ \vdots & \vdots & & \vdots \\ t_1 I_n & t_2 I_n & \cdots & t_k I_n \end{pmatrix},$$

where I_n denotes the n-dimensional identity matrix.
As an immediate consequence we get

$$\begin{aligned} \mathbb{E}[Z] &= \mu, \\ \mathbb{C}ov\left[Z^j; Z^m\right] &= \mathbb{E}\left[(Z^j - \mu^j)(Z^m - \mu^m)\right] = c^{j,m}. \end{aligned}$$

Moreover, $(W_t)_{t\in T}$ has **independent increments***, i.e., for all $0 \le t_1 \le \ldots \le t_k$ the random variables*

$$W_{t_1}, W_{t_2} - W_{t_1}, W_{t_3} - W_{t_2}, \dots, W_{t_k} - W_{t_{k-1}}$$

are independent.

Proof. See Øksendal (2006, p. 13). □

Remark 1.2.6.

(i) In the literature (cf. Kloeden & Platen (2000)), a Wiener process starting at x_0 is frequently introduced as a stochastic process with normally distributed and independent increments satisfying

$$\mathbb{E}\left[W_{t_{i+1}} - W_{t_i}\right] = 0; \quad \mathbb{C}\mathrm{ov}\left[W_{t_{i+1}} - W_{t_i}\right] = (t_{i+1} - t_i) I_n.$$

As Theorem 1.2.5 implies that $\mathbb{C}\mathrm{ov}\left[W_{t_{i+1}} - W_{t_i}\right] = (t_{i+1} - t_i) I_n$, we see that our definition via densities contains the approach of Kloeden & Platen (2000).

(ii) Let $(W_t)_t = (W_t^1, \dots, W_t^n)$ be a n-dimensional Wiener process. By Theorem 1.2.5 the covariance matrix $\mathbb{C}\mathrm{ov}[W_t] = \left(\mathbb{C}\mathrm{ov}\left[W_t^i; W_t^j\right]\right)_{i,j}$ of W_t for each $t \in \mathbb{R}$ is then

$$\mathbb{C}\mathrm{ov}[W_t] = t I_n. \tag{1.2.4}$$

The one-dimensional processes $W_t^1, \dots, W_t^n$ are themselves by definition one-dimensional Wiener processes and again by Theorem 1.2.5 Gaussian processes. As normally distributed random variables are independent iff they are uncorrelated, equation (1.2.4) tells us that the one-dimensional processes $W_t^1, \dots, W_t^n$ are independent.

The Theorem of Wick-Isserlis, that we shall formulate in Theorem 2.3.4 of Chapter 2, states that in the case of a normally distributed random variable X the determination of higher central moments can be reduced to the determination of its second moments. This theorem also implies that a Wiener process $(W_t)_t$ satisfies for all $0 \leq s, t < \infty$

$$\mathbb{E}\left[\|W_t - W_s\|_2^4\right] = n(n+2)|t-s|^2.$$

Thus we see that a Wiener process satisfies the assumptions of Kolmogorov's continuity theorem, Theorem 1.1.22, i.e., a Wiener process has a continuous version. On the other hand, it can be shown that a Wiener process has a.s. nowhere differentiable sample paths, cf. Kloeden & Platen (2000, p. 74).

Let a Wiener process $(W_t)_t$ be given. From now on we denote by $\mathcal{F}_t$ the smallest σ-algebra such that $(W_s)_{0\leq s\leq t}$ is measurable. $(\mathcal{F}_t)_{0\leq t<\infty}$ thus defines a filtration that we call the **filtration associated to** $(W_t)_t$.

Theorem 1.2.7. *Let $(W_t)_t$ be a Wiener process with associated filtration $(\mathcal{F}_t)_t$. Then $(W_t)_t$ is a martingale with respect to $(\mathcal{F}_t)_t$.*

Proof. We have to show that the equation

$$\mathbb{E}\left[W_t|\mathcal{F}_s\right] = W_s$$

holds for each $0 \leq s < t$. By Remark 1.2.6.(i) the random variable $W_t - W_s$ is independent of W_s. Thus by Theorem 1.1.17.(iv) we conclude

$$\mathbb{E}\left[W_t - W_s|\mathcal{F}_s\right] = \mathbb{E}\left[W_t - W_s\right] = 0,$$

where the last equality holds again by Remark 1.2.6.(i). By Theorem 1.1.17.(ii) we further conclude

$$\begin{aligned}\mathbb{E}\left[W_t|\mathcal{F}_s\right] &= \mathbb{E}\left[W_t - W_s|\mathcal{F}_s\right] + \mathbb{E}\left[W_s|\mathcal{F}_s\right] \\ &= \mathbb{E}\left[W_t - W_s\right] + \mathbb{E}\left[W_s|\mathcal{F}_s\right] \\ &= 0 + W_s = W_s.\end{aligned}$$

□

Finally, we now state a theorem that explains the dominant role of Gaussian random variables. We shall need in Chapter 2 to estimate the operator $\mathbb{E}[\cdot]$:

Theorem 1.2.8. (**Central limit theorem**, cf. Kloeden & Platen (2000, Proposition 5.9)) Let $X_1, X_2, \ldots$ be independent and identically distributed real valued random

variables with $\mathbb{E}[X_i] = 0$ and $\mathbb{E}[X_i^2] = 1$, and let ζ be a normally distributed random variable with mean 0 and variance 1. Then

$$Z_n := \frac{1}{\sqrt{n}} \sum_{k=1}^{n} X_i$$

converges in distribution to ζ, i.e., $\lim_{n\to\infty} F_{Z_n}(x) = F_\zeta(x)$ for all points x for which F_ζ is continuous at x.

Proof. See KALLENBERG (2002, p. 90). □

1.3 Diffusion process

In this section we define diffusion processes that are special Markov processes. Since most of the processes handled in this exposition are in fact diffusion processes we discuss them and their connections to PDEs in detail.

Let $X = (X_t)_{t\in[0,\infty)}$ be a Markov process assuming values in $\mathbb{R}^n$. For $0 \leq s < t$, $B \in \mathcal{B}(\mathbb{R}^n)$ and $x \in \mathbb{R}^n$ we define the **transition probabilities** by

$$\mathbb{P}(t, B; s, x) := \mathbb{P}\left(X_t \in B | X_s = x\right).$$

In some cases the stochastic process X has a family of density functions and we are able to define transition densities as in Remark 1.1.26. Then we can write

$$\mathbb{P}(t, B; s, x) = \int_B p(t, y|s, x) d\lambda(y).$$

Using the Markov property it can be shown that the so called **Chapman-Kolmogorov equation** holds, cf. GARDINER (1990, pp. 43–44):

$$P(t, B; s, x) = \int_{\mathbb{R}} P(t, B; \tau, z) P(\tau, dz; s, x) \tag{1.3.1}$$

for $0 \leq s \leq \tau \leq t$ and $x \in \mathbb{R}^n$, which simplifies in the case of the existence of transition densities to

$$p(t, y|s, x) = \int_{\mathbb{R}} p(t, y|\tau, z) p(\tau, z|s, x) d\lambda(z)$$

for $0 \leq s \leq \tau \leq t$ and $x, y \in \mathbb{R}^n$. The integral in (1.3.1) is an improper Lebesgue-Stieltjes integral where we used the notation $P(\tau, dz; s, x)$ to indicate that the Lebesgue-Stieltjes integral is defined with the integrator function $z \mapsto P(\tau, z; s, x)$.

Definition 1.3.1 (cf. KLOEDEN & PLATEN (2000, p. 68))**.** We call a Markov process $X = (X_t)_{t\in[0,\infty)}$ a **diffusion process** if the following limits exist for any $\epsilon > 0, s \geq 0, x \in$

$\mathbb{R}^n$:

$$\lim_{t\downarrow s}\frac{1}{t-s}\int_{\|y-x\|_2>\epsilon} P(t,dy;s,x)=0, \tag{1.3.2a}$$

$$\lim_{t\downarrow s}\frac{1}{t-s}\int_{\|y-x\|_2\leq\epsilon}(y-x)P(t,dy;s,x)=a(s,x) \tag{1.3.2b}$$

and

$$\lim_{t\downarrow s}\frac{1}{t-s}\int_{\|y-x\|_2\leq\epsilon}(y-x)(y-x)^\top P(t,dy;s,x)=B(s,x)B(s,x)^\top \tag{1.3.2c}$$

where $D(s,x)=B(s,x)B(s,x)^\top$ has to be symmetric positive definite. We call a the **drift vector** and D the **diffusion matrix**.

Remark 1.3.2.

(i) By letting $\epsilon\to\infty$ it is possible to prove that (1.3.2b) and (1.3.2c) imply

$$a^i(s,x)=\lim_{t\downarrow s}\frac{1}{t-s}\mathbb{E}\left[X_t^i-X_s^i|X_s=x\right],$$
$$D^{i,j}(s,x)=\lim_{t\downarrow s}\frac{1}{t-s}\mathbb{E}\left[(X_t^i-X_s^i)(X_t^j-X_s^j)|X_s=x\right],$$

where X_s^i denotes the ith component of X_s, and $a(s,x)=(a^i(s,x))$, $D(s,x)=(D^{i,j}(s,x))$ for $i,j=1,\dots n$.

(ii) If X is a Markov process with a family of transition densities, equations (1.3.2a), (1.3.2b) and (1.3.2c) simplify to

$$\lim_{t\downarrow s}\frac{1}{t-s}\int_{\|y-x\|_2>\epsilon} p(t,y|s,x)d\lambda(y)=0, \tag{1.3.3a}$$

$$\lim_{t\downarrow s}\frac{1}{t-s}\int_{\|y-x\|_2\leq\epsilon}(y-x)p(t,y|s,x)d\lambda(y)=a(s,x) \tag{1.3.3b}$$

and

$$\lim_{t\downarrow s}\frac{1}{t-s}\int_{\|y-x\|_2\leq\epsilon}(y-x)(y-x)^\top p(t,y|s,x)d\lambda(y)=B(s,x)B(s,x)^\top. \tag{1.3.3c}$$

(iii) It is possible to weaken the condition upon the diffusion matrix D to be only symmetric positive semidefinite rather than symmetric positive definite, as suggested in Gardiner (1990, Section 3.5.2). We shall turn to that point later in Section 1.5.3.

If a diffusion process $X=(X_t)_t$ has a family of transition densities $p(t,y|s,x)$, it is possible to show that they satisfy the **Kolmogorov forward equation**, also known as

Fokker-Planck equation, cf. GARDINER (1990, Section 3.4):

$$\begin{aligned} &\frac{\partial p}{\partial t} + \sum_{i=1}^{n} \frac{\partial}{\partial y_i} \left(a^i(t,y)p \right) - \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2}{\partial y_i \partial y_j} \left(D^{i,j}(t,y)p \right) = 0, \\ \iff: &\frac{\partial p}{\partial t} + \nabla_y \cdot (a(t,y)p) - \frac{1}{2} \nabla_y^2 : (D(t,y)p) = 0 \end{aligned} \tag{1.3.4}$$

with s and x in p fixed and with the initial condition

$$\lim_{t \downarrow s} p(s,x;t,y) = \delta(x-y).$$

We can also regard (1.3.4) in operator form: We define the operator $\mathcal{G}$ for $u \in C^2(\mathbb{R}^{n+1})$ by

$$(\mathcal{G}(u))(t,y) := -\nabla_y \left(a(t,y)u(t,y) \right) + \frac{1}{2} \nabla_y^2 : \left(D(t,y)u(t,y) \right).$$

$\mathcal{G}$ is elliptic because D is symmetric positive definite. Equation 1.3.4 then transforms to

$$\frac{\partial p}{\partial t} - \mathcal{G}p = 0$$

with s and x fixed. On $L^2(\mathbb{R}^n)$ the operator $\mathcal{G}$ can be seen as the formal adjoint of the elliptic operator $\mathcal{L}$ defined by

$$(\mathcal{L}(u))(s,x) = \sum_{i=1}^{n} a^i(s,x) \frac{\partial u}{\partial x_i}(s,x) + \frac{1}{2} \sum_{i,j}^{n} D^{i,j}(s,x) \frac{\partial^2 u}{\partial x_i \partial x_j}(s,x),$$

i.e., $\mathcal{G} = \mathcal{L}^*$ which can be verified by intergration by parts. It can be shown by similar techniques as in the proof of the Kolmogorov forward equation, cf. GARDINER (1990, Section 3.4), that the transition densities $p(t,y|s,x)$ of a diffusion process fulfill the so called **Kolmogorov backward equation**:

$$\frac{\partial p}{\partial s} + \mathcal{L}p = 0 \tag{1.3.5}$$

with t and y fixed.

A Wiener process $(W_t)_t$ has the transition densities

$$p(t,y|s,x) = \frac{1}{(2\pi(t-s))^{n/2}} \exp\left(-\frac{\|y-x\|_2^2}{2(t-s)} \right),$$

which exactly correspond to (1.2.1) with $t-s$ as the time parameter. Markov processes with the property that their transition densities only depend on the time difference $t-s$ rather than on the specific values of s and t are called **time homogeneous**. It can

be shown that a Wiener process indeed satisfies the Markov property (1.1.11) by direct calculation using its transition densities. Similarly, a Wiener process satisfies equations (1.3.2a) – (1.3.2c), resp. (1.3.3a) – (1.3.3c) with drift vector 0 and diffusion matrix I_n.

1.4 Stochastic integration

In this section we are interested in giving sense to a stochastic integral

$$\int_S^T \Delta(t,\omega) dW_t(\omega) \tag{1.4.1}$$

where $(W_t)_t$ is a given Wiener process and $(\Delta(t,\cdot))_t$ a – yet – arbitrary stochastic process. Before we concretize the definition of a stochastic integral we give the following motivation: Given a finite partition $\Pi = \{t_1 = S, \ldots, t_n = T\}$ of the interval $[S,T]$, it is possible to approximate $(\Delta(t,\cdot))_t$ by a step function

$$(t,\omega) \mapsto \sum_{i=1}^{n-1} \Delta(t_i^*,\omega) \mathbb{1}_{[t_i,t_{i+1})}(t); \quad t_i^* \in [t_i, t_{i+1}). \tag{1.4.2}$$

We can then define (1.4.1) as the limit (in a sense that still has to be explained) of $\sum_{i=1}^{n-1} \Delta(t_i^*,\omega)\left(W_{t_{i+1}} - W_{t_i}\right)$ as $\max_i |t_{i+1} - t_i| =: \|\Pi\| \to 0$.
But unlike to Lebesgue-Stieltjes integrals – which are defined in that way – the choice of t_i^* in (1.4.2) is crucial in the case of stochastic integrals, cf. Øksendal (2006, Example 3.1.1). Two choices of t_i^* are the most common:

1) $t_i^* = t_i$, i.e., the left border point of the interval $[t_i, t_{i+1})$ leads to the **Itô integral** that we denote by
$$\int_S^T \Delta(t,\omega) dW_t(\omega). \tag{1.4.3}$$

2) $t_i^* = \frac{t_{i+1}+t_i}{2}$, i.e., the mid point of the interval $[t_i, t_{i+1})$ leads to the **Stratonovich integral** that we denote by
$$\int_S^T \Delta(t,\omega) \circ dW_t(\omega). \tag{1.4.4}$$

In general the two integrals give us different real numbers for fixed $\omega \in \Omega$ but we shall see in Subsection 1.5.4 in which way they can be seen to be equivalent.

We demonstrate in detail the construction of a stochastic integral by means of the Itô integral guided by ØKSENDAL (2006, Chapter 3) and KLOEDEN & PLATEN (2000, Chapter 3, Section 2-4). In Subsection 1.4.5 we briefly state the definition of the Stratonovich integral but slightly different to how we motivated the construction above. However in the case of sufficiently smooth functions both ways shall turn out to be equivalent.

1.4.1 Itô integral

In this subsection we discus the construction of the Itô integral for real-valued processes. We shall treat vector-valued Itô integrals in Section 1.4.3. I.e., throughout this subsection we assume that a one-dimensional Wiener process $(W_t)_t$ starting at 0 on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and the filtration $(\mathcal{F}_t)_t$ generated by $(W_t)_t$ are given.

We now define the class of functions for which we construct the Itô integral:

Definition 1.4.1 (cf. ØKSENDAL (2006, Definition 3.1.4))**.** We denote by $\mathcal{V} = \mathcal{V}(S,T)$ the set of functions

$$f(t,\omega) : [0,\infty) \times \Omega \to \mathbb{R},$$

such that

(i) $(t,\omega) \mapsto f(t,\omega)$ is $\mathcal{B}([0,\infty)) \times \mathcal{F}$-measurable,

(ii) for all $t \in [0,\infty)$ the function $\omega \mapsto f(t,\omega)$ is $\mathcal{F}_t$-adapted,

(iii) $\mathbb{E}\left[\int_S^T f(t,\omega)^2 d\lambda(t)\right] < \infty$, i.e., $f \in L^2([S,T] \times \Omega)$.

Definition 1.4.2 (cf. ØKSENDAL (2006, p. 26))**.** We call a function $\phi \in \mathcal{V}$ **elementary** if there exists a partition $\Pi := \{t_1 = S, t_2, \ldots, t_n = T\}$ of $[S,T]$ such that

$$\phi(t,\omega) = \sum_i^{n-1} e_i(\omega) \mathbb{1}_{[t_i, t_{i+1})},$$

where e_i are $\mathcal{F}_{t_i}$-measurable random variables. We denote the set of elementary functions of $\mathcal{V}$ by $\mathcal{V}_{elem}$.

For an elementary function ϕ we define the Itô integral over $[S,T]$ as the random variable

$$\mathcal{I}[\phi](\omega) := \int_S^T \phi(t,\omega) dW_t(\omega) := \sum_i^{n-1} e_i(\omega) \left(W_{t_{i+1}}(\omega) - W_{t_i}(\omega)\right).$$

The following observation plays an important key role in the extension of the Itô integral to the whole class $\mathcal{V}$:

Lemma 1.4.3 (**Itô isometry**; cf. ØKSENDAL (2006, Lemma 3.1.5))**.** *Let $\phi \in \mathcal{V}$ be bounded and elementary. Then*

$$\mathbb{E}\left[\left(\int_S^T \phi(t,\cdot)dW_t(\cdot)\right)^2\right] = \mathbb{E}\left[\int_S^T \phi(t,\cdot)^2 d\lambda(t)\right].$$

Proof. See ØKSENDAL (2006, p. 26). □

Remark 1.4.4. Theorem 1.4.3 states that the Itô integral $\mathcal{I}[\cdot]$ defines an isometry as a mapping between subsets of $L^2([S,T]\times\Omega)$ and $L^2(\Omega)$:

$$\|\mathcal{I}[\phi]\|_{L^2(\Omega)} = \|\phi\|_{L^2([S,T]\times\Omega)}; \quad \forall \phi \in \mathcal{V} \text{ and } \phi \text{ elementary.}$$

To extend the definition of the Itô integral to the whole class $\mathcal{V}$ we proceed in three steps:

Proposition 1.4.5 (**first step**; cf. ØKSENDAL (2006, p. 27))**.** *The set $\mathcal{V}_{elem}$ is* **dense** *in $\mathcal{V}_{cont}^{\infty} := \left\{g \in \mathcal{V} \,\middle|\, g(\cdot,\omega) \text{ is bounded and continuous for each } \omega \in \Omega\right\}$ w.r.t. $\|\cdot\|_{L^2([S,T]\times\Omega)}$, i.e., for all $g \in \mathcal{V}_{cont}^{\infty}$ there exists a sequence $\phi_n \in \mathcal{V}_{elem}$ such that*

$$\|g - \phi_n\|_{L^2([S,T]\times\Omega)} \to 0 \quad \text{for } n \to \infty.$$

Proof. See ØKSENDAL (2006, p. 27). □

Proposition 1.4.6 (**second step**; cf. ØKSENDAL (2006, p. 27))**.** *The set $\mathcal{V}_{cont}^{\infty}$ is* **dense** *in $\mathcal{V}^{\infty} := \left\{h \in \mathcal{V} \,\middle|\, h \text{ is bounded}\right\}$ w.r.t. $\|\cdot\|_{L^2([S,T]\times\Omega)}$, i.e., for all $h \in \mathcal{V}^{\infty}$ there exists a sequence $g_n \in \mathcal{V}_{cont}^{\infty}$ such that*

$$\|h - g_n\|_{L^2([S,T]\times\Omega)} \to 0 \quad \text{for } n \to \infty.$$

Proof. See ØKSENDAL (2006, p. 27). □

Proposition 1.4.7 (**third step**; cf. ØKSENDAL (2006, p. 28))**.** *The set $\mathcal{V}^{\infty}$ is* **dense** *in $\mathcal{V}$ w.r.t. $\|\cdot\|_{L^2([S,T]\times\Omega)}$, i.e., for all $f \in \mathcal{V}$ there exists a sequence $h_n \in \mathcal{V}^{\infty}$ such that*

$$\|f - h_n\|_{L^2([S,T]\times\Omega)} \to 0 \quad \text{for } n \to \infty.$$

Proof. See ØKSENDAL (2006, p. 28). □

The preceding steps show that the set $\mathcal{V}_{elem}$ is dense in $\mathcal{V}$ w.r.t. $\|\cdot\|_{L^2([S,T]\times\Omega)}$. We are now able to extend the definition of the Itô integral to the whole class $\mathcal{V}$:

Definition 1.4.8. Let $f \in \mathcal{V}$ be arbitrary and $\phi_n \in \mathcal{V}_{elem}$ be a sequence such that

$$\|f - \phi_n\|_{L^2([S,T]\times\Omega)} \to 0 \quad \text{for } n \to \infty.$$

We then define the Itô integral of f to be the limit

$$\mathcal{I}[f](\omega) := \int_S^T f(t,\omega)dW_t(\omega) := \lim_{n\to\infty} \int_S^T \phi_n(t,\omega)dW_t \tag{1.4.5}$$

where the limit is considered w.r.t. $\|\cdot\|_{L^2(\Omega)}$.

Proposition 1.4.9. *The Itô integral of $f \in \mathcal{V}$ is* **well-defined**, *i.e., the limit in* (1.4.5) *exists and is independent of the choice of $(\phi_n)_n$.*

Proof. By Proposition 1.4.5-1.4.7 there exists a sequence $\phi_n \in \mathcal{V}_{elem}$ with $\phi_n \to f$ in $L^2([S,T]\times\Omega)$. By the Itô isometry, Lemma 1.4.3, the sequence $\left(\int_S^T \phi_n(t,\cdot)dW_t(\cdot)\right)_n$ is a Cauchy sequence in $L^2(\Omega)$. As $L^2(\Omega)$ is complete, the limit on the right side of (1.4.5) exists and is unique. Any other sequence $\tilde{\phi}_n \in \mathcal{V}_{elem}$ with

$$\left\|f - \tilde{\phi}_n\right\|_{L^2([S,T]\times\Omega)} \to 0 \quad \text{for } n \to \infty$$

satisfies $\left\|\phi_n - \tilde{\phi}_n\right\|_{L^2([S,T]\times\Omega)} \to 0$. Using the Itô isometry again, we see that the limit of (1.4.5) is indeed independent of the choice of $\phi_n \in \mathcal{V}_{elem}$. □

By construction of the Itô integral we can deduce the following results as a corollary

Corollary 1.4.10 (**Itô isometry**, cf. ØKSENDAL (2006, Corollary 3.1.7))**.** *Let $f \in \mathcal{V}$ be arbitrary, then*

$$\|\mathcal{I}[f]\|_{L^2(\Omega)} = \|f\|_{L^2([S,T]\times\Omega)},$$

i.e.,

$$\mathbb{E}\left[\left(\int_S^T f(t,\cdot)dW_t(\cdot)\right)^2\right] = \mathbb{E}\left[\int_S^T f(t,\cdot)^2 d\lambda(t)\right].$$

1.4.2 Properties of the Itô integral

In this subsection we summarize several main properties of the Itô integral.

Theorem 1.4.11 (cf. ØKSENDAL (2006, p. 30))**.** *Let $f, g \in \mathcal{V}(0,T)$ and $0 \le S < U < T$. Then*

(i) $\int_S^T f(t,\omega)dW_t(\omega) = \int_S^U f(t,\omega)dW_t(\omega) + \int_U^T f(t,\omega)dW_t(\omega)$ *w.p.1,*

(ii) $\int_S^T (cf+g)dW_t = c\int_S^T f dW_t + \int_S^T g dW_t$ *w.p.1 for all $c \in \mathbb{R}$,*

(iii) $\mathbb{E}\left[\int_S^T f dW_t\right] = 0$,

(iv) $\int_S^T f dW_t$ *is* $\mathcal{F}_T$*-measurable.*

Proof. See ØKSENDAL (2006, p. 30). □

Theorem 1.4.12 (cf. ØKSENDAL (2006, Theorem 3.2.5))**.** *Let* $f \in \mathcal{V}(0,T)$ *be arbitrary. Then there exists a version of*

$$\left(\int_0^t f(s,\cdot)dW_s(\cdot)\right)_{0\leq t\leq T}$$

that is sample path continuous.

Proof. See ØKSENDAL (2006, p. 32). □

Remark 1.4.13. Theorem 1.4.12 assures that $\left(\int_0^t f(s,\cdot)dW_s(\cdot)\right)_{0\leq t\leq T}$ has a t-continuous version. From now on we shall always assume that

$$\left(\int_0^t f(s,\cdot)dW_s(\cdot)\right)_{0\leq t\leq T}$$

represents a t-continuous version of the Itô integral.

Theorem 1.4.14 (cf. ØKSENDAL (2006, Corollary 3.2.6))**.** *Let* $f \in \mathcal{V}(0,T)$ *for all* $T > 0$*. Then the stochastic process* $(M_t)_t$ *defined by*

$$M_t(\omega) := \int_0^t f(t,\omega)dW_t(\omega)$$

is a martingale with respect to $(\mathcal{F}_t)_t$*.*

Proof. See ØKSENDAL (2006, p. 33). □

1.4.3 Extension of Itô integral

As stated in ØKSENDAL (2006, Section 3.3) the Itô integral as defined above can be extended to a wider class of functions by weakening conditions (ii) and (iii) of Definition 1.4.1:

Instead of considering $(\mathcal{F}_t)_t$ the filtration *generated* by $(W_t)_t$, we can weaken condition (ii) to

(ii)' There exists a filtration $(\mathcal{A}_t)_t$ such that

(a) $(W_t)_t$ is a martingale with respect to $(\mathcal{A}_t)_t$ and

(b) for all $t \in [0, \infty)$ the function $\omega \mapsto f(t, \omega)$ is $\mathcal{A}_t$-adapted.

This means that the filtration $(\mathcal{A}_t)_t$ fulfills $\mathcal{F}_t \subseteq \mathcal{A}_t$ and thus contains possibly more information than $\mathcal{F}_t$. However $(W_t)_t$ is still a martingale w.r.t. $(\mathcal{A}_t)_t$, i.e., $(\mathcal{A}_t)_t$ does not contain "too much" information. As stated in ØKSENDAL (2006, p. 34) the same results of the last two subsections follow if we define the Itô integral with (ii)'. This weakening allows us to consider Itô integrals of the form

$$\int W_t^2 dW_t^1$$

where $(W_t^1)_t$ and $(W_t^2)_t$ are two *independent* Wiener processes. We are now able to define multidimensional Itô integrals:

Definition 1.4.15. *(cf. ØKSENDAL (2006, Definition 3.3.1)) Let* $(W_t)_t = (W_t^1, W_t^2, \ldots, W_t^n)$ *be a* n*-dimensional Wiener process and* $(\mathcal{A}_t)_t$*. We denote by* $\mathcal{V}_{\mathcal{A}_t}^{m \times n}(S, T)$ *the set of all* $m \times n$ *matrices* $v = (v^{i,j}(t, \omega))$*, where every* v_{ij} *is a stochastic process that satisfies conditions* (i) *and* (iii) *of Definition 1.4.1 and further* (ii)' *above with respect to a filtration* $(\mathcal{A}_t)_t$*.*
For $v \in \mathcal{V}_{\mathcal{A}_t}^{m \times n}(S, T)$ *we define the Itô integral*

$$\int_S^T v dW_t = \int_S^T \begin{pmatrix} v^{1,1} & \cdots & v^{1,n} \\ \vdots & & \vdots \\ v^{m,1} & \cdots & v^{m,n} \end{pmatrix} \begin{pmatrix} dW_t^1 \\ \vdots \\ dW_t^1 \end{pmatrix} \tag{1.4.6}$$

to be the vector whose i*th component is the following sum of (extended) one-dimensional Itô integrals*

$$\sum_{j=1}^{n} \int_S^T v^{i,j}(t, \omega) dW_s^j(\omega). \tag{1.4.7}$$

It is also possible to weaken condition (iii) of Definition 1.4.1: Instead we can require

(iii)' $\mathbb{P}\left(\int_S^T f(t, \omega)^2 d\lambda(t) < \infty\right) = 1$.

Definition 1.4.16 (cf. ØKSENDAL (2006, Definition 3.3.2))**.** We denote by $\mathcal{W}_{\mathcal{A}_t}(S, T)$ the set of all functions

$$f(t, \omega) : [0, \infty) \times \Omega \to \mathbb{R}$$

that satisfy condition (i) of Definition 1.4.1 and further condition (ii)' and (iii)' above.

Let $(W_t)_t$ be a given one-dimensional Wiener process. It is possible to show – cf. ØKSENDAL (2006, p. 35) – that for each $f \in \mathcal{W}_{\mathcal{A}_t}$ there is a sequence of elementary functions $\phi \in \mathcal{W}_{\mathcal{A}_t}$ such that $\lim_{n \to \infty} \mathbb{P}(\int_0^t |\phi_n - f| d\lambda(s) < \epsilon) = 1$ for all $\epsilon \in \mathbb{R}$, i.e., $\int_0^t |\phi_n - f| d\lambda(s) \to 0$ **in**

probability. In this case $\int_0^t \phi_n(s,\omega)dW_s$ converges in probability to a random variable where the limit is independent of the sequence ϕ_n. Thus it is possible to define for all $f \in \mathcal{W}_{\mathcal{A}_t}$

$$\int_0^t f(s,\omega)dW_s = \lim_{n\to\infty} \phi_n(s,\omega)dW_s,$$

where the limit is considered in probability. For $f \in \mathcal{V}_{\mathcal{H}} \subseteq \mathcal{W}_{\mathcal{A}_t}$ this definition agrees with Definition 1.4.15.
For $f \in \mathcal{W}_{\mathcal{A}_t}$ there exists similar to Theorem 1.4.12 a continuous version. Nevertheless, the Itô integral defined on $\mathcal{W}_{\mathcal{A}_t}$ is in general *not* a martingale. For $f \in \mathcal{W}_{\mathcal{A}_t}$, there is a sequence of martingales that converges in probability to the Itô integral of f, cf. Kloeden & Platen (2000, p. 90).

1.4.4 Itô–Doeblin formula

In this subsection we introduce the Itô–Doeblin formula that can be seen as the chain rule of stochastic calculus. We frequently make use of the Itô–Doeblin formula in the following sections and in the chapters of numerical methods and computer experiments.

We start with the definition of an important concept in stochastic calculus:

Definition 1.4.17 (cf. Shreve (2004, Definition 3.4.1); cf. Øksendal (2006, p. 19))**.** Let $X = (X_t)_t, Y = (Y_t)_t$ be real-valued stochastic processes on $(\Omega, \mathcal{F}, \mathbb{P})$. The **cross variation** of X and Y is defined as the real number

$$[X,Y](T,\omega) := \lim_{\|\Pi\|\to 0} \sum_{j=1}^{n-1} (X(t_{j+1},\omega) - X(t_j,\omega)) \cdot (Y(t_{j+1},\omega) - Y(t_j,\omega)), \tag{1.4.8}$$

where $\Pi = \{t_1, \ldots, t_n\}$; $0 = t_1 < t_2 < \ldots < t_n = T$ is a partition, $\|\Pi\| := \max_i \{t_{i+1} - t_i\}$ and the limit in (1.4.8) is regarded as limit in probability. The **quadratic variation** of X is defined as $[X,X](T,\omega)$, i.e.,

$$[X,X](T,\omega) := \lim_{\|\Pi\|\to 0} \sum_{j=1}^{n-1} (X(t_{j+1},\omega) - X(t_j,\omega))^2. \tag{1.4.9}$$

Remark 1.4.18.

(i) Let $f : [0,\infty) \to \mathbb{R}$ be continuously differentiable. If we conceive f as a the stochastic process defined by $(t,\omega) \mapsto f(t)$, it is easy to show that $[f,f](T,\omega) = 0$ for all $T \geq 0$ and $\omega \in \Omega$. That is the reason why quadratic variation is of no importance in ordinary calculus. Theorem 1.4.19 shows that a Wiener process has *non-zero*

quadratic variation. One might say that a main difference between ordinary and stochastic calculus is the concept of quadratic variation.

(ii) The quadratic variation of a stochastic process is itself again a stochastic process.

Theorem 1.4.19 (cf. SHREVE (2004, Theorem 3.4.3)). *Let $(W_t)_t = (W_t^1, \ldots, W_t^m)$ be a m-dimensional Wiener process. Then $[W^i, W^j](T) = T \cdot \delta_{ij}$ almost surely and for all $T \geq 0$, where δ_{ij} denotes the Kronecker Delta.*

Proof. See SHREVE (2004, p. 102) for the case $i = j$. This proof shows that (1.4.9) converges in **mean square**, i.e., in $L^2(\Omega)$. As mean square convergence implies convergence in probability the proof can be used for our definition, too.
See SHREVE (2004, p. 165) for the case $i \neq j$. □

Remark 1.4.20. There exists also a converse of Theorem 1.4.19 which is related to the recognition of a Wiener process and is known as **Lévy's theorem**, cf. SHREVE (2004, Theorem 4.6.4). We state it in the one-dimensional version: Let $(M_t)_t$ be a martingale w.r.t. a filtration $\mathcal{A}_t$ satisfying $M_0 = 0$ a.s.. If $(M_t)_t$ has continuous sample paths and satisfies $[M, M](t) = t$, then $(M_t)_t$ is a Wiener process.

Theorem 1.4.21 (cf. SHREVE (2004, Theorem 4.3.1)). *Let $\Delta \in \mathcal{V}_{\mathcal{A}_t}^{1\times 1}$ and $\mathcal{I}[\Delta](t,\omega) := \int_0^t \Delta(s,\omega) dW_s(\omega)$. Then*

$$[\mathcal{I}[\Delta], \mathcal{I}[\Delta]](t,\omega) = \int_0^t \Delta(s,\omega)^2 d\lambda(s). \tag{1.4.10}$$

Proof. See SHREVE (2004, p. 131) for a proof for *elementary* $\Delta \in \mathcal{V}_{\mathcal{A}_t}^{1\times 1}$. The **standard machinery** (cf. SHREVE (2004, proof of Theorem 1.5.1)) extends the result to the whole class $\mathcal{V}_{\mathcal{A}_t}^{1\times 1}$. □

Remark 1.4.22. We could write $[W, W](T) = T$ *formally* as **differentials**

$$dW_t \, dW_t = dt,$$

meaning that a Wiener process accumulates quadratic variation at rate one per unit time. By using this formal notation we could also rewrite (1.4.10) in differential form

$$\Delta_t dW_t \, dW_t = \Delta_t^2 dt. \tag{1.4.11}$$

Similarly we could rewrite $[W^i, W^j](T) = 0$ formally in differential form as

$$dW_t^i \, dW_t^j = 0.$$

Proposition 1.4.23. *Let $(W_t)_t$ be a one-dimensional Wiener process and $T \geq 0$. Then*

$$\lim_{\|\Pi\|\to 0} \sum_{j=1}^{n-1} \left(W(t_{j+1},\omega) - W(t_j,\omega)\right)(t_{j+1}-t_j) = 0, \tag{1.4.12}$$

$$\lim_{\|\Pi\|\to 0} \sum_{j=1}^{n-1} (t_{j+1}-t_j)^2 = 0, \tag{1.4.13}$$

where $\Pi = \{t_1,\ldots,t_n\}$, $0 = t_1 < t_2 < \ldots < t_n = T$ is a partition and the limit in (1.4.12) *and* (1.4.13) *holds as limit in probability and as limit in mean-square sense, i.e., w.r.t. the norm of $L^2(\Omega)$.*

Proof. We only give a proof for (1.4.12), since (1.4.13) is proved analogously.
We set

$$Q_\Pi(\omega) := \sum_{j=1}^{n-1} \left(W(t_{j+1},\omega) - W(t_j,\omega)\right)(t_{j+1}-t_j)$$

and deduce by Theorem 1.2.5 that

$$\begin{aligned} \mathbb{E}[Q_\Pi] &= 0, \\ \mathbb{V}\mathrm{ar}\left[\left(W(t_{j+1},\cdot) - W(t_j,\cdot)\right)(t_{j+1}-t_j)\right] &= (t_{j+1}-t_j)^2. \end{aligned}$$

Thus we conclude by the independence of the increments $W(t_{j+1},\cdot) - W(t_j,\cdot)$

$$\begin{aligned} \mathbb{V}\mathrm{ar}[Q_\Pi] = \sum_{j=1}^{n-1} \mathbb{V}\mathrm{ar}\left[\left(W(t_{j+1},\omega) - W(t_j,\omega)\right)(t_{j+1}-t_j)\right] = \sum_{j=1}^{n-1} (t_{j+1}-t_j)^2 \leq \\ \leq \|\Pi\| \cdot T. \end{aligned}$$

As $\|\Pi\| \to 0$ we see that $\mathbb{V}\mathrm{ar}[Q_\Pi] = \mathbb{E}\left[(Q_\Pi)^2\right] \to 0$, i.e., we proved convergence in $L^2(\Omega)$. As convergence in mean-square sense also implies convergence in probability the proof is finished. □

Remark 1.4.24. Equations (1.4.12) and (1.4.13) can be formally rewritten in differential form as

$$\begin{aligned} dW_t\, dt &= 0, \\ dt\, dt &= 0. \end{aligned}$$

Definition 1.4.25 (cf. Kloeden & Platen (2000, p. 96))**.** Let $(W_t)_t$ be a given m-dimensional Wiener process and let $e = (e^i)_i : [0,T]\times\Omega \to \mathbb{R}^n$, $f = (f^{i,j})_{i,j} : [0,T]\times\Omega \to \mathbb{R}^{n\times m}$ be two given stochastic processes satisfying $\sqrt{|e^i|}$, $f^{i,j} \in \mathcal{W}_{\mathcal{A}_t}$ for $i = 1,\ldots,n$,

$j = 1, \ldots, m$. Then we call the stochastic process $(X_t)_{t\in[0,T]}$ defined w.p.1 by

$$X_t(\omega) - X_s(\omega) = \int_s^t e(u,\omega)d\lambda(u) + \int_s^t f(u,\omega)dW_u(\omega) \tag{1.4.14}$$

for any $0 \leq s \leq t \leq T$ a **stochastic differential** or **Itô process**.

Remark 1.4.26.

(i) We can rewrite (1.4.14) formally as

$$dX_t = e(t,\omega)dt + f(t,\omega)dW_t(\omega), \tag{1.4.15}$$

having in mind that (1.4.15) is just a shorter symbolic way of writing (1.4.14) with – if not stated otherwise – $0 \leq s \leq t \leq T$.

(ii) Our assumptions upon e and f assure that the integrals in (1.4.14) exist w.p.1. The conditions upon e can also be weakened, cf. Øksendal (2006, Definition 4.1.1).

Let $(X_t)_t$ be an Itô process and $U : [0,T] \times \mathbb{R}^n \to \mathbb{R}$ be a sufficiently smooth function. We are interested in smooth changes of $(X_t)_t$ under U and want to rewrite them again as an Itô process, i.e., we want to compute $dU(t, X_t)$ and interpret $dU(t, X_t)$ as an Itô process of the form (1.4.15):

Theorem 1.4.27 (**Itô–Doeblin formula**; cf. Kloeden & Platen (2000, p. 97))**.** *Let $(X_t)_t$ be an Itô process as in* (1.4.14) *with m-dimensional Wiener process $(W_t)_t$ and coefficients $e_t(\cdot) := (e_t^i(\cdot))_i := (e^i(t,\cdot))_i$, $f_t(\cdot) := (f_t^{i,j}(\cdot))_{i,j} := (f^{i,j}(t,\cdot))_{i,j}$, $i = 1, \ldots, n$, $j = 1, \ldots, m$ and let $U : [0,T]\times\mathbb{R}^n \to \mathbb{R}$ be a C^2 map $(t,x) \mapsto U(t,x)$. Then the stochastic process*

$$Y_t(\omega) = U(t, X_t(\omega))$$

is itself an Itô process satisfying

$$\begin{aligned} dY_t = &\left(\frac{\partial U}{\partial t} + \sum_{i=1}^n e_t^i \frac{\partial U}{\partial x_i} + \frac{1}{2}\sum_{i,k=1}^n \sum_{j=1}^m f_t^{i,j} f_t^{k,j} \frac{\partial^2 U}{\partial x_i \partial x_k} \right) dt + \\ &+ \sum_{j=1}^m \sum_{i=1}^n f_t^{i,j} \frac{\partial U}{\partial x_i} dW_t^j, \end{aligned} \tag{1.4.16}$$

where the partial derivatives of U are evaluated at (t, X_t). A shorter way of writing (1.4.16) *would be*

$$dY_t = \left(\frac{\partial U}{\partial t} + e_t^\top \nabla U + \frac{1}{2} tr\left(f_t f_t^\top \nabla(\nabla U) \right) \right) dt + \nabla U^\top f_t dW_t,$$

cf. Kloeden & Platen *(2000, p. 97), where* ∇ *denotes the gradient operator with respect to the spatial variable* x, $\top$ *denotes the vector or matrix transpose and tr is the trace of the inscribed matrix.* $\nabla(\nabla U)$ *thus denotes the Hessian matrix of* U.

Proof. See Kloeden & Platen (2000, pp. 92–95) for the one-dimensional case $U : [0,T] \times \mathbb{R} \to \mathbb{R}$. As the key ingredients of the proof are just the Taylor expansion of U and the results of Theorem 1.4.19, Proposition 1.4.23, the proof can by easily extended to the higher dimensional case. □

Remark 1.4.28.

(i) The Itô–Doeblin formula can be extended to vector-valued functions U. Let $U : [0,T] \times \mathbb{R}^m \to \mathbb{R}^p$ be a C^2 map $(t,x) \mapsto U(t,x) := (U_1(t,x), \ldots, U_p(t,x))$. We can deduce a formula for $dU(t, X_t)$ by applying the Itô–Doeblin formula to each component U_i, $i = 1, \ldots, p$.

(ii) Another way of writing (1.4.16) with $X_t = (X_t^1, \ldots, X_t^n)$ would be

$$dY_t = \frac{\partial U}{\partial t}(t, X_t)dt + \sum_{i=1}^{n} \frac{\partial U}{\partial x_i}(t, X_t)dX_t^i + \frac{1}{2}\sum_{i,k=1}^{n} \frac{\partial^2 U}{\partial x_i \partial x_k}(t, X_t)dX_t^i \, dX_t^k,$$

where $dX_t^i \, dX_t^k$ denotes the cross variation of $(X_t^i)_t$ and $\left(X_t^k\right)_t$. The formal notation that we have introduced in Remark 1.4.22 and 1.4.24 proves to be handy in calculating $dX_t^i \, dX_t^k$:

$$\begin{aligned} dX_t^i \, dX_t^k &= (e^i dt + \sum_{j=1}^{m} f^{ij} dW_t^j)(e^k dt + \sum_{l=1}^{m} f^{k,l} dW_t^l) \\ &= e^i e^k \underbrace{dt \, dt}_{=0} + \sum_{j,l=1}^{m} f^{i,j} f^{k,l} \underbrace{dW_t^l \, dW_t^j}_{=\delta_{lj} dt} + \sum_{j=1}^{m} f^{i,j} \underbrace{dt \, dW_t^j}_{=0} + \sum_{l=1}^{m} f^{k,l} \underbrace{dt \, dW_t^l}_{=0} \\ &= \sum_{j=1}^{m} f^{i,j} f^{k,j} dt. \end{aligned}$$

To justify the calculations that we have just made, one could compute directly $\left[X_t^i, X_t^k\right] = dX_t^i \, dX_t^k$ using Theorem 1.4.19 and Proposition 1.4.23.

(iii) Let $U(t,x) = \frac{1}{2}x^2$ for $x \in \mathbb{R}$ and $(W_t)_t$ be a one-dimensional Wiener process. We derive from the Itô–Doeblin formula for $X_t = W_t$

$$dU(W(t)) = \frac{1}{2}(W_t^2 - W_s^2) = \int_s^t W_u dW_u + \frac{1}{2}\int_s^t du,$$

and thus $\int_s^t W_u dW_u = \frac{1}{2}(W_t^2 - W_s^2) - \frac{1}{2}(t-s)$. From the perspective of ordinary calculus the term $-\frac{1}{2}(t-s)$ on the right side is amazing as it does not appear in the case of normal Lebesgue-Stieltjes integration.

1.4.5 Stratonovich integral

In this subsection we briefly formulate the definition of the Stratonovich integral and remark the Stratonovich version of the Itô–Doeblin formula, which resembles the rules of deterministic calculus.

For a given one-dimensional Wiener process $(W_t)_t$ we have motivated at the beginning of Section 1.4 the Stratonovich integral

$$\int_S^T \Delta(t,\omega) \circ dW_t(\omega)$$

of a stochastic process $(\Delta_t)_t$ to be the (mean-square) limit of

$$\sum_{i=1}^{n-1} \Delta\left(\frac{1}{2}(t_{i+1}+t_i),\omega\right)\left(W_{t_{i+1}} - W_{t_i}\right) \tag{1.4.17}$$

as $\|\Pi\| \to 0$ for a partition $\Pi = \{S = t_1, \ldots, T = t_n\}$. However, since (1.4.17) is not easy to work with, we now give another definition of the Stratonovich integral that shall turn out to be equivalent to (1.4.17) for integrands with continuously differentiable sample paths:

Definition 1.4.29 (cf. Kloeden & Platen (2000, Section 3.5))**.** Let $\Delta \in \mathcal{V}_{\mathcal{A}_t}^{1\times 1}(S,T)$ and $(W_t)_t$ be a one-dimensional Wiener process. We define the Stratonovich integral

$$\int_S^T \Delta(t,\omega) \circ dW_t(\omega)$$

to be the mean-square limit – i.e., w.r.t to the norm in $L^2(\Omega)$ – of

$$\int_S^T \Delta(t,\omega) \circ dW_t(\omega) := \lim_{\|\Pi\|\to 0} \sum_{i=1}^{n-1} \left(\frac{1}{2}\Delta\left(t_i,\omega\right) + \frac{1}{2}\Delta\left(t_{i+1},\omega\right)\right)\left(W_{t_{i+1}} - W_{t_i}\right).$$

For $\Delta \in \mathcal{V}_{\mathcal{A}_t}^{m\times n}(S,T)$ and n-dimensional Wiener process $(W_t)_t$ we define the Stratonovich integral in the same was as in (1.4.6), resp. (1.4.7).

We now show that this definition is equivalent to our motivation (1.4.17) for a stochas-

tic process Δ that has w.p.1 continuously differentiable sample paths and satisfies

$$\sup_{t' \in [S,T]} \left\| \frac{\partial \Delta}{\partial t}(t', \cdot) \right\|_{L^2(\Omega)} \leq C$$

for some constant $C \geq 0$. In this case we obtain by Taylor's formula

$$\Delta\left(\frac{1}{2}(t_{i+1} + t_i), \omega\right) = \frac{1}{2}\Delta\left(t_i, \omega\right) + \frac{1}{2}\Delta\left(t_{i+1}, \omega\right) + \frac{1}{2}\left(\frac{\partial \Delta}{\partial t}(\xi_{i,1}, \omega) - \frac{\partial \Delta}{\partial t}(\xi_{i,2}, \omega)\right)(t_{i+1} - t_i)$$

for appropriate $\xi_{i,1}, \xi_{i,2} \in [t_i, t_{i+1}]$. By our assumption and Proposition 1.4.23 we deduce

$$\begin{aligned}
&\left\| \sum_{i=1}^{n-1} \left(\frac{\partial \Delta}{\partial t}(\xi_{i,1}, \omega) - \frac{\partial \Delta}{\partial t}(\xi_{i,2}, \omega)\right)(t_{i+1} - t_i)\left(W_{t_{i+1}} - W_{t_i}\right) \right\|_{L^2(\Omega)} \leq \\
&\leq 2C \sum_{i=1}^{n-1} \left\| (t_{i+1} - t_i)\left(W_{t_{i+1}} - W_{t_i}\right) \right\|_{L^2(\Omega)} = \\
&\overset{(*)}{=} 2C \left\| \sum_{i=1}^{n-1} (t_{i+1} - t_i)\left(W_{t_{i+1}} - W_{t_i}\right) \right\|_{L^2(\Omega)} = \\
&\overset{1.4.23}{=} 0
\end{aligned}$$

where we have used in $(*)$ the independence of the increments and applied Theorem 1.4.23 componentwisely. I.e., Definition 1.4.29 and our motivation (1.4.17) are equivalent.

Remark 1.4.30. Let h be a function such that $(h(W_t))_t \in \mathcal{V}_{\mathcal{A}_t}^{1\times 1}(0,T)$ and let U be an anti-derivative of h, i.e., $U'(x) = h(x)$. Then for the Stratonovich integral the rules of classical calculus apply (cf. Kloeden & Platen (2000, p. 101)):

$$\int_0^T h(W_t) \circ dW_t = U(h(W_T)) - U(h(W_0)).$$

1.5 Stochastic differential equations

For given $b : [t_0, t] \times \mathbb{R}^n \to \mathbb{R}^n, \sigma : [t_0, t] \times \mathbb{R}^n \to \mathbb{R}^{n\times m}$, $t_0 \in [0, T]$ and given random variable X_{t_0} on $(\Omega, \mathcal{F}, \mathbb{P})$, an **Itô stochastic differential equation**, short **SDE**, is an equation

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t \tag{1.5.1}$$

that we interpret as the integral equation

$$X_t = X_{t_0} + \int_{t_0}^{t} b(u, X_u)d\lambda(u) + \int_{t_0}^{t} \sigma(u, X_u)dW_u. \tag{1.5.2}$$

The stochastic integral on the right side of the equation is an Itô integral. Similarly we define a **Stratonovich stochastic differential equation** as

$$dX_t = b(t, X_t) + \sigma(t, X_t) \circ dW_t, \tag{1.5.3}$$

interpreting it as the integral equation

$$X_t = X_{t_0} + \int_{t_0}^{t} b(u, X_u)d\lambda(u) + \int_{t_0}^{t} \sigma(u, X_u) \circ dW_u,$$

where the stochastic integral on the right side of the equation is a Stratonovich integral. We call the function b the **drift (vector)** and σ the **diffusion (matrix)**.
The random variable X_{t_0} is called the **initial value** and is known. We are looking for a stochastic process $(X_t)_t$ that satisfies equation (1.5.1), resp. (1.5.3).

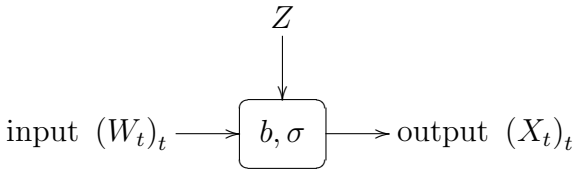
Similarly to our approach to stochastic integration we first focus on Itô stochastic differential equations. We discuss the difference between weak and strong solutions and derive an existence and uniqueness theorem for Itô SDEs. We also discuss connections between diffusion processes and solutions of Itô SDEs. In the last subsection we shall handle Stratonovich stochastic differential equations and state their relation to Itô SDEs since this shall lead immediately to an existence and uniqueness theorem for Stratonovich SDEs, too.

1.5.1 Weak and strong solutions

For a SDE (1.5.1) there are two concepts of a solution, we distinguish between so called "weak" and "strong" solutions, depending if we are free to choose the Wiener process $(W_t)_t$ in (1.5.1) and a filtration $\mathcal{A} := (\mathcal{A}_t)_t$ such that the solution is adapted to $\mathcal{A}$, or if we have to find a solution for each given Wiener process such that the solution is adapted to the filtration $(\mathcal{F}_t)_t$ *generated* by $(W_t)_t$:

Definition 1.5.1 (cf. Øksendal (2006, p. 72))**.** Let $(W_t)_t$ be an arbitrary Wiener process on $(\Omega, \mathcal{F}, \mathbb{P})$, the filtration $(\mathcal{F}_t)_t$ generated by $(W_t)_t$ be given and $b : [t_0, t] \times \mathbb{R}^n \to \mathbb{R}^n, \sigma : [t_0, t] \times \mathbb{R}^n \to \mathbb{R}^{n\times m}$, $t_0 \in [0, T]$. Also, let Z be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ independent of W_t for all $t \geq 0$. We call a stochastic process $X = (X_t)_t$ a **strong solution** of (1.5.1) with initial value $X_{t_0} = Z$, if it satisfies (1.5.1) with $X_{t_0} = Z$ and if X is $\mathcal{F}_t^Z$-adapted, where $\mathcal{F}_t^Z$ denotes the filtration generated by Z and W_s for $0 \leq s \leq t$.

We thus call $X = (X_t)_t$ a strong solution if we are given a Wiener process in advance and are able to construct a $\mathcal{F}_t^Z$-adapted stochastic process $(X_t)_t$. A strong solution X can be understood as the "output" of a dynamical system described by the drift and diffusion coefficients b, σ, whose "input" is $(W_t)_t$ and the initial value Z, cf. KARATZAS & SHREVE (2004, p. 286):



The *principle of causality* for dynamical systems, cf. KARATZAS & SHREVE (2004, p. 286), requires the solution X_t at fixed time t to depend only on Z and the values of the input $(W_s)_{0 \leq s \leq t}$. This principle finds its mathematical expression in the fact that X is $\mathcal{F}_t^Z$-adapted, cf. KARATZAS & SHREVE (2004, p. 286).

Definition 1.5.2 (cf. ØKSENDAL (2006, p. 72))**.** Let $b : [t_0, t] \times \mathbb{R}^n \to \mathbb{R}^n, \sigma : [t_0, t] \times \mathbb{R}^n \to \mathbb{R}^{n \times m}$, $t_0 \in [0, T]$ and let Z be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. We call a stochastic process $X = (X_t)_t$ a **weak solution** of (1.5.1) with initial value $X_{t_0} = Z$, if there is a Wiener process $(W_t)_t$ and a filtration $(\mathcal{A}_t)$ such that condition (ii)' on page 26 holds for $f(t, \omega) = X(t, \omega)$ and X satisfies (1.5.1) with $X_{t_0} = Z$.

A weak solution $(X_t)_t$ need not be $\mathcal{F}_t^Z$-adapted, as $\mathcal{F}_t^Z \subseteq \mathcal{A}_t$. However, whatever extra information is needed to compute X_t at time t must be independent of $W_u - W_t$ for $u > t$, because $(W_t)_t$ is still a $\mathcal{A}_t$-martingale. Clearly every strong solution is also a weak solution.

We give the following example of a one-dimensional SDE that has only a weak solution but no strong solution. It is known as the **Tanaka equation**, cf. ØKSENDAL (2006, Example 5.3.2), KARATZAS & SHREVE (2004, Example 5.3.5):

$$dX_t = \operatorname{sgn}(X_t) dW_t; \quad X_0 = 0, \tag{1.5.4}$$

where sgn denotes the signum function, i.e., $\operatorname{sgn}(x) = 1$ for $x \geq 0$ and $\operatorname{sgn}(x) = -1$ for $x < 0$. Let $\left(\widetilde{W}\right)_t$ be a one-dimensional Wiener process and $\left(\widetilde{\mathcal{F}_t}\right)_t$ be the filtration generated by $\left(\widetilde{W}\right)_t$. We define

$$Y_t = \int_0^t \operatorname{sgn}(\widetilde{W_s}) d\widetilde{W_s},$$

and it can be shown, cf. ØKSENDAL (2006, Tanka Formula, 4.3.12), that $(Y_t)_t$ has to be measurable w.r.t. $(\mathcal{G}_t)_t$, where $(\mathcal{G}_t)_t$ denotes the filtration generated by $\left(|\widetilde{W_t}|\right)_t$. For

any measurable real-valued function f that can attain with positive probability negative values, we get $\mathcal{H}_{|f|} \subsetneqq \mathcal{H}_f$. Thus we conclude that $\mathcal{G}_t$ is strictly contained in $\widetilde{\mathcal{F}_t}$.
Now if we suppose that $X = (X_t)_t$ is a strong solution of (1.5.4), then because of $dX_t dX_t = dt$ we can conclude by Remark 1.4.20 that $(X_t)_t$ is a Wiener process. Let $(\mathcal{M}_t)_t$ denote the filtration generated by $(X_t)_t$. Then we can rewrite (1.5.4) as

$$dW_t = \operatorname{sgn}(X_t) dX_t$$

and by the above arguments applied to $Y_t = W_t$ and $\widetilde{W_t} = X_t$, we conclude that the filtration $(\mathcal{F}_t)_t$ generated by W_t is strictly contained in $(\mathcal{M}_t)_t$, which contradicts that X is a strong solution. Hence there are no strong solutions of (1.5.4).
On the other hand we can choose $X = (X_t)_t$ to be *any* one-dimensional Wiener process $(W_t)_t$ and denote by $(\mathcal{F}_t)_t$ the filtration generated by $(W_t)_t$. Then we define

$$\widetilde{W_t} = \int_0^t \operatorname{sgn}(W_s) dW_s = \int_0^t \operatorname{sgn}(X_s) dX_s.$$

Again we see that X satisfies the SDE

$$dX_t = \operatorname{sgn}(X_t) d\widetilde{W_t}; \quad X_0 = 0., \tag{1.5.5}$$

and by Remark 1.4.20 we conclude that $\left(\widetilde{W_t}\right)_t$ is a Wiener process. If we denote by $\left(\widetilde{\mathcal{F}_t}\right)_t$ the filtration generated by $\left(\widetilde{W_t}\right)_t$, we deduce by the above argument that $\widetilde{\mathcal{F}_t} \subsetneqq \mathcal{F}_t$. Hence X is a weak solution of (1.5.4).

1.5.2 Existence and uniqueness

In this subsection we formulate an existence and uniqueness result for Itô stochastic differential equations. For a stochastic process that is a solution of (1.5.1) it might not be clear what we mean by uniqueness. We give the following uniqueness definition that is based on sample path equivalence:

Definition 1.5.3 (cf. Kloeden & Platen (2000, p. 128))**.** Let X_t and $\tilde{X}_t$ two solutions of (1.5.1) with initial value X_{t_0} that are w.p.1 sample path continuous. We call the solution of the SDE (1.5.1) with initial value X_{t_0} on $[t, T]$ **pathwise unique** or **strongly unique** if any two such solutions X_t, $\tilde{X}_t$ are **sample path equivalent** on $[t, T]$, i.e.,

$$\mathbb{P}\left(\sup_{t_0 \le t \le T} \left| X_t - \tilde{X}_t \right| \right) = 0.$$

Remark 1.5.4.

(i) We shall see that under reasonable assumptions upon the coefficients b and σ there is always a continuous version of a solution $X = (X_t)_t$ of the SDE (1.5.1). Basically we only have to assure that the integrals on the right side of (1.5.1) make sense, then Theorem 1.4.12 assures the existence of a sample path continuous version of X.

(ii) There is also a weaker form of uniqueness: We call the solution of (1.5.1) **weakly unique** if any two solutions X_t and $\tilde{X}_t$ are identical in law, i.e., they have the same finite dimensional distributions. However throughout this exposition we refer to uniqueness as in Definition 1.5.3.

W.l.o.g. we can set t_0 of (1.5.1) to be 0. The proof of the uniqueness works roughly the same way as for ordinary differential equations: We require a Lipschitz condition for the spatial coordinates of b and σ. By using the Gronwall inequality we are able to prove the uniqueness. The proof of existence needs more effort but again works the same way as in the case of ordinary differential equations.
We state the following existence and uniqueness result following Øksendal (2006, Section 5.2). Kloeden & Platen (2000, Section 4.5) deduce an existence and uniqueness result under roughly the same assumptions. The only difference lies in the measurability condition upon the initial value X_0. Whereas Øksendal (2006) only requires X_0 to be independent of the prescribed Wiener process $(W_t)_t$, Kloeden & Platen (2000) require X_0 to be $\mathcal{F}_0$ measurable. As $W_0 = 0$ w.p.1 it is $\{\Omega, \emptyset\}$-measurable, thus X_0 is itself a.s. constant and thus independent of the Wiener process $(W_t)_t$. As the requirements of Øksendal (2006) are less restricting, we state the theorem in this version:

Theorem 1.5.5. (**Existence and uniqueness theorem for Itô SDEs**, cf. Øksendal (2006, Theorem 5.2.1)) Let $(W_t)_t$ be a given m-dimensional Wiener process on $(\Omega, \mathcal{F}, \mathbb{P})$ and let $T \geq 0$ and $b : [0,T] \times \mathbb{R}^n \to \mathbb{R}^n$, $\sigma : [0,T] \times \mathbb{R}^n \to \mathbb{R}^{n \times m}$ be measurable functions satisfying a linear growth bound

$$\|b(t,x)\|_2 + \|\sigma(t,x)\|_2 \leq C(1 + \|x\|_2); \quad \forall x \in \mathbb{R}^n, t \in [0,T] \tag{1.5.6}$$

for some constant $C \geq 0$ – where we set $\|\sigma(t,x)\|_2 = \sum_{i,j} \sigma^{i,j}(t,x)^2$ – and also a Lipschitz condition for the spatial coordinates

$$\|b(t,x) - b(t,y)\|_2 + \|\sigma(t,x) - \sigma(t,y)\|_2 \leq D \|x - y\|_2; \quad \forall x, y \in \mathbb{R}^n, t \in [0,T] \tag{1.5.7}$$

for some constant $D \geq 0$. Also let $Z : \Omega \to \mathbb{R}^n$ be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ that is independent of the σ-algebra $\mathcal{F}_\infty^{(m)}$ generated by $(W_t)_t$ for $s \geq 0$ and such that $Z \in L^2(\Omega)$.

Then the Itô stochastic differential equation

$$dX_t = b(t, X_t) + \sigma(t, X_t)dW_t; \quad 0 \leq t \leq T, X_0 = Z \tag{1.5.8}$$

has a unique sample path continuous solution $X = (X_t)_{0 \leq t \leq T}$ with the property that X is adapted to the filtration $\mathcal{F}_t^Z$ generated by Z and $(W_s)_{0 \leq s \leq t}$ and

$$\mathbb{E}\left[\int_0^T \|X_t\|_2^2 \, d\lambda(t)\right] < \infty. \tag{1.5.9}$$

Proof. See ØKSENDAL (2006, pp. 68–71). □

Remark 1.5.6.

(i) Equation (1.5.6) assures that the solution $(X_t)_{0 \leq t \leq T}$ really exists for all $T \geq 0$: It prevents each path $(X_t(\omega))_{0 \leq t \leq T}$, $\omega \in \Omega$ to "explode" in finite time.

(ii) Equation (1.5.7) assures the uniqueness of the solution $(X_t)_{0 \leq t \leq T}$.

(iii) Equation (1.5.9) implies by the Theorem of Fubini, Theorem 1.1.8, that the solution $X = (X_t)_{0 \leq t \leq T}$ fulfills $X \in L^2([0,T] \times \Omega)$. Because $\mathbf{1} : (t, \omega) \mapsto 1$ is also in $L^2([0,T] \times \Omega)$, we deduce that $X \in L^1([0,T] \times \Omega)$. Thus the unique solution X of (1.5.8) satisfies

$$\mathbb{E}\left[\int_0^T \|X_t\|_2 \, d\lambda(t)\right] = \int_0^T \mathbb{E}\left[\|X_t\|_2\right] d\lambda(t).$$

(iv) The solution X of (1.5.8) is in fact a *strong* solution since it can be constructed for each Wiener process $(W_t)_t$ and as it is adapted to $\mathcal{F}_t^Z$.

1.5.3 Connection to diffusion processes

We now want to study the connections between the solution of a one-dimensional SDE

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t \tag{1.5.10}$$

and diffusion processes. We want to find conditions such that the solution of (1.5.10) is in fact a diffusion process and vice versa.

Theorem 1.5.7. (**strong solution as diffusion process**, cf. KLOEDEN & PLATEN (2000, Theorem 4.6.1)) Let b and σ of (1.5.10) and the initial value Z satisfy the conditions of Theorem 1.5.5, i.e., there exists a unique strong solution $(X_t)_t$ of (1.5.10) on $[t_0, T]$ with initial value $X_{t_0} = Z$. Additionally let b and σ be continuous functions. Then $(X_t)_t$ is a diffusion process with drift b and diffusion coefficient σ^2.

Proof. See Kloeden & Platen (2000, Theorem 4.6.1). □

Remark 1.5.8.

(i) It is also possible to prove Theorem 1.5.7 in the case of a higher dimensional SDE (1.5.1) if both coefficients $b : [t_0, T] \times \mathbb{R}^n \to \mathbb{R}^n$, $\sigma : [t_0, T] \times \mathbb{R}^n \to \mathbb{R}^{n\times m}$ are sufficiently smooth, e.g. they satisfy the assumptions of Theorem 1.5.5, cf. Kloeden & Platen (2000, Section 4.8). Then $(X_t)_t$ is a diffusion process with drift vector b and diffusion matrix $D = \sigma\sigma^\top$. Generally, one has to be aware that the diffusion matrix D is only positive semi-definite rather than positive definite as required in Definition 1.3.1. As mentioned in Remark 1.3.2, it is possible to weaken the definition of a diffusion process to a process with positive semi-definite diffusion matrix. If the transition probabilities have transition densities, they satisfy the Kolmogorov forward equation (1.3.4) and Kolmogorov backward equation (1.3.5). A positive semi-definite diffusion matrix may lead to singularities of the transition densities, cf. Kloeden & Platen (2000, p. 152).

(ii) Theorem 1.5.7 states that a strong solution of a SDE with continuous drift and diffusion is a diffusion process.

We now want to turn to the converse problem: Let $(Y_t)_t$ be a one-dimensional diffusion process with drift b and diffusion coefficient σ^2. We try to find a SDE (1.5.10) such that $(Y_t)_t$ is a solution of it. On a superficial level we can of course just take *any* Wiener process $(W_t)_t$ and consider the SDE

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t$$

with initial value $X_{t_0} = Y_{t_0}$. Provided that the coefficients b and σ are sufficiently smooth, there is a unique solution $(X_t)_t$ of this SDE for each Wiener process. Then $(X_t)_t$ has the same probability law as $(Y_t)_t$, i.e., the induced measures $\mathbb{P}_{X_t}$ and $\mathbb{P}_{Y_t}$ are the same. But in general, the sample paths of $(Y_t)_t$ are not equivalent to $(X_t)_t$. To guarantee the sample path equivalence, a specific Wiener process $(W_t)_t$ has to be constructed from the given drift and diffusion coefficients:

Theorem 1.5.9. (**diffusion process as weak solution**, cf. Kloeden & Platen (2000, Theorem 4.7.1)) Let $(Y_t)_{t\in[0,T]}$ be a one-dimensional diffusion process with drift coefficient $b(t, y)$ and strictly positive diffusion coefficient $(\sigma(t, y))^2$ satisfying for all $y \in \mathbb{R}$ and $t \in [0, T]$ the following conditions:

(a) $b(t, y)$ is continuous and

$$|b(t, y)| \leq K(1 + |y|)$$

for some constant $K \geq 0$.

(b) $\sigma(t,y)$ is continuous, $\frac{1}{\sigma(t,y)}$ is bounded and the partial derivatives $\frac{\partial\sigma}{\partial t}$, $\frac{\partial\sigma}{\partial y}$ are continuous and bounded.

(c) There exists a function $\psi(y) > 1 + |y|$ such that

(i) $\sup_{0\leq t\leq T} \mathbb{E}\left[\psi(Y_t)\right] < \infty$,

(ii) $\mathbb{E}\left[|Y_t - Y_s| \,|Y_s = y\right] + \mathbb{E}\left[|Y_t - Y_s|^2 \,|Y_s = y\right] \leq (t-s)\psi(y)$,

(iii) $\mathbb{E}\left[|Y_t| \,|Y_s = y\right] + \mathbb{E}\left[|Y_t|^2 \,|Y_s = y\right] \leq \psi(y)$.

We define a stochastic process $\left(\widetilde{W}_t\right)_t$ by

$$g(t,y) = \int_0^y \frac{1}{\sigma(t,x)} d\lambda(x), \tag{1.5.11a}$$

$$\bar{b}(t,z) = \left(\frac{\partial g}{\partial t} + b\frac{\partial g}{\partial y} + \frac{1}{2}\sigma^2\frac{\partial^2 g}{\partial y^2}\right)\left(t, g^{-1}(t,z)\right), \tag{1.5.11b}$$

$$Z_t = g(t, Y_t), \tag{1.5.11c}$$

$$\widetilde{W}_t = Z_t - Z_0 - \int_0^t \bar{b}(s, Z_s) d\lambda(s), \tag{1.5.11d}$$

where $y = g^{-1}(t,z)$ is the inverse of $z = g(t,y)$.
Then $\left(\widetilde{W}_t\right)_t$ is a Wiener process and and $(Y_t)_t$ is a solution of the stochastic differential equation

$$dY_t = b(t, Y_t)dt + \sigma(t, Y_t)d\widetilde{W}_t. \tag{1.5.12}$$

Proof. See Kloeden & Platen (2000, Theorem 4.7.1). □

Remark 1.5.10.

(i) The tricky task of Theorem 1.5.9 is to show that the process $(Z_t)_t$ defined by (1.5.11c) is a diffusion process with drift $\bar{b}$ and diffusion coefficient 1, and that $\left(\widetilde{W}_t\right)_t$ is indeed a Wiener process. Then equation (1.5.11d) is equivalent to the SDE

$$dZ_t = \bar{b}(t, Z_t)dt + 1d\widetilde{W}_t. \tag{1.5.13}$$

If we apply the Itô–Doeblin formula to $Z_t = g(t, Y_t)$ and compare the dt and $d\widetilde{W}_t$ term to (1.5.13) we see that $(Y_t)_t$ is indeed a solution of (1.5.12).

(ii) Theorem 1.5.9 states that the diffusion process $(Y_t)_t$ is a weak solution of an SDE.

1.5.4 Comparison of Itô and Stratonovich SDEs

In this subsection we discuss the relation between Itô and Stratonovich SDEs. This connection also implies an existence and uniqueness result for Stratonovich SDEs.

For a Stratonovich SDE

$$dX_t = b(t, X_t) + \sigma(t, X_t) \circ dW_t \tag{1.5.14}$$

there might be some ambiguity how we define the Stratonovich integral $\sigma(t, X_t) \circ dW_t$ for a function $\sigma = \sigma(t, x)$ and a stochastic process $(X_t)_t$: We follow Kloeden & Platen (2000, Section 4.9) and define $\sigma(t, X_t) \circ dW_t$ for a *one-dimensional* Wiener process $(W_t)_t$ to be the mean-square limit

$$\int_0^T \sigma(t, X_t) \circ dW_t := \lim_{\|\Pi\| \to 0} \sum_{i=1}^{n-1} \sigma\left(t_i, \frac{1}{2}(X_{t_i} + X_{t_{i+1}})\right)\left(W_{t_{i+1}} - W_{t_i}\right). \tag{1.5.15}$$

We define the multi-dimensional version of (1.5.15) in the same way as in (1.4.6) and (1.4.7). To assure the existence of this limit we need $(t, \omega) \mapsto \sigma(t, X_t(\omega)) \in \mathcal{V}_{\mathcal{A}_t}^{m \times n}$ and require σ to be continuous in t and continuously differentiable w.r.t. x, cf. Kloeden & Platen (2000, Section 4.9). The following theorem states the connection between Itô and Stratonovich SDEs:

Theorem 1.5.11. *Let $(X_t)_t$ be a solution of the one-dimensional Itô SDE*

$$dX_t = \bar{b}(t, X_t) + \sigma(t, X_t) dW_t, \tag{1.5.16}$$

with b and σ continuous. Further let $h(t, x) : [0, T] \times \mathbb{R} \to \mathbb{R}$ be continuous in t and continuously differentiable w.r.t. x and let $(t, \omega) \mapsto h(t, X_t(\omega)) \in \mathcal{V}_{\mathcal{A}_t}^{1 \times 1}$. Then

$$\int_0^T h(t, X_t) \circ dW_t = \int_0^T h(t, X_t) dW_t + \frac{1}{2} \int_0^T \sigma(t, X_t) \frac{\partial h}{\partial x}(t, X_t) d\lambda(t). \tag{1.5.17}$$

Proof. See Kloeden & Platen (2000, pp. 156–157). □

Remark 1.5.12.

(i) Theorem 1.5.11 can be extended analogously to higher dimensions (cf. Kloeden & Platen (2000, p. 159)): Let $(X_t)_t$ be a solution of (1.5.16) with n-dimensional Wiener process $(W_t)_t$, m-dimensional drift $\bar{b}$ and $m \times n$-dimensional diffusion $\sigma = \left(\sigma^{j,k}\right)$. Further let $h(t, \mathbf{x}) : [0, T] \times \mathbb{R}^n \to \mathbb{R}^{m \times n}$, $h = \left(h^{j,k}\right)$ be continuous in t and continuously differentiable w.r.t. $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Also let $(t, \omega) \mapsto$

$h(t, X_t(\omega)) \in \mathcal{V}_{\mathcal{A}_t}^{m\times n}$. We define a m-dimensional vector valued function $c = (c^i)$ by

$$c^i(t, \mathbf{x}) := \frac{1}{2}\sum_{j=1}^{m}\sum_{k=1}^{n}\sigma^{j,k}(t, \mathbf{x})\frac{\partial h^{i,k}}{\partial x_j}(t, \mathbf{x}); \quad \text{for } i = 1, \dots, m.$$

Then the Itô and Stratonovich integrals of $h(t, X_t)$ are related by

$$\int_0^T h(t, X_t) \circ dW_t = \int_0^T h(t, X_t)dW_t + \int_0^T c(t, X_t)d\lambda(t). \tag{1.5.18}$$

(ii) Equations (1.5.17) resp. (1.5.18) also imply that the Stratonovich integral is in general not a martingale because (1.5.18) implies

$$\mathbb{E}\left[\int_0^T h(t, X_t) \circ dW_t\right] = \mathbb{E}\left[\int_0^T c(t, X_t)d\lambda(t)\right],$$

which is in general not equal to 0.

Theorem 1.5.11 now gives us the possibility to relate Itô and Stratonovich SDEs: Let $X = (X_t)_t \in \mathbb{R}^m$ satisfy the Itô SDE (1.5.16) with n-dimensional Wiener process, by setting $h \equiv \sigma$ we see that X satisfies the Stratonovich SDE

$$dX_t = b(t, X_t) + \sigma(t, X_t) \circ dW_t$$

with modified m-dimensional drift

$$b^i(t, \mathbf{x}) = \bar{b}^i(t, \mathbf{x}) - \frac{1}{2}\sum_{j=1}^{m}\sum_{k=1}^{n}\sigma^{j,k}(t, \mathbf{x})\frac{\partial \sigma^{i,k}}{\partial x_j}(t, \mathbf{x}); \text{ for } i = 1, \dots, m.$$

If we start with a Stratonovich integral of the form (1.5.14) the corresponding Itô SDE is

$$dX_t = \bar{b}(t, X_t) + \sigma(t, X_t)dW_t \tag{1.5.19}$$

with modified m-dimensional drift

$$\bar{b}^i(t, \mathbf{x}) = b^i(t, \mathbf{x}) + \frac{1}{2}\sum_{j=1}^{m}\sum_{k=1}^{n}\sigma^{j,k}(t, \mathbf{x})\frac{\partial \sigma^{i,k}}{\partial x_j}(t, \mathbf{x}); \text{ for } i = 1, \dots, m.$$

This connection implies the same existence and uniqueness result for Stratonovich SDEs, provided the drift and diffusion coefficients of (1.5.19) satisfy the conditions of Theorem 1.5.5.

Chapter 2

Numerical Methods

In this chapter we introduce numerical methods of solving Itô stochastic differential equations. In Section 2.1 we give a brief overview of numerical approaches to stochastic differential equations. As we are interested to find so called weak approximations of the solution X of an Itô stochastic differential equation, we focus in Section 2.2 and 2.3 on the development of time discrete approximations, i.e., one-step methods, and the so called Gaussian approximation. The one-step methods that we shall discuss are based on truncation of the so called Itô–Taylor expansion, which can be seen as a direct implementation of Taylor approximations for ordinary differential equations. Theorem 2.2.17 proves the asserted convergence order of truncated Itô–Taylor expansions as well as the one-step methods that we shall use in Chapter 3. The basic idea of Gaussian approximation is to consider the solution X of a SDE as being Gaussian. If the drift and diffusion coefficients are of a specific type, namely expolynomials, higher moments of X can be reduced to the determination of the first and second moments of X. This approach via Gaussian approximation leads to an ordinary differential equation, ODE, whose solution approximates the first and second moments of the solution of X. Whereas the first and second moments computed by one-step methods converge in theory to the first and second moments of X as the stepsize Δ reaches 0, the Gaussian approximation is in general not exact, even if the corresponding ODE is solved accurately. Nonetheless, we shall see in Chapter 3 that the approach via Gaussian approximation works well if the noise intensity is moderate and the time interval for the solution is not too large.

By Subsection 1.5.4 of Chapter 1 we can also apply the numerical methods of this Chapter to Stratonovich SDEs by converting it appropriately into an Itô SDE.

Throughout this chapter, we mainly follow Kloeden & Platen (2000, Chapter 5, 9, 15) and Honerkamp (1990, Chapter 10).

2.1 Overview of numerical solutions of SDEs

In the literature, there are several different approaches to numerical solution of stochastic differential equations. To begin, we mention the following idea of Boyce (1978), to conceive a SDE as a "random ordinary differential equation"

$$\frac{\partial X}{\partial t}(t) = K\left(t, X(t), \gamma(t, X_t)\right),$$

where K is a deterministic function and γ is a stochastic process. The solution X is then determined via appropriate discretization of the random input and Monte Carlo methods. According to Kloeden & Platen (2000) this approach proves to be rather inefficient for SDEs as it does not comprise the special structure of the SDE, namely the characterization via its drift and diffusion coefficients. Kushner & Dupuis (1992) proposed to approximate the solution X of a SDE by finite state Markov chains that can be handled via their (finite dimensional) transition matrices. Kloeden & Platen (2000) argue that this approach becomes inefficient in higher dimension and unbounded domains, as the transition matrices contain a considerable amount of superfluous information that has to be repeatedly reused and computed.
Another approach is based on the so called Feynman–Kac formula, cf. Karatzas & Shreve (2004, Theorem 7.6), which states a direct connection between n-dimensional stochastic differential equations and n-dimensional partial differential equations. For simplicity we state the one-dimensional version: We consider the one-dimensional Itô SDE

$$dX_t = b(t, X_t)dt + \sigma(t,X_t)dW_t, \tag{2.1.1}$$

and define for a given Borel-measurable function $h : \mathbb{R} \to \mathbb{R}$ the function $g(t,x) = \mathbb{E}\left[h(X_T)|X_t = x\right]$, see (1.1.8) for the definition of the conditional expectation. Then g satisfies the (deterministic) partial differential equation

$$g_t(t,x) + b(t,x)g_x(t,x) + \frac{1}{2}\sigma(t,x)^2 g_{xx}(t,x) = 0 \tag{2.1.2}$$

and the terminal condition $g(T,x) = h(x)$. The PDE (2.1.2) corresponds to the Kolmogorov backward equation, see (1.3.5). The numerical solution g is then a so called weak approximation of X as it approximates the expectation $\mathbb{E}\left[h(X_T)|X_t = x\right]$. Weak approximations of X can thus be found by solving the corresponding PDE. This approach is very promising in lower dimensions as it solves various initial value problems of the SDE at the same time. Further, SDEs with high diffusion coefficients, which usually imply the necessity of simulating a larger sample in the approach via one-step methods – see below for details –, can be solved conveniently via Feynman–Kac.

The famous Black–Scholes–Merton equation of Black & Scholes (1973) and Merton (1973) can also be traced back to the Feynman–Kac formula. In finance, numerical solution of the PDE (2.1.2) is thus of great importance, a survey of numerical solution to the Black–Scholes–Merton equation can be found for example in Leentvaar (2003).

If we are primarily interested in numerical solution of a PDE like (2.1.2), we might get into problems if the dimension is too high, since the complexity grows exponentially with the dimension. According to Feynman–Kac we can also find weak approximations of the SDE (2.1.1) as they are solutions of (2.1.2) via $g(t,x) = \mathbb{E}[h(X_T)|X_t = x]$. To do so, we can compute various realizations of one-step methods of (2.1.1) to approximate $\mathbb{E}[h(X_T)|X_t = x]$. The main advantage of one-step methods over the approach via PDEs lies in the face that the complexity of solving an SDE via one-step methods grows only linearly with the dimension. See for example Buchmann & Petersen (2003) for a survey of this approach.

A widely applicable approach of solving SDEs according to Kloeden & Platen (2000) seems to be the direct simulation of sample paths of the solution X via one-step methods. As the weak approximation of X usually involves a large number of simulations of sample paths, they are particularly suited for computers with parallel architecture. Further, one-step methods seem to be the only way to directly approximate the sample paths, which is also known as strong approximation and which is of importance in visualizations of the solution X. The construction of weak approximations is based on the truncation of the so called Itô–Taylor expansion and usually involves derivatives of the drift and diffusion coefficients. To avoid the use of derivatives, one can also apply so called Runge–Kutta schemes, which are usually not direct heuristic implementations of the Runge–Kutta schemes for ODEs. Schemes of high convergence order get very complex and are only simple to sketch in specific cases, for example in the case of scalar noise. An increase of the convergence order can be easily achieved by extrapolating the results of a lower-order one-step method, see Kloeden & Platen (2000, Section 15.3). There are also implementations of linear multi-step methods that are especially useful if the stepsize in each step is constant, cf. for example Pulch (2007, pp. 87–89). In the case of high noise intensity the number of realizations usually has to be increased considerably to get reliable results. So called variance reduction methods allow the reduction of the variance of the realizations, and are either based on a measure transformation via the Theorem of Girsanov, cf. Shreve (2004, Theorem 5.2.3), or on the theory of Monte-Carlo intergration, see for both approaches Kloeden & Platen (2000, Chapter 16).

Last we mention an approach via perturbation theory and functional integrals, which are of particular interest in physics. In this case the stochastic behavior of the solution X of a SDE is represented by so called Feynman graphs. Its building blocks are – among others

– so called vertex function. By neglecting all higher vertex function we presume implicitly that the solution X is Gaussian and get the so called Gaussian approximation, also known as mean field approximation. By assuming that X is Gaussian we can reduce a SDE with so called expolynomial coefficients to a system of ordinary differential equations, where the solution of the ODE approximates the first and second moments of X. If the noise intensity is not too high, the Gaussian approximation usually gives satisfactory results, see for example Honerkamp (1990, Section 10.3.2). By implying more vertex function we get the so called Schwinger–Dyson approximation that also captures a potential non-Gaussian behavior. Whereas both the Gaussian and Schwinger–Dyson approximation do not accurately reproduce the first and second moments, we see that the numerical effort of solving the corresponding ODE is much lower than of one-step methods for SDEs. Archambeau et al. (2007) demonstrated another approach of Gaussian process approximation that is based on approximation of the induced measure of the solution over paths. The idea was demonstrated in the case of additive noise and is favorable for dynamical models of dimension $\mathcal{O}(10^6)$, which arise for example in weather prediction models.

2.2 Time discretizations and approximations

In this section we introduce time discrete approximations also called one-step methods for Itô SDEs. We only discuss one-step methods that are based on truncation of the Itô–Taylor expansion, that is introduced in Section 2.2.3 along with multiple Itô integrals. Further, we introduce the notions of weak consistency and convergence, which can be seen as direct implementations of the respective terms from the theory of numerical solution to ordinary differential equations. The main Theorem 2.2.17 proves the convergence order of the numerical shemes that we shall use.

Let a multi dimensional Itô SDE

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t \tag{2.2.1}$$

with initial value $X(t_0) = X_{t_0} \in \mathbb{R}^n$ be given. We assume that the drift and diffusion coefficients $b = (b^k)_k : [t_0, t] \times \mathbb{R}^n \to \mathbb{R}^n, \sigma = (\sigma^{k,j})_{k,j} : [t_0, t] \times \mathbb{R}^n \to \mathbb{R}^{n\times m}$ satisfy the conditions of Theorem 1.5.5, i.e., there exists a unique solution $X = (X_t)_t$ of (2.2.1). We also assume that there is a filtration $(\mathcal{A}_t)_t$ given such that the m-dimensional Wiener process $(W_t)_t$ is adapted to $(\mathcal{A}_t)_t$ and a martingale w.r.t. $(\mathcal{A}_t)_t$.
We want to approximate the solution $(X_t)_t$ of (2.2.1) on a given interval $[t_0, T]$ by a so called **time discrete approximation**, similarly to one-step methods for ordinary differential equations.

Definition 2.2.1 (cf. Kloeden & Platen (2000, p. 321))**.** For a given interval $[t_0, T]$ and number $\delta > 0$ that we call the **maximum step size** we define

$$(\tau)_\delta = \left\{ \tau_n \,\middle|\, n = 0, 1, \dots \right\}$$

as a sequence of time instants $\left\{ \tau_n \,\middle|\, n = 0, 1, \dots \right\}$ with random variables τ_n satisfying

$$t_0 = \tau_0 < \tau_1 < \ldots < \tau_n < \ldots < \infty \quad \text{w.p.1},$$

$$\sup_n (\tau_{n+1} - \tau_n) \leq \delta \quad \text{w.p.1}$$

and

$$n_t \leq \infty \quad \text{w.p.1 for all } t > 0,$$

where we set $n_t := \max \left\{ n = 0, 1, \dots \,\middle|\, \tau_n \leq t \right\}$. We also assume that the random variables τ_{n+1} are $\mathcal{A}_{\tau_n}$-measurable and $\tau_{n_T} = T$. Then we call $(\tau)_\delta$ a **time discretization** of $[t_0, T]$.

Remark 2.2.2.

(i) The restriction upon τ_{n+1} to be $\mathcal{A}_{\tau_n}$-measurable implies that the **step size**, or **step length** $\Delta_n := \tau_{n+1} - \tau_n$ may only depend on the information available at time τ_n. If we build a numerical approximation based on a time discretization $(\tau)_\delta$, i.e., we associate to each point τ_n an approximation Y_{τ_n} such that $Y_{\tau_n} \approx X_{\tau_n}$, this assumption implies that the step size Δ_n itself can only depend on our observations made up to time τ_n and we cannot "look into the future".

(ii) A simple choice of $(\tau)_\delta$ would be a time discretization of constant step size $\delta = \Delta_n$ for all $n = 0, \dots, n_T$.

Definition 2.2.3 (cf. Kloeden & Platen (2000, p. 322))**.** Let $(Y_t)_{t \in [t_0, T]}$ be a stochastic process that is **right sample path continuous with left side limits**, i.e., Definition 1.1.20 holds with right side limits and for each $t \in (t_0, T]$ we define

$$A_t = \left\{ \omega \in \Omega \,\middle|\, \forall \text{ sequences } (s_n)_n \nearrow t, \text{the sequence } \left(X(s_n, \omega)\right)_n \text{ is Cauchy} \right\}$$

and further have that $A := \bigcup_t A_t \in \mathcal{F}$ and $\mathbb{P}(A) = 1$. Further let $\delta > 0$ and $(\tau)_\delta$ be a time discretization. We call $(Y_t)_t$ a **time discrete approximation** with maximum step size δ based on $(\tau)_\delta$, if Y_{τ_n} is $\mathcal{A}_{\tau_n}$-measurable and $Y_{\tau_{n+1}}$ can be expressed as a function of $Y_{\tau_0}, Y_{\tau_1}, \dots, Y_{\tau_n}, \tau_0, \tau_1, \dots, \tau_{n+1}$ and a finite number ℓ of $\mathcal{A}_{\tau_{n+1}}$-measurable random variables $Z_{n+1,j}$ for $j = 1, \dots, \ell$ and each $n = 0, 1, \dots$.

Remark 2.2.4.

(i) Let $Y := (Y_t)_t$ be a time discrete approximation based on the time discretization $(\tau)_\delta$. To emphasize the dependence of Y on the maximal step size δ, we sometimes also write $Y^\delta := Y$, $(Y_{t,\delta})_t := (Y_t)_t$.

(ii) The time discrete approximation Y_{τ_n} should not involve more information than is available at time τ_n, which is covered by the restriction that Y_{τ_n} is $\mathcal{A}_{\tau_n}$-measurable. The value of $Y_{\tau_{n+1}}$ depends only on the previous computations Y_{τ_i}, the step sizes Δ_i for $i = 0, \ldots, n$ and a finite number of random variables that should generate the noise within one step.

If we compute approximations Y_{τ_n} to X at time instants τ_n for $n = 0, 1, \ldots$, we can always construct a stochastic process $(Y_t)_t$ that is right sample path continuous with left side limits – thus satisfying Definition 2.2.3 – by either **piecewise constant interpolation**

$$Y(t) = Y_{\tau_{n_t}},$$

or **linear interpolation**

$$Y(t) = Y_{\tau_{n_t}} + \frac{t - \tau_{n_t}}{\tau_{n_t+1} - \tau_{n_t}} \left(Y_{\tau_{n_t+1}} - Y_{\tau_{n_t}} \right).$$

Nonetheless we note that we shall concentrate on the values of the time discrete approximation at the time discretization times τ_n, as the solution of the SDE (2.2.1) usually inherits the irregularity of the Wiener process $(W_t)_t$, e.g. the nondifferentiability of its sample paths, and since piecewise constant and linear interpolation do not capture these properties, cf. Kloeden et al. (1994, p. 111).

A simple example of a time discrete approximation for (2.2.1) is the **Euler–Maruyama** approximation, which is any stochastic process $(Y_t)_t$ that is right sample path continuous with left side limits satisfying the following iterative scheme

$$Y_{\ell+1}^k = Y_\ell^k + b^k(\tau_\ell, Y_\ell)\Delta_\ell + \sum_{j=1}^{m} \sigma^{k,j}(\tau_\ell, Y_\ell)\Delta W_\ell^j \tag{2.2.2}$$

for $k = 1, \ldots, n$, where the we have used

$$\begin{aligned} Y_\ell &:= Y(\tau_\ell), \\ \Delta_\ell &= (\tau_{\ell+1} - \tau_\ell), \\ \Delta W_\ell^j &:= \left(W_{\tau_{\ell+1}}^j - W_{\tau_\ell}^j \right), \end{aligned}$$

and denoted by Y_ℓ^k the kth component of Y_ℓ. In the numerical scheme (2.2.2) we have to generate normally distributed random increments ΔW_ℓ^j, $j = 1, \dots, m$ of the Wiener process $(W_t)_t = (W_t^1, \dots, W_t^m)$. From Theorem 1.2.5 we know that the increments are independent random variables satisfying

$$\begin{aligned} \mathbb{E}\left[\Delta W_\ell^j\right] &= 0, \\ \mathbb{V}\mathrm{ar}\left[\Delta W_\ell^j\right] &= \Delta_\ell \end{aligned} \tag{2.2.3}$$

for $j = 1, \dots, m$.

2.2.1 Weak convergence

In this subsection we clarify in which way we want a time discrete approximation $Y = (Y_t)_t$ of an SDE (2.2.1) to approximate the solution $X = (X_t)_t$. In applications it is often desirable to approximate the moments of X at a fixed time instant T, i.e.,

$$\mathbb{E}\left[p(X_T)\right]$$

where $p : \mathbb{R}^n \to \mathbb{R}$ is a polynomial. In general it is of interest to extend the approximation of $\mathbb{E}\left[g(X_T)\right]$ to functions g that are sufficiently smooth and have together with their derivatives polynomial growth, so we define by $\mathcal{C}_P^l(\mathbb{R}^n, \mathbb{R})$ the space of l times continuously differentiable functions $g : \mathbb{R}^m \to \mathbb{R}$ which together with their partial derivatives up to and including order l have polynomial growth, i.e., there exist constants $K > 0$ and $r \in \{1, 2, \dots\}$ such that

$$\left|\partial^i g(x)\right| \leq K\left(1 + \|x\|_2^{2r}\right)$$

for all $x \in \mathbb{R}^n$ and any partial derivative ∂^i of order $i \leq l$. We are now able to give the following definition:

Definition 2.2.5 (cf. Kloeden & Platen (2000, p. 327))**.** Let $Y^\delta = (Y_{t,\delta})_t$ be a time discrete approximation based on a time discretization $(\tau)_\delta$ with maximum step size δ and $T \in (\tau)_\delta$. We say that Y^δ **converges weakly with order** $\beta > 0$ to X at time T as $\delta \downarrow 0$ if for each $g \in \mathcal{C}_P^l(\mathbb{R}^n, \mathbb{R})$ there exists a constant $C > 0$ independent of δ and a $\delta_0 < \infty$ such that

$$\left|\mathbb{E}\left[g(X_T)\right] - \mathbb{E}\left[g(Y_{T,\delta})\right]\right| \leq C\delta^\beta \tag{2.2.4}$$

for each $\delta \in (0, \delta_0)$.

Remark 2.2.6. Instead of (2.2.4) we could require

$$\mathbb{E}\left[|X_T - Y_{T,\delta}|\right] \leq C\delta^\beta$$

to hold which is known as **strong convergence**, cf. KLOEDEN & PLATEN (2000, Chapter 9.6). Strong convergence requires a pathwise approximation and implies weak convergence in the case of $g(x) = x$ by

$$|\mathbb{E}[X_T] - \mathbb{E}[Y_{T,\delta}]| = |\mathbb{E}[X_T - Y_{T,\delta}]| \leq \mathbb{E}[|X_T - Y_{T,\delta}|].$$

Throughout this exposition we shall consider the weak approximation of time discrete approximations and refer the reader interested in strong convergence to KLOEDEN & PLATEN (2000, Chapter 10-13).

From a practical point of view it is not evident how we can evaluate the expectation $\mathbb{E}[Y_T]$ in (2.2.4) as it requires us to know the time discrete approximation Y_T for all $\omega \in \Omega$. Instead we can simulate numerically N trajectories $Y_{T,1} := Y_T(\omega_1), Y_{T,2} := Y_T(\omega_2), \ldots, Y_{T,N} := Y_T(\omega_N)$ of Y_T and compute the **sample mean**

$$\mathbb{E}[Y_t] \approx \frac{1}{N} \sum_{i=1}^{N} Y_{T,i}.$$

The law of large numbers Theorem 1.1.12 ensures the a.s. convergence of $\frac{1}{N}\sum_{i=1}^{N} Y_{T,i}$ to $\mathbb{E}[Y_T]$ as $N \to \infty$. In computer experiments we usually set $100 \leq N \leq 15000$ and it is desirable to construct a confidence interval for the real error $\epsilon := |\mathbb{E}[X_T] - \mathbb{E}[Y_T]|$. To do so, we arrange the simulation of the trajectories into M batches of N simulations and by writing $Y_{T,i,j}$ as the outcome of the ith simulation of the jth batch we can compute

$$\hat{\epsilon}_j := \frac{1}{N} \sum_{i=1}^{N} Y_{T,i,j} - \mathbb{E}[X_T]$$

for $j = 1, \ldots, M$. We can now estimate the mean of the batch averages

$$\hat{\epsilon} := \frac{1}{M} \sum_{j=1}^{M} \hat{\epsilon}_j,$$

to compute the **sample variance**

$$\hat{\sigma}_\epsilon^2 := \frac{1}{M-1} \sum_{j=1}^{M} (\hat{\epsilon}_j - \hat{\epsilon})^2.$$

By the central limit theorem, Theorem 1.2.8, the batch averages $\hat{\epsilon}_j$ are approximately Gaussian for large N and KLOEDEN & PLATEN (2000, p. 312) suggest to use $N = 100$

to achieve a Gaussian behavior. To compute a $100(1-\alpha)\%$ confidence interval

$$(\hat{\epsilon} - \Delta\hat{\epsilon}, \hat{\epsilon} + \Delta\hat{\epsilon})$$

for the real error ϵ, we set

$$\Delta\hat{\epsilon} = t_{1-\alpha,M-1}\sqrt{\frac{\hat{\sigma}_\epsilon^2}{M}}$$

by using the so called **Student t-distribution**, cf. Kloeden & Platen (2000, Chapter 1.9), which can be applied because $\hat{\epsilon}_j$ are approximately Gaussian. $t_{1-\alpha,M-1}$ is determined by the Student t-distribution and can be looked up in tables, cf. Kloeden & Platen (2000, Table 1.9.1). In the computer experiments of Chapter 3 we shall use $\alpha = 0.1$, i.e., we construct a 90% confidence interval for ϵ, and $M = 20$. In this case we need

$$t_{1-\alpha,M-1} = t_{0.9,19} = 1.73.$$

2.2.2 Weak consistency

As in the case of ordinary differential equations, cf. Melenk (2009, Chapter 2.2) Stoer & Bulirsch (2000, Chapter 7), we have to require certain basic conditions on a time discrete approximation $Y := (Y_t)_t$ such that it converges to the solution $X := (X_t)_t$ at a fixed time instant T in the sense of Definition 2.2.5:

Definition 2.2.7 (cf. Kloeden & Platen (2000, pp. 327–328))**.** We call a time discrete approximation $(Y_{t,\delta})_t$ based on a time discretization $(\tau)_\delta$ with maximum stepsize δ **weakly consistent** for the SDE (2.2.1), if there exists a nonnegative function $\delta \mapsto c(\delta)$ such that

$$\mathbb{E}\left[\left|\mathbb{E}\left[\frac{Y_{n+1,\delta} - Y_{n,\delta}}{\Delta_n}|\mathcal{A}_{\tau_n}\right] - b(\tau_n, Y_{n,\delta})\right|^2\right] \leq c(\delta) \tag{2.2.5}$$

and

$$\mathbb{E}\left[\left|\mathbb{E}\left[\frac{1}{\Delta_n}\left(Y_{n+1,\delta} - Y_{n,\delta}\right)\left(Y_{n+1,\delta} - Y_{n,\delta}\right)^\top |\mathcal{A}_{\tau_n}\right] - \sigma(\tau_n, Y_{n,\delta})\sigma(\tau_n, Y_{n,\delta})^\top\right|^2\right] \leq c(\delta) \tag{2.2.6}$$

for all fixed values $Y_{n,\delta} = y$ and $n = 0, 1, \ldots$, where we read both inequalities for each component in the case of a higher dimensional system.

Remark 2.2.8.

(i) In the absence of randomness and noise, condition (2.2.5) is equivalent to the definition of the consistency of a one-step method for solving the ordinary differential

equation

$$\frac{\partial X}{\partial t}(t) = b(t, X(t)),$$

cf. Melenk (2009, Lemma 2.7) or Stoer & Bulirsch (2000, p. 118).

(ii) Condition (2.2.6) requires the variance of the increment of the time discrete approximation to be close to the variance of the solution X of the SDE (2.2.1).

Using the independence of the increments ΔW_n and $\mathcal{A}_{\tau_n}$ it is not hard to see that the Euler–Maruyama approximation (2.2.2) is weakly consistent. We can even change the Euler–Maruyama scheme to the so called **simplified Euler–Maruyama** approximation

$$Y_{\ell+1}^k = Y_\ell^k + b^k(\tau_\ell, Y_\ell)\Delta_\ell + \sum_{j=1}^{m} \sigma^{k,j}(\tau_\ell, Y_\ell)\xi_\ell^j \Delta_\ell^{1/2} \tag{2.2.7}$$

for $k = 1, \ldots, n$, where ξ_ℓ^j are independent two-point distributed random variables with $\mathbb{P}(\xi_\ell^j = \pm 1) = 0.5$. Similarly to the Euler–Maruyama scheme the simplified Euler–Maruyama scheme is also weakly consistent. We shall see later that it is possible and often necessary to exchange the random variables of a time discrete approximation by simpler random variables, without influencing the order of convergence.

The next theorem states similarly to ordinary differential equations that a weakly consistent scheme is weakly convergent. The theorem is stated in the one-dimensional autonomous case, i.e., $n, m = 1$ and $b(t, x) = b(x)$, $\sigma(t, x) = \sigma(x)$. According to Kloeden & Platen (2000, p. 328) it also holds in much more general situations.

Theorem 2.2.9 (cf. Kloeden & Platen (2000, Theorem 9.7.4))**.** *Suppose that the autonomous coefficients b and σ of (2.2.1) for $n, m = 1$ are four times continuously differentiable with polynomial growth and uniformly bounded derivatives. Let Y^δ be a weakly consistent time discrete approximation with* equidistant *time steps $\Delta_n = \delta$ and initial value $Y_{0,\delta} = X(0)$ which satisfies the moment bounds*

$$\mathbb{E}\left[\max_n |Y_{n,\delta}|^{2q}\right] \leq K\left(1 + \mathbb{E}\left[|X(0)|^{2q}\right]\right)$$

for $q = 1, 2, \ldots$ and

$$\mathbb{E}\left[\frac{1}{\Delta_n}|Y_{n+1,\delta} - Y_{n,\delta}|^6\right] \leq c(\delta)$$

for $n = 0, 1, 2, \ldots$, where c is defined as in Definition 2.2.7. Then Y^δ converges weakly to X.

Proof. See Kloeden & Platen (2000, pp. 329–331). □

2.2.3 Stochastic Taylor expansion

In this subsection so called stochastic Taylor expansions are derived and investigated. They can be seen as a generalization of the deterministic Taylor expansion and serve as a basis for the numerical one-step methods that we shall use.

Before we go into detail, we demonstrate the idea of the stochastic Taylor expansion for the SDE (2.2.1) that we write as

$$X_t = X_{t_0} + \int_{t_0}^{t} b(s, X_s) d\lambda(s) + \int_{t_0}^{t} \sigma(s, X_s) dW_s, \tag{2.2.8}$$

with a m-dimensional Wiener process $W_s = (W_s^1, W_s^2, \ldots, W_s^m)$. By introducing the following operators

$$L^0 := \frac{\partial}{\partial t} + \sum_{k=1}^{n} b^k \frac{\partial}{\partial x^k} + \frac{1}{2} \sum_{k,l=1}^{n} \sum_{j=1}^{m} \sigma^{k,j} \sigma^{l,j} \frac{\partial^2}{\partial x^k \partial x^l}, \tag{2.2.9}$$

$$L^j := \sum_{k=1}^{n} \sigma^{k,j} \frac{\partial}{\partial x^k} \tag{2.2.10}$$

for $j = 1, \ldots, m$, we can derive for any twice continuously differentiable function $f : [0, \infty) \times \mathbb{R}^n \to \mathbb{R}$ using the Itô–Doeblin formula Theorem 1.4.27

$$f(t, X_t) = f(t_0, X_{t_0}) + \int_{t_0}^{t} L^0 f(s, X_s) d\lambda(s) + \sum_{j=1}^{m} \int_{t_0}^{t} L^j f(s, X_s) dW_s^j.$$

If we apply this result to the functions $b^k, \sigma^{k,j}$ in Equation (2.2.8) we derive for each $k = 1, \ldots, n$

$$\begin{aligned} X_t^k = X_{t_0}^k & \\ + \int_{t_0}^{t} & \left(b^k(t_0, X_{t_0}) + \int_{t_0}^{s} L^0 b^k(z, X_z) d\lambda(z) + \sum_{j=1}^{m} \int_{t_0}^{s} L^j b^k(z, X_z) dW_z^j \right) d\lambda(s) \\ + \sum_{l=1}^{m} \int_{t_0}^{t} & \left(\sigma^{k,l}(t_0, X_{t_0}) + \int_{t_0}^{s} L^0 \sigma^{k,l}(z, X_z) d\lambda(z) + \sum_{j=1}^{m} \int_{t_0}^{s} L^j \sigma^{k,l}(z, X_z) dW_z^j \right) dW_s^l, \end{aligned}$$

which is known as the **Itô–Taylor expansion of** X of order 1. We notice that the terms $b^k(t_0, X_{t_0})$ and $\sigma^{k,l}(t_0, X_{t_0})$ are constant and can thus be taken out of the integrals with respect to s and W_s^j. We also see that multiple Itô integrals appear, either involving integration with respect to time s and a Wiener process W_s^l or involving twice integration with respect to Wiener processes W_s^j, W_s^l. We shall introduce and handle multiple Itô integrals in the next subsection by using multi-indices. At last, it can be seen that we

can apply the Itô–Doeblin formula once again to each term involved in $L^i b^k$, $L^i \sigma^{k,l}$ for $i = 0, \dots, m$ to extend the Itô–Taylor expansion to a higher order: In that case we need higher order derivatives of the drift and diffusion coefficient and have to treat more complex multiple Itô integrals.
There is also a **Stratonovich–Taylor expansion** if we regard (2.2.8) in the Stratonovich version rather than the Itô version and if we apply to the coefficients b^k, $\sigma^{k.l}$ the chain rule instead of the Itô–Doeblin formula, cf. Remark 1.4.30. The Stratonovich–Taylor expansion usually has a simpler structure than the Itô–Taylor expansion. Nonetheless we only treat the latter as it shall serve us better for the construction of numerical one-step methods for the Itô SDE (2.2.8).

Multiple Itô integrals

Before we can treat multiple Itô integrals we need some notation to work with multi-indices. The notation and the definitions follow Kloeden & Platen (2000, Chapter 5.2, 5.4).

We call a row vector

$$\alpha = (j_1, j_2, \dots, j_l),$$

where $j_i \in \{0, 1, \dots, m\}$, $i = 1, \dots, l$ and $m \in \mathbb{N}$, a **multi-index** of length

$$l = l(\alpha) \in \{1, 2, \dots\}.$$

We denote by v the multi-index of length 0, i.e., $l(v) = 0$, and we write $n(\alpha)$ for the number of components of α which are equal to 0, e.g., $n((2,3,0)) = 1$, $n((0,0,1)) = 2$. We shall denote the set of all multi-indices by $\mathcal{M}$, i.e.,

$$\mathcal{M} := \{(j_1, j_2, \dots, j_l) \,|\, j_i \in \{0, 1, \dots, m\}, i \in \{1, \dots, l\}, \text{for } l \in \mathbb{N}\} \cup \{v\}. \tag{2.2.11}$$

We now introduce two operations on $\mathcal{M}$ that we shall use later: For $\alpha \in \mathcal{M}$ we denote by $-\alpha$ and $\alpha-$ the multi-index in $\mathcal{M}$ that is obtained by deleting the first respectively the last component of α. For example $-(0,0,1) = (0,1)$ and $(1,1,2)- = (1,1)$.
To introduce the Itô–Taylor expansion of a given order β we need the following definition: We call a subset $\mathcal{N} \subseteq \mathcal{M}$ a **hierarchical set** if $\mathcal{N}$ is nonempty and if the multi-indices in $\mathcal{N}$ are uniformly bounded in length, i.e.,

$$\sup_{\alpha \in \mathcal{N}} l(\alpha) < \infty,$$

and if for each $\alpha \in \mathcal{N} \backslash \{v\}$ the set $\mathcal{N}$ also satisfies

$$-\alpha \in \mathcal{N}.$$

Thus if a multi-index α belongs to an hierarchical set, then $-\alpha$ also belongs to the same hierarchical set. For example, the sets

$$\{v\}, \quad \{v, (0), (1)\}, \quad \{v, (0), (1), (0,0), (1,0), (1,1)\}$$

are hierarchical sets. When we intend to form the Itô–Taylor expansion related to an hierarchical set $\mathcal{N}$ we also have to consider the **remainder set** $\mathcal{R}(\mathcal{N})$ of $\mathcal{N}$ that contains all subsequent multi-indices, i.e.,

$$\mathcal{R}(\mathcal{N}) := \{\alpha \in \mathcal{M} \backslash \mathcal{N} \,\big|\, -\alpha \in \mathcal{N}\}.$$

In the case of $m = 1$ in equation (2.2.11), we find for example

$$\mathcal{R}\left(\{v\}\right) = \{(0), (1)\}, \quad \mathcal{R}\left(\{v, (0)\}\right) = \{(0,0), (1,0)\}.$$

The most important hierarchical set we shall use to construct the Itô–Taylor expansion of order β is

$$\Gamma_\beta := \{\alpha \in \mathcal{M} \,\big|\, l(\alpha) \leq \beta\}. \tag{2.2.12}$$

We are now able to define multiple Itô integrals. To begin we shall define three sets of adapted right continuous stochastic processes $f = (f_t)_{t \geq 0} = (f_t)_{t \geq 0}$ assuming values in $\mathbb{R}$ with left hand limits: We denote by $\mathcal{H}_v$ the set of such stochastic processes satisfying

$$\mathbb{P}\left(f(t, \omega) < \infty\right) = 1$$

for each $t \geq 0$. By $\mathcal{H}_{(0)}$ we denote the set of all such stochastic processes satisfying

$$\mathbb{P}\left(\int_0^t f(s, \omega) d\lambda(s) < \infty\right) = 1 \tag{2.2.13}$$

for each $t \geq 0$. The set $\mathcal{H}_{(1)}$ shall contain all those processes f satisfying

$$\mathbb{P}\left(\int_0^t f(s, \omega)^2 d\lambda(s) < \infty\right) = 1 \tag{2.2.14}$$

for each $t \geq 0$. We also write for each $j \in \{2, 3, \ldots, m\}$

$$\mathcal{H}_{(j)} := \mathcal{H}_{(1)}. \tag{2.2.15}$$

Definition 2.2.10 (Kloeden & Platen (2000, pp. 168–169))**.** Let $\alpha = (j_1, j_2, \ldots, j_l) \in \mathcal{M}$ be a multi-index and let $f = (f_t)_{t \geq 0}$ be a right continuous stochastic process with left hand limits assuming values in $\mathbb{R}$. For $t_0, t \geq 0$ we define the **multiple Itô integral** $\mathcal{I}_\alpha [f(\cdot)]_{t_0,t}$ recursively by

$$\mathcal{I}_\alpha [f(\cdot)]_{t_0,t} := \begin{cases} f(t) & : l = 0, \\ \int_{t_0}^t \mathcal{I}_{\alpha-} [f(\cdot)]_{t_0,s} \, d\lambda(s) & : l \geq 1 \text{ and } j_l = 0, \\ \int_{t_0}^t \mathcal{I}_{\alpha-} [f(\cdot)]_{t_0,s} \, dW_s^{j_l} & : l \geq 1 \text{ and } j_l \geq 1, \end{cases} \tag{2.2.16}$$

given that the stochastic process $\left(\mathcal{I}_{\alpha-} [f(\cdot)]_{t_0,t}\right)_{t \geq 0}$ satisfies

$$\left(\mathcal{I}_{\alpha-} [f(\cdot)]_{t_0,\cdot}\right)_{t \geq 0} \in \mathcal{H}_{(j_l)}, \tag{2.2.17}$$

with the convention that $(j_l) := v$ if $\alpha = v$.

For $\alpha = (j_1, j_2, \ldots, j_l) \in \mathcal{M}$ with $l(\alpha) \geq 2$ we define the set $\mathcal{H}_\alpha$ rescursively to consist of all adapted right continuous processes $f = (f_t)_{t \geq 0}$ with left hand limits for which equation (2.2.17) holds.

Remark 2.2.11.

(i) The assumptions that f is an adapted right continuous stochastic process with left hand limits such that equation (2.2.17) holds assure that the two integrals, respectively the random variable $f(t)$ appearing on the right side of (2.2.16) exist w.p.1. Condition (2.2.13) implies that the Lebesgue integral exists w.p.1 and condition (2.2.14) implies because of condition (iii)' on page 27 that the Itô integral exists. Thus for each $f \in \mathcal{H}_\alpha$ the multiple Itô integral $\mathcal{I}_\alpha [f(\cdot)]_{t_0,t}$ exists.

(ii) The definition of a multiple Itô integral can be extended to a stochastic process $f := (f_t)_{t \geq 0} = \left((f_t^k)_{k=1,\ldots,n}\right)_{t \geq 0}$ that assumes values in $\mathbb{R}^n$ by applying the definition to each component $\left(f_t^k\right)_{t \geq 0}$ for $k = 1, \ldots, n$.

We now consider the following examples to illustrate the terminology:

$$\begin{aligned}
\mathcal{I}_v\left[f(\cdot)\right]_{t_0,t} &= f(t),\\
\mathcal{I}_{(0)}\left[f(\cdot)\right]_{t_0,t} &= \int_{t_0}^{t} f(s)d\lambda(s),\\
\mathcal{I}_{(1)}\left[f(\cdot)\right]_{t_0,t} &= \int_{t_0}^{t} f(s)dW_s^1,\\
\mathcal{I}_{(1,0)}\left[f(\cdot)\right]_{t_0,t} &= \int_{t_0}^{t}\int_{t_0}^{s_2} f(s_1)dW_{s_1}^1 d\lambda(s_2),\\
\mathcal{I}_{(1,0,2)}\left[f(\cdot)\right]_{t_0,t} &= \int_{t_0}^{t}\int_{t_0}^{s_3}\int_{t_0}^{s_2} f(s_1)dW_{s_1}^1 d\lambda(s_2)dW_{s_3}^2,
\end{aligned}$$

for appropriate processes f.

Later we shall often use the special case $f(t) \equiv \mathbf{1}$, $t_0 = 0$ with $\mathbf{1} : \Omega \to \mathbb{R}$, $\mathbf{1}(\omega) = 1$ and write for convenience

$$\mathcal{I}_{\alpha,t_0,t} := \mathcal{I}_\alpha\left[\mathbf{1}\right]_{t_0,t}, \tag{2.2.18}$$

$$\mathcal{I}_{\alpha,t} := \mathcal{I}_\alpha\left[\mathbf{1}\right]_{0,t}. \tag{2.2.19}$$

Itô–Taylor expansion

We are now going to define functions that we need later to denote the coefficients of the Itô–Taylor expansion. Let $\alpha = (j_1, \ldots, j_l)$ be a multi-index and $h := l(\alpha) + n(\alpha)$. For any function $f \in \mathcal{C}^h([0,\infty) \times \mathbb{R}^n, \mathbb{R})$ we define recursively the **Itô coefficient functions** using the operators (2.2.9) and (2.2.10)

$$f_\alpha := \begin{cases} f & : l = 0,\\ L^{j_1} f_{-\alpha} & : l \geq 1. \end{cases} \tag{2.2.20}$$

By requiring $f \in \mathcal{C}^h([0,\infty) \times \mathbb{R}^n, \mathbb{R})$ we see that the definition of f_α in (2.2.20) makes sense, i.e, all derivatives on the right side of (2.2.20) exist. In the general case of $m, n \geq 1$ the coefficient functions f_α can contain complicated terms involving multiple derivatives with respect to time $t \in [0,\infty)$ and the spatial coordinates $x \in \mathbb{R}^n$. The following example holds for the one-dimensional autonomous case $m = n = 1$, i.e., $b(t,x) = b(x)$, $\sigma(t,x) = \sigma(x)$, and $f(t,x) = x$:

$$f_{(0)} = b; \quad f_{(1)} = \sigma; \quad f_{(1,1)} = \sigma\left(\frac{\partial\sigma}{\partial x}\right); \quad f_{(0,1)} = b\left(\frac{\partial\sigma}{\partial x}\right) + \frac{1}{2}\sigma^2\left(\frac{\partial^2\sigma}{\partial x^2}\right).$$

We are now able to state the main theorem of Itô–Taylor expansions that can be seen

as an analogon to deterministic Taylor expansions:

Theorem 2.2.12 (cf. Kloeden & Platen (2000, Theorem 5.5.1))**.** *Let $(X_t)_{t_0 \le t \le T}$ be the solution of (2.2.8) and $t_0 \le \rho \le \tau \le T$. Let $\mathcal{N} \subseteq \mathcal{M}$ be an hierarchical set and let $f : [0, \infty) \times \mathbb{R}^n \to \mathbb{R}$. Then the Itô–Taylor expansion*

$$f(\tau, X_\tau) = \sum_{\alpha \in \mathcal{N}} \mathcal{I}_\alpha \left[f_\alpha(\rho, X_\rho) \right]_{\rho,\tau} + \sum_{\alpha \in \mathcal{R}(\mathcal{N})} \mathcal{I}_\alpha \left[f_\alpha(\cdot, X_\cdot) \right]_{\rho,\tau} \tag{2.2.21}$$

holds w.p.1, provided all derivatives of f, b^k and $\sigma^{k,l}$ for $k = 1, \dots, n$, $l = 1, \dots, m$ and all the multiple Itô integrals appearing in (2.2.21) exist.

Proof. See Kloeden & Platen (2000, pp. 186–187). □

Remark 2.2.13.

(i) Equation (2.2.21) can be extended to functions f assuming values in $\mathbb{R}^n$ by applying (2.2.21) to each component f^k, $k = 1, \dots, n$.

(ii) To construct numerical one-step methods we shall usually apply equation (2.2.21) to $f(t, x) = x$, $t_0 = \rho = 0$ and $\mathcal{N} = \Gamma_\beta$ for $\beta \ge 1$, where Γ_β is defined as in (2.2.12), i.e., we set $Y_0 = X_0$ and define recursively

$$Y_{i+1} := \sum_{\alpha \in \Gamma_\beta} \mathcal{I}_\alpha \left[f_\alpha(i\Delta, Y_i) \right]_{i\Delta,(i+1)\Delta} .$$

(iii) We see that the integrands appearing in the multiple Itô integrals within

$$\sum_{\alpha \in \mathcal{N}} \mathcal{I}_\alpha \left[f_\alpha(\rho, X_\rho) \right]_{\rho,\tau}$$

are in fact constant and can thus be extracted from the respective multiple Itô integral. In the case of $\rho = t_0 = 0$ we can thus deduce by using the notation of (2.2.19)

$$\mathcal{I}_\alpha \left[f_\alpha(0, X_0) \right]_{0,\tau} = f_\alpha(0, X_0) \mathcal{I}_{\alpha,\tau}$$

for each $\alpha \in \mathcal{N}$. The challenge in constructing numerical one-step methods lies in the realization or approximation of the multiple Itô integrals $\mathcal{I}_{\alpha,\tau}$.

As indicated above we now consider the **truncated weak Itô–Taylor expansion**

$$\eta_\beta(t) := \sum_{\alpha \in \Gamma_\beta} \mathcal{I}_\alpha \left[f_\alpha(0, X_0) \right]_{0,t} , \tag{2.2.22}$$

i.e., with $t_0 = \rho = 0$, $\mathcal{N} = \Gamma_\beta$ and for $f(t,x) = x$, where we assume that all necessary derivatives and integrals exist for all $\alpha \in \Gamma_\beta \cup \mathcal{R}(\Gamma_\beta)$. We are interested in the order of the weak error

$$\sup_{0\leq t\leq T} |\mathbb{E}\left[g(X_t) - g(\eta_\beta(t))|\mathcal{A}_0\right]|,$$

where $(\mathcal{A}_t)_{t\geq 0}$ is the filtration introduced at the beginning of this chapter on page 48, and $g \in \mathcal{C}_P^{2(\beta+1)}(\mathbb{R}^n, \mathbb{R})$ is arbitrary. We follow Kloeden & Platen (2000) and state the following theorem in the autonomous version $b^k(t,x) = b^k(x)$, $\sigma^{k,l}(t,x) = \sigma^{k,l}(x)$. The non-autonomous version can always be obtained from the autonomous version in the following way: We set

$$Z_t = \begin{pmatrix} t \\ X_t \end{pmatrix}; \quad \tilde{b}(z) = \begin{pmatrix} 1 \\ b(z) \end{pmatrix}; \quad \tilde{\sigma}(z) = \begin{pmatrix} 0 \\ \sigma(z) \end{pmatrix};$$

where $z = (t,x)$ and the first row of $\tilde{\sigma}(z)$ should contain a $1 \times m$-dimensional vector of zeros. Then we apply the theorem to the SDE

$$dZ_t = \tilde{b}(Z_t)dt + \tilde{\sigma}(Z_t)dW_t; \quad Z_0 = (0, X_0)$$

and obtain the same results for X_t.

Theorem 2.2.14 (cf. Kloeden & Platen (2000, Proposition 5.11.1))**.** *Let $\beta \in \{1, 2, \dots\}$ and $T \in (0, \infty)$ be given and suppose that $b^k = b^k(x)$ and $\sigma^{k,l} = \sigma^{k,l}(x)$ belong to $\mathcal{C}_P^{2(\beta+1)}(\mathbb{R}^n, \mathbb{R})$ and satisfy Lipschitz conditions and linear growth bounds for $k = 1, \dots, n$ and $l = 1, \dots, m$. Then for each $g \in \mathcal{C}_P^{2(\beta+1)}(\mathbb{R}^n, \mathbb{R})$ there exist constants $K > 0$ and $r \in \{1, 2, \dots\}$ such that*

$$\sup_{0\leq t\leq T} |\mathbb{E}\left[g(X_t) - g(\eta_\beta(t))|\mathcal{A}_0\right]| \leq K\left(1 + |X_0|^{2r}\right) T^{\beta+1}.$$

Proof. See Kloeden & Platen (2000, pp. 219–221). □

Remark 2.2.15. If we start at X_0, i.e., at the correct initial value of the SDE (2.2.1) for $t_0 = 0$ and perform one step of length Δ of our one-step method $(Y_t)_t$ using (2.2.22), i.e., we set

$$Y_1 := \eta_\beta(\Delta); \quad Y_0 = X_0,$$

then Theorem 2.2.14 states that we make an error of $\mathcal{O}(\Delta^{\beta+1})$. This cannot yet be interpreted that the one-step method which we obtain by repeatedly applying (2.2.22) is

weakly convergent with order $\beta + 1$, since we have to set in the next step

$$Y_2 := \sum_{\alpha \in \Gamma_\beta} \mathcal{I}_\alpha \left[f_\alpha(\Delta, Y_1) \right]_{\Delta, 2\Delta},$$

rather than

$$Y_2 = \sum_{\alpha \in \Gamma_\beta} \mathcal{I}_\alpha \left[f_\alpha(\Delta, X_\Delta) \right]_{\Delta, 2\Delta}.$$

I.e., for the next step we cannot apply Theorem 2.2.14 as the initial value Y_1 is in general not equal to X_Δ. We shall see later in Theorem 2.2.17 that the one-step methods we get by repeatedly applying the scheme above are weakly convergent of order β, rather than $\beta + 1$.

As indicated when we have introduced the simplified Euler scheme (2.2.7) as opposed to the Euler scheme (2.2.2), we have some freedom to choose the random variables that should approximate the multiple Itô integrals. Thus if we replace the multiple Itô integral $\mathcal{I}_{\alpha,t} = \mathcal{I}_\alpha \left[\mathbf{1} \right]_{0,t}$ in (2.2.22) by other random variables $\tilde{\mathcal{I}}_{\alpha,t}$ we obtain

$$U_\beta(t) := \sum_{\alpha \in \Gamma_\beta} f_\alpha(0, X_0) \tilde{\mathcal{I}}_{\alpha,t}.$$

The following corollary tells us conditions such that the weak error between U_β and X_t is still of the same order as in Theorem 2.2.14:

Corollary 2.2.16 (cf. Kloeden & Platen (2000, Corollary 5.12.1))**.** *Suppose under the assumptions of Theorem 2.2.14 that for $t \in [0, T]$ and $\beta \in \{1, 2, \dots\}$ there exists a constant $L > 0$ such that the moment condition*

$$\left| \mathbb{E} \left[\prod_{k=1}^{l} \mathcal{I}_{\alpha_k,t} - \prod_{k=1}^{l} \tilde{\mathcal{I}}_{\alpha_k,t} \middle| \mathcal{A}_0 \right] \right| \leq L t^{\beta+1} \tag{2.2.23}$$

holds for all choices of multi-indices $\alpha_k \in \Gamma_\beta \backslash \{v\}$ with $k = 1, \dots, l$ and $l = 1, \dots, 2\beta + 1$. Then for each $g \in \mathcal{C}_P^{2(\beta+1)}(\mathbb{R}^n, \mathbb{R})$ there exist constants $K > 0$ and $r \in \{1, 2, \dots\}$ such that

$$\sup_{0 \leq t \leq T} \left| \mathbb{E} \left[g(X_t) - g(U_\beta(t)) | \mathcal{A}_0 \right] \right| \leq K \left(1 + |X_0|^{2r} \right) T^{\beta+1}.$$

Proof. See Kloeden & Platen (2000, p. 221; proof of Proposition 5.11.1). □

We now cite suggestions of Kloeden & Platen (2000) for $\tilde{\mathcal{I}}_{\alpha,\Delta}$ that satisfy condition (2.2.23) of Corollary 2.2.16 in the case of $\beta = 1, 2, 3$, cf. Kloeden & Platen (2000, pp.

225–226): In all of these cases we set

$$\tilde{\mathcal{I}}_{(0),\Delta} = \Delta; \quad \tilde{\mathcal{I}}_{(0,0),\Delta} = \frac{1}{2}\Delta^2; \quad \tilde{\mathcal{I}}_{(0,0,0),\Delta} = \frac{1}{3!}\Delta^3. \tag{2.2.24}$$

For $\beta = 1, m = 1, 2, \ldots$ and $j \in \{1, 2, \ldots, m\}$ we can set:

$$\tilde{\mathcal{I}}_{(j),\Delta} = \Delta\tilde{W}^j, \tag{2.2.25}$$

where $\Delta\tilde{W}^j$ are independent normally distributed Gaussian random variables with mean 0 and variance Δ, *or* independent two-point distributed random variables satisfying

$$\mathbb{P}\left(\Delta\tilde{W}^j = \pm\sqrt{\Delta}\right) = \frac{1}{2}. \tag{2.2.26}$$

For $\beta = 2, m = 1, 2, \ldots$ and $j_1, j_2 \in \{1, 2, \ldots, m\}$ we set:

$$\tilde{\mathcal{I}}_{(j_1),\Delta} = \Delta\tilde{W}^{j_1}, \quad \tilde{\mathcal{I}}_{(0,j_1),\Delta} = \tilde{\mathcal{I}}_{(j_1,0),\Delta} = \frac{1}{2}\Delta \cdot \Delta\tilde{W}^{j_1}, \tag{2.2.27}$$

$$\tilde{\mathcal{I}}_{(j_1,j_2),\Delta} = \frac{1}{2}\left(\Delta\tilde{W}^{j_1} \cdot \Delta\tilde{W}^{j_2} + V_{j_1,j_2}\right), \tag{2.2.28}$$

where $\Delta\tilde{W}^{j_*}$ are independent normally distributed Gaussian random variables with mean 0 and variance Δ, *or* independent three-point distributed random variables satisfying

$$\mathbb{P}\left(\Delta\tilde{W}^{j_*} = \pm\sqrt{3\Delta}\right) = \frac{1}{6}, \quad \mathbb{P}\left(\Delta\tilde{W}^{j_*} = 0\right) = \frac{2}{3}, \tag{2.2.29}$$

and V_{j_1,j_2} are independent two-point distributed random variables with

$$\mathbb{P}\left(V_{j_1,j_2} = \pm\Delta\right) = \frac{1}{2} \tag{2.2.30}$$

for $j_2 = 1, \ldots, j_1 - 1$,

$$V_{j_1,j_1} = -\Delta \tag{2.2.31}$$

and

$$V_{j_2,j_1} = -V_{j_1,j_2} \tag{2.2.32}$$

for $j_2 = j_1 + 1, \ldots, m$ and $j_1 = 1, \ldots, m$. We see that in the case of $m = 1$ the random variable V_{j_1,j_2} is not required and can be set to $-\Delta$.

For $\beta = 3$ and $m = 1$, i.e., in the case of **scalar noise** – the noise is generated by only

one Wiener process – we can set

$$\tilde{\mathcal{I}}_{(1),\Delta} = \Delta\tilde{W}, \tag{2.2.33}$$

$$\tilde{\mathcal{I}}_{(0,1),\Delta} = \Delta \cdot \Delta\tilde{W} - \Delta\tilde{Z}, \quad \tilde{\mathcal{I}}_{(1,0),\Delta} = \Delta\tilde{Z}, \tag{2.2.34}$$

$$\tilde{\mathcal{I}}_{(1,1),\Delta} = \frac{1}{2}\left(\left(\Delta\tilde{W}\right)^2 - \Delta\right), \tag{2.2.35}$$

$$\tilde{\mathcal{I}}_{(0,0,1),\Delta} = \tilde{\mathcal{I}}_{(0,1,0),\Delta} = \tilde{\mathcal{I}}_{(1,0,0),\Delta} = \frac{1}{6}\Delta^2 \cdot \Delta\tilde{W}, \tag{2.2.36}$$

$$\tilde{\mathcal{I}}_{(1,1,0),\Delta} = \tilde{\mathcal{I}}_{(1,0,1),\Delta} = \tilde{\mathcal{I}}_{(0,1,1),\Delta} = \frac{1}{6}\Delta \cdot \left(\left(\Delta\tilde{W}\right)^2 - \Delta\right), \tag{2.2.37}$$

$$\tilde{\mathcal{I}}_{(1,1,1),\Delta} = \frac{1}{6}\Delta\tilde{W} \cdot \left(\left(\Delta\tilde{W}\right)^2 - 3\Delta\right), \tag{2.2.38}$$

where $\Delta\tilde{W}$ and $\Delta\tilde{Z}$ are *correlated* Gaussian random variables, $\Delta\tilde{W}$ with mean 0 and variance Δ, $\Delta\tilde{Z}$ with mean 0 and variance $\frac{1}{3}\Delta^3$ and covariance

$$\mathbb{C}\mathrm{ov}\left[\Delta\tilde{W}; \Delta\tilde{Z}\right] = \mathbb{E}\left[\Delta\tilde{W} \cdot \Delta\tilde{Z}\right] = \frac{1}{2}\Delta^2.$$

2.2.4 Weak Taylor schemes

In this subsection we introduce several one-step methods called weak Taylor schemes, that are based on the Itô–Taylor expansion of the last subsection and that are of different convergence order. At the end of this subsection we formulate a convergence theorem for general weak Taylor schemes that also proves the asserted weak convergence order of the introduced schemes.

Before we formulate new numerical schemes based on Itô–Taylor expansions, we first state the connection between the beforehand introduced Euler–Maruyama scheme

$$Y_{\ell+1}^k = Y_\ell^k + b^k(\ell \cdot \Delta, Y_\ell)\Delta + \sum_{j=1}^m \sigma^{k,j}(\ell \cdot \Delta, Y_\ell)\Delta W^j,$$

as well as the simplified Euler–Maruyama scheme

$$Y_{\ell+1}^k = Y_\ell^k + b^k(\ell \cdot \Delta, Y_\ell)\Delta + \sum_{j=1}^m \sigma^{k,j}(\ell \cdot \Delta, Y_\ell)\xi_\ell^j\Delta^{1/2}$$

and the Itô–Taylor expansions of the last subsection. Both equations above are formulated for $k = 1, \dots, n$ with *constant* step size $\Delta_n = \Delta$. Also, ΔW^j are independent Gaussian random variables with mean 0 and variance Δ and ξ_ℓ^j are independent two-point distributed random variables with $\mathbb{P}(\xi_\ell^j = \pm 1) = 0.5$.
By comparing the Euler–Maruyama scheme with (2.2.22) we see that the Euler–Maruyama scheme is indeed the truncated Itô–Taylor expansion for $\beta = 1$. The simplified Euler–Maruyama scheme emerges from the Euler–Maruyama scheme by replacing the normally distributed random variables ΔW^j with the two-point distributed random variables ξ_ℓ^j, thus by applying Corollary 2.2.16 with the random variables given in (2.2.24) and (2.2.26).

By implying all the double Itô integrals in our numerical scheme, i.e, setting $\beta = 2$, we get a higher order scheme: The truncated Itô–Taylor expansion with *constant* step size Δ then takes the form

$$\begin{aligned} Y_{\ell+1}^k =& Y_\ell^k + b^k \Delta + \frac{1}{2} L^0 b^k \Delta^2 \\ &+ \sum_{j=1}^m \left(\sigma^{k,j} \Delta W^j + L^0 \sigma^{k,j} \mathcal{I}_{(0,j)} + L^j b^k \mathcal{I}_{(j,0)} \right) \\ &+ \sum_{j_1, j_2 = 1}^m L^{j_1} \sigma^{k,j_2} \mathcal{I}_{(j_1,j_2)}, \end{aligned}$$

with $Y_0 = X(0)$ and $k = 1, \dots, n$, cf. Kloeden & Platen (2000, p. 466). To assist the reader we have omitted the function argument $(\ell \cdot \Delta, Y_\ell)$ for the functions appearing on the right side and also the integration boundaries $\ell \cdot \Delta \leq (\ell + 1) \cdot \Delta$ in the subscript of the multiple Itô integrals. In general the double Itô integrals $\mathcal{I}_{(0,j)}$, $\mathcal{I}_{(j,0)}$, $\mathcal{I}_{(j_1,j_2)}$ are not easy to generate directly, but by Corollary 2.2.16 we can exchange them by simpler random variables, e.g. by random variables satisfying (2.2.24) and (2.2.27)-(2.2.32). In that case we get the **order 2.0 weak Taylor scheme**

$$\begin{aligned} Y_{\ell+1}^k =& Y_\ell^k + b^k \Delta + \frac{1}{2} L^0 b^k \Delta^2 \\ &+ \sum_{j=1}^m \left(\sigma^{k,j} + \frac{1}{2} \Delta \left(L^0 \sigma^{k,j} + L^j b^k \right) \right) \Delta \tilde{W}^j \\ &+ \frac{1}{2} \sum_{j_1, j_2 = 1}^m L^{j_1} \sigma^{k,j_2} \left(\Delta \hat{W}^{j_1} \cdot \Delta \hat{W}^{j_2} + V_{j_1,j_2} \right), \end{aligned} \tag{2.2.39}$$

where $k = 1, \dots, n$, cf. Kloeden & Platen (2000, pp. 466–467), which is of weak convergence order 2.

The truncated Itô–Taylor expansion of order $\beta = 3$ with constant step size Δ takes

the form

$$\begin{aligned} Y_{\ell+1}^k =& Y_\ell^k + b^k\Delta + \frac{1}{2}L^0 b^k \Delta^2 \\ &+ \sum_{j=1}^m \left(\sigma^{k,j}\Delta W^j + L^0\sigma^{k,j}\mathcal{I}_{(0,j)} + L^j b^k \mathcal{I}_{(j,0)}\right) \\ &+ \sum_{j_1,j_2=1}^m L^{j_1}\sigma^{k,j_2}\mathcal{I}_{(j_1,j_2)} + \sum_{j_1,j_2=0}^m L^{j_1}L^{j_2}b^k\mathcal{I}_{(j_1,j_2,0)} \\ &+ \sum_{j_1,j_2=0}^m \sum_{j_3=1}^m L^{j_1}L^{j_2}b^{k,j_3}\mathcal{I}_{(j_1,j_2,j_3)}. \end{aligned}$$

In the case of scalar noise $m = 1$ we can substitute the complicated multiple Itô integrals by simpler random variables such as in (2.2.33)-(2.2.38) and get the **scalar noise order 3.0 weak Taylor scheme**

$$\begin{aligned} Y_{\ell+1}^k =& Y_\ell^k + b^k\Delta + \sigma^k\Delta\tilde{W} + \frac{1}{2}L^1\sigma^k\left(\left(\Delta\tilde{W}\right)^2 - \Delta\right) \\ &+ L^1 b^k\Delta\tilde{Z} + \frac{1}{2}L^0 b^k\Delta^2 + L^0\sigma^k\left(\Delta\tilde{W}\cdot\Delta - \Delta\tilde{Z}\right) \\ &+ \frac{1}{6}\left(L^0L^0\sigma^k + L^0L^1b^k + L^1L^0b^k\right)\Delta\tilde{W}\cdot\Delta^2 \\ &+ \frac{1}{6}\left(L^1L^1b^k + L^1L^0\sigma^k + L^0L^1\sigma^k\right)\left(\left(\Delta\tilde{W}\right)^2 - \Delta\right)\cdot\Delta \\ &+ \frac{1}{6}L^0L^0b^k\Delta^3 + \frac{1}{6}L^1L^1\sigma^k\left(\left(\Delta\tilde{W}\right)^2 - 3\Delta\right)\cdot\Delta\tilde{W}, \end{aligned} \tag{2.2.40}$$

where $\left(\sigma^k\right)_k := \left(\sigma^{k,1}\right)_k$ and $k = 1,\ldots,n$, cf. Kloeden & Platen (2000, pp. 468–469), which is of weak convergence order 3.

The following theorem implies the asserted weak convergence order of the numerical schemes that we have introduced before. We formulate it in the autonomous version as is comprises the nonautonomous version.

Theorem 2.2.17 (cf. Kloeden & Platen (2000, Theorem 14.5.2))**.** *Let Y^δ be a time discrete approximation of the autonomous version of the SDE* (2.2.8) *with solution X corresponding to a time discretization $(\tau)_\delta$, such that all moments of the initial value X_0 exist, i.e.,*

$$\mathbb{E}\left[\|X_0\|_2^i\right] < \infty$$

for $i = 1, 2, \ldots$, and such that $Y_{0,\delta}$ converges weakly with order β to X_0 as $\delta \downarrow 0$ for some fixed $\beta = 1, 2, \ldots$. We also suppose that the drift and diffusion coefficient of (2.2.8) *are*

Lipschitz continuous with components

$$b^k, \sigma^{k,j} \in \mathcal{C}_P^{2(\beta+1)}(\mathbb{R}^n, \mathbb{R})$$

for all $k = 1, \dots, n$ *and* $j = 1, \dots, m$ *and satisfy the linear growth bound*

$$\|b(x)\|_2 + \|\sigma(x)\|_2 \leq L(1 + \|x\|_2)$$

for all $x \in \mathbb{R}^n$*, where* $L < \infty$ *and* $\|\sigma(x)\|_2 = \sum_{i,j} \sigma^{i,j}(x)^2$*. In addition, we suppose that for each* $p = 1, 2, \dots$ *there exist constants* $K < \infty$ *and* $r \in \{1, 2, \dots\}$*, which do not depend on* δ*, such that for each* $q \in \{1, \dots, p\}$

$$\mathbb{E}\left[\max_{0 \leq n \leq n_T} \|Y_{n,\delta}\|_2^{2q} \,\middle|\, \mathcal{A}_0\right] \leq K\left(1 + \|Y_{0,\delta}\|_2^{2r}\right) \tag{2.2.41}$$

and

$$\mathbb{E}\left[\|Y_{n+1,\delta} - Y_{n,\delta}\|_2^{2q} \,|\mathcal{A}_{\tau_n}\right] \leq K\left(1 + \max_{0 \leq k \leq n} \|Y_{k,\delta}\|_2^{2r}\right)(\tau_{n+1} - \tau_n)^q \tag{2.2.42}$$

for $n = 0, 1, \dots, n_T - 1$*, and such that*

$$\begin{aligned} &\left|\mathbb{E}\left[\prod_{h=1}^{l}\left(Y_{n+1,\delta}^{p_h} - Y_{n,\delta}^{p_h}\right) - \prod_{h=1}^{l}\left(\sum_{\alpha \in \Gamma_\beta \setminus \{v\}} f_\alpha^{p_h}(\tau_n, Y_{n,\delta}) \mathcal{I}_{\alpha,\tau_n,\tau_{n+1}}\right) \,\middle|\, \mathcal{A}_{\tau_n}\right]\right| \\ &\quad \leq K\left(1 + \max_{0 \leq k \leq n_T} \|Y_{k,\delta}\|_2^{2r}\right)\delta^\beta(\tau_{n+1} - \tau_n) \end{aligned} \tag{2.2.43}$$

for all $n = 0, 1, \dots, n_T - 1$ *and* $(p_1, \dots, p_l) \in \{1, \dots, n\}^l$*, where* $l = 1, \dots, 2\beta + 1$ *and* $Y_{n,\delta}^{p_h}$ *denotes the* p_h*th component of* $Y_{n,\delta}$ *and we have used the notation of* (2.2.18)*. Then the time discrete approximation converges weakly of order* β *as* $\delta \downarrow 0$ *to* X *at time* T*.*

Proof. See Kloeden & Platen (2000, pp. 477–480). □

Remark 2.2.18.

(i) Throughout the computer experiments of this exposition we shall set X_0 to be constant and $Y_{0,\delta} = X_0$. I.e., the conditions at the beginning of Theorem 2.2.17 are satisfied.

(ii) Conditions (2.2.41)-(2.2.42) are regularity conditions for the time discrete approximation and make the weak convergence criterion meaningful, cf. Kloeden & Platen (2000, p. 475).

(iii) The most important condition is (2.2.43). In case of schemes that are based on truncation of the Itô–Taylor expansion and direct simulation of the multiple Itô integrals $\mathcal{I}_{\alpha,\tau_n,\tau_{n+1}}$ this condition is satisfied since the left side then turns to be 0.

We now want to show that the numerical schemes we have introduced before satisfy the crucial condition (2.2.43). As they satisfy the regularity conditions (2.2.41)-(2.2.42), cf. Kloeden & Platen (2000, pp. 582–583), we have then shown that under the assumptions of Theorem 2.2.17 the numerical schemes converge in the asserted order. First of all our numerical schemes use a nonrandom time discretization of constant step size $\Delta_n = \Delta \leq \delta$. We can thus write $\tau_n = n \cdot \Delta$.
As already mentioned in the previous remark, the Euler–Maruyama scheme satisfies (2.2.43) as it is a direct implementation of the truncated Itô–Taylor expansion of order $\beta = 1$ and thus the left side of (2.2.43) is 0. As regards the simplified Euler–Maruyama scheme, the order 2.0 weak Taylor scheme and the order 3.0 weak Taylor scheme, we have exchanged the complicated multiple Itô integrals $\mathcal{I}_{\alpha,\tau_n,\tau_{n+1}}$ by simpler random variables $\tilde{\mathcal{I}}_{\alpha,\tau_n,\tau_{n+1}}$ – given in (2.2.24)-(2.2.38) – that satisfy condition (2.2.23) of Corollary 2.2.16. By a linear shift of the time variable t we can rewrite condition (2.2.23) as

$$\left| \mathbb{E}\left[\prod_{k=1}^{l} \mathcal{I}_{\alpha_k,n\Delta,(n+1)\Delta} - \prod_{k=1}^{l} \tilde{\mathcal{I}}_{\alpha_k,n\Delta,(n+1)\Delta} \middle| \mathcal{A}_{\tau_n} \right] \right| \leq L\delta^{\beta+1}. \tag{*}$$

We now want to show that (*) implies condition (2.2.43). To assist the reader we drop the argument of the Itô coefficients $f_\alpha^{p_h}(n\Delta, Y_{n,\delta})$ and keep only the subindex α of $\tilde{\mathcal{I}}_{\alpha,n\Delta,(n+1)\Delta}$ and $\mathcal{I}_{\alpha,n\Delta,(n+1)\Delta}$. We then compute

$$\begin{aligned} H_1 &:= \left| \mathbb{E}\left[\prod_{h=1}^{l} \left(Y_{n+1,\delta}^{p_h} - Y_{n,\delta}^{p_h} \right) - \prod_{h=1}^{l} \left(\sum_{\alpha \in \Gamma_\beta \setminus \{v\}} f_\alpha^{p_h} \mathcal{I}_\alpha \right) \middle| \mathcal{A}_{\tau_n} \right] \right| \\ &= \left| \mathbb{E}\left[\prod_{h=1}^{l} \left(\sum_{\alpha \in \Gamma_\beta \setminus \{v\}} f_\alpha^{p_h} \tilde{\mathcal{I}}_\alpha \right) - \prod_{h=1}^{l} \left(\sum_{\alpha \in \Gamma_\beta \setminus \{v\}} f_\alpha^{p_h} \mathcal{I}_\alpha \right) \middle| \mathcal{A}_{\tau_n} \right] \right|. \end{aligned} \tag{2.2.44}$$

We expand both products within the conditional expectation and factor the Itô coefficients appearing in both sums. Then we get a sum with each summand consisting of $l+1$ products

$$f_{\alpha_1}^{p_{h_1}} \cdots f_{\alpha_l}^{p_{h_l}} \cdot \left(\prod_{k=1}^{l} \tilde{\mathcal{I}}_{\alpha_k} - \prod_{k=1}^{l} \mathcal{I}_{\alpha_k} \right)$$

for any choice of $\alpha_1, \ldots, \alpha_l \in \Gamma_\beta \setminus \{v\}$. Each Itô coefficient $f_\alpha^{p_h}$ involves the drift, diffusion and their derivatives, which are by assumption of Theorem 2.2.17 of polynomial growth. Since $Y_{n,\delta}$ is $\mathcal{A}_{\tau_n}$ measurable, we see that so is $f_{\alpha_1}^{p_{h_1}}(n\Delta, Y_{n,\delta}) \cdots f_{\alpha_l}^{p_{h_l}}(n\Delta, Y_{n,\delta})$.

By applying (ii) of Theorem 1.1.17 we conclude

$$\begin{aligned} H_1 &\leq \sum \left| f_{\alpha_l}^{p_{h_l}} \cdots f_{\alpha_l}^{p_{h_l}} \right| \left| \mathbb{E}\left[\prod_{k=1}^{l} \tilde{\mathcal{I}}_{\alpha_k} - \prod_{k=1}^{l} \mathcal{I}_{\alpha_k} \middle| \mathcal{A}_{\tau_n} \right] \right| \\ &\leq \sum K \left(1 + \|Y_{n,\delta}\|_2^{2r}\right) \left| \mathbb{E}\left[\prod_{k=1}^{l} \tilde{\mathcal{I}}_{\alpha_k} - \prod_{k=1}^{l} \mathcal{I}_{\alpha_k} \middle| \mathcal{A}_{\tau_n} \right] \right| \end{aligned}$$

where the sum is taken over all summands that appear if we expand the products of (2.2.44) and we have applied the polynomial growth of the Itô coefficients, see p. 51 for the definition of polynomial growth. The constants K and r can be chosen to depend only on β, so we can further deduce

$$H_1 \leq K \left(1 + \max_{0 \leq n \leq n_T} \|Y_{n,\delta}\|_2^{2r}\right) \sum \left| \mathbb{E}\left[\prod_{k=1}^{l} \tilde{\mathcal{I}}_{\alpha_k} - \prod_{k=1}^{l} \mathcal{I}_{\alpha_k} \middle| \mathcal{A}_{\tau_n} \right] \right|.$$

Since $l = 1, \ldots, 2\beta + 1$, we are able to apply equation (*) and conclude

$$H_1 \leq \tilde{K} \left(1 + \max_{0 \leq n \leq n_T} \|Y_{n,\delta}\|_2^{2r}\right) \delta^{\beta+1},$$

which is condition (2.2.43). Thus the numerical schemes we have introduced before do have the asserted convergence order.

2.3 Gaussian approximations

In this section we want to discuss the so called Gaussian approximation that we apply to a special class of Itô differential equations. We shall consider the SDE

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \tag{2.3.1}$$

where the drift b and diffusion coefficient σ are build of so called expolynomials that we introduce in Definition 2.3.1. In this case there exists to each SDE of that type an ordinary differential equation with a solution $Y = Y(t)$ which we call Gaussian approximation. The solution Y shall then be an approximation to the first and second moments of the solution X.

We first start with an example where the Gaussian approximation is exact: Let the following scalar linear SDE be given

$$dX_t = \gamma X_t dt + \delta X_t dW_t, \tag{2.3.2}$$

with $\gamma, \delta \in \mathbb{R}$, $X(0) = c \in \mathbb{R}$ and one-dimensional Wiener process $(W_t)_t$. Using the Itô–Doeblin formula we can further determine the time development of the second moment, namely

$$\begin{aligned} d((X_t)^2) &= 2X_t dX_t + dX_t dX_t \\ &= (2\gamma(X_t)^2 + \delta^2(X_t)^2)dt + 2\delta(X_t)^2 dW_t. \end{aligned} \tag{2.3.3}$$

By applying the expectation operator $\mathbb{E}[\cdot]$ on both sides of (2.3.2) and (2.3.3), we further deduce by applying the martingale property of the Itô integral, see Theorem 1.4.14, that

$$\begin{aligned} d\mathbb{E}[X_t] &= \gamma \mathbb{E}[X_t]\, dt, \\ d\mathbb{E}\left[(X_t)^2\right] &= (2\gamma + \delta^2)\mathbb{E}\left[(X_t)^2\right] dt, \end{aligned} \tag{2.3.4}$$

where we were allowed to interchange integration by Remark 1.5.6.(iii). We note from (2.3.4) that the determination of the first and second moment of the solution of (2.3.2) reduced to the determination of the solution $(y(t), Y(t))$ of the (independent) system of ordinary differential equations

$$\begin{aligned} \frac{\partial y}{\partial t} &= \gamma y(t), \\ \frac{\partial Y}{\partial t} &= (2\gamma + \delta^2) Y(t), \end{aligned}$$

where $y(t) = \mathbb{E}[X_t]$, $Y(t) = \mathbb{E}[(X_t)^2]$. $(y(t), Y(t))$ is the Gaussian approximation of (2.3.2) and determines exactly the first and second moments of X.

We now define the class of functions that we shall consider for the drift b and diffusion coefficient σ:

Definition 2.3.1. We denote by $\mathcal{P}_{ex}(\mathbb{R}^n, \mathbb{R})$ the set of all **expolynomials**, i.e., the set of all functions $g : \mathbb{R}^n \to \mathbb{R}$ that can be written as

$$g(x) = \exp\left(\alpha^\top x\right) p(x) = \exp\left(\sum_{i=1}^{n} \alpha^i x^i\right) p(x)$$

for all $x \in \mathbb{R}^n$, where $\alpha \in \mathbb{R}^n$ and $p \in \mathcal{P}(\mathbb{R}^n, \mathbb{R})$.

From now on we require each component of the drift $b = (b^k)$ and diffusion coefficient $\sigma = (\sigma^{k,j})$ of (2.3.1) to satisfy $b^k, \sigma^{k,j} \in \mathcal{P}_{ex}$ for $k = 1, \ldots, n$, $j = 1, \ldots, m$. We now introduce new notation as it proves convenient in the upcoming discussion:

Definition 2.3.2. Let Z be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ assuming values in $\mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$ be a Borel-measurable function. Further, let $C \in \mathbb{R}^{n \times n}$ be a positive definite

matrix. We then define

$$\mathbb{E}_C\left[f(Z)\right] := \frac{1}{\sqrt{\det(2\pi C)}} \int_{\mathbb{R}^n} \exp\left(-\frac{z^\top C^{-1} z}{2}\right) f(z) d\lambda(z).$$

Remark 2.3.3. Let Z be a normally distributed random variable assuming values in $\mathbb{R}^n$ with positive definite covariance matrix $C := \mathbb{C}\mathrm{ov}\left[Z\right]$ and mean $\mu \in \mathbb{R}^n$. We then find for f as in Definition 2.3.2

$$\mathbb{E}\left[f(Z)\right] = \mathbb{E}_C\left[f(Z+\mu)\right].$$

This motivates the following *definition* in case of a normally distributed random variable Z with general – and not necessarily positive definite – covariance matrix $C := \mathbb{C}\mathrm{ov}\left[Z\right]$ and mean $\mu \in \mathbb{R}^n$:

$$\mathbb{E}_C\left[f(Z+\mu)\right] := \mathbb{E}\left[f(Z)\right].$$

If Z is an arbitrary random variable with mean μ and covariance matrix $C := \mathbb{C}\mathrm{ov}\left[Z\right]$, we suppose

$$\mathbb{E}\left[f(Z)\right] \approx \mathbb{E}_C\left[f(Z+\mu)\right]$$

to hold, given that Z behaves similarly to a normally distributed random variable.

The basic idea of the Gaussian approximation is to consider the solution X as being Gaussian and then to utilize certain properties of Gaussian random variables. One property that we shall frequently use is that for Gaussian random variables the determination of higher moments can be reduced to the determination of the second moments, which is known as the Theorem of Wick-Isserlis:

Theorem 2.3.4 (**Theorem of Wick-Isserlis**). *Let $Z = (Z^1, \ldots, Z^n)$ be a normally distributed random variable assuming values in $\mathbb{R}^n$ with $\mathbb{E}\left[Z\right] = 0$ and with covariance matrix $C = (C^{i,j})_{i,j} := \mathbb{C}ov[Z]$. Further let (ℓ_i) be a finite sequence in $\{1, \ldots, n\}$ of length N. Then*

$$\mathbb{E}_C\left[Z^{\ell_1} Z^{\ell_2} \cdots Z^{\ell_N}\right] = 0; \quad \textit{for } N \textit{ odd} \tag{2.3.5}$$

i.e., all higher odd moments vanish and

$$\mathbb{E}_C\left[Z^{\ell_1} Z^{\ell_2} \cdots Z^{\ell_N}\right] = \sum \prod \mathbb{E}\left[C^{\ell_i,\ell_j}\right]; \quad \textit{for } N \textit{ even}, \tag{2.3.6}$$

where the notation $\sum\prod$ means summing over all distinct ways of partitioning $Z^{\ell_1}, Z^{\ell_2}, \ldots, Z^{\ell_N}$ into pairs.

Proof. See Wick (1950) and Isserlis (1918). □

Remark 2.3.5.

(i) The idea of the proof of 2.3.4 is easy to sketch: A normally distributed random variable Z has the characteristic function mentioned in Definition 1.2.3 and Remark 1.2.4 relates derivatives of the characteristic function to higher moments of Z.

(ii) Let X be a normally distributed random variable with mean μ. As we can apply Theorem 2.3.4 to the normally distributed random variable $Z := X - \mu$, we can also determine higher moments of X.

We now give two examples how to reduce the determination of higher moments of normally distributed random variables to the the determination of the second moments: Let $Z = (X, Y)$ satisfy the assumptions of Theorem 2.3.4. We compute the following expectations:

- $\mathbb{E}_C[X^2Y^2] = \mathbb{E}[XXYY]$: All distinct ways of partitioning X, X, Y, Y into pairs are $P_1 = \{(X,X),(Y,Y)\}$, $P_2 = \{(X,Y),(X,Y)\}$, $P_3 = P_2 = \{(X,Y),(X,Y)\}$, so we deduce
$$\mathbb{E}_C\left[X^2Y^2\right] = \mathbb{E}_C\left[X^2\right]\mathbb{E}_C\left[Y^2\right] + 2\mathbb{E}_C[XY]^2.$$

- $\mathbb{E}_C[X^4] = \mathbb{E}_C[XXXX]$: All distinct ways of partitioning X, X, X, X into pairs are $P_1 = P_2 = P_3 = \{(X,X),(X,X)\}$. Thus we deduce
$$\mathbb{E}_C\left[X^4\right] = 3\mathbb{E}_C\left[X^2\right]^2.$$

The following theorem states how we can determine the expectation of an expolynomial of a normally distributed random variable.

Theorem 2.3.6. *Let $X = (X^1, \ldots, X^n)$ be a normally distributed random variable with mean $\mu \in \mathbb{R}^n$ and covariance matrix $C := \mathbb{C}ov[X]$. Further let $\alpha \in \mathbb{R}^n$ and $p \in \mathcal{P}(\mathbb{R}^n, \mathbb{R})$. Then*

$$\mathbb{E}\left[\exp\left(\alpha^\top X\right)p(X)\right] = \exp\left(\alpha^\top\mu + \frac{\alpha^\top C\alpha}{2}\right)\mathbb{E}_C\left[p\left(X + C\alpha + \mu\right)\right]. \tag{2.3.7}$$

We can determine the expectation $\mathbb{E}_C[p(X + C\alpha + \mu)]$ by using Theorem 2.3.4.

Proof. We first assume that C is positive definite, i.e., X has a density function

$$\varphi(X) = \frac{1}{\sqrt{\det(2\pi C)}}\exp\left(-\frac{(x-\mu)^\top C^{-1}(x-\mu)}{2}\right).$$

We deduce using Definition 1.1.4, $K := \frac{1}{\sqrt{\det(2\pi C)}}$ and $C = C^\top$ that $g(X)$ defined by $g(X) := \exp\left(\alpha^\top X\right) p(X)$ has the expectation

$$\begin{aligned}
\mathbb{E}\left[g(X)\right] &= K \int \exp(\alpha^\top x) p(x) \exp\left(-\frac{(x-\mu)^\top C^{-1}(x-\mu)}{2}\right) d\lambda(x) \\
&= K \int p(x) \exp\left(-\frac{(x-\mu)^\top C^{-1}(x-\mu) - 2\alpha^\top x}{2}\right) d\lambda(x) \\
&= K \int \exp\left(\frac{2\alpha^\top \mu + \alpha^\top C\alpha}{2}\right) \\
&\qquad \cdot p(x) \exp\left(-\frac{\left(x-\mu-C\alpha\right)^\top C^{-1}\left(x-\mu-C\alpha\right)}{2}\right) d\lambda(x) \\
&= K \exp\left(\alpha^\top \mu + \frac{\alpha^\top C\alpha}{2}\right) \\
&\qquad \cdot \int p(x) \exp\left(-\frac{\left(x-(\mu+C\alpha)\right)^\top C^{-1}\left(x-(\mu+C\alpha)\right)}{2}\right) d\lambda(x).
\end{aligned}$$

By using the affine transformation $\tilde{x} = x - C\alpha$, we further deduce

$$\begin{aligned}
\mathbb{E}\left[g(X)\right] &= \exp\left(\alpha^\top \mu + \frac{\alpha^\top C\alpha}{2}\right) \int p\left(\tilde{x} + C\alpha\right) \varphi(\tilde{x}) d\lambda(\tilde{x}) \\
&= \exp\left(\alpha^\top \mu + \frac{\alpha^\top C\alpha}{2}\right) \mathbb{E}\left[p\left(Z + C\alpha\right)\right] \\
&= \exp\left(\alpha^\top \mu + \frac{\alpha^\top C\alpha}{2}\right) \mathbb{E}_C\left[p\left(Z + C\alpha + \mu\right)\right].
\end{aligned}$$

For general positive semidefinite matrix C, we define $C_\epsilon := C + \epsilon I_n$ for $\epsilon > 0$ and note that C_ϵ is positive definite. Further we define by X_ϵ the normally distributed random variable with mean μ and covariance matrix C_ϵ. By the above we find

$$\mathbb{E}\left[\exp\left(\alpha^\top X_\epsilon\right) p(X_\epsilon)\right] = \exp\left(\alpha^\top \mu + \frac{\alpha^\top C_\epsilon \alpha}{2}\right) \mathbb{E}_{C_\epsilon}\left[p\left(X_\epsilon + C_\epsilon\alpha + \mu\right)\right]. \tag{2.3.8}$$

We know by Definition 1.2.3 that the characteristic function $\phi_\epsilon : u \mapsto \mathbb{E}\left[\exp\left(iu^\top X_\epsilon\right)\right]$ of X_ϵ is given by

$$\phi_\epsilon(u) = \exp\left(-\frac{1}{2} u^\top \left(C_\epsilon u\right) + iu^\top \mu\right).$$

As ϕ_ϵ is continuous in C_ϵ, we deduce that $\phi_\epsilon(u) \to \phi(u)$ for all $u \in \mathbb{R}^n$ as $\epsilon \to 0$, where ϕ denotes the characteristic function of X. This implies that the distribution of X_ϵ converges to the distribution of X, cf. Kallenberg (2002, p. 86), which further implies that $\mathbb{E}\left[f(X_\epsilon)\right] \to \mathbb{E}\left[f(X)\right]$ as $\epsilon \to 0$ for any continuous and *bounded* function f,

cf. Kallenberg (2002, p. 65). As we want to show $\mathbb{E}[f(X_\epsilon)] \to \mathbb{E}[f(X)]$ in the case of the unbounded functions $f = g$ and $f = p$, we are not yet finished. But in order to have $\mathbb{E}[f(X_\epsilon)] \to \mathbb{E}[f(X)]$ for a measurable and not necessarily bounded function f, it suffices to show that there exists a $p > 1$ and a constant M independent of ϵ such that $\|f(X_\epsilon)\|_{L^p} \leq M < \infty \ \forall \epsilon$, cf. Van der Vaart (2000, Theorem 2.20) and Kallenberg (2002, p. 67). We first show directly that $\mathbb{E}[p(X_\epsilon)] \to \mathbb{E}[p(X)]$ in the case of a polynomial p and then further deduce that g satisfies $\|g(X_\epsilon)\|_{L^p} \leq M < \infty \ \forall \epsilon$ with $p = 2$:
From Remark 1.2.4 we know that

$$\mathbb{E}\left[\left(X_\epsilon^1\right)^{m_1} \cdots \left(X_\epsilon^n\right)^{m_n}\right] = \prod_{j=1}^{n} (-i)^{m_j} \frac{\partial^{m_j} \phi_\epsilon}{\partial u_j^{m_j}}(0)$$

for any $m_1, \ldots, m_n \in \mathbb{N}$. By the definition of ϕ_ϵ we deduce that $\frac{\partial^{m_j} \phi_\epsilon}{\partial u_j^{m_j}}(0)$ depends continuously on C_ϵ and thus $\frac{\partial^{m_j} \phi_\epsilon}{\partial u_j^{m_j}}(0) \to \frac{\partial^{m_j} \phi}{\partial u_j^{m_j}}(0)$ as $\epsilon \to 0$. By the above we conclude $\mathbb{E}\left[(X_\epsilon^1)^{m_1} \cdots (X_\epsilon^n)^{m_n}\right] \to \mathbb{E}\left[(X^1)^{m_1} \cdots (X^n)^{m_n}\right]$ as $\epsilon \to 0$ and in general $\mathbb{E}[p(X_\epsilon)] \to \mathbb{E}[p(X)]$ for a polynomial p. Because of $\mathbb{E}_{C_\epsilon}[f(X_\epsilon + \mu)] = \mathbb{E}[f(X_\epsilon)]$ for any f, this further implies

$$\exp\left(\frac{\alpha^\top C_\epsilon \alpha}{2}\right) \mathbb{E}_{C_\epsilon}\left[p\left(X_\epsilon + C_\epsilon \alpha + \mu\right)\right] \to \exp\left(\frac{\alpha^\top C \alpha}{2}\right) \mathbb{E}_C\left[p\left(X + C\alpha + \mu\right)\right],$$

as $\epsilon \to 0$. Using (2.3.8), we further deduce

$$\|g(X_\epsilon)\|_{L^2}^2 = \exp\left(2\alpha^\top \mu + 2\alpha^\top C_\epsilon \alpha\right) \mathbb{E}_{C_\epsilon}\left[\left(p\left(X_\epsilon + 2C_\epsilon \alpha + \mu\right)\right)^2\right].$$

As we have already shown that the right side of the last equation converges for any $\alpha \in \mathbb{R}^n$ and any polynomial, we see that there exists a constant M independent of ϵ such that $\|g(X_\epsilon)\|_{L^2}^2 \leq M < \infty$ for all ϵ. By the above reasoning, this implies $\mathbb{E}[g(X_\epsilon)] \to \mathbb{E}[g(X)]$. Thus

$$\mathbb{E}\left[\exp\left(\alpha^\top X\right) p(X)\right] = \exp\left(\alpha^\top \mu + \frac{\alpha^\top C \alpha}{2}\right) \mathbb{E}_C\left[p\left(X + C\alpha + \mu\right)\right].$$

□

Remark 2.3.7.

(i) Using the notation of Definition 2.3.2 we can rewrite Equation (2.3.7) as

$$\mathbb{E}_C\left[\exp\left(\alpha^\top \left(X + \mu\right)\right) p(X + \mu)\right] = \exp\left(\alpha^\top \mu + \frac{\alpha^\top C \alpha}{2}\right) \mathbb{E}_C\left[p\left(X + C\alpha + \mu\right)\right].$$

(ii) Let X satisfy the assumptions of Theorem 2.3.6. Theorem 2.3.6 and Theorem

2.3.4 imply that the determination of an expolynomial of X can be reduced to the determination of its mean μ and its second moments $C(t)$.

Given a SDE (2.3.1) with expolynomial drift and diffusion coefficients we determine the Gaussian approximation in the following way:

(i) Let $\mu(t) := \mathbb{E}[X_t]$ be the mean of $X = (X_t) = (X_t^1, \dots, X_t^n)$ and $C(t) := \mathbb{C}\mathrm{ov}[X_t]$ the covariance matrix of X. We determine the dynamics of the second moments $C(t)$ by using the Itô–Doeblin formula, Theorem 1.4.27. The drift and diffusion coefficients of the SDEs of the second moments again consist of expolynomials. We get the following system of SDEs:

$$\begin{aligned}
dX_t &= b(t, X_t)dt + \sigma(t, X_t)dW_t, \\
d\left((X_t^1)^2\right) &= 2X_t^1 dX_t^1 + dX_t^1 dX_t^1, \\
d\left(X_t^1 X_t^2\right) &= X_t^1 dX_t^2 + X_t^2 dX_t^1 + dX_t^1 dX_t^2, \\
&\vdots \\
d\left(X_t^1 X_t^n\right) &= X_t^1 dX_t^n + X_t^n dX_t^1 + dX_t^1 dX_t^n, \\
d\left((X_t^2)^2\right) &= 2X_t^2 dX_t^2 + dX_t^2 dX_t^2, \\
d\left(X_t^2 X_t^3\right) &= X_t^2 dX_t^3 + X_t^3 dX_t^2 + dX_t^2 dX_t^3, \\
&\vdots \\
d\left(X_t^2 X_t^n\right) &= X_t^2 dX_t^n + X_t^n dX_t^2 + dX_t^2 dX_t^n, \\
&\vdots \\
d\left(X_t^n X_t^{n-1}\right) &= X_t^n dX_t^{n-1} + X_t^{n-1} dX_t^n + dX_t^n dX_t^{n-1}, \\
d\left((X_t^n)^n\right) &= 2X_t^n dX_t^n + dX_t^n dX_t^n.
\end{aligned} \tag{2.3.9}$$

(ii) We then apply the expectation operator to each equation in (2.3.9). The Itô integrals vanish by the martingale property. Futher, we are allowed to interchange integration w.r.t. the probability measure $\mathbb{P}$ and w.r.t. the Lebesgue measure λ by Remark 1.5.6.(iii). In total, Equation (2.3.9) changes to

$$\begin{aligned}
d\mathbb{E}\left[X_t^i\right] &= \mathbb{E}\left[m^i(X_t)\right] dt, \\
d\mathbb{E}\left[X_t^j X_t^k\right] &= \mathbb{E}\left[n^{j,k}(X_t)\right] dt,
\end{aligned}$$

for $i, j, k = 1, \dots, n$, where $m^i, n^{j,k} : \mathbb{R}^n \to \mathbb{R}$ are appropriately chosen functions,

which are expolynomials by the last point. This system can also be written as

$$\begin{aligned} d\mu^i &= \mathbb{E}\left[m^i(X_t)\right] dt, \\ &\approx \mathbb{E}_C\left[m^i(X_t+\mu)\right] dt, \\ dC^{j,k} &= \mathbb{E}\left[n^{j,k}(X_t)\right] dt - d\left(\mu^j\mu^k\right) \\ &\approx \mathbb{E}_C\left[n^{j,k}(X_t+\mu)\right] dt - d\left(\mu^j\mu^k\right), \end{aligned} \tag{2.3.10}$$

where we have used in line 2 and 4 that we suppose that X_t behaves similarly to a normally distributed random variable with mean $\mu(t)$ and covariance matrix $C(t) = \mathbb{C}\mathrm{ov}[X_t]$, see Remark 2.3.3. To assist the reader we have omitted the argument t in (2.3.10), as well as in the following equations.

(iii) We first assume that the solution X of (2.3.1) is Gaussian with mean $\mu(t)$ and covariance matrix $C(t)$. In this case the sign $\approx$ in Equation (2.3.10) turns into an equal sign $=$, see Remark 2.3.3. Using $m = (m^i)_i$ and $n = (n^{j,k})_{j,k}$ we can thus rewrite equation (2.3.10) as

$$\begin{aligned} d\mu &= \mathbb{E}_C\left[m(X_t+\mu)\right] dt, \\ dC &= \mathbb{E}_C\left[n(X_t+\mu)\right] dt - d\left(\mu\mu^\top\right), \end{aligned} \tag{2.3.11}$$

As m and n contain only expolynomials we can reduce the determination of $\mathbb{E}_C[m(X_t+\mu)]$ and $\mathbb{E}_C[n(X_t+\mu)]$ to the determination of the mean $\mu(t)$ and the second moments $C(t)$ of X by using Theorem 2.3.4 and Theorem 2.3.6. We can thus evaluate the coefficients involving the expectation of an expolynomial in terms of the first and second moments of X. By writing $y(t) = \mu(t)$, $Y(t) = C(t) + \left(\mu\mu^\top\right)(t)$ we thus get the following *closed* system of ordinary differential equations

$$\begin{aligned} \frac{\partial y}{\partial t} &= \tilde{m}(y, Y), \\ \frac{\partial Y}{\partial t} &= \tilde{n}(y, Y), \end{aligned} \tag{2.3.12}$$

where $\tilde{m}$ and $\tilde{n}$ result from m and n according to Theorem 2.3.4 and Theorem 2.3.6.

(iv) In general, X is not Gaussian but we find the following relation between $y(t)$, $Y(t)$ of (2.3.12) and $\mu(t)$, $C(t)$:

$$\begin{aligned} y &\approx \mathbb{E}[X_t] = \mu, \\ Y &\approx \left(\mathbb{E}\left[X_t^j X_t^k\right]\right)_{j,k} = C + \mu\mu^\top. \end{aligned}$$

Remark 2.3.8. In the above we have deduced a differential equation in terms of $y(t) \approx \mu(t)$ and $Y(t) \approx C(t) + \left(\mu\mu^\top\right)(t)$. Of course it is also possible to state a differential equation in terms of $y(t) \approx \mu(t)$ and $\tilde{Y}(t) \approx C(t)$. The latter approach can sometimes be numerically more stable than the first.

As an example we consider the following one-dimensional SDE

$$dX_t = \left(g(X_t)^3 - mX_t\right) dt + \sigma dW_t, \tag{2.3.13}$$

with $g, m, \sigma \in \mathbb{R}$, cf. Honerkamp (1990, p. 232). By applying the Itô–Doeblin formula we further deduce

$$\begin{aligned} d\left((X_t)^2\right) &= 2X_t dX_t + dX_t dX_t, \\ &= (2g(X_t)^4 - 2m(X_t)^2 + \sigma^2)dt + 2X_t dW_t, \end{aligned} \tag{2.3.14}$$

and by applying the expectation operator on both sides of (2.3.13) and (2.3.14), we get

$$\begin{aligned} d\mathbb{E}\left[X_t\right] &= \left(g\mathbb{E}\left[(X_t)^3\right] - m\mathbb{E}\left[X_t\right]\right) dt, \\ d\left(\mathbb{E}\left[(X_t)^2\right]\right) &= (2g\mathbb{E}\left[(X_t)^4\right] - 2m\mathbb{E}\left[(X_t)^2\right] + \sigma^2)dt. \end{aligned}$$

We compute the higher moments $\mathbb{E}\left[(X_t)^3\right]$ and $\mathbb{E}\left[(X_t)^4\right]$ by using Theorem 2.3.4 but having in mind that we have to apply it to the normalized random variables $(X_t - \mathbb{E}\left[X_t\right])^3$ and $(X_t - \mathbb{E}\left[X_t\right])^4$:

$$\begin{aligned} \mathbb{E}\left[(X_t)^3\right] &= 3\mathbb{E}\left[X\right]\mathbb{E}\left[(X_t)^2\right] - 2\mathbb{E}\left[X_t\right]^3, \\ \mathbb{E}\left[(X_t)^4\right] &= 3\mathbb{E}\left[(X_t)^2\right]^2 - 2\mathbb{E}\left[X_t\right]^4. \end{aligned}$$

In total we get the following system of ordinary differential equations

$$\begin{aligned} \frac{\partial y}{\partial t} &= g\left(3y \cdot Y - 2y^3\right) - my, \\ \frac{\partial Y}{\partial t} &= 2g\left(3Y^2 - 2y^4\right) - 2mY + \sigma^2, \end{aligned}$$

with $y(t) \approx \mathbb{E}\left[X_t\right] \in \mathbb{R}$ and $Y(t) \approx \mathbb{E}\left[(X_t)^2\right] \in \mathbb{R}$. If we follow the idea of Remark 2.3.8 and want to find a differential equation w.r.t. $\tilde{Y}(t) \approx C(t)$, we deduce from the ODEs for y and Y by $\tilde{Y} = Y - y^2$

$$\begin{aligned} \frac{\partial y}{\partial t} &= g\left(3y \cdot Y - 2y^3\right) - my, \\ \frac{\partial \tilde{Y}}{\partial t} &= 6g\left(\tilde{Y}^2 - 3\tilde{Y}y^2 + 2y^4\right) - 2m\left(\tilde{Y} - 2y^2\right) + \sigma^2. \end{aligned}$$

In Section 3.2.2 of Chapter 3 we shall investigate Equation (2.3.13) numerically and compare the results of the Gaussian approximation with the results of the Euler–Maruyama method (2.2.2).

Remark 2.3.9. The assumption that the solution X is Gaussian is in general not true. It only holds for Ornstein–Uhlenbeck processes, i.e., for process with linear drift term and constant diffusion term, cf. Gardiner (1990, pp. 74–77) or Protter (1990, p. 243). Nonetheless, we shall see in Chapter 3 that – though not exact – we can still attain good results via Gaussian approximation. The Gaussian approximation is known in the context of **perturbation theory** and is also called the **mean field approximation**. Perturbation theory of a polynomial Itô SDE (2.3.1), i.e., the drift and diffusion coefficients are polynomials, results in so called Feynman graphs that describe the stochastic character of the solution X. One of the building blocks of such Feynman graphs are so called vertex functions. In the case of Gaussian approximation they are set to 0 if they excess a certain order. By involving higher order vertex functions it is possible to imply a non Gaussian character of the solution X, for example by using the so called **Schwinger–Dyson** approximation. We refer the reader interested in perturbation theory to Honerkamp (1990, Chapter 9-10).

Chapter 3

Computer Experiments

This chapter treats the MATLAB implementation of the one-step methods that we have introduced in Section 2.2 of Chapter 2, as well as a comparison between the Gaussian approximation of Section 2.3 of Chapter 2 and the one-step schemes. We first discuss the generation of random numbers in MATLAB, based on the random variables that are used by the one-step schemes of Section 2.2 of Chapter 2. Thereafter, we treat four examples of expolynomial SDEs, each of different structure with one or two-dimensional noise, additive or multiplicative noise and assuming values in $\mathbb{R}$ or $\mathbb{R}^2$. The first three SDEs are purely polynomial, the fourth SDE also contains real expolynomials. To each SDE we apply the Gaussian approximation as well as a one-step scheme for comparison and then discuss the results.
The first and last SDE of Subsection 3.2.1 and 3.2.4 are constructed examples and are only of theoretical interest. The second and third SDE of Subsection 3.2.2 and 3.2.3 are taken from HONERKAMP (1990, p. 232) and KLOEDEN et al. (1994, p. 219) and are also of practical interest.

3.1 Random numbers in Matlab

The numerical one step schemes that we have introduced in Section 2.2 of Chapter 2 require the generation of pseudo random numbers. In this section we shortly summarize how we can generate pseudo random numbers in MATLAB. Throughout this section we denote by $\Delta_\ell = \Delta$ the stepsize of the ℓth step.

Euler–Maruyama scheme

For the Euler–Maruyama scheme, see (2.2.2), we have to generate normally distributed random numbers with mean 0 and variance Δ_ℓ, see (2.2.3). In MATLAB we can generate n_t-many independent normally distributed m-dimensional pseudo-random vectors

satisfying these properties with the command

```
1 √Δ_ℓ*randn(m,n_t)
```

Simplified Euler–Maruyama scheme

The simplified Euler–Maruyama scheme, see (2.2.7), requires the generation of independent two-point distributed random variables ξ_ℓ^j satisfying $\mathbb{P}(\xi_\ell^j = \pm 1) = 0.5$. In MATLAB we can generate n_t-many such two-point distributed m-dimensional pseudo-random vectors with the command

```
1 2*randi(2,m,n_t)-3
```

Order 2.0 weak Taylor scheme

For the order 2.0 weak Taylor scheme, see (2.2.39), we have to generate a two-point distributed random variable V_{j_1,j_2} that satisfies $\mathbb{P}\left(V_{j_1,j_2} = \pm\Delta\right) = \frac{1}{2}$, which is done similarly as for the simplified Euler–Maruyama scheme with `randi`. Moreover, the order 2.0 weak Taylor scheme requires the generation of independent random variables $\Delta\tilde{W}^j$ that are either normally distributed with mean 0 and variance Δ – which is done similarly as for the Euler–Maruyama scheme with `randn` – or independent three-point distributed random variables satisfying condition (2.2.29). In MATLAB we can generate the m such random variables with the command

```
1 randsample([-sqrt(3*Δ),0,sqrt(3*Δ)],m,true,[1/6 2/3 1/6])
```

In our implementation of the order 2.0 weak Taylor scheme, see p. 119, we choose the generation of the three-point distributed random variables.

Order 3.0 weak Taylor scheme

The order 3.0 weak Taylor scheme requires the generation of two normally distributed random variables $\Delta\tilde{W}$, ΔZ, each with mean 0 and variance Δ, respectively $\frac{1}{3}\Delta^3$. $\Delta\tilde{W}$ and ΔZ are correlated by $\mathbb{E}\left[\Delta\tilde{W}\Delta Z\right] = \frac{1}{2}\Delta^2$. In MATLAB m pairs of Gaussian random numbers satisfying these conditions can be generated by defining the covariance matrix `SIGMA`$= \mathbb{C}\mathrm{ov}\left[\Delta\tilde{W}; \Delta\tilde{Z}\right]$

```
1 SIGMA=[Δ,1/2*Δ^2;1/2*Δ^2,1/3*Δ^3]
```

and then using the command

```
2 mvnrnd([0 0],SIGMA,m)
```

3.2 One-step methods vs. Gaussian approximation

In this section we compare in various examples the results of the Gaussian approximation, see Section 2.3 of Chapter 2, with the results of the one-step methods that we have introduced in Section 2.2 of Chapter 2. To do so, we shall only consider stochastic differential equations with expolynomial drift and diffusion coefficients, see Definition 2.3.1. The examples handle SDEs that are of dimension $n = 1, 2$, either with additive or multiplicative noise.

As regards the Gaussian approximation, the corresponding system of ordinary differential equations is solved using the MATLAB solver `ode45` for non-stiff ordinary differential equations. As we only handle SDEs of dimension $n = 1, 2$, the corresponding ODE of the Gaussian approximation is of dimension 2, respectively 5. We shall use the following notation that is slightly different to the one we have used in Section 2.2 of Chapter 2, but better suited for the numerical implementation: Let $Y_t = (Y_t^1, Y_t^2)$, resp. $Y_t = (Y_t^1, Y_t^2, Y_t^3, Y_t^4, Y_t^5)$, be the solution of the corresponding ODE of the Gaussian approximation. Y_t shall then approximate the first and second moments of the solution $(X_t) \in \mathbb{R}^{n=1,2}$ of the SDE in the following way:

for $n = 1$	for $n = 2$
$Y_t^1 \approx \mathbb{E}[X_t] \qquad Y_t^2 \approx \mathbb{E}\left[(X_t)^2\right]$	$Y_t^1 \approx \mathbb{E}\left[X_t^1\right] \quad Y_t^3 \approx \mathbb{E}\left[\left(X_t^1\right)^2\right]$ $Y_t^2 \approx \mathbb{E}\left[X_t^2\right] \quad Y_t^4 \approx \mathbb{E}\left[X_t^1 X_t^2\right]$ $Y_t^5 \approx \mathbb{E}\left[\left(X_t^2\right)^2\right]$

where we write $X_t = (X_t^1, X_t^2)$ in the case of $n = 2$.

To simplify the comparison between the Gaussian approximation and a one-step scheme, we further use the following notation: If X denotes the solution of the respective SDE, we denote by $\mathbb{E}[X_t^i]$ [Gaussian approximation], resp. $\mathbb{E}\left[X_t^j X_t^k\right]$ [Gaussian approximation], the approximation of $\mathbb{E}[X_t^i]$, resp. $\mathbb{E}\left[X_t^j X_t^k\right]$, computed via the Gaussian approximation. Similarly, we denote by $\mathbb{E}[X_t^i]$ [one-step method], $\mathbb{E}\left[X_t^j X_t^k\right]$ [one-step method] the approximation of $\mathbb{E}[X_t^i]$, resp. $\mathbb{E}\left[X_t^j X_t^k\right]$, computed by the respective one-step method, e.g. $\mathbb{E}[X_t^i]$ [Euler–Maruyama] denoting the approximation of $\mathbb{E}[X_t^i]$ computed by the Euler–Maruyama scheme (2.2.2).

As to the specific MATLAB functions that we use in the examples, we refer the reader to the Appendix where the programs are described in more detail.

3.2.1 Explicitly solvable SDE

In this subsection we consider a two-dimensional SDE with solution $X = (X_t^1, X_t^2)$ for which we can explicitly determine $\mathbb{E}[X_t^1]$, but $\mathbb{E}[X_t^1]$ is not solved accurately by the

Gaussian approximation method. We consider the following two-dimensional SDE with multiplicative one-dimensional noise

$$\begin{aligned} dX_t^1 &= \left(X_t^2\right)^3 dt + \alpha X_t^1 dW_t, \\ dX_t^2 &= \beta X_t^2 dt + \gamma X_t^2 dW_t, \end{aligned} \tag{3.2.1}$$

with initial value $X_0 = (0.1, 1)$, where $\alpha, \beta, \gamma \in \mathbb{R}$ and $t \in [0, 1]$. This SDE is a rather constructed example but we can see that the Gaussian approximation does not accurately determine the first moment $\mathbb{E}\left[X_t^1\right]$, as we have to apply the Theorem of Wick and Isserlis, Theorem 2.3.4, to determine $\mathbb{E}\left[\left(X_t^2\right)^3\right]$. Further, we find that the (linear) SDE for X_t^2 is independent of X_t^1 and that the solution of X_t^2 is given by

$$X_t^2 = 1 \cdot \exp\left(\left(\beta - \frac{1}{2}\gamma^2\right) t - \gamma W_t\right).$$

Using the Itô–Doeblin formula, Theorem 1.4.27, the dynamics of $\left(X_t^2\right)^3$ are given by

$$d\left(\left(X_t^2\right)^3\right) = 3\left(\beta + \gamma^2\right)\left(X_t^2\right)^3 dt + 3\gamma\left(X_t^2\right)^3 W_t.$$

In the following we are only interested to find an approximation of $\mathbb{E}\left[X_t^1\right]$, because in this case we can compute the exact value of $\mathbb{E}\left[X_t^1\right]$ by using the dynamics of $\left(X_t^2\right)^3$. We then find

$$\mathbb{E}\left[X_t^1\right] = \int_0^t 1^3 \exp\left(3\left(\beta + \gamma^2\right) s\right) d\lambda(s) + \mathbb{E}\left[X_0^1\right] = \frac{\exp\left(3(\beta + \gamma^2)t\right) - 1}{3\left(\beta + \gamma^2\right)} + 0.1.$$

Knowing the exact value of $\mathbb{E}\left[X_t^1\right]$ we want to perform the following three numerical experiments: Setting $\alpha = 10^{-6}$ very low and $\beta = 1$ we shall discuss in the first experiment the influence of γ on the maximum error of $\mathbb{E}\left[X_t^i\right]$ [Gaussian approximation] for $t \in [0, 1]$. In the second experiment we compare numerically for α, β as above and $\gamma = 0.25$ whether the Euler–Maruyama method, see (2.2.2), or the simplified Euler–Maruyama method, see (2.2.7), attain better results in approximating $\mathbb{E}\left[X_t^1\right]$ with respect to the cputime that they require. In the third and last experiment, we show numerically the convergence order of the order 3.0 weak Taylor scheme, see (2.2.40), in approximating $\mathbb{E}\left[X_t^1\right]$ at $t = 1$ and compare the results with the values of the Gaussian approximation. Also, we want to revisit the results of both Euler schemes of the second experiment and compare them to the required cputime of the Taylor scheme.

As regards the Gaussian approximation, the corresponding system of ordinary differ-

ential equations is given by

$$\begin{aligned}
\frac{\partial Y^1}{\partial t} =& \left(Y^2\right)^3 + 3Y^2\left(Y^5 - \left(Y^2\right)^2\right), \\
\frac{\partial Y^2}{\partial t} =& \beta Y^2, \\
\frac{\partial Y^3}{\partial t} =& Y^1\left(Y^2\right)^3 + 3\left(Y^2\right)^2\left(Y^4 - Y^1Y^2\right) + 3Y^1Y^2\left(Y^5 - \left(Y^2\right)^2\right) + \alpha^2 Y^3 \\
&+ 3\left(Y^5 - \left(Y^2\right)^2\right)\left(Y^4 - Y^1Y^2\right), \\
\frac{\partial Y^4}{\partial t} =& \left(Y^2\right)^4 + 6\left(Y^2\right)^2\left(Y^5 - \left(Y^2\right)^2\right) + 3\left(Y^5 - \left(Y^2\right)^2\right)^2 + \left(\beta + \alpha\gamma\right)Y^4, \\
\frac{\partial Y^5}{\partial t} =& 2\beta Y^5 + \gamma^2 Y^5,
\end{aligned} \tag{3.2.2}$$

where we have dropped the argument t to assist the reader. Further, we get the following relations $Y^1(t) \approx \mathbb{E}\left[X_t^1\right]$, $Y^2(t) \approx \mathbb{E}\left[X_t^2\right]$, $Y^3(t) \approx \mathbb{E}\left[\left(X_t^1\right)^2\right]$, $Y^4(t) \approx \mathbb{E}\left[X_t^1X_t^2\right]$, $Y^5(t) \approx \mathbb{E}\left[\left(X_t^2\right)^2\right]$ and the corresponding initial value for (3.2.2) is given by $Y(0) = (0.1, 1, (0.1)^2, 1 \cdot 0.1, 1)$.

1st experiment: Influence of γ on $\mathbb{E}\left[X_t^1\right]$ [Gaussian approximation]

Setting $\alpha = 10^{-6}$, $\beta = 1$, we want to discuss the influence of γ in (3.2.1) and (3.2.2) on $\mathbb{E}\left[X_t^1\right]$ [Gaussian approximation] for $t \in [0, 1]$. The relevant equations in (3.2.2) for the determination of $\mathbb{E}\left[X_t^1\right]$ are the ODEs for $\frac{\partial Y^1}{\partial t}$, $\frac{\partial Y^2}{\partial t}$ and $\frac{\partial Y^5}{\partial t}$. As the term γ appears in those equations only squared, we consider w.l.o.g. only positive values for γ. We solve (3.2.2) for $\gamma = \frac{i}{20}$, $i = 0, \ldots, 20$ and compute the maximum relative and maximum absolute error between $\mathbb{E}\left[X_t^1\right]$ and $\mathbb{E}\left[X_t^1\right]$ [Gaussian approximation] for $t = \frac{i}{8}$, $i = 0, \ldots, 8$. All those steps are executed by `gaussian_approximation_explicit` that we call with the argument `0:0.125:1`.

Figure 3.1 shows the result of the numerical experiment. Before we start the discussion, it is important to mention that the specific time instants t, t' for which the maximum absolute and maximum relative error between $\mathbb{E}\left[X_t^1\right]$ [Gaussian approximation] and $\mathbb{E}\left[X_t^1\right]$ are computed for fixed γ, are usually *not* equal. Whereas the maximum absolute error is always attained at $t = 1$, the maximum relative error is sometimes attained at $t' < 1$. According to Figure 3.1 the maximum relative and absolute error of $\mathbb{E}\left[X_t^1\right]$ [Gaussian approximation] for $\gamma = 0$ were around 10^{-7} and 10^{-6} respectively. In the noise-free case $\gamma = 0$, the system (3.2.2) turns into a simple quadrature problem and (3.2.2) should determine $\mathbb{E}\left[X_t^1\right]$ accurately. Solving (3.2.2) explicitly via quadrature – using the MATLAB function `quad` – leads to a maximum relative and absolute error of around 10^{-8}. We can further see from Figure 3.1 that the maximum relative error is

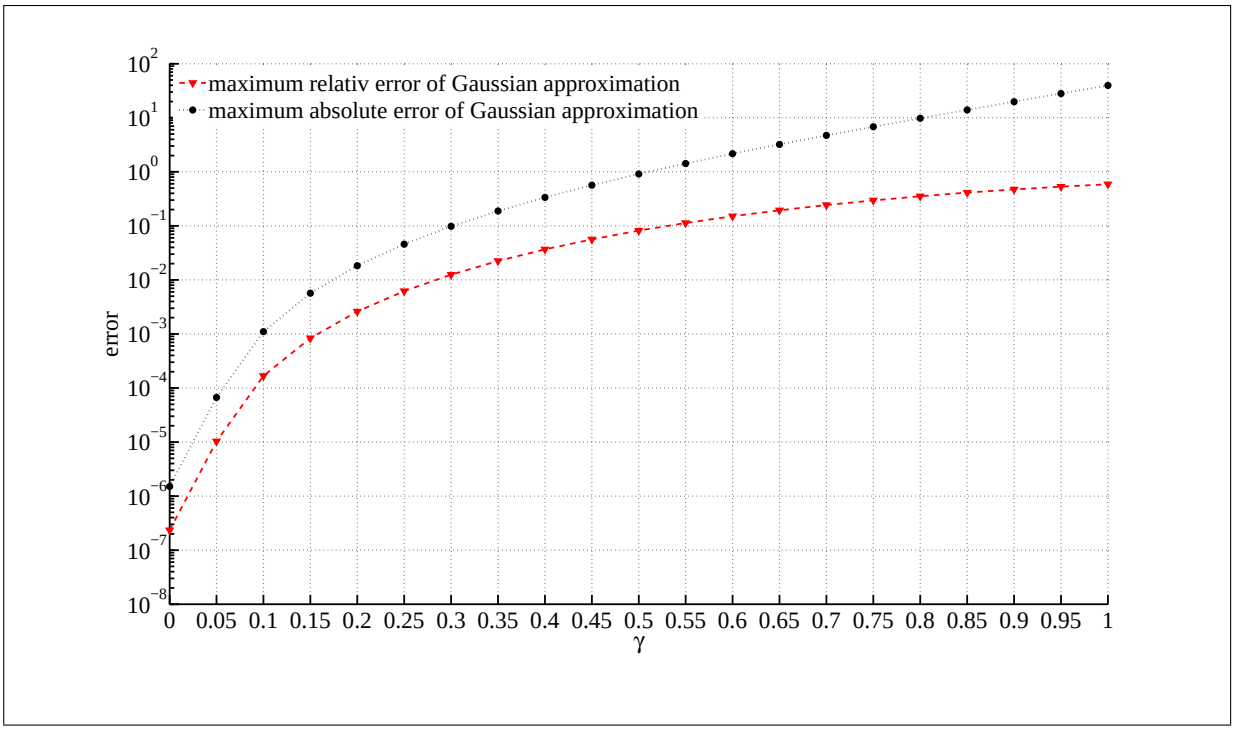


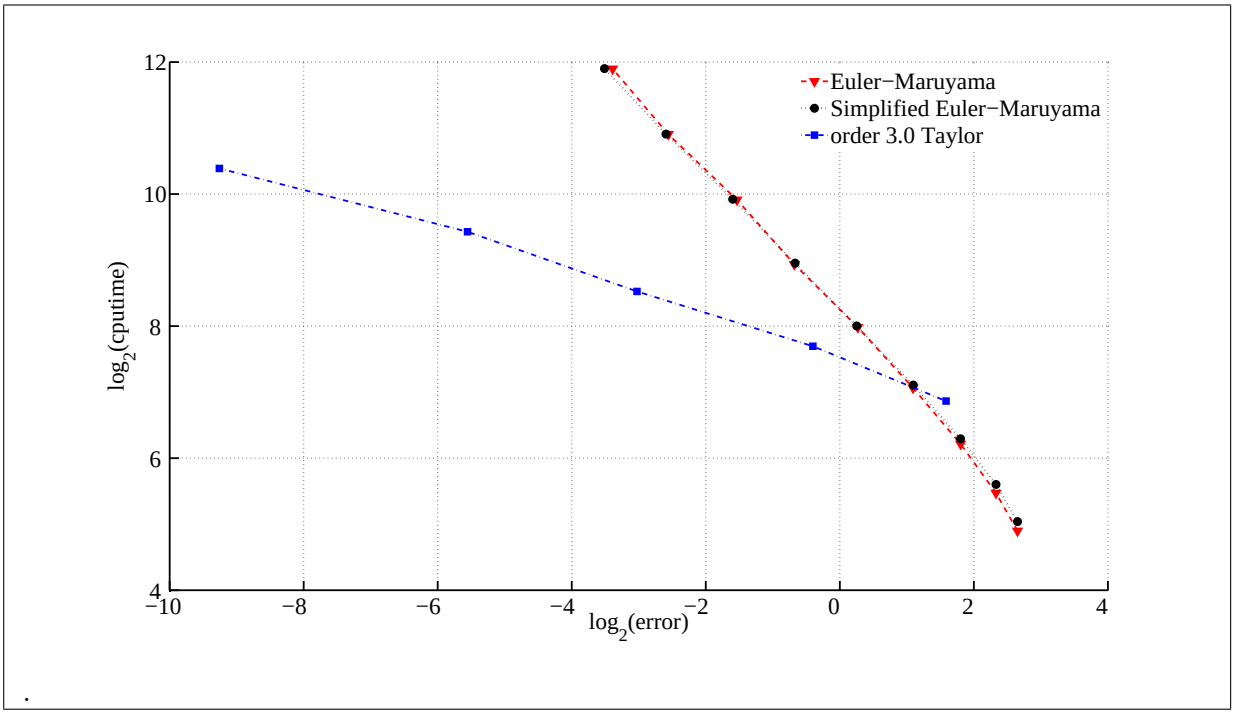
Figure 3.1: Maximum relative error (dashed line) between $\mathbb{E}[X_t^1]$ [Gaussian approximation] and $\mathbb{E}[X_t^1]$ and maximum absolute error (dotted line) between $\mathbb{E}[X_t^1]$ [Gaussian approximation] and $\mathbb{E}[X_t^1]$, both versus γ and for $t = \frac{i}{8}$, $i = 0, \ldots, 8$. α is set to 10^{-6} (although irrelevant for the computation of $\mathbb{E}[X_t^1]$ via Gaussian approximation) and β is set to 1.

moderate for low noise $\gamma \leq 0.25$. Starting with $\gamma \geq 0.3$, the relative error exceeds 1% and is unsatisfactory. We thus see from Figure 3.1 that the Gaussian approximation of (3.2.1) works fine in the case of low noise, but fails if the noise intensity is higher. We can take these results as a guideline for general SDEs with multiplicative noise.

2nd experiment: Euler–Maruyama vs. simplified Euler–Maruyama

We know from Theorem 2.2.17 that both the Euler–Maruyama scheme, see (2.2.2), and the simplified Euler–Maruyama scheme, see (2.2.7), have weak convergence order 1.0. Whereas the Euler–Maruyama scheme explicitly resembles the truncated Itô–Taylor expansion, see Subsection 2.2.3 of Chapter 2, the simplified Euler–Maruyama method uses simpler two-point distributed random variables – generated by `randi` – instead of normally distributed random variables, that are generated by `randn`. On first sight we might assume that the simplified Euler–Maruyama method requires less cputime than the Euler–Maruyama method to attain a certain error limit.
We want to show numerically if our hypothesis is right. We set $\alpha = 10^{-6}$, $\beta = 1$, $\gamma = 0.25$ and want to determine the error of $\mathbb{E}\left[X_t^1\right]$ [Euler–Maruyama] as well as $\mathbb{E}\left[X_t^1\right]$ [Simplified Euler–Maruyama] at $t = 1$ using equidistant stepsizes $\Delta = \frac{1}{2^i}$ for $i = 0, \ldots, 8$. Further we calculate $M = 20$ batches of $N = 15000$ realizations to minimize the confidence interval. Allthough we have to set N quite high since the noise generated by $\gamma = 0.25$ is also quite intense, we would also attain good results for $N \approx 8000$. The large value for N is due to the order 3.0 weak Taylor scheme that we use in the third experiment: As we want to compare the results of all schemes we have to use the same N for each. We have set α that low and barely noticeable, because we only want the noise to be generated by γ, as only γ influences $\mathbb{E}\left[X_t\right]$. Setting α to a higher value requires a higher number or realizations. All the previously mentioned steps are executed by `gaussian_approximation_explicit` that we call with the argument `0:1`.

In Figure 3.2 we plot the required cputime of each method versus the respective error on a double $\log_2$-scale. Figure 3.2 also contains the data of the order 3.0 weak Taylor scheme, that we use in the third experiment. We shall discuss the results of the order 3.0 weak Taylor scheme in the next experiment. For comparison we note that the absolute error of $\mathbb{E}\left[X_t^1\right]$ [Gaussian approximation] at $t = 1$ is around $\approx 2^{-4.44}$ and the required cputime is about $\approx 2^{-6.64}$. We do not plot the results of the Gaussian approximation in Figure 3.2 as this would distort the figure.
As each marker (triangle and circle) in Figure 3.2 denotes the result for a specific stepsize $\Delta = \frac{1}{2^i}$, $i = 0, \ldots, 8$, we see that reducing the stepsize by 50% leads to a doubling of required cputime. We note that there is barely a difference between the error of the Euler–Maruyama scheme and the simplified Euler–Maruyama scheme for a fixed stepsize,



.

Figure 3.2: Cputime versus abs. error of $\mathbb{E}[X_t^1]$ [Euler–Maruyama] (dashed line), abs. error of $\mathbb{E}[X_t^1]$ [Simplified Euler–Maruyama] (dotted line) and abs. error of $\mathbb{E}[X_t^1]$ [order 3.0 Taylor] (dash-dot line) at $t = 1$ for (3.2.1) with $\alpha = 10^{-6}$, $\beta = 1$, $\gamma = 0.25$. Each triangle and circle states the result for the respective stepsize $\Delta = \frac{1}{2^i}$, $i = 0, \ldots, 8$. Each square states the result for the respective stepsize $\Delta = \frac{1}{2^i}$, $i = 0, \ldots, 4$. For comparison: Absolute error of $\mathbb{E}[X_t^1]$ [Gaussian approximation] at $t = 1$: $\approx 2^{-4.44}$; Required cputime of $\mathbb{E}[X_t^1]$ [Gaussian approximation]: $\approx 2^{-6.64}$.

and that the convergence order of both schemes is comparable. Further, the required cputime of the Euler–Maruyama scheme is roughly the same as for the simplified Euler–Maruyama scheme, which contradicts our hypothesis at the beginning.
This has the following reason: Starting with MATLAB 5, `randn` generates normally distributed random variables by a sophisticated table lookup algorithm, cf. MOLER (2004, Chapter 9). This algorithm is based on the work of MARSAGLIA & TSANG (1984) and generates normally distributed pseudorandom numbers as fast as uniformly distributed pseudorandom numbers. As all pseudorandom number generators in MATLAB are based on uniformly distributed pseudorandom number generators, we see that `randn` actually generates normally distributed pseudorandom numbers as fast as `randi` generates two-point distributed pseudorandom numbers.
Further, we want to analyze the convergence order of the Euler–Maruyama and simplified Euler-Maruyama scheme in Figure 3.2. As the theoretical convergence order is 1.0 we expect the quotient between the error attained with stepsize Δ and the error attained with stepsize 2Δ to be $\frac{1}{2}$. If the computed quotient is above $\frac{1}{2}$ the convergence order within one step is *below* 1.0, if it is below $\frac{1}{2}$ it is *above* 1.0. The computed quotients are given in the next table, where the first column tells the $\log_2$ of the stepsize Δ, as described above.

$\log_2$(stepsize)	Quotient Euler	Quotient simpl. Euler
-1	0.799815	0.799744
-2	0.693941	0.693102
-3	0.610796	0.613202
-4	0.565287	0.557400
-5	0.518707	0.528846
-6	0.552479	0.524173
-7	0.491396	0.501649
-8	0.562013	0.529346

As shown in the above table, both schemes require a step size of at least 2^{-4} to get close to a convergence order of 1.0. Similarly to Figure 3.2 we also see that the order of convergence is comparable for both schemes.

We thus conclude that the difference between the Euler–Maruyama scheme and simplified Euler–Maruyama scheme is mainly of theoretical interest, as in practice there is no saving of time if we use the simplified Euler–Maruyama scheme. If normally distributed pseudo-random numbers are generated by an algorithm that is less efficient, e.g. the Polar Marsaglia method – cf. KLOEDEN & PLATEN (2000, p. 13) –, the simplified Euler–Maruyama scheme is of course to be preferred.

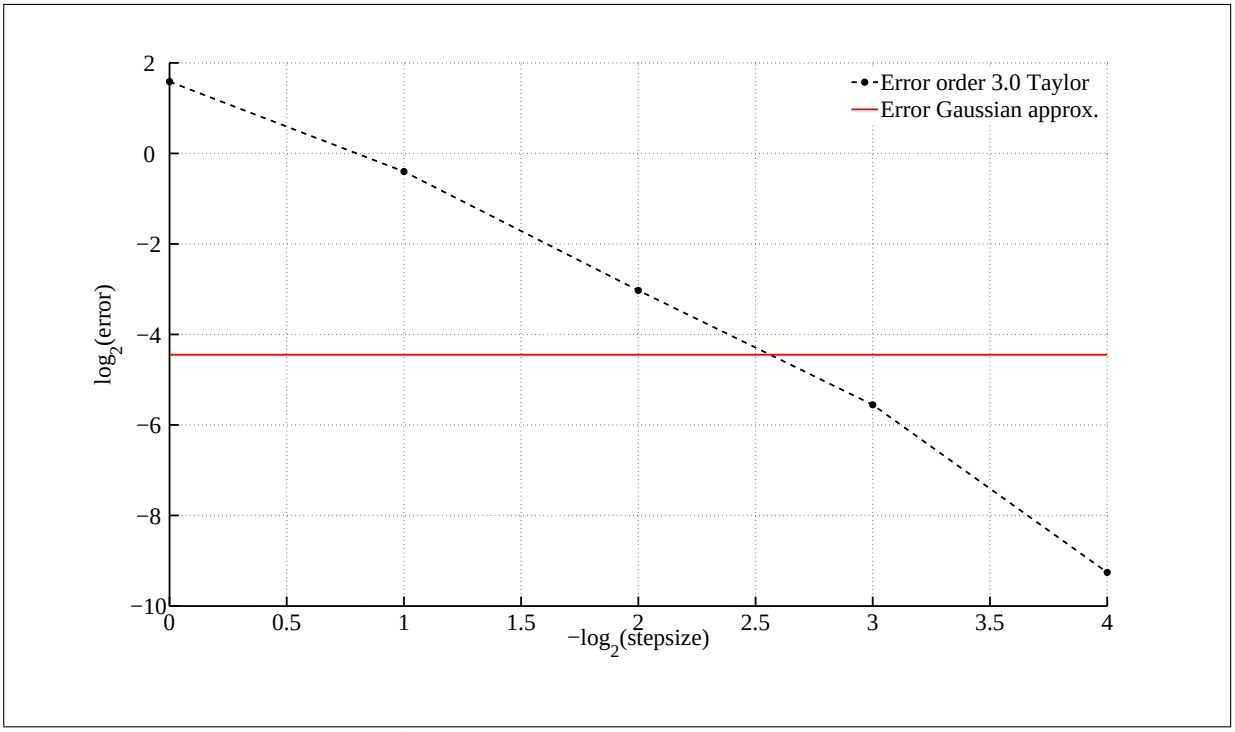


Figure 3.3: Error of $\mathbb{E}\left[X_t^1\right]$ [order 3.0 Taylor] (dashed line) at $t = 1$ for (3.2.1) with $\alpha = 10^{-6}$, $\beta = 1$ $\gamma = 0.25$ and stepsizes $\Delta = \frac{1}{2^i}$, $i = 0, \ldots, 4$. For comparison: Error of $\mathbb{E}\left[X_t^1\right]$ [Gaussian approximation] (solid line) at $t = 1$.

3rd experiment:

In this experiment we want to compare the results for the computation of $\mathbb{E}\left[X_t^1\right]$ via the order 3.0 weak Taylor scheme to the results of the Gaussian approximation at $t = 1$. From Theorem 2.2.17 we know that the order 3.0 weak Taylor scheme has the convergence order 3.0. We also want to show if we attain the same convergence order numerically.
Similarly to the last experiments we set $\alpha = 10^{-6}$ very low, $\beta = 1$ and $\gamma = 0.25$. We have chosen $\gamma = 0.25$, because we have seen in the first experiment that the maximum relative error of the Gaussian approximation is below 1% and thus still competitive. As indicated in the second experiment, we have to compute $M = 20$ batches of $N = 15000$ realizations since the Taylor scheme required this large number of realizations to attain reliable results and low confidence intervals. We simulate the Taylor scheme using stepsizes $\Delta = \frac{1}{2^i}$, $i = 0, \ldots, 4$, since we shall see that they are small enough to beat the Gaussian approximation. For each stepsize we determine the error of $\mathbb{E}\left[X_t^1\right]$ [order 3.0 Taylor] at $t = 1$ and further compute for the lowest stepsize $\Delta = \frac{1}{2^4}$ the error of $\mathbb{E}\left[X_t^1\right]$ [order 3.0 Taylor] at $t = \frac{i}{8}$, $i = 0, \ldots, 8$. All the mentioned steps are again executed by the function `gaussian_approximation_explicit` that we call with the argument `0:0.125:1`.
Figure 3.3 shows the error of $\mathbb{E}\left[X_t^1\right]$ [order 3.0 Taylor] at $t = 1$ for the stepsizes $\Delta = \frac{1}{2^i}$, $i = 0, \ldots, 4$, as well as the error of $\mathbb{E}\left[X_t^1\right]$ [Gaussian approximation] at $t = 1$. We can

see from Figure 3.3 that the Taylor scheme attains better results than the Gaussian approximation if $\Delta \geq \frac{1}{2^3}$. Moreover, Figure 3.2 shows that the order 3.0 weak Taylor scheme is more efficient than both Euler–Maruyama schemes, i.e., it attains (roughly) the same result in less time. To check the convergence order we compute in the same way as in the second experiment the quotient between the error obtained with stepsize Δ and the error obtained with stepsize 2Δ. In the case of a third order scheme this quotient should be $\frac{1}{8} = 0.125$. The computed quotients are given in the following table, where the first column indicates the $\log_2$ of the stepsize Δ.

$\log_2$(stepsize)	Quotient Taylor
-1	0.252207
-2	0.162180
-3	0.173344
-4	0.076764

We can see from this table that the convergence order of the order 3.0 weak Taylor scheme changes quite rapidly. At the beginning, the Taylor scheme comes along with some overhead and attains a convergence order between 2.0 and 2.5. The quotient of the last step corresponds to an even higher convergence order of approximately 3.7. If we fit the $\log_2$-scaled data points of Figure 3.3 in the least square sense with an affine linear curve – using `polyfit` – and determine its slope we get ≈ 2.7. Since another realizations of this experiment brought roughly the same results, we suspect the overhead within the first steps to have decreased the total convergence order of theoretical 3.0 to ≈ 2.7.
We now want to compare the result of $\mathbb{E}[X_t^1]$ [order 3.0 Taylor] with stepsize $\Delta = \frac{1}{2^4}$ to $\mathbb{E}[X_t^1]$ [Gaussian approximation] at $t = \frac{i}{8}$, $i = 0, \ldots, 8$. Figure 3.4 shows the plot of each error versus time t. In Figure 3.4 we also include the 90% confidence intervals around the error of $\mathbb{E}[X_t^1]$ [order 3.0 Taylor], that are determined in the same way as described in Subsection 2.2.1 of Chapter 2. We deduce from Figure 3.4 that $\mathbb{E}[X_t^1]$ [Gaussian approximation] lies within the 90% confidence intervals for $t \leq 0.5$ but leaves them for $t \geq 0.5$. We also see that the value 0 always lies within the 90% confidence intervals, i.e., the confidence intervals correctly estimate $\mathbb{E}[X_t^1]$. Taking into account that the maximum relative error of the Gaussian approximation is still below 1%, see the first experiment, and that the required cputime to attain this error is substantially less than for the Taylor scheme, see the second experiment and Figure 3.3, we except nonetheless the approximations of $\mathbb{E}[X_t^1]$ computed by the Gaussian approximation.

In summary, we have seen that we attain better results using higher order Taylor schemes than the Gaussian approximation. On the other hand the computation via higher order Taylor scheme requires considerably more cputime: As indicated in the second experiment, both Euler schemes obtain good and reliable results, i.e., with low confidence

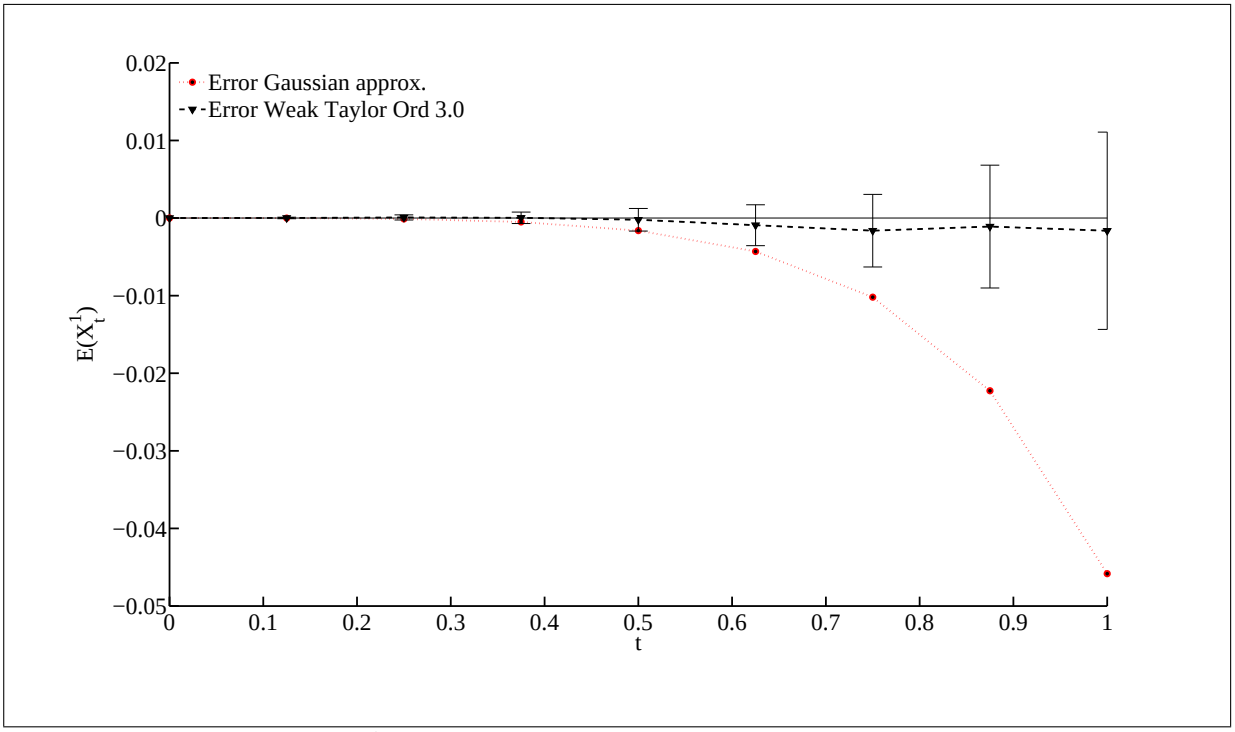


Figure 3.4: Error of $\mathbb{E}\left[X_t^1\right]$ [order 3.0 Taylor] (dashed line) with 90% confidence intervals computed for stepsize $\Delta = \frac{1}{2^4}$ and error of $\mathbb{E}\left[X_t^1\right]$ [Gaussian approximation] (dotted line), both for (3.2.1) with $\alpha = 10^{-6}$, $\beta = 1$ $\gamma = 0.25$.

intervals, even for a low number of realizations $N \approx 8000$. The Taylor scheme requires a higher number N of realizations within each batch to attain low confidence intervals. We performed the experiment using a noise coefficient $\gamma = 0.25$ that guarantees according to the first experiment a maximum relative error of the Gaussian approximation of below 1%. If we increase the noise intensity, the Gaussian approximation becomes unreliable, but in this case we also have to increase the number of simulations N to keep the confidence intervals of the one-step methods low. Finally, we note that we have always chosen α to be very low. An increase of α does not effect the results of the Gaussian approximation but it does influence the approximations computed via the one-step method. If we increase α we also have to rise the number of realizations N. I.e., in the case of high noise we have to consider other approximations based on perturbation theory, e.g. the Schwinger–Dyson method, see end of Section 2.3 of Chapter 2, or variance reduction methods for one-step methods, see Section 2.1 of Chapter 2.

3.2.2 One-dimensional SDE with additive noise

We consider the following one-dimensional SDE with additive noise

$$dX_t = (0.4X_t^3 - 3X_t)dt + dW_t; \quad X_0 = \frac{1}{2}, \tag{3.2.3}$$

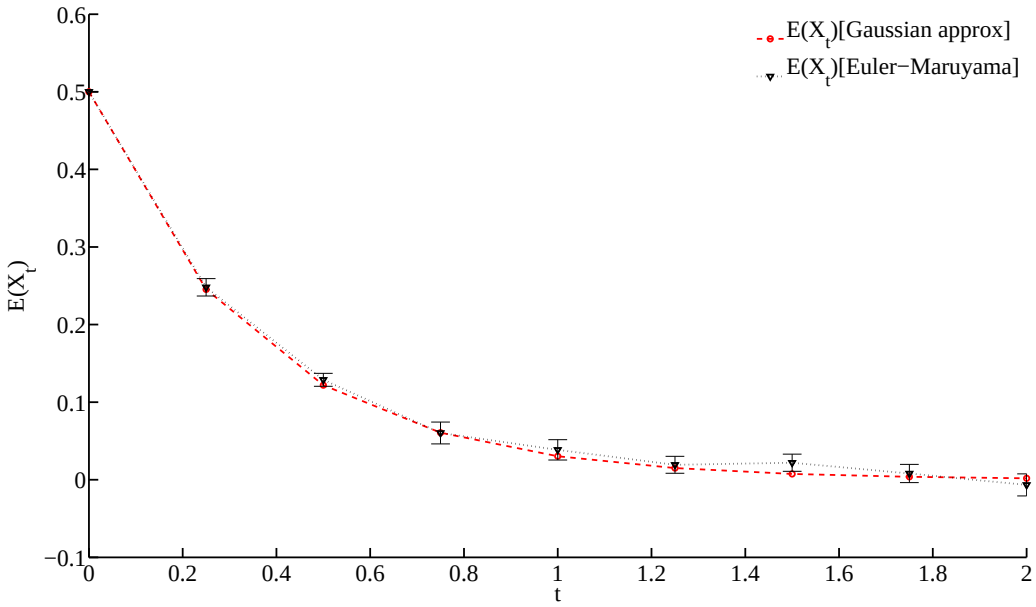
which corresponds to equation (2.3.13) with $g = 0.4$, $m = 3$ and $\sigma = 1$. We want to find weak approximations for the solution X of (3.2.3) on the time interval $t \in [0, 2]$ and compute therefore approximations of $\mathbb{E}[X_t]$ and $\mathbb{E}\left[(X_t)^2\right]$ via Gaussian approximation and the Euler–Maruyama method, see Equation (2.2.2). We then compare the results of the Gaussian approximation and Euler–Maruyama approximation at the time instants $\frac{i}{4}$, for $i = 0, \ldots, 8$.
For the Gaussian approximation we already determined at the end of Section 2.3 of Chapter 2 the corresponding system of ordinary differential equations to be

$$\begin{aligned} \frac{\partial Y^1}{\partial t}(t) &= 0.4\left(3Y^1(t)\cdot Y^2(t) - 2\left(Y^1(t)\right)^3\right) - 3Y^1(t), \\ \frac{\partial Y^2}{\partial t}(t) &= 0.8\left(3\left(Y^2(t)\right)^2 - 2\left(Y^1(t)\right)^4\right) - 6Y^2(t) + 1. \end{aligned} \tag{3.2.4}$$

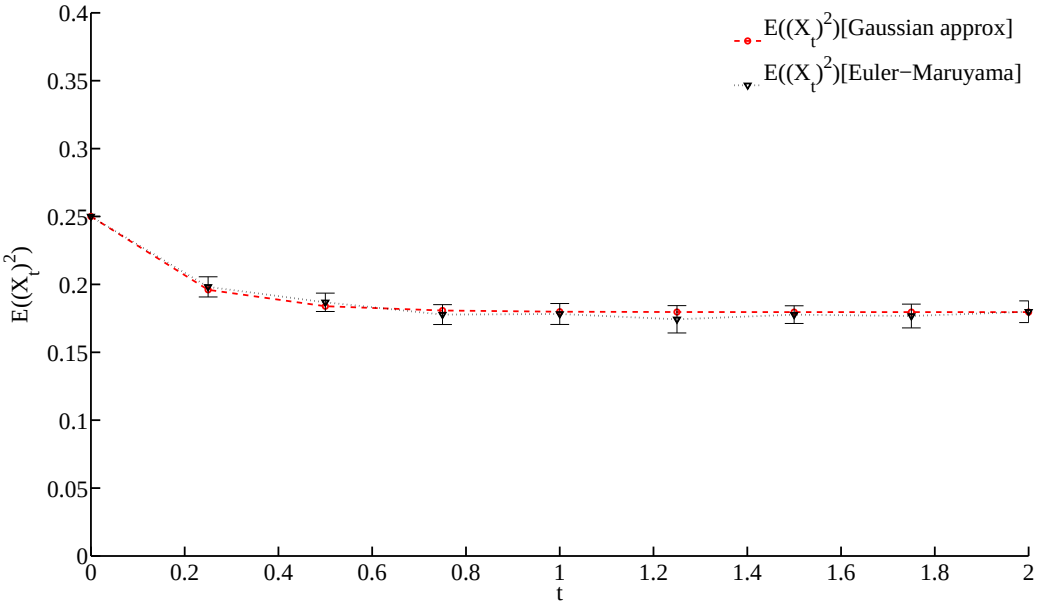
Because of $Y^1(t) \approx \mathbb{E}[X_t]$ and $Y^2(t) \approx \mathbb{E}\left[(X_t)^2\right]$ we set the initial value for this ODE to $Y(0) = (\frac{1}{2}, \frac{1}{4})$.
As regards the Euler-Maruyama method (2.2.2), we approximate the first and second moments of X on a 90% confidence interval using $M = 20$ batches of each $N = 200$ realizations. We perform the Euler–Maruyama method using a constant step size of $\Delta = 2^{-9}$. All previously mentioned steps are executed by the function `gaussian_approximation_ex1` and since we want to compare the Gaussian approximation to the Euler–Maruyama approximation at the time instants $\frac{i}{4}$, $i = 0, \ldots, 8$, we call it by handling the argument `0:0.25:2`.

Figure 3.5 shows the first and second moments computed via Gaussian approximation and the Euler–Maruyama method at the time instants $t = \frac{i}{4}$, for $i = 0, \ldots, 8$. The error bars in Figure 3.5 are centered at the values of the Euler–Maruyama approximation and indicate the 90% confidence interval. We see from Figure 3.5 that both the first and second moment of the Euler–Maruyama method are close to the result of the Gaussian approximation, in the sense that the values of the first and second moment computed by the Gaussian approximation almost always lie within the 90% confidence interval.

By subtracting from each data point of Figure 3.5 the corresponding value of the Euler–Maruyama method, we get the difference between the Gaussian approximation and Euler–Maruyama approximation. Figure 3.6 shows the resulting plots. In Figure 3.6 we have also included the 90% confidence intervals centered at 0. If the value of the difference between the Gaussian and Euler–Maruyama approximation lies within the confidence interval at t, this implies that the Gaussian approximation lies in the specific confidence interval at time instant t in Figure 3.5. We thus see from Figure 3.5 that the second moments computed by the Gaussian approximation always lie within the 90% confidence interval around the Euler–Maruyama approximation. Hence, we conclude that



(a) approximation of $\mathbb{E}\left[(X_t)\right]$



(b) approximation of $\mathbb{E}\left[(X_t)^2\right]$

Figure 3.5: First and second moments of Gaussian approximation (dashed line) and Euler–Maruyama approximation (dotted line) with 90% confidence interval computed for (3.2.3).

we can easily accept the values of the Gaussian approximation at a confidence level of 90%. The first moments computed by the Gaussian approximation also always lie in the 90% confidence interval, with the exception of the time instant $t = 1.5$. By using the specific values at $t = 1.5$

t	$\mathbb{E}[X_t]$ [Gauss]	$\mathbb{E}[X_t]$ [Euler] $\pm$90% conf. interval
1.50	0.007518	0.021926 ± 0.011096

we can further calculate

$$c_{1.5} := \frac{0.007518 - (0.021926 - 0.011096)}{2 \cdot 0.011096} \approx -0.1492$$

to be the distance of the first moment computed by the Gaussian approximation to the 90% confidence interval around the first moment computed by the Euler–Maruyama approximation, relative to the length of the confidence interval. As this relative distance $c_{1.5}$ is quite moderate we can easily accept the value of the Gaussian approximation at the time instant $t = 1.5$ without changing considerably the confidence level.

Last we note that we considered the computer experiment on the time interval $[0, 2]$ as the first moments $\mathbb{E}[X_t]$ computed via Gaussian approximation and the Euler–Maruyama method approach fast zero. In fact, $X_t = 0$ is a stable fixed point of the deterministic version of (3.2.3), i.e., for $\sigma = 0$ in (2.3.13). As we expect the stochastic solution of (3.2.3) to behave similar to the deterministic solution, most of the trajectories of the solution $X = (X_t)$ converge to 0 as t increases and only very few of them leave the stable region around 0. Hence, it suffices to consider the time interval $[0, 2]$.

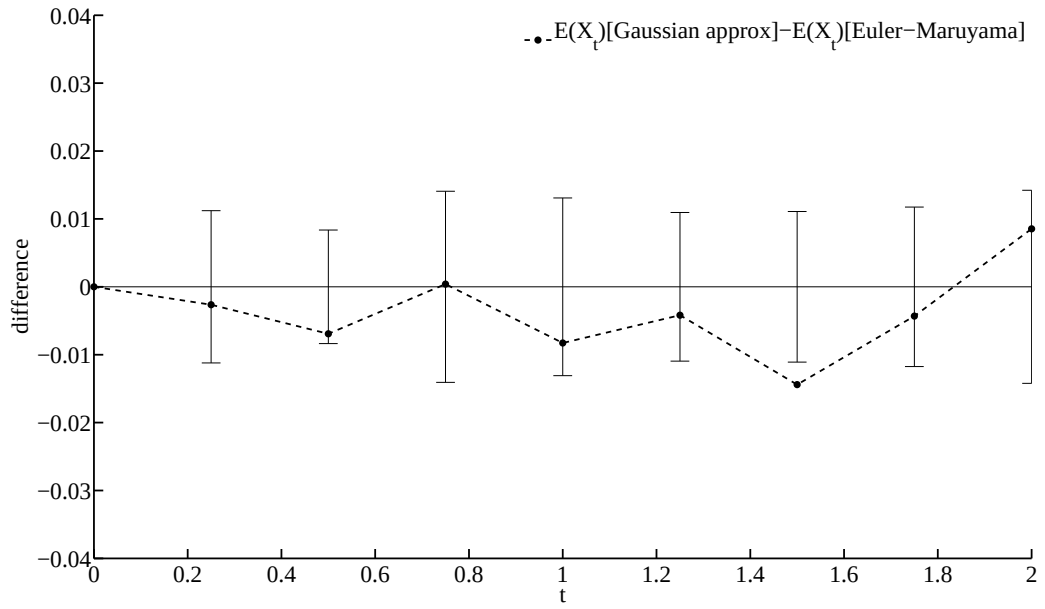
3.2.3 Two-dimensional SDE with one dimensional additive noise: Bonhoeffer-Van der Pol oscillator

In this subsection we consider the so called **Bonhoeffer–Van der Pol oscillator** with additive noise

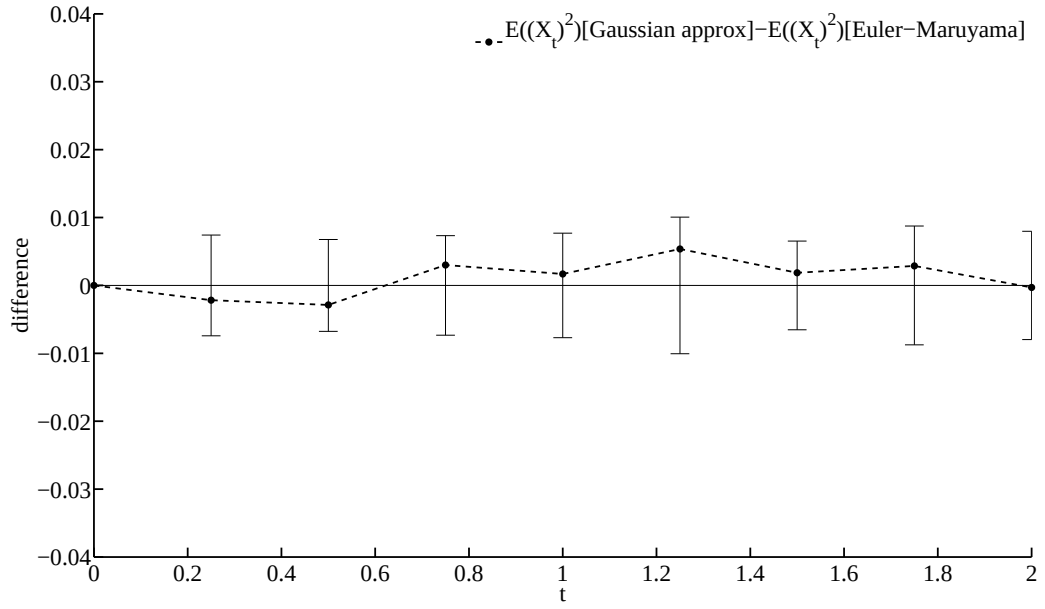
$$\begin{aligned} dX_t^1 &= c\left(X_t^1 + X_t^2 - \frac{1}{3}\left(X_t^1\right)^3 + z\right)dt + \sigma dW_t, \\ dX_t^2 &= -\frac{1}{c}\left(X_t^1 + bX_t^2 - a\right)dt, \end{aligned} \tag{3.2.5}$$

where $a, b, c, z, \sigma \in \mathbb{R}$ and $(W_t)_t$ is a one-dimensional Wiener process, cf. Kloeden et al. (1994, pp. 219–222). We consider the initial value $X_0 = (-1.9, 1.2)$ and the case $a = 0.7$, $b = 0.8$, $c = 3.0$, $z = -0.34$, because then in the noise-free case, i.e., for $\sigma = 0$, (3.2.5) has a so called limit cycle, see Figure 3.7.

The Bonhoeffer–Van der Pol equation is a 2-dimensional simplification of the 4-dimensional



(a) difference $\mathbb{E}\left[(X_t)\right]$ (Gaussian approximation)-$\mathbb{E}\left[(X_t)\right]$ (Euler–Maruyama)



(b) difference $\mathbb{E}\left[(X_t)^2\right]$ (Gaussian approximation) $-$ $\mathbb{E}\left[(X_t)^2\right]$ (Euler–Maruyama)

Figure 3.6: Difference between Gaussian approximation and Euler–Maruyama approximation in Figure 3.5. 90% confidence intervals are centered at 0.

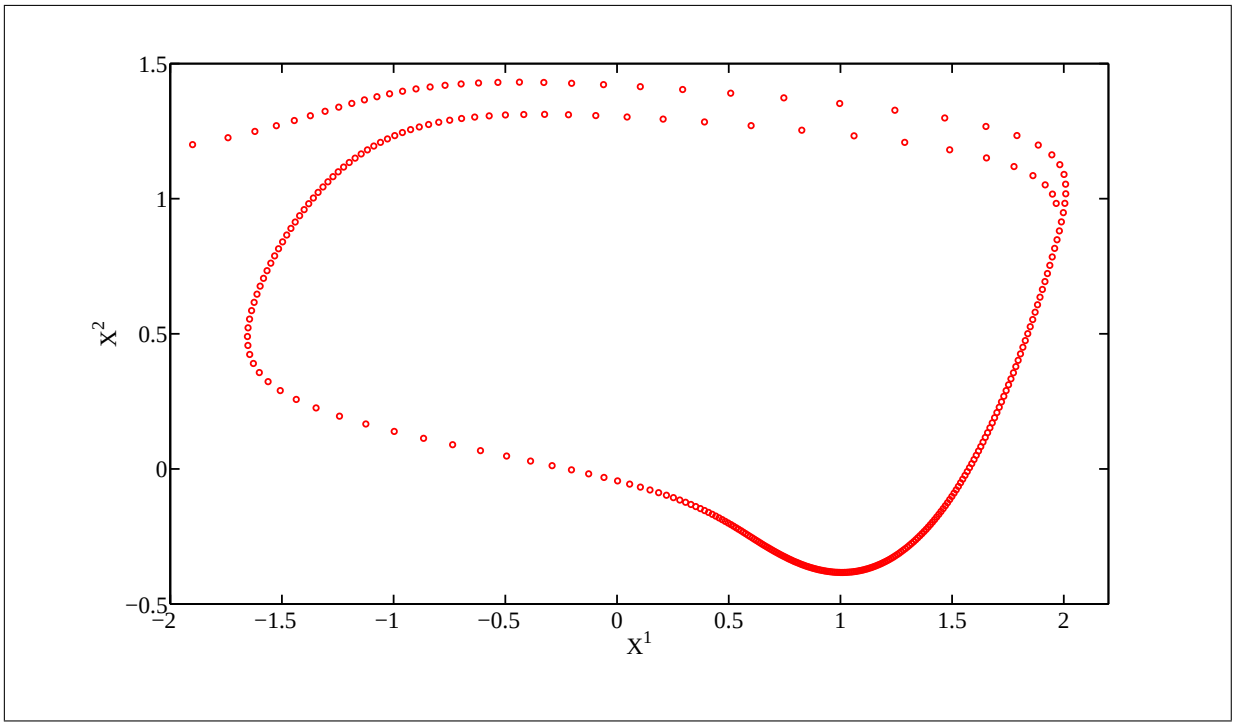


Figure 3.7: The image of the deterministic solution of (3.2.5) on $[0, 15]$, i.e., for $\sigma = 0$. The time distance between each circle is $t = 0.05$.

Hodgkin and Huxley equations to model the firing of a neuron, cf. KLOEDEN et al. (1994, p. 219). X_t^1 is the negative of the membrane voltage, X_t^2 is the permability of the membrane. By including small additive noise in the component X_t^1 with $\sigma > 0$, it is possible to simulate the effects of membrane imperfections and the firing of nearby neurons, cf. KLOEDEN & PLATEN (2000, p. 220).

Before we treat the numerical solution of the SDE (3.2.5) we first consider the noise-free case $\sigma = 0$ in which (3.2.5) turns into an autonomous ODE. Figure 3.7 shows the image of the deterministic solution of the Bonhoeffer–Van der Pol oscillator with the initial condition $X_0 = (-1.9, 1.2)$ on the time interval $[0, 15]$. First, we can clearly see the limit circle to which the solution converges. As the time distance between each plotted circle is $t = 0.05$, we can also estimate the speed of the image (X^1, X^2) of the solution. We thus see that the solution (X^1, X^2) is quite slow in the (approximate) region around $(X^1, X^2) \in [0.5, 1.5] \times [-0.1, -0.4]$ and quite fast around $(X^1, X^2) \in [-1.5, 0] \times [0, 0.25]$. In fact, the deterministic Bonhoeffer–Van der Pol oscillator has a stable fixed point approximately at $F := (0.960, -0.325) \in [0.5, 1.5] \times [-0.1, -0.4]$. Since the Bonhoeffer–Van der Pol oscillator is autonomous we also underline that the displacement of the solution (X^1, X^2) is only influenced by its current position.

If we add small additive noise $\sigma = 0.05$ we expect the stochastic solution of (3.2.5) to behave similarly to the deterministic Bonhoeffer–Van der Pol oscillator. For example we expect the stochastic solution of (3.2.5) to have a *noisy* limit circle and to behave similarly

in the change of speed. In comparing two numerical schemes at fixed time instants we have to take in account that the approximations can also rapidly vary their speed, similarly to the exact solution.

We now want to solve (3.2.5) numerically on the time interval $[0, 30]$ using Gaussian approximation and the weak order 2.0 Taylor scheme. The corresponding ODE of the Gaussian approximation is determined in the same way as explained in Section 2.3 of Chapter 2. We then find

$$\begin{aligned}
\frac{\partial Y^1}{\partial t} &= \frac{c}{3}\left(3Y^1 + 3Y^2 + 2\left(Y^1\right)^3 - 3Y^1Y^3 + 3z\right),\\
\frac{\partial Y^2}{\partial t} &= -\frac{1}{c}\left(Y^1 + bY^2 - a\right),\\
\frac{\partial Y^3}{\partial t} &= 2cY^3 + 2cY^4 + \frac{4}{3}c\left(Y^1\right)^4 - 2c\left(Y^3\right)^2 + 2czY^1 + \sigma^2,\\
\frac{\partial Y^4}{\partial t} &= \frac{1}{3c}\Big(3c^2Y^4 + 3c^2Y^5 + 2c^2Y^2\left(Y^1\right)^3 - 3c^2Y^4Y^3 + 3c^2zY^2\\
&\qquad\qquad - 3Y^3 - 3bY^4 + 3aY^1\Big),\\
\frac{\partial Y^5}{\partial t} &= -\frac{2}{c}\left(Y^4 + bY^5 - aY^2\right),
\end{aligned} \tag{3.2.6}$$

where we have dropped the argument t to alleviate notation. Further, we get the following relations $Y^1(t) \approx \mathbb{E}\left[X_t^1\right]$, $Y^2(t) \approx \mathbb{E}\left[X_t^2\right]$, $Y^3(t) \approx \mathbb{E}\left[\left(X_t^1\right)^2\right]$, $Y^4(t) \approx \mathbb{E}\left[X_t^1X_t^2\right]$, $Y^5(t) \approx \mathbb{E}\left[\left(X_t^2\right)^2\right]$ and the corresponding initial value of (3.2.6) is given by $Y(0) = ((-1.9), 1.2, (-1.9)^2, (-1.9)\cdot 1.2, (1.2)^2)$. In calculating (3.2.6) we have to apply twice Theorem 2.3.4, namely to calculate $\mathbb{E}\left[\left(X_t^1\right)^3 X_t^2\right]$ and $\mathbb{E}\left[\left(X_t^1\right)^4\right]$. From (3.2.6) we see that the additive noise coefficient σ only appears in the differential equation for Y^3. As $\sigma = 0.05$ and thus $\sigma^2 = 0.0025$, the contribution of σ is quite low and (3.2.6) is in fact very close to the deterministic Bonhoeffer–Van der Pol equation, that we would get by setting in (3.2.6) $\sigma = 0$.

To underline the assumptions that we have made above on the behavior of the stochastic solution of (3.2.5), namely the inheritance of the properties of the deterministic solution, we plot in Figure 3.8 the Gaussian approximation of (3.2.6) and 5 realizations of the order 2.0 weak Taylor scheme with stepsize $\Delta = 2^{-9}$ and in Figure 3.9 another 6 realizations of the order 2.0 weak Taylor scheme with stepsize $\Delta = 2^{-9}$, all on the time interval $[0, 30]$. Each row corresponds to one numerical scheme, in the first column we plot the component X^1 versus X^2 – for the Gaussian approximation Y^1 versus Y^2 –, in the second column X^1 – resp. Y^1 – versus time t and in the third column X^2 – resp. Y^2 – versus time t. All plots are generated by calling the program `gaussian_approximation_ex2` with the argument `0:2^(-4):30`, check the Appendix for details. Although the resolution of the

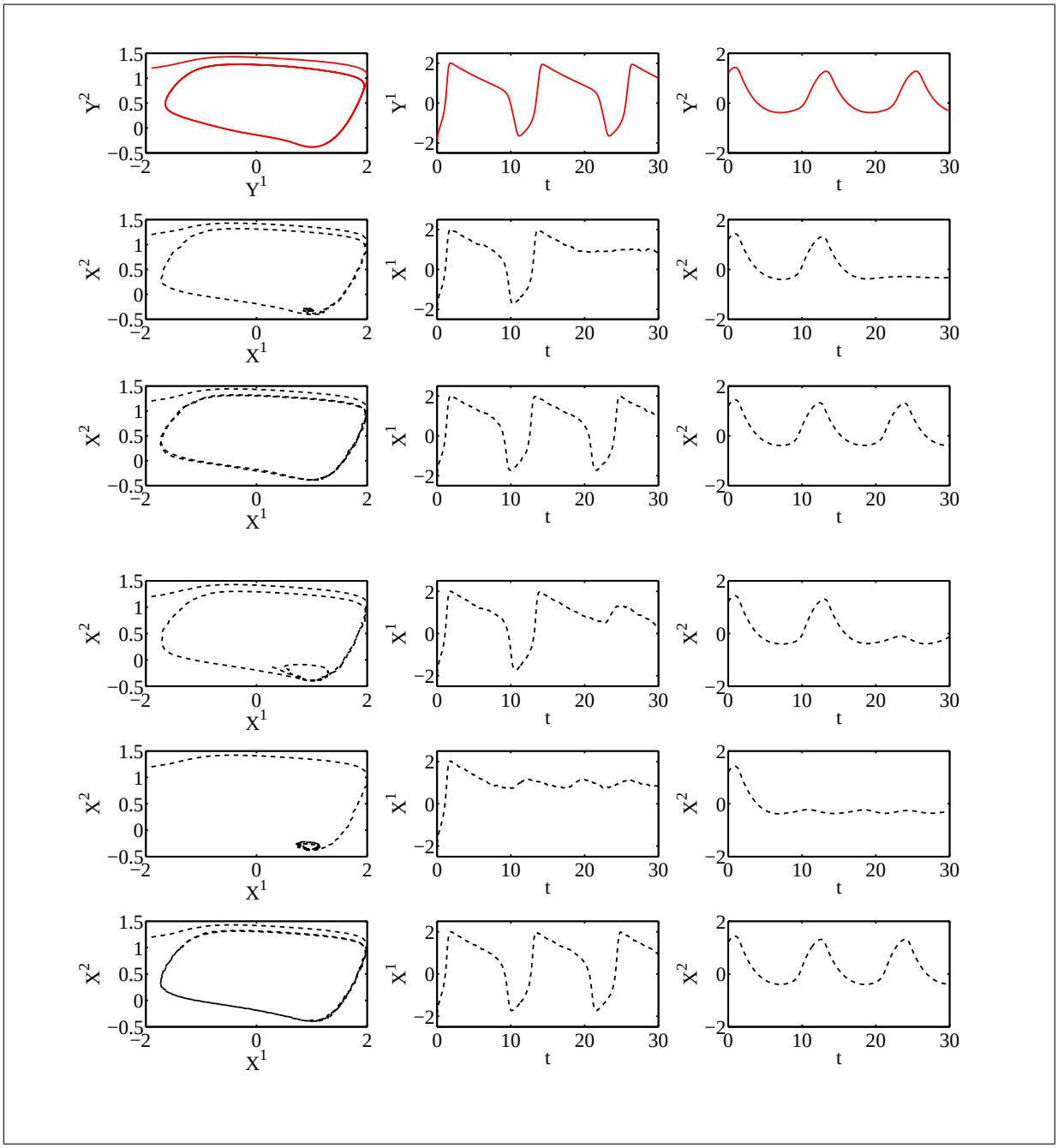


Figure 3.8: In the first row (solid line): Gaussian approximation of (3.2.5) with $\sigma = 0.05$; In row 2 - 6 (dashed line): 5 realizations of the order 2 weak Taylor scheme for (3.2.5) with $\sigma = 0.05$; In the first column: plot of first component X^1 – resp. Y^1 – versus second component X^2 – resp. Y^2 –; In the second column: plot of X^1 – resp. Y^1 – versus time t; In the third column: plot of X^2 –resp. Y^2 – versus time t.

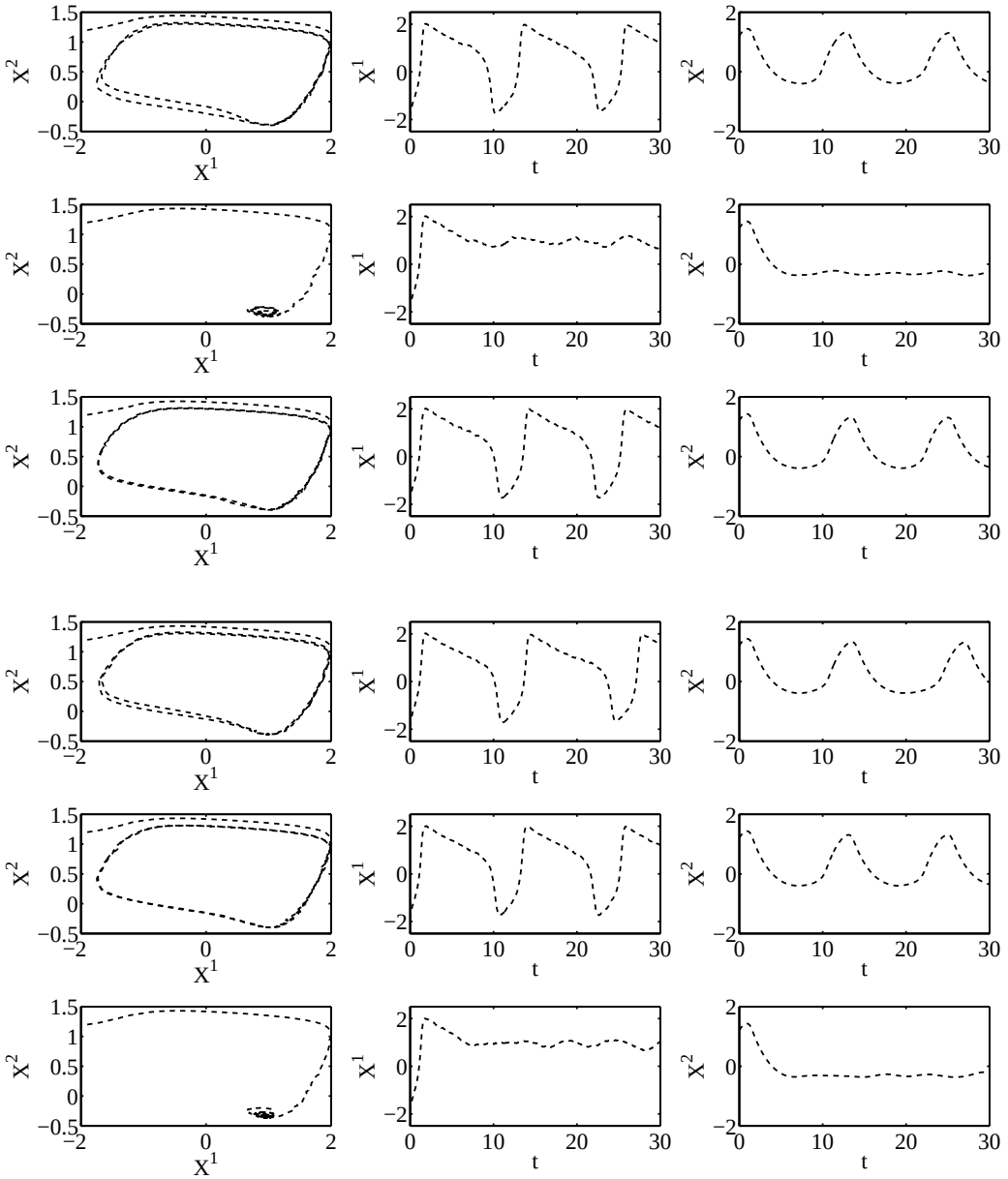


Figure 3.9: The same as Figure 3.8 but in row 1 - 6 (dashed line) another 6 realizations of the order 2.0 weak Taylor scheme for (3.2.5) with $\sigma = 0.05$.

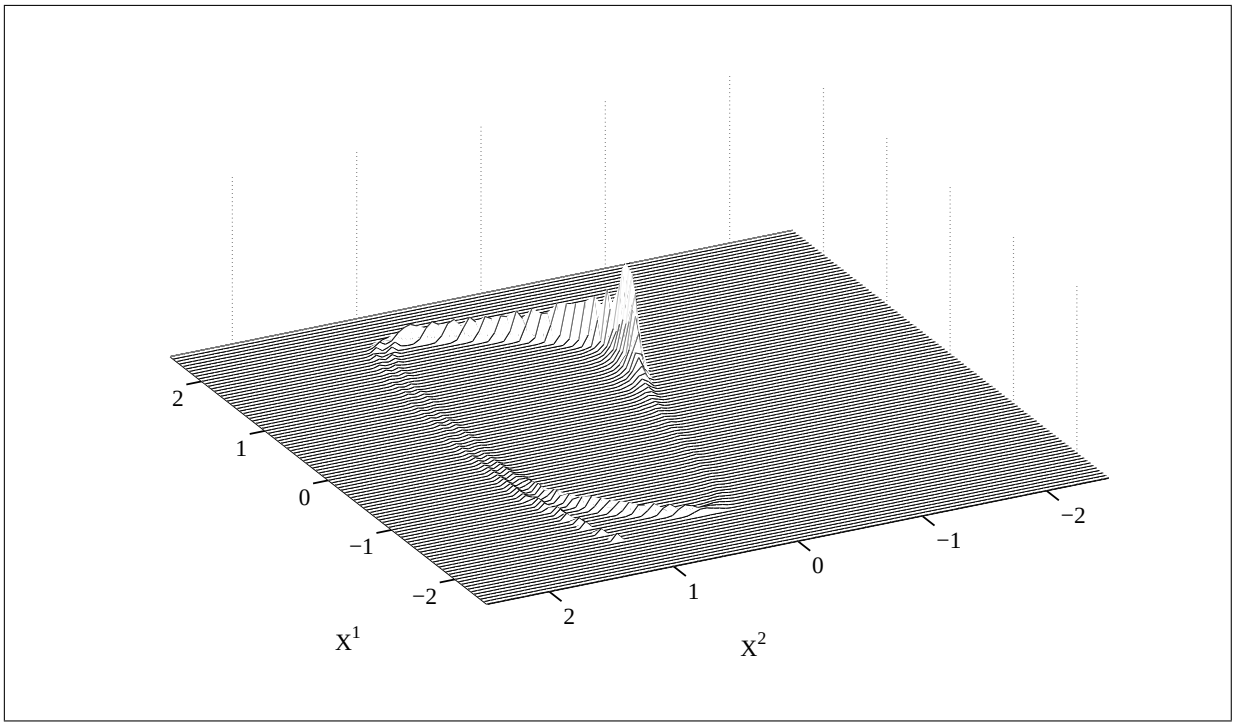


Figure 3.10: A frequency histogram for 200 realizations of the order 2.0 weak Taylor scheme for (3.2.5) with $\sigma = 0.05$.

plots in Figure 3.8 and 3.9 is not very high, we can see the following: The Gaussian approximation behaves very similar to the deterministic Bonhoeffer–Van der Pol oscillator, the plots of Y^1 versus time t, resp. Y^2 versus time t also indicate a very rapid change of the speed of Y^1 and Y^2. As indicated above we observe this similar behavior to the deterministic solution because the system of ordinary differential equation (3.2.6) is in fact very close to the deterministic system.

As regards the realizations of the order 2.0 weak Taylor scheme we observe two different kinds of behavior: Either the image of the realization moves on roughly the same path as the deterministic Bonhoeffer–Van der Pol oscillator and rotates on a noisy limit circle. Or the image of the realization gets stuck around the fixed point F of the deterministic system: As we can deduce from the plots of X^1 versus t and X^2 versus t, the realizations – once stuck around F – circulate slightly around $(1, -0.3 \pm 0.1)$ but do not return back to the limit cycle. On the other hand no realization really entered far into the interior part of the noisy limit cycle, say around the point $(X^1, X^2) = (0, 0.5)$ or $(1, 0.5)$.

Of course we have only considered a very small sample of realizations of the order 2.0 weak Taylor scheme, far too small to draw conclusions on the behavior of the expectation of the stochastic solution $X = (X^1, X^2)$ of (3.2.5). To sketch the behavior of more sample paths we partition the set $[-2.5, 2.5] \times [-2.5, 2.5] \ni (X^1, X^2)$ into a grid of squares, each of size 0.05. We then perform 200 realizations of the Taylor scheme with step size $\Delta = 2^{-9}$ and count the number of sample path iterates that fall into each grid cell. At the end we

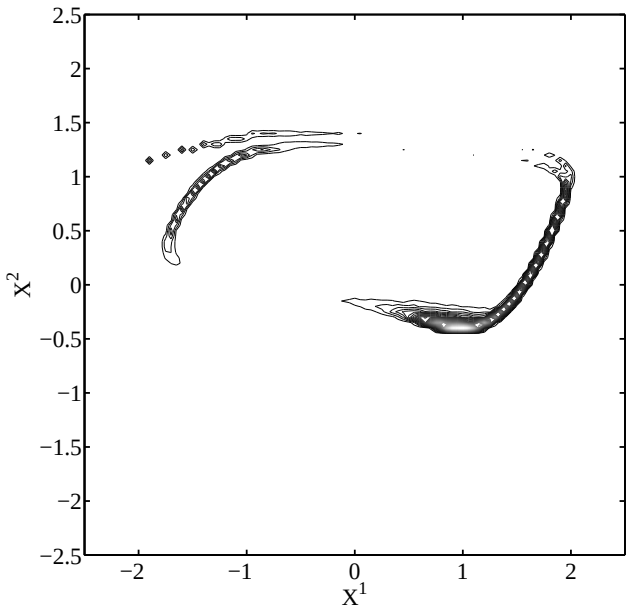


Figure 3.11: The contour plot corresponding to Figure 3.10.

divide each number by the total number of iterates. All those steps are performed by the command `gaussian_approximation_ex2(0:2^(-4):30)`. Figure 3.10 shows a frequency histogram obtained by this procedure, Figure 3.11 shows the contour plot of Figure 3.10. We deduce from Figure 3.10 that a considerable number of iterates rotate on a noisy limit cycle and indeed the highest number of iterates falls into the region near to the deterministic fixed point F. On Figure 3.11 we also see that a considerable number of iterates not only passed the fixed point like in the deterministic setting, but also stayed and probably rotated around F. In a stochastic system it is not impossible for a realization that has left once the noisy limit cycle to get back to the noisy limit cycle. But as F is a stable fixed point for the deterministic Bonhoeffer–Van der Pol equation we expect the probability of a realization getting back to the noisy limit cycle to be small.

We finally compare the result of the Gaussian approximation and the order 2.0 weak Taylor scheme on a 90% confidence interval at time instants $t = \frac{i}{2^4}$, $i = 0, \ldots, 30 \cdot 2^4$: We calculate $M = 20$ batches of $N = 200$ realizations of the Taylor scheme using a step size of $\Delta = 2^{-9}$ by the command `gaussian_approximation_ex2(0:2^(-4):30)`. Figure 3.12 shows the difference between the first moments of the Gaussian approximation and the Taylor scheme. Figure 3.13 and 3.14 show the difference between the second moments of the Gaussian approximation and the order 2.0 weak Taylor scheme. To assist the reader, we have only plotted the largest confidence interval of each plot. The largest confidence interval is of the size ≈ 0.045 and thus very small.

From the figures we see that the difference between the Gaussian approximation and

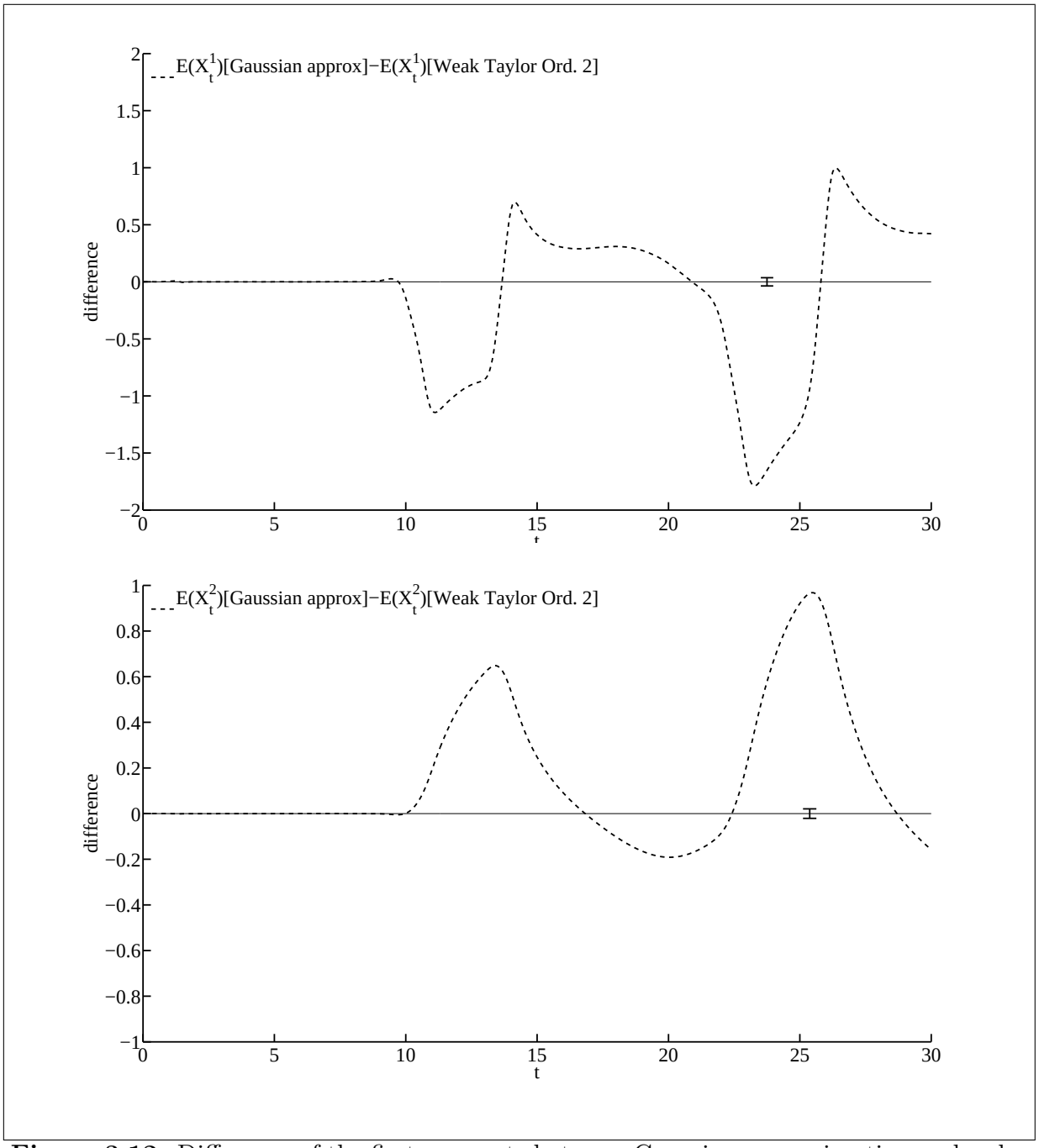


Figure 3.12: Difference of the first moments between Gaussian approximation and order 2.0 weak Taylor scheme. Only the largest confidence interval in each plot is plotted.

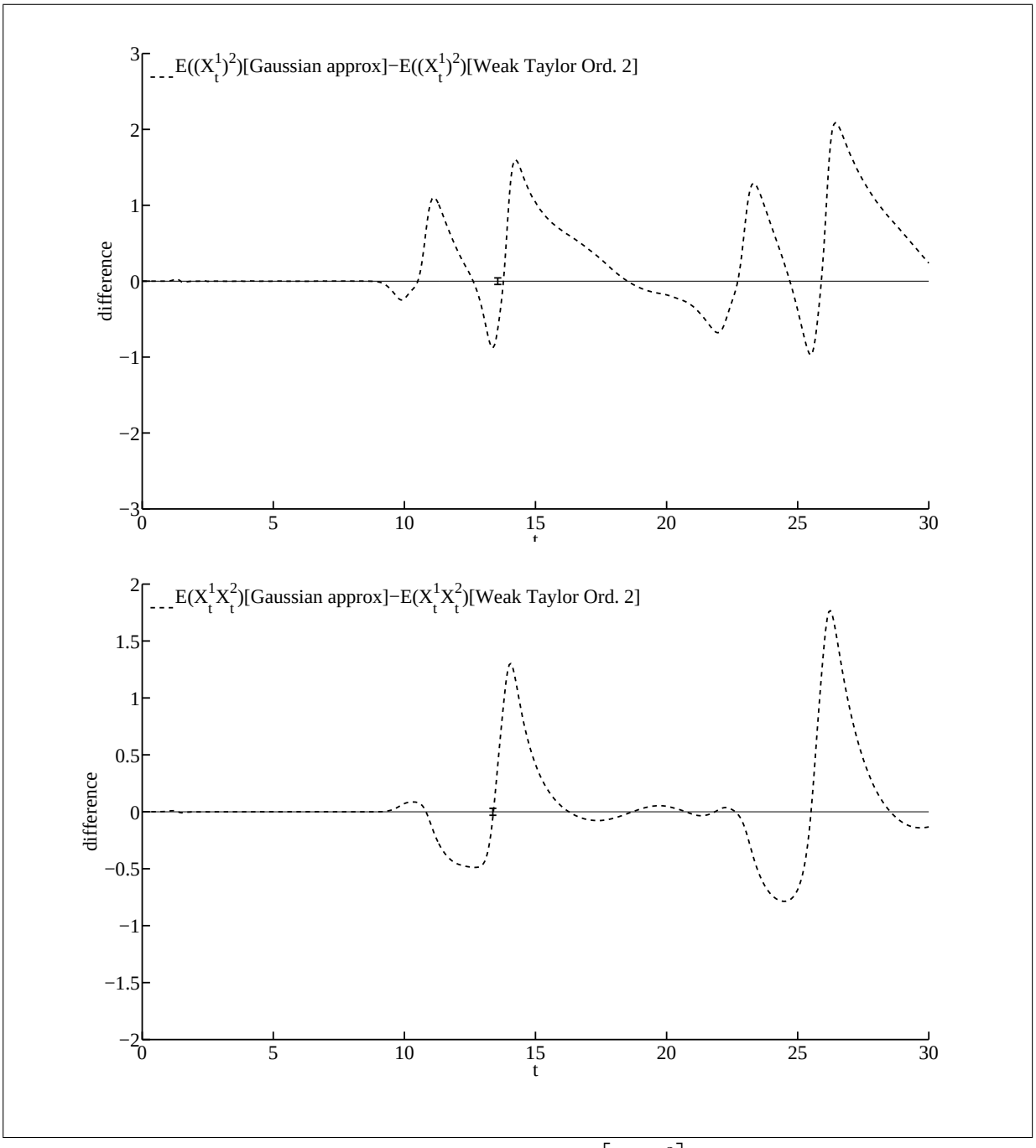


Figure 3.13: Difference of the second moments $\mathbb{E}\left[\left(X_t^1\right)^2\right]$, $\mathbb{E}\left[X_t^1 X_t^2\right]$ between Gaussian approximation and order 2.0 weak Taylor scheme. Only the largest confidence interval in each plot is plotted.

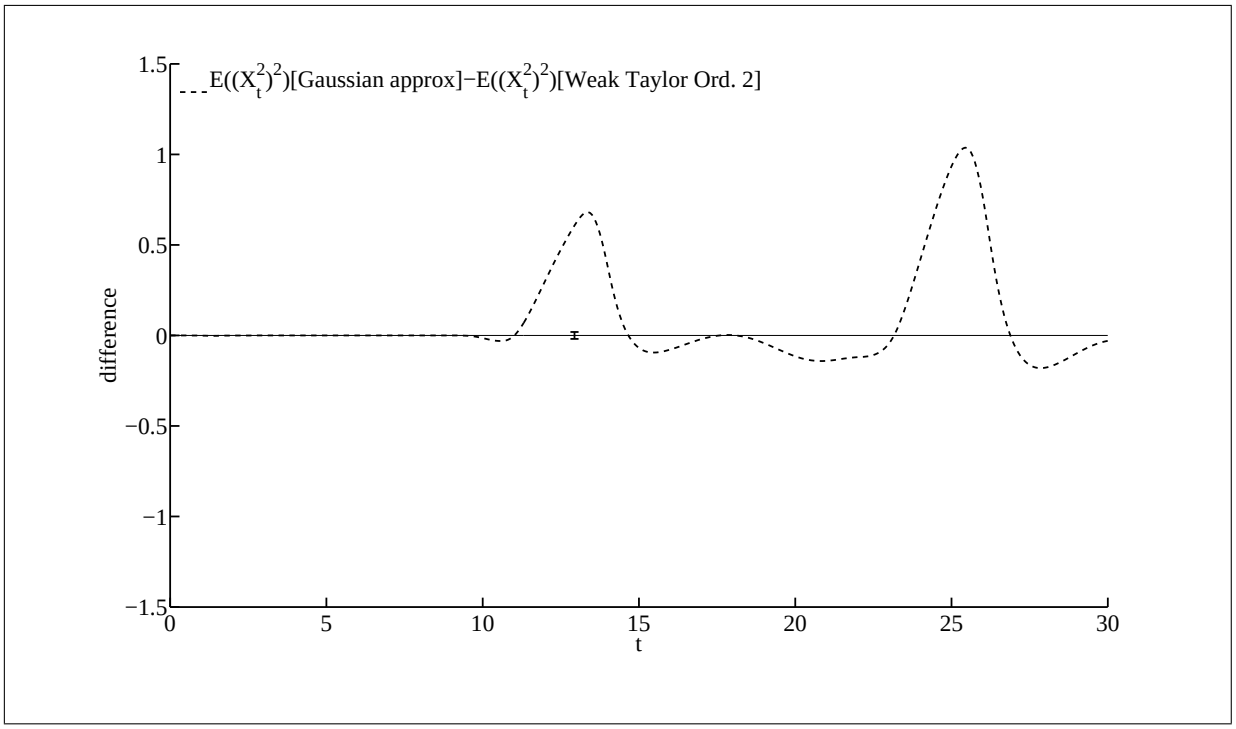


Figure 3.14: Difference of the first moment second moment $\mathbb{E}\left[\left(X_t^2\right)^2\right]$ between Gaussian approximation and order 2.0 weak Taylor scheme. Only the largest confidence interval in each plot is plotted.

the Taylor scheme becomes very large as $t \geq 10$. The values oscillate along time and the maximum difference within each plot increases along time. As indicated above we also have to take in account that temporarily small jumps are possible due to a change of speed of the trajectories. Nonetheless we can not trace back the observed behavior to a rapid change of speed of the trajectories, since the peaks of differences are detectable along a large time interval.

Figure 3.15 shows the plots of $\mathbb{E}\left[X_t^1\right]$ versus $\mathbb{E}\left[X_t^2\right]$, once computed via Gaussian approximation and once by the Taylor scheme. Figure 3.16 shows the plots of $\mathbb{E}\left[X_t^1\right]$ versus time t and $\mathbb{E}\left[X_t^2\right]$ versus time t, again in each case for the Gaussian approximation and the Taylor scheme. We see from both figures that as time increases, $\left(\mathbb{E}\left[X_t^1\right], \mathbb{E}\left[X_t^2\right]\right)$ computed by the Taylor scheme enters the center region of the limit cycle and circulates closer and closer to the deterministic fixed point F. We can give the following reason: From above we know that most of the realizations of the order 2.0 weak Taylor scheme either stay on a noisy limit cycle or circulate in a small region around F, but almost none enters the center region of the limit cycle. As the expectation $\mathbb{E}\left[\cdot\right]$ takes the average over all realizations and thus mostly the average over those two extreme behaviors, $\left(\mathbb{E}\left[X_t^1\right], \mathbb{E}\left[X_t^2\right]\right)$ can lie in the center region of the limit cycle, even if it is improbable that a considerable number of realizations enters the very same region. $\left(\mathbb{E}\left[X_t^1\right], \mathbb{E}\left[X_t^2\right]\right)$ computed by the Taylor scheme

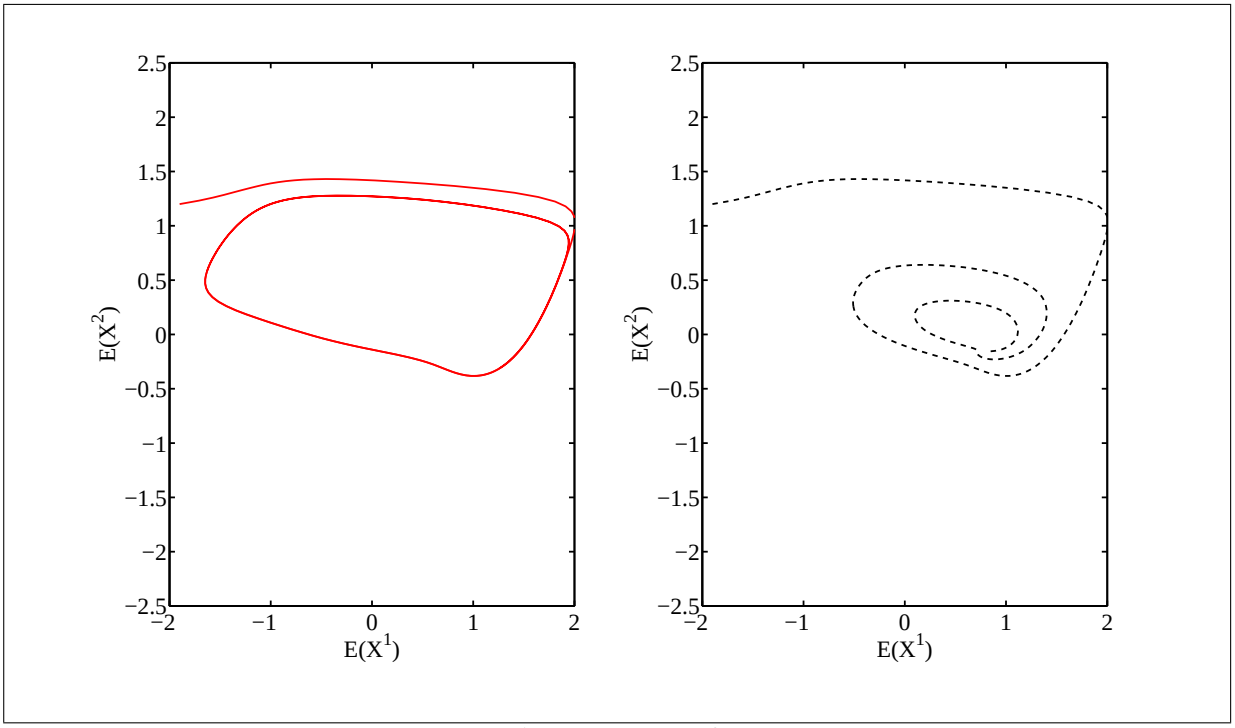


Figure 3.15: Left (solid line): $\mathbb{E}[X_t^1]$ versus $\mathbb{E}[X_t^2]$ computed by Gaussian approximation; right (dashed line): $\mathbb{E}[X_t^1]$ versus $\mathbb{E}[X_t^2]$ computed by order 2.0 weak Taylor scheme. Confidence intervals are omitted.

circulates closer and closer to F, as the deterministic fixed point F is stable and thus it is more probable for a realization to get stuck near F than to get back to the noisy limit cycle. Hence as time increases the number of realizations circulating in a region around F rises and $(\mathbb{E}[X_t^1], \mathbb{E}[X_t^2])$ gets closer to F, which explains the observed behavior.

The Gaussian approximation fails in capturing the special time evolution of the stochastic system (3.2.5) as the system of ODEs (3.2.6) is too close to the deterministic system. One might use another approximation method, such as the Schwinger–Dyson approximation, to capture a higher degree of non-Gaussian stochastic behavior. Last, we note that the observed results are not due to discretization errors of the order 2.0 weak Taylor scheme, namely to a too large step size $\Delta = 2^{-9}$. The computation of $M \cdot N = 20 \cdot 100$ realizations on the time interval $[0, 30]$ with stepsize $\Delta = 2^{-9}$ took about 230 minutes on a roughly up to date personal computer. A decrease of the step size on a smaller time interval such as $[0, 15]$ did not lead to a significant change of the results.

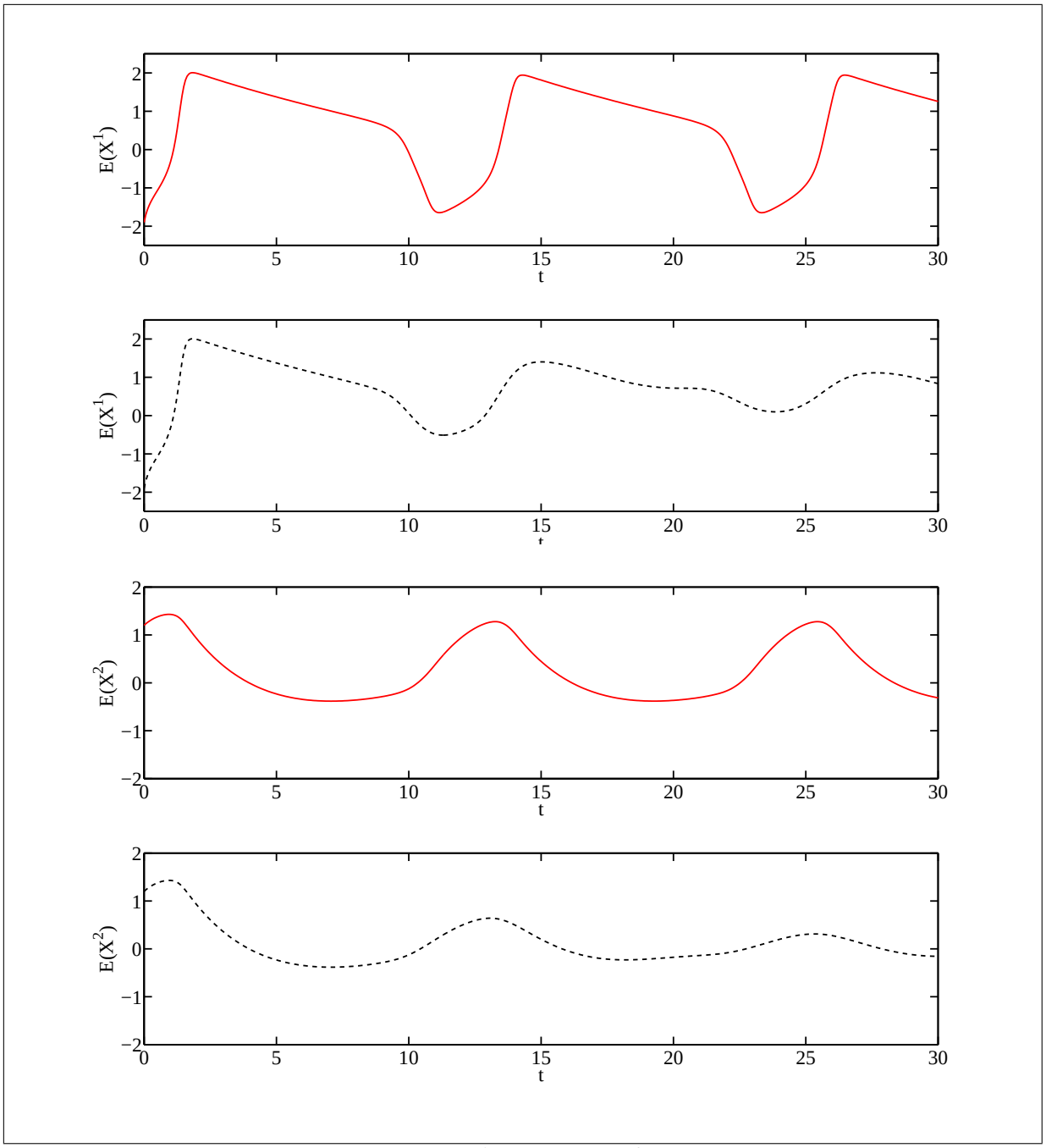


Figure 3.16: Row 1,3 (solid line): $\mathbb{E}[X_t^1]$, resp. $\mathbb{E}[X_t^2]$, versus time computed by Gaussian approximation; row 2,4 (dashed line): $\mathbb{E}[X_t^1]$, resp. $\mathbb{E}[X_t^2]$, versus time computed by order 2.0 weak Taylor scheme. Confidence intervals are omitted.

3.2.4 One-dimensional SDE with two-dimensional multiplicative noise

We consider the following one-dimensional SDE with two-dimensional multiplicative noise

$$dX_t = \exp(\alpha X_t)\, X_t dt + \beta X_t dW_t^1 + \gamma X_t dW_t^2; \quad X_0 = 0.1, \tag{3.2.7}$$

where $\alpha, \beta, \gamma \in \mathbb{R}$ and $W_t = (W_t^1, W_t^2)$ is a two-dimensional Wiener process. We want to find weak approximations of (3.2.7) on the time interval $t \in [0, 5]$. Therefore, we compute approximations of $\mathbb{E}[X_t]$ and $\mathbb{E}\left[(X_t)^2\right]$ using Gaussian approximation and the order 2.0 weak Taylor scheme, see (2.2.39). We then compare the results at the time instants $t = \frac{i}{2}$, $i = 0, \ldots, 10$.

We first have a look at the deterministic version of (3.2.7), i.e., for $\beta, \gamma = 0$. We find that $X_t = 0$ is an unstable fixpoint of the respective ODE. Further, the positive or negative growth of a solution of the ODE is only influenced by the positive or negative sign of the solution. In any case, a solution starting at $X_0 \neq 0$ converges to $\pm\infty$, depending on the $\operatorname{sgn}(X_0)$.
As regards the SDE (3.2.7), adding multiplicative noise considerably modifies the behavior of the stochastic solution $X = (X_t)$. I.e., as opposed to additive noise where the noise contribution is constant, the contribution of multiplicative noise increases linearly w.r.t. the solution X. To prevent a high variance of the solution X, which also produces a high variance of the realizations computed by the order 2.0 weak Taylor scheme, we have to inhibit most of the sample paths of X to grow too fast. Thus, to control the growth of the sample paths of X, we set $\alpha = -1$ to be negative and $\beta = 0.25$, $\gamma = 0.1$ to be moderate.

The corresponding ODE for the Gaussian approximation is determined in the same way as explained in Section 2.3 of Chapter 2 and takes the form

$$\begin{aligned}
\frac{\partial Y^1}{\partial t} =& \exp\left(\frac{1}{2}\left(Y^2 - \left(Y^1\right)^2\right) - Y^1\right) \cdot \left(\left(Y^1\right)^2 - Y^2 + Y^1\right), \\
\frac{\partial Y^2}{\partial t} =& \frac{1}{16} Y^2 + \frac{1}{100} Y^2 + 2\exp\left(\frac{1}{2}\left(Y^2 - \left(Y^1\right)^2\right) - Y^1\right) \\
& \cdot \left(\left(Y^2 - \left(Y^1\right)^2\right)^2 + \left(Y^1\right)^2 + \left(1 - 2Y^1\right)\left(Y^2 - \left(Y^1\right)^2\right)\right),
\end{aligned} \tag{3.2.8}$$

where we have dropped the argument t to assist the reader, $Y^1(t) \approx \mathbb{E}[X_t]$, $Y^2(t) \approx \mathbb{E}\left[(X_t)^2\right]$ and the initial value is given by $Y(0) = (0.1, 0.1^2)$.
As regards the order 2.0 weak Taylor scheme, we want to solve (3.2.7) using a constant step size $\Delta = 2^{-9}$. We calculate $M = 20$ batches of each $N = 500$ realizations to determine a 90% confidence intervals. We choose N that high to minimize the length of the

confidence intervals. Choosing a lower value for N resulted in the computer experiments in unsatisfactorily large confidence intervals. We want to compare the results of the Taylor scheme and the Gaussian approximation at $t = \frac{i}{2}$, $i = 0, \ldots, 10$. All previously mentioned steps are performed by the program `gaussian_approximation_ex3`, which we call with the parameter `0:2^(-1):5`.

Figure 3.17 shows the first and second moments computed by the Gaussian approximation and the Euler-Maruyama method at the time instants $t = \frac{i}{2}$, $i = 0, \ldots, 10$. We see that the result of the Taylor scheme and the Gaussian approximation are very close and the computed values for $\mathbb{E}[X_t], \mathbb{E}[X_t^2] \in [0.1, 2.5]$ are quite moderate. Figure 3.18 shows the difference between the Gaussian approximation and the Taylor scheme, both for $\mathbb{E}[X_t]$ and $\mathbb{E}\left[(X_t)^2\right]$. As regards the approximation of the first moment $\mathbb{E}[X_t]$, we see that the computation of $\mathbb{E}[X_t]$ via Gaussian approximation works very well on $[0, 4]$, i.e., the values of the Gaussian approximation always lie in the 90% confidence interval around the computed values of the order 2.0 weak Taylor scheme. We see that the value $\mathbb{E}[X_t]$ [Gaussian approximation] does not lie in the corresponding confidence intervals for $t = 4.5, 5$, reaching at $t = 5$ an absolute distance to $\mathbb{E}[X_t]$ [Weak Taylor Ord. 2] of ≈ -0.02. Similar to Section 3.2.2, we compute from the values

t	$\mathbb{E}[X_t]$ [Gauss]	$\mathbb{E}[X_t]$ [Taylor] $\pm$90% conf. interval
4.50	1.288564	1.300578 ± 0.009178
5.00	1.449324	1.466783 ± 0.011661

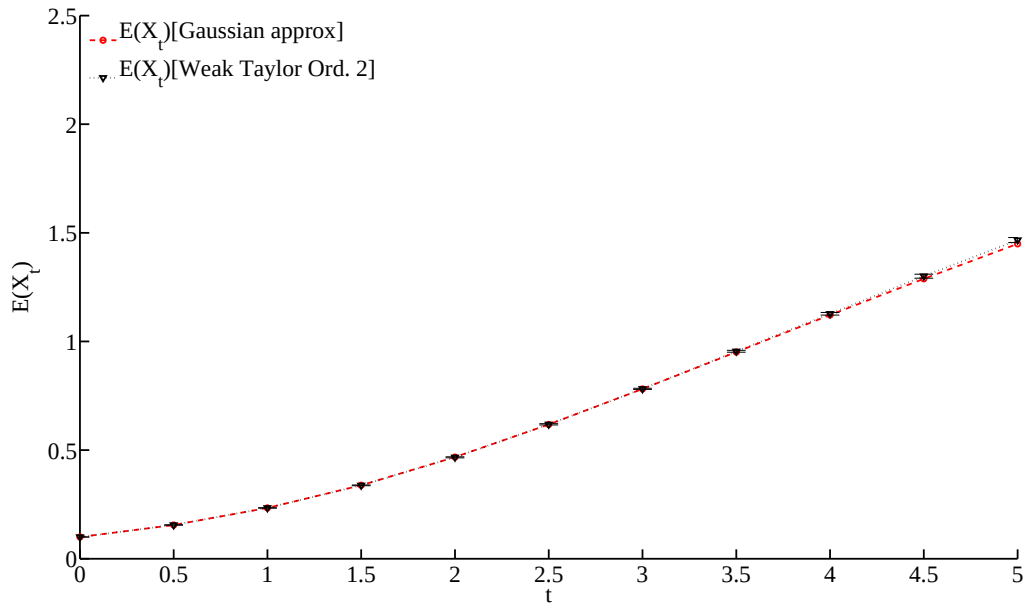
of Figure 3.17, that the distances $c_{4.5}$, c_5 of the the values of the Gaussian approximation to the respective confidence intervals at $t = 4.5, 5$, relative to the length of the confidence interval, are given by

$$c_{4.5} = \frac{1.288564 - (1.300578 - 0.009178)}{2 \cdot 0.009178} \approx -0.1545,$$
$$c_5 = \frac{1.449324 - (1.466783 - 0.011661)}{2 \cdot 0.011661} \approx -0.2486.$$

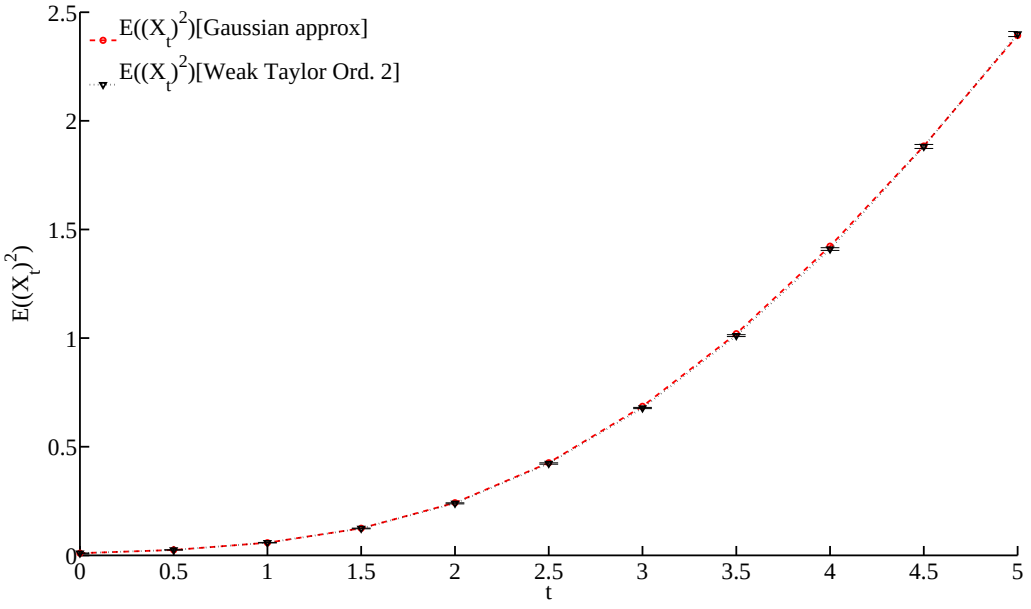
Whereas $c_{4.5}$ is still moderate at around -15%, we find c_5 at about -25% a bit high. The relative distances r_5 between $\mathbb{E}[X_t]$ [Gaussian approximation] and $\mathbb{E}[X_t]$ [Weak Taylor Ord. 2] at $t = 5$ is computed by

$$r_5 = \frac{1.449324 - 1.466783}{1.466783} \approx -0.0119$$

and thus very low. We can easily accept both approximations $\mathbb{E}[X_t]$ [Gaussian approximation] at $t = 4.5$ and $t = 5$, since the confidence level is not changed considerably at $t = 4.5$. For $t = 5$ we have to accept a more considerable change of the confidence level, but as the relative difference to the accepted value of the Taylor approximation is very



(a) approximation of $\mathbb{E}\left[(X_t)\right]$



(b) approximation of $\mathbb{E}\left[(X_t)^2\right]$

Figure 3.17: First and second moments of Gaussian approximation (dashed line) and order 2.0 weak Taylor scheme (dotted line) with 90% confidence interval computed for (3.2.7).

low, the "error" we make by accepting the value $\mathbb{E}[X_t]$ [Gaussian approximation] at $t = 5$ is moderate.

As regards the approximation of $\mathbb{E}\left[(X_t)^2\right]$, we find that the values of the Gaussian approximation always lie within the 90% confidence intervals around the Taylor scheme. We also see from Figure 3.17 that the lengths of the 90% confidence intervals increase very fast on the time interval $[3.5, 5]$. Whereas the respective confidence intervals around $\mathbb{E}[X_t]$ [Weak Taylor Ord. 2] also increase but remain moderate, i.e., the interval at $t = 5$ has double the size of the interval at $t = 3.5$, the confidence interval around $\mathbb{E}\left[(X_t)^2\right]$ [Weak Taylor Ord. 2] at $t = 5$ almost quadruples comparing to the interval at $t = 3.5$. To give reasons for the increase of the confidence intervals, we compute the dynamics of $(X_t)^2$ to be given by

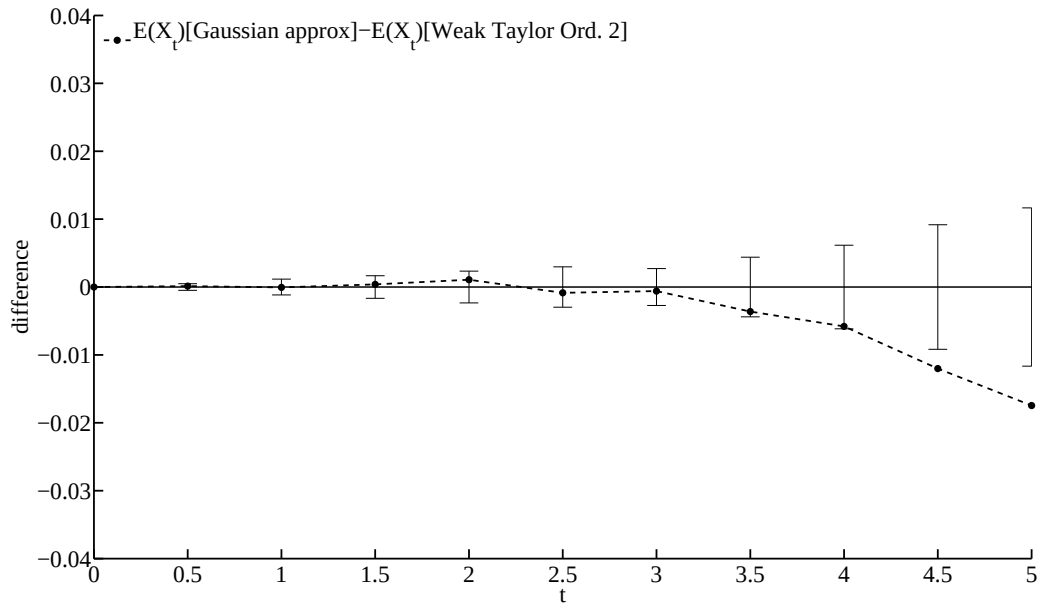
$$d\left((X_t)^2\right) = \left(2\exp(\alpha X_t) + \beta^2 + \gamma^2\right)(X_t)^2\,dt + 2\beta\,(X_t)^2\,dW_t^1 + 2\gamma\,(X_t)^2\,dW_t^2. \quad (3.2.9)$$

Comparing (3.2.9) with the dynamics of X_t in (3.2.7), we see that the relative noise contribution for $(X_t)^2$ is double as high than for X_t. As soon as a considerable number of realizations $X_t(\omega_i)$ – with $\omega_i \in \Omega$ – satisfies $|X_t(\omega_i)| \geq 1$, we find because of $(X_t(\omega_i))^2 \geq |X_t(\omega_i)| \geq 1$ also the absolute noise contribution for $(X_t)^2$ to be higher than for X_t. As the noise contribution also influences the variance of the realizations of the Taylor scheme, it also has an impact on the length of the confidence intervals. We thus expect the confidence intervals to increase considerably as soon as a considerable number of realizations $X_t(\omega_i)$ of the Taylor scheme satisfies $|X_t(\omega_i)| \geq 1$.
We can see from Figure 3.17 that $\mathbb{E}\left[(X_t)^2\right]$ [Weak Taylor Ord. 2] ≥ 1 for $t \geq 3.5$, i.e., either a lot of realizations $X_t(\omega_i)$ of the Taylor scheme satisfied $(X_t(\omega_i))^2 \geq 1$ or just a few realizations satisfied $(X_t(\omega_i))^2 \geq 1$ but in this case even $(X_t(\omega_i))^2 \gg 1$. We can thus justify the considerable increase of the confidence intervals around $\mathbb{E}\left[(X_t)^2\right]$ [Weak Taylor Ord. 2] for $t \geq 3.5$, by the rise of the noise contribution in (3.2.9).

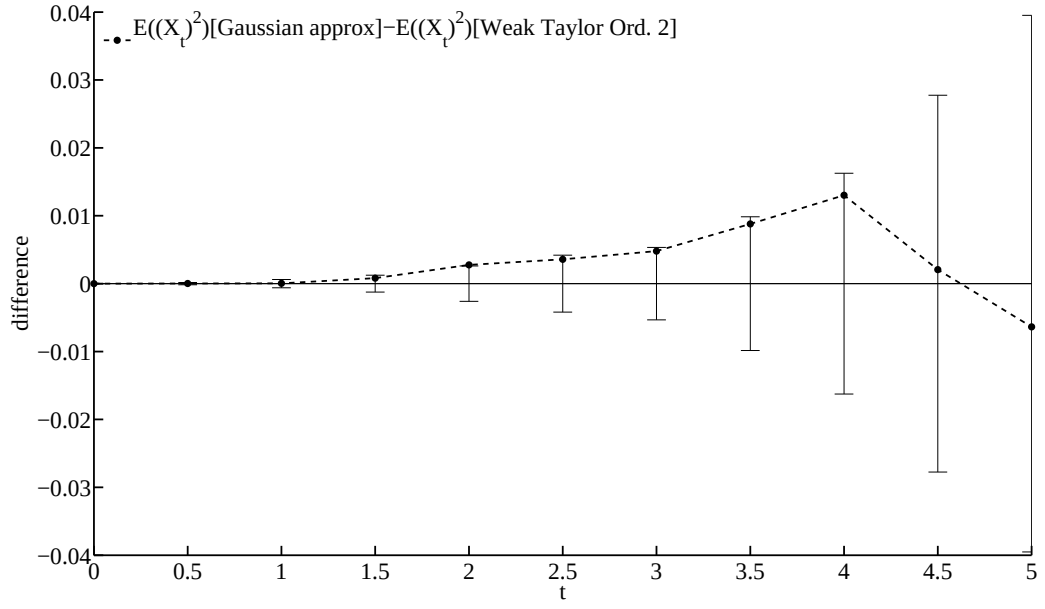
We have seen that the Gaussian approximation generates good approximations of $\mathbb{E}[X_t]$ and $\mathbb{E}\left[(X_t)^2\right]$. Greater diffusion coefficients β and γ than the ones we have chosen, lead to a higher variance of the realizations of the Taylor scheme. In order to keep in this case the size of the confidence intervals reasonable, one might apply variance reduction methods, see Section 2.1 of Chapter 2.

3.3 Conclusion

We have seen that the Gaussian approximation can give satisfactory results in the approximation of the moments of an expolynomial SDE if the noise intensity is not too high. The first experiment in the first example of an explicitly solvable SDE demonstrated that



(a) difference $\mathbb{E}[(X_t)]$(Gaussian approximation)-$\mathbb{E}[(X_t)]$(Weak Taylor Ord 2.)



(b) difference $\mathbb{E}\left[(X_t)^2\right]$(Gaussian approximation)-$\mathbb{E}\left[(X_t)^2\right]$(Weak Taylor Ord 2.)

Figure 3.18: Difference between first and second moment of Gaussian approximation and order 2.0 weak Taylor scheme in Figure 3.17. 90% confidence intervals are centered at 0.

the Gaussian approximation works only well in the case of moderate and low noise. If we increase the noise in all of the experiment, the Gaussian approximation might lead to worse results. On the other hand, in the case of higher noise we also face problems with the one-step methods of Section 2.2 of Chapter 2: We have to rise the number of realizations N of the respective one-step method to obtain reliable results, i.e., to obtain small confidence intervals. This means that both numerical methods are not satisfactorily applicable for high noise systems and we rather apply variance reduction methods for the one-step schemes, see Section 2.1 of Chapter 2, and rather use e.g. the Schwinger–Dyson approximation than the Gaussian approximation, see Section 2.3 of Chapter 2, to solve those systems satisfactorily.

Except for the Bonhoeffer–Van der Pol oscillator of Subsection 3.2.3, we have considered a rather small time interval for each SDE. If we increase the time interval we have to face on one hand higher discretization errors of the one-step schemes and on the other hand worse results of the Gaussian approximation as the stochastic behavior of the solution might resemble less and less the behavior of a Gaussian process. The latter has been one reason why the Gaussian process has failed in the moment approximation of the Bonhoeffer–Van der Pol oscillator.

After all, it is remarkable how well the Gaussian approximation computed the moments of an expolynomial SDE with low noise on a moderate time interval. Taking into account that the various realizations of the one-step methods took approximately 180 to 300 minutes and that the MATLAB function `ode45` usually solved the corresponding ODE of the Gaussian approximation within a few seconds, the Gaussian approximation is far more efficient.

Appendix

The Appendix contains all MATLAB functions that are used in Chapter 3. All functions have been tested with MATLAB R2008b (7.7.0.471).

Main functions in the chapter of computer experiments

This section contains all main functions that are called in Chapter 3.

gaussian_approximation_explicit

Matlab program header

```
function [result_1st,result_2nd,result_3rd]=gaussian_approximation_explicit(time_vector,
    filename)
%[result_1st,result_2nd,result_3rd]=gaussian_approximation_explicit(time_ve
%ctor,filename)
%
%Description: Computes E(X_t) given by the SDE
%dX_t=Y_t^2dt+alpha*X_tdW_t          X(0)=0.1
%dY_t=beta*Y_t dt+gamma*Y_t dW_t     Y(0)=1
%on t\in[0,1] in three different experiments. E(X_t) is explictly solvable.
%1st experiment:         computes E(X_t) using Gaussian approximation,
%                        alpha=10^(-6), beta=1 and gamma=0:0.05:1;
%                        the maximum rel. and abs. error at time_vector is
%                        returned in result_1st (see below for details)
%2nd experiment:         computes E(X_t) at t=1 using Euler--Maruyama and
%                        simplified Euler--Maruyama method, alpha=10^(-6),
%                        beta=1, gamma=0.25, and stepsizes
%                        2.^(0:-1:-8); the maximum abs. error for each
%                        stepsize is returned in result_2nd (see below for
%                        details)
%3rd experiment:         computes E(X_t) at t=1 using the order 3.0 weak
%                        Taylor scheme, alpha=10^(-6), beta=1, gamma=0:0.25
%                        and stepsizes 2.^(0:-1:-4); the
%                        maximum abs. error for each stepsize is returned in
%                        result_3rd (see below for details). Further the
%                        error of the Taylor scheme using stepsize 2^(-4) is
%                        compared to the Gaussian approximation at
%                        time_vector in a plot
```

```
27 %
28 % INPUT:
29 % time_vector           vector of EQUIDISTANT time instants starting at 0
30 %                       for which E(X_t) in the first and third experiment
31 %                       is computed
32 % filename              if filename is not empty then a latex
33 %                       table containing the return values
34 %                       is written into filename
35 %
36 % OUTPUT:
37 % result_1st            2xlength(time_vector) matrix that contains the
38 %                       absolute error of the Gaussian approximation at
39 %                       time_vector in the first row and the relative error
40 %                       of the Gaussian approximation at time_vector at
41 %                       time_vector in the second row
42 %
43 % result_2nd            2x9 vector that contains the
44 %                       absolute error of the Euler scheme attained
45 %                       for each of the 9 stepsizes 2.^(0:-1:-8) in the
46 %                       first row and the resp. absolute error of the simpl.
47 %                       Euler scheme in the second row
48 %
49 % result_3rd            1x5 vector that contains the absolute error of the
50 %                       order 3.0 weak Taylor scheme attained for each of
51 %                       the 5 stepsizes 2.^(0:-1:-4)
```

Requires

- eulermaruyama
- eulermaruyama_simpl
- weak_taylor_3ord
- set_plot_pref
- write_data

gaussian_approximation_ex1

Matlab program header

```
1 function [euler_compute_moments,ode_compute_moments,difference]=
      gaussian_approximation_ex1(time_vector,filename)
2 %[euler_compute_moments,ode_compute_moments,difference]=
3 %gaussian_approximation_ex1(time_vector,filename)
4 %
5 %Description: Computes 1st and 2nd moment of
6 %dX_t=(gX^3_t-mX_t)dt+dW_t; X(0)=initial_value
7 %at equidistant time_vector (e.g., [0 0.25 0.5 0.75 1])
8 %using the Euler-Maruyama scheme and Gaussian approximation
9 %
```

```
% INPUT:
% time_vector           vector of EQUIDISTANT time instants starting at 0
%                       for which the first and second moments should be
%                       computed
% filename              if filename is not empty then a Latex
%                       table containing the moments
%                       euler_compute_moments and ode_compute_moments
%                       is written into filename
%
% OUTPUT:
% euler_compute_moments 2xlength(time_vector) matrix that contains the
%                       first and second moments at time_vector computed by
%                       Euler method with stepsize
%                       2^(stepsize_exponents(end)) as follows:
%                       euler_compute_moments=[E(X);E(X^2)];
%                       where
%                       E(X)=[E(X(time_vector(1))),E(X(time_vector(2))),E(X
%                       (time_vector(3))), etc.];
%                       etc.
%
% ode_compute_moments   2xlength(time_vector) matrix that contains the
%                       first and second moments at time_vector computed by
%                       Gaussian approximation as follows:
%                       ode_compute_moments=[E(X);E(Y);E(X^2);E(XY);E(Y^2
%                       )];
%
% difference            2xlength(time_vector) matrix that contains
%                       euler_compute_moments-ode_compute_moments in the
%                       first two rows, i.e., the difference between
%                       euler_compute_moments and
%                       ode_compute_moments
```

Requires

- eulermaruyama
- set_plot_pref
- write_data

gaussian_approximation_ex2

Matlab program header

```
function [taylor_compute_moments,ode_compute_moments,difference]=
    gaussian_approximation_ex2(time_vector,filename)
%[taylor_compute_moments,ode_compute_moments,difference]=gaussian_approximation_ex2(
    time_vector,filename)
%
%Description: Computes first and second moments of Bonhoeffer-Van der Pol equation
%dX_t=c*(X_t+Y_t-1/3*(X_t)^3+z)dt+sigma*dW_t;   X(0)=initial_value(1)
%dY_t=-1/c*(X_t+b*Y_t-a)dt;                     Y(0)=initial_value(2)
```

```
7  %at time_vector (e.g., [0 0.25 0.5 0.75 1])
8  %using the weak Taylor order 2 scheme and Gaussian approximation
9  %
10 % INPUT:
11 % time_vector            vector of EQUIDISTANT time instants starting at 0
12 %                        for which the first and second moments should be
13 %                        computed
14 % filename                if filename is not empty then a latex
15 %                        table containing the moments
16 %                        taylor_compute_moments and ode_compute_moments
17 %                        is written into filename
18 %
19 % OUTPUT:
20 % taylor_compute_moments 5xlength(time_vector) matrix that contains the
21 %                        first and second moments at time_vector computed by
22 %                        the Weak Taylor Ord. 2 method as follows:
23 %                        taylor_compute_moments=[E(X);E(Y);E(X^2);E(XY);E(Y^2
24 %                        )];
25 %                        where
26 %                        E(X)=[E(X(time_vector(1))),E(X(time_vector(2))),E(X
27 %                        (time_vector(3))), etc.];
28 %                        E(Y)=[E(Y(time_vector(1))),E(Y(time_vector(2))),E(Y
29 %                        (time_vector(3))), ...];
30 %                        etc.
31 %
32 % ode_compute_moments    5xlength(time_vector) matrix that contains the
33 %                        first and second moments at time_vector computed by
34 %                        Gaussian approximation as follows:
35 %                        ode_compute_moments=[E(X);E(Y);E(X^2);E(XY);E(Y^2
36 %                        )];
37 %
38 % difference             5xlength(time_vector) matrix that contains
39 %                        ode_compute_moments-taylor_compute_moments i.e.
40 %                        the difference between ode_compute_moments and
41 %                        taylor_compute_moments
```

Requires

- `weak_taylor_2ord`
- `set_plot_pref`
- `write_data`

Comments

- The order 2.0 Weak Taylor scheme, see (2.2.39), is performed using three-point distributed random variables satisfying condition (2.2.29).
- To get plots like in Figure 3.8 and 3.9, set at the beginning of the function below the comment *%setting problem values* the default values

```
52 M_max=20;
53 N_max=200;
```

to

```
52 M_max=1;
53 N_max=11;
```

or set `N_max` to any other number of desired realizations of the order 2.0 weak Taylor scheme.

- All plots other than Figure 3.7 are computed in the case of `M_max>1`. The confidence intervals make only sense if `M_max=20`.

gaussian_approximation_ex3

Matlab program header

```
function [taylor_compute_moments,ode_compute_moments,difference]=
    gaussian_approximation_ex3(time_vector, filename)
%[taylor_compute_moments,ode_compute_moments,difference]=gaussian_approxima
%tion_ex3(time_vector, filename)
%
%Description: Computes first and second moments of
%dX_t=exp(alpha X_t)*X_t dt+beta*X_t dW_t1+gamma*X_t dW_t2;   X(0)=initial
%at time_vector (e.g., [0 0.25 0.5 0.75 1])
%using the weak Taylor order 2 scheme method and Gaussian approximation
%
% INPUT:
% time_vector           vector of EQUIDISTANT time instants starting at 0
%                       for which the first and second moments should be
%                       computed
% filename              if filename is not empty then a Latex
%                       table cotaining the moments
%                       taylor_compute_moments and ode_compute_moments
%                       is written into filename
%
% OUTPUT:
% taylor_compute_moments 2xlength(time_vector) matrix that contains the
%                       first and second moments at time_vector computed by
%                       Weak Taylor Ord. 2 method method as follows:
%                       taylor_compute_moments=[E(X);E(X^2)];
%                       where
%                       E(X)=[E(X(time_vector(1))),E(X(time_vector(2))),E(X
%                       (time_vector(3))), etc.];
%                       etc.
%
% ode_compute_moments   2xlength(time_vector) matrix that contains the
%                       first and second moments at time_vector computed by
%                       Gaussian approximation as follows:
```

```
%                       ode_compute_moments=[E(X);E(Y);E(X^2);E(XY);E(Y^2
%                       )];
%
% difference            2xlength(time_vector) matrix that contains
%                       taylor_compute_moments-ode_compute_moments i.e.
%                       the difference between taylor_compute_moments and
%                       ode_compute_moments
```

Requires

- `weak_taylor_2ord`
- `set_plot_pref`
- `write_data`

Comments

- The order 2.0 Weak Taylor scheme, see (2.2.39), is performed using three-point distributed random variables satisfying condition (2.2.29).

One-step solver

This section contains all one-step solvers for SDEs that are used by the main functions of the last section.

Euler–Maruyama scheme

This function implements the Euler–Maruyama scheme with equidistant stepsize, see (2.2.2).

Matlab program header

```
function y=eulermaruyama(initial, t0, drift,diffusion,steps,steplength)
%y=eulermaruyama(initial, t0, drift,diffusion,steps,steplength)
%
%Performs Euler-Maruyama method for general d-dimensional SDE with
%m-dimensional noise
%dX_t=drift(t,x)dt+diffusion(t,x)dW_t; X(t0)=initial
%i.e. drift(t,x)\in\mathbf{R}^d and diffusion(t,x)\in\mathbf{R}^{d\times m}
% INPUT:
% initial=          initial value X(t0) for SDE
% t0=               starting point t0
% drift=            function handle to d-dimensional drift term for SDE
% diffusion=        function handle to d\times m-dimensional diffusion term for SDE
% steps=            number of steps to be perfomed
```

```
% steplength=        steplength of each step to be performed
%
% OUTPUT
% y=                 d\times(steps+1) dimensional matrix containing values of
%                    the Euler-Maruyama method at time
%                    instants t0:steplength:steplength*steps, i.e.
%                    y(:,i)=approximation of weak_taylor_2ord at
%                    t0+steplength*(i-1)
```

Simplified Euler–Maruyama scheme

This function implements the simplified Euler–Maruyama scheme with equidistant stepsize, see (2.2.7).

Matlab program header

```
\subsection*{Euler--Maruyama scheme}
\subsubsection*{Matlab program header}
\begin{scriptsize}
  \begin{lstlisting}{eulermaruyama}
function y=eulermaruyama(initial, t0, drift,diffusion,steps,steplength)
%y=eulermaruyama(initial, t0, drift,diffusion,steps,steplength)
%
%Performs Euler-Maruyama method for general d-dimensional SDE with
%m-dimensional noise
%dX_t=drift(t,x)dt+diffusion(t,x)dW_t; X(t0)=initial
%i.e. drift(t,x)\in\mathbf{R}^d and diffusion(t,x)\in\mathbf{R}^{d\times m}
% INPUT:
% initial=          initial value X(t0) for SDE
% t0=               starting point t0
% drift=            function handle to d-dimensional drift term for SDE
% diffusion=        function handle to d\times m-dimensional diffusion term for SDE
% steps=            number of steps to be perfomed
% steplength=       steplength of each step to be performed
%
% OUTPUT
% y=                d\times(steps+1) dimensional matrix containing values of
%                   the Euler-Maruyama method at time
%                   instants t0:steplength:steplength*steps, i.e.
%                   y(:,i)=approximation of weak_taylor_2ord at
%                   t0+steplength*(i-1)
```

Order 2.0 weak Taylor scheme

This function implements the order 2.0 weak Taylor scheme with equidistant stepsize, see (2.2.39). The required random variables $\Delta\tilde{W}$ are three-point distributed satisfying condition (2.2.29).

Matlab program header

```
function y=weak_taylor_2ord(initial, t0, drift,diffusion, op1_drift, op1_diffusion, steps
    ,steplength)
%y=weak_taylor_2ord(initial, t0, drift,diffusion, op1_drift, op1_diffusion,
%steps,steplength)
%
%Performs simplified order 2.0 Weak Taylor Scheme for general d-dimensional
%SDE with m-dimensional noise
%dX_t=drift(t,x)dt+diffusion(t,x)dW_t; X(t0)=initial
%i.e. drift(t,x)\in\mathbf{R}^d and diffusion(t,x)\in\mathbf{R}^{d\times m}
% Simplified in this case means that the multi-dimensional ito integrals are
% approximated by simple two point and three point random variables as
% opposed to normally distributed random variables.
% The scheme is described in Kloeden Platen, Chapter 14.2, it involves
% terms containing the diff. operators L0, Li (i=1...m) as described in
% Kloeden Platen, eq. (10.1.1), (10.1.3). Those terms containing one
% operator -- like L0(drift) or L1(diffusion) -- are given in op1_drift
% resp. op1_diffusion, see below for details.
% INPUT:
% initial=          initial value X(t0) for SDE
% t0=               starting point t0
% drift=            function handle to d-dimensional drift term for SDE
% diffusion=        function handle to d\times m-dimensional diffusion term
%                   for SDE
% op1_drift=        function handle to matrix ((m+1)\times d) valued function
%                   containing the terms L0(drift), L1(drift), L2(drift),
%                   ... Lm(drift) as follows:
%                   op1_drift(t,x)_{i+1,j}=Li(drift_{j})(t,x) where the
%                   subindices {i+1,j} and {j} denote the {i+1,j}-th resp.
%                   {j}-th component of the matrix op1_drift(t,x) and
%                   vector drift_{j}.
% op1_diffusion=    function handle to matrix ((m+1)\times (d*m)) valued
%                   function containing the terms L0(diffusion),
%                   L1(diffusion), L2(diffusion), ...Lm(diffusion) as follows:
%                   Each row op1_diffusion(t,x)_{i+1,:} contains the terms
%                   Li(diffusion_{k,j})(t,x) for k=1...d; j=1...m, i.e. the
%                   operator Li applied to each component of the diffusion
%                   matrix diffusion(t,x). The terms within one row
%                   op1_diffusion(t,x)_{i+1,:} are ordered LEXICOGRAPHICALLY
%                   according to the index {k,j} of the diffusion matrix
%                   diffusion(t,x), i.e.
%                   op1_diffusion(t,x)_{i+1,:}=[Li(diffusion_{1,1})(t,x),Li(d
%                   iffusion_{1,2})(t,x),...,Li(diffusion_{1,m})(t,x),
%                   Li(diffusion_{2,1})(t,x),...,Li(diffusion_{2,m})(t,x),
%                   ...
%                   Li(diffusion_{d,1})(t,x),...,Li(diffusion_{d,m})(t,x)];
% steps=            number of steps to be perfomed
% steplength=       steplength of each step to be performed
%
% OUTPUT
% y=                d\times(steps+1) dimensional matrix containing values of
%                   the simplified Order 2.0 weak taylor scheme at time
%                   instants t0:steplength:steplength*steps, i.e.
%                   y(:,i)=approximation of weak_taylor_2ord at
%                   t0+steplength*(i-1)
```

Comments

- See (2.2.9) and (2.2.10) for the definition of the operators that are required for calling `weak_taylor_2ord`.

Order 3.0 weak Taylor scheme

This function implements the order 3.0 weak Taylor scheme with equidistant stepsize, see (2.2.40). This scheme only works for scalar noise.

Matlab program header

```
function y=weak_taylor_3ord(initial, t0, drift,diffusion, op1_drift, op1_diffusion,
    op2_drift, op2_diffusion, steps,steplength)
%y=weak_taylor_3ord(initial, t0, drift,diffusion, op1_drift,
%op1_diffusion, op2_drift, op2_diffusion, steps,steplength)
%
%Performs order 3.0 Weak Taylor scheme for n-dimensional SDE with scalar
%noise
%dX_t=drift(t,x)dt+diffusion(t,x)dW_t; X(t0)=initial
%i.e. drift(t,x)\in\mathbf{R}^n and diffusion(t,x)\in\mathbf{R}^{n\times 1}
% The scheme involves terms containing the diff. operators L0 and L1 as
% described in Kloeden Platen, eq. (10.1.1), (10.1.3). Those terms
% containing just one operator -- like L0(drift) or L1(diffusion) -- are
% given in op1_drift resp op1_diffusion, see below for details.
% Those terms containg two operators --like L0(L1(drift)) or L0(L0(drift)) --
% are given in op2_drift resp. op2_diffusion.
%
% INPUT:
% initial=          column vector X(t0), i.e., initial value of SDE
% t0=               starting point t0
% drift=            function handle to scalar drift term for SDE
% diffusion=        function handle to scalar diffusion term for SDE
% op1_drift=        function handle to row-vector valued function containing
%                   the terms L0(drift) and L1(drift), i.e.
%                   op1_drift(t,x)=[L0(drift)(t,x),L1(drift)(t,x)]; where
%                   Li(drift)=Li(drift^1),Li(drift^2),...,Li(drift^n)
% op1_diffusion=    function handle to row-vector valued function containing
%                   [L0(diffusion)(t,x),L1(diffusion)(t,x)]
% op2_drift=        function handle to row-vector valued function containing
%                   the terms involving twice either the operator L0 or L1
%                   and the drift. The terms are ordered lexicographically,
%                   i.e. op2_drift(t,x)=[L0L0(drift)(t,x), L0L1(drift)(t,x),
%                   L1L0(drift)(t,x),L1L1(drift)(t,x)]
% op2_diffusion=    function handle to row-vector valued function containing
%                   the terms involving twice either the operator L0 or L1
%                   and the diffusion. The vector components are ordered
%                   lexicographically.
% steps=            number of steps to be perfomed
% steplength=       steplength of each step to be performed
%
% OUTPUT
% y=                n*(steps+1) matrix containing values of the Order 3.0
```

```
41 %                    weak taylor scheme at time instants
42 %                    t0:steplength:steplength*steps
```

Comments

- See (2.2.9) and (2.2.10) for the definition of the operators that are required for calling `weak_taylor_3ord`.

Additional functions

This section contains additional functions that are used to properly layout the plots and to write data to a file with name `filename`.

write_data

All main functions in the chapter of computer experiments require `write_data` to write the computed data to `filename`.

set_plot_pref

All main functions in the chapter of computer experiments require `set_plot_pref` to set the plot preferences of each generated plot.

Bibliography

C. Archambeau, D. Cornford, M. Opper, and J. Shawe-Taylor. Gaussian process approximations of stochastic differential equations. In *Journal of Machine Learning Research Workshop and Conference Proceedings*, vol. 1, pp. 1–16, 2007.

F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of political economy*, 81(3):637–654, 1973.

W.E. Boyce. Approximate solution of random ordinary differential equations. *Advances in Applied Probability*, 10:172–184, 1978.

FM Buchmann and WP Petersen. Solving Dirichlet problems numerically using the Feynman–Kac representation. *Bit Numerical Mathematics*, 43(3):519–540, 2003.

C.W. Gardiner. *Handbook of stochastic methods*. Springer, New York, 1990.

J. Honerkamp. *Stochastische Dynamische Systeme: Konzepte, numerische Methoden, Datenanalyse*. VCH, Weinheim, 1990.

L. Isserlis. On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, 12(1-2):134–139, 1918.

O. Kallenberg. *Foundations of Modern Probability*. Springer, New York, 2002.

I. Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer, New York, 2004.

P.E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer, Berlin, 2000.

P.E. Kloeden, E. Platen, and H. Schurz. *Numerical Solution of SDE Through Computer Experiments*. Springer, Berlin, 1994.

H.J. Kushner and P. Dupuis. *Numerical methods for stochastic control problems in continuous time*. Springer, Berlin, 1992.

C.C.W. Leentvaar. Numerical solution of the black-scholes equation with a small number of grid points. Master's thesis, Delft University of Technology, Delft, 2003.

G. Marsaglia and W.W. Tsang. A fast, easily implemented method for sampling from decreasing or symmetric unimodal density functions. *SIAM Journal on scientific and statistical computing*, 5:349, 1984.

J.M. Melenk. Vorlesungsaufzeichnungen Numerik von gewöhnlichen Differentialgleichungen SS 2009, 2009. URL `http://www.math.tuwien.ac.at/~melenk/teach/num_DGL_SS09/`.

R.C. Merton. Theory of rational option pricing. *The Bell Journal of Economics and Management Science*, 4(1):141–183, 1973.

C. Moler. *Numerical Computing with MATLAB*. SIAM, Philadelphia, 2004. URL `http://www.mathworks.com/moler/chapters.html`.

B.K. Øksendal. *Stochastic differential equations: an introduction with applications*. Springer, Berlin, 2006.

P.E. Protter. *Stochastic Integration and Differential Equations*. Springer, New York, 1990.

R. Pulch. Vorlesungsaufzeichnungen zu Stochastische Differentialgleichungen: Theorie und Numerik, Wintersemester 2006/07, 2007. URL `http://www-num.math.uni-wuppertal.de/Lectures/Stdgl_ws06/skript_stochdgl.pdf`.

M. Rao. Brownian Motion and Classical Potential Theory. *Lecture Notes Series*, 47, Aarhus University, 1977.

S.E. Shreve. *Stochastic Calculus for Finance II: Continuous Time Models*. Springer, New York, 2004.

J. Stoer and R. Bulirsch. *Numerische Mathematik 2*. Springer, Berlin, 2000.

D.W. Stroock and S.R.S. Varadhan. *Multidimensional Diffusion Processes*. Springer, New York, 1979.

A.W. Van der Vaart. *Asymptotic statistics*. Cambridge University Press, Cambridge, 2000.

G.C. Wick. The evaluation of the collision matrix. *Physical Review*, 80(2):268–272, 1950.

Curriculum Vitae

Name:	Johannes Peter Temme
Birth:	20^{th} of January, 1985, in Bad Schwalbach, Germany
Citizenship:	Germany
High school diploma:	Bundesgymnasium Wien 23, Vienna Bilingual Schooling, 1230 Vienna, Draschestraße 90-92, 3^{rd} of June, 2003
Diploma studies in Mathematics:	fall term 2004 – spring term 2009 nr. of terms: 10
Other studies:	Mathematics in Science and Technology (Bac.): University of Technology, Vienna; since fall term 2007
Tutor for Mathematics:	2008: University of Technology, Vienna 2008 – 2009: University of Natural Resources and Applied Life Sciences, Vienna 2009: University of Vienna